



Program and Table of Contents

| | |
|--|-----|
| Helpful Information..... | 5 |
| Industry Support..... | 6 |
| About Us..... | 7 |
| Conference Faculty..... | 8 |
| Scripps Continuing Education Annual Course Listing | 206 |

Thursday, March 1, 2012

| | | |
|-----------|--|-----------|
| 7:30 a.m. | Registration and Breakfast | |
| 8:15 a.m. | Welcome and Opening Remarks..... | 13 |
| | Chris Van Gorder, FACHE, President and Chief Executive Officer, Scripps Health | |
| | Eric J. Topol, MD, Director, Scripps Translational Science Institute | |
| 8:30 a.m. | How Did Sequencing Our Genomes Change Our Lives?..... | 15 |
| | Joe, Retta, Noah and Alexis Beery | |

Morning Session: Changing Medical Practice – Pharmacogenomics

Moderators: Samuel Levy, PhD and Eric Topol, MD

| | | |
|------------|--|-----------|
| 9:00 a.m. | Pharmacogenetic testing to optimize treatment of patients with coronary artery disease..... | 23 |
| | Matthew J. Price, MD | |
| 9:20 a.m. | Exome sequencing to understand pharmacogenomics..... | 24 |
| | Samuel Levy, PhD | |
| 9:40 a.m. | Hypertension pharmacogenomics – Advances and Challenges..... | 25 |
| | Julie Johnson, PharmD | |
| 10:00 a.m. | <i>Break, View Exhibits and Networking</i> | |
| 10:20 a.m. | Predicting Adverse Events with Commonly Used Drugs..... | 26 |
| | Michael R. Hayden MB, ChB, PhD | |
| 10:40 a.m. | Pharmacogenomics to optimize treatment of childhood leukemia..... | 27 |
| | William E. Evans, PharmD | |
| 11:00 a.m. | Morning Panel Discussion/Q&A | |
| | Morning Session Speakers and Moderators | |

Hot Topics in Genomic Medicine

Moderators: Evan Eichler, PhD and Sarah Murray, PhD

| | | |
|------------|--|-----------|
| 11:25 a.m. | *****Incorporating Genomic Medicine into the Delivery System..... | 29 |
| | Reed Tuckson, MD | |
| 11:50 a.m. | DNA Transistors and Handheld Sequencing..... | 31 |
| | Leila Shepherd, PhD | |
| 12:15 p.m. | <i>Lunch, View Exhibits and Networking</i> | |
| 1:00 p.m. | Social Networks and Evolution..... | 32 |
| | James Fowler | |



1:30 p.m. **Genetic Basis of Autism and Developmental Delay**.....33
Evan Eichler, PhD

Afternoon Session I: Changing Medical Practice – Chronic Diseases

Moderators: Nicholas Schork, PhD and Bradley Patay, MD

1:50 p.m. **Genomic Diagnosis of Pediatric Conditions**.....34
Hakon Hakonarson, MD, PhD

2:15 p.m. **Infrastructure for clinical genomics**.....35
Isaac Kohane , MD, PhD

2:40 p.m. **Genes, pathways, environment and clinical implications in autoimmune diabetes**.....36
John A. Todd, FRS, PhD

3:05 p.m. **Afternoon Session I Panel Discussion/Q&A**
Afternoon Session I Speakers and Moderators

3:25 p.m. *Break, View Exhibits and Networking*

Afternoon Session II: Changing Medical Practice – Rare Idiopathic Diseases

Moderators: Eric Topol, MD and Nicholas Schork, PhD

3:45 p.m. **Personalized Medicine: Whole genome sequencing in the clinic**.....43
Howard J. Jacob, PhD

4:10 p.m. **Exome Sequencing in Mendelian Neurodevelopmental Disorders**.....45
Joseph G. Gleeson, MD

4:35 p.m. **Efficient detection of causative mutations for rare diseases: Rethinking clinical practice**.....46
Stanley F. Nelson, MD

5:00 p.m. **Whole-genome sequencing and the detection of disease-causing mutations**.....56
Lynn Jorde, PhD

5:25 p.m. **Afternoon Session II Panel Discussion/Q&A**
Afternoon Session I Speakers and Moderators

5:45 p.m. *Wrap Up & Adjourn*

5:45 – 7:30 p.m.

Thursday Evening Welcome Reception at Scripps Seaside Forum

Location: 8610 Kennel Way

La Jolla, California 92037

Friday, March 02, 2012

7:30 a.m. Registration and Breakfast

8:00 a.m. **The Creative Destruction of Medicine**.....94
Eric J. Topol, MD

Morning Session: New Approaches

Moderators: Aravinda Chakravarti, PhD and Hakan Sakul, PhD

8:45 a.m. **Understanding and Predicting Sudden Cardiac Death: The View from Genes**.....95
Aravinda Chakravarti, PhD

9:05 a.m. **Revisiting Lamarck and Mendel: Transgenerational genetic effects on health and disease**.....96
Joseph Nadeau, PhD



| | |
|------------|--|
| 9:30 a.m. | Healthy vs. Unhealthy Genomes: Individual and Population Perspectives..... 106 Nicholas Schork, PhD |
| 9:55 a.m. | Clinical Development Out of Academia Based on Genetic Stratification..... 108 Hakon Hakonarson MD, PhD |
| 10:20 a.m. | Morning Panel Discussion/Q&A Morning Session Speakers and Moderators |
| 10:45 a.m. | <i>Break, View Exhibits and Networking</i> |

Broad Perspectives

Moderators: Eric Topol, MD

| | |
|------------|--|
| 11:05 a.m. | The Collision: What Happens When Medicine Gets Genetic?..... 109 Matthew Herper |
| 11:30 a.m. | All Genomes Are Dysfunctional: Gene-Disrupting Variants in Healthy Individuals..... 111 Daniel MacArthur |
| 11:55 a.m. | Thinking Inside the Box - The Regulatory Science Paradox..... 118 Issam Zineh, PharmD, MPH |
| 12:20 p.m. | Panel Discussion/Q&A 5th Annual Scripps Genomic Medicine Award Presentation |
| 12:30 p.m. | <i>Lunch, View Exhibits and Networking</i> |

Afternoon Session I: Changing Medical Practice – Cancer

Moderators: Elaine Mardis, PhD and Jeffrey Trent, PhD

| | |
|-----------|---|
| 1:15 p.m. | Rx/Dx co-development: the Xalkori story..... 128 Hakan Sakul, PhD |
| 1:40 p.m. | Cancer Genomes and Personalized Medicine..... 129 Thomas J. Hudson, MD |
| 2:05 p.m. | Using Biointelligence to Search the Cancer Genome: Knowledge Recovery Efforts to Move Towards Precision Medical Genomics (PMG)..... 130 Jeffrey M. Trent, PhD |
| 2:30 p.m. | A Relentless Molecular Pursuit Approach to Take Out Pancreatic Cancer..... 131 Daniel D. Von Hoff, MD |
| 2:55 p.m. | Cancer Genome Evolution..... 132 Elaine R. Mardis, PhD |
| 3:20 p.m. | Afternoon Panel I Discussion/Q&A Afternoon Session I Speakers and Moderators |
| 3:45 p.m. | <i>Break, View Exhibits and Networking</i> |

Afternoon Session II: Diseases-in-a-Dish

Moderators: Samuel Levy, PhD and Fred Gage, PhD

| | |
|-----------|---|
| 4:05 p.m. | Modeling Human Brain Disease with patient specific iPS cells..... 133 Fred H. Gage, PhD |
|-----------|---|



4:30 p.m. **The RASopathies: Studying inherited human disease via iPS.....167**
Bruce D. Gelb, MD

4:55 p.m. **Clinical Challenges of iPSC Therapy.....177**
Joseph C. Wu, MD, PhD

5:20 p.m. **Afternoon Session II Panel Discussion/Q&A**
Afternoon Session II Speakers and Moderators

5:45 p.m. *Wrap Up & Adjourn*





Helpful Information

Parking and Transportation during the conference

There will be a looping shuttle on Thursday, Mar. 1 and Friday, Mar. 2 between The Estancia Hotel & Spa, The La Jolla Shores Hotel, Kellogg Park and Scripps Seaside Forum. The shuttle will run from 6:30 a.m. to 7:45 p.m. on Thursday and from 6:30 a.m. to 6:15 p.m. on Friday.

Parking is available in the following locations:

- Parking is available in the underground parking lot at The La Jolla Shores Hotel (\$5/day, \$14 overnight) and at Kellogg Park (free). Street parking is also available.
- There will be a looping shuttle on Thursday, Mar. 1 and Friday, Mar. 2 between Estancia La Jolla Hotel & Spa, La Jolla Shores Hotel, Kellogg Park and Scripps Seaside Forum.
- You may also park and walk to/from La Jolla Shores Hotel and Kellogg Park to Scripps Institution of Oceanography (approximately 15 minutes).
- UCSD Employees/Faculty/Students are encouraged to park in a UCSD parking lot and take the SIO shuttle to/from Scripps Seaside Forum.

Important Note for those who park in The La Jolla Shores Hotel underground parking structure: Please see the front desk and mention you are attending the Scripps Genomic Medicine Conference in order to receive a parking pass.

Thursday Evening Welcome Reception

The welcome reception will take place on Thursday from 5:45 p.m. to 7:30 p.m. at Scripps Seaside Forum. Refreshments and hors d'oeuvres will be provided.

The Future of Genomic Medicine Online Learning Module

It is highly encouraged that conference attendees view the **Scripps Genomics Primer Online Learning Module** before the conference. [Please Click here to access the FoGMV Online Learning Module.](#) *This component is designed to help conference attendees in three ways:*

- To provide online resources for furthering genomics knowledge;
- To provide reference materials and tutorial guides to prepare participants for the 2 day 'immersion' into the field of genomics;
- To make available topic-specific publications to prepare participants for the main discussion themes of the conference.

Conference Syllabus

The **conference syllabus** can be downloaded at: <http://www.scripps.org/events/future-of-genomic-medicine-v>

Internet Access during the conference

Wireless Internet is available during the conference. Upon accessing the internet you will be directed to a page that requests your email. Upon entering your information you will be registered automatically.

Social Networking

Visit our Scripps Health Facebook 2012 Future of Genomic Medicine IV events page at: <http://on.fab.me/kQoPiX> to interact with colleagues before arriving to the conference. This is a great opportunity to connect with other attendees in your specialty from all over the country.

We are on Twitter! Connect with like-minded members on topics of interest and expand your professional contacts. Follow the conversation with the following hash tag: #FOGM12.

Conference Evaluation

Your feedback is very important! Please be sure to complete the online conference evaluation <http://www.surveymonkey.com/s/futureofgenomicmedicineV> following your participation in the conference.

Questions?

Please contact Scripps Conference Services & CME at: (P) 858-652-5400 (E) Med.Edu@Scrippshealth.org





Industry Support

We would like to thank the following companies for providing generous support for this educational activity. Their support helps make this conference possible.

Life Technologies

Gen Probe

The Medicines Company

Gilead Sciences, Inc.

Johnson & Johnson

Oracle

Roche

UnitedHealth Group

Agilent

Complete Genomics

Quest Diagnostics

Genomic Health

AstraZeneca

Illumina

Pathway Genomics

UCLA Clinical Genomics Research Center

Novartis



About Scripps Translational Science Institute and Scripps Health

Scripps Translational Science Institute



SCRIPPS GENOMIC MEDICINE
A COLLABORATION OF SCRIPPS HEALTH
AND THE SCRIPPS RESEARCH INSTITUTE



Scripps Translational Science Institute (STSI), a collaborative, bridge initiative between the Scripps Health organization and The Scripps Research Institute, is transforming and integrating clinical and basic research by promoting interdisciplinary omics-based translational research efforts which advance the field of individualized medicine. STSI is supported by the NIH flagship program, Clinical and Translational Science Award (CTSA). Besides Scripps, there are collaborations with the Sanford-Burnham Institute for Medical Research, Children's Hospital of Philadelphia, Genomics Institute of the Novartis Research Foundation, J. Craig Venter Institute, Salk Institute for Biological Studies, San Diego State University, San Diego Supercomputer Center and The Neurosciences Institute on the San Diego Mesa. The Institute emphasizes all three dimensions of translation: *i*) traditional bench-to-bedside, *ii*) bedside-to-bench and back-to-bedside, and *iii*) bedside to the community medicine. STSI harnesses the excitement of today's scientific advances to catapult tomorrow's preventions, betterment of health, and the training of future leaders of academic medicine.

For more information on The Scripps Translational Science Institute (STSI), visit www.stsiweb.org.

Scripps Health



Scripps Health, a nonprofit health care system based in San Diego, California, includes more than 2,600 affiliated physicians and over 13,000 employees at five acute-care hospital campuses, home health care services and a network of clinics, physician offices and outpatient centers. Scripps is committed to contributing to the future of medicine through its clinical research trials, graduate medical education and continuing medical education programs. More information can be found at www.scripps.org.

For more information on Scripps Health, visit www.scripps.org.





Conference Faculty

Course Director

Eric J. Topol, MD

Director, Scripps Translational Science Institute
Chief Academic Officer, Scripps Health
Professor of Genomics, The Scripps Research Institute
The Gary and Mary West Chair of Innovative Medicine
La Jolla, California

Planning Committee

Samuel Levy, PhD

Director of Genomic Sciences
Scripps Translational Science Institute
Scripps Health and The Scripps Research Institute
La Jolla, California

Sarah S. Murray, PhD

Director, Genetics
Scripps Translational Science Institute
Scripps Health and The Scripps Research Institute
La Jolla, California

Nicholas J. Schork, PhD

Director of Research, Scripps Genomic Medicine
Scripps Health
Director of Biostatistics and Bioinformatics
The Scripps Translational Science Institute
Professor, Department of Molecular and Experimental Medicine, The Scripps Research Institute
La Jolla, California



Conference Faculty

Joe, Retta, Alexis & Noah Beery

*Note: Joe Beery is the Senior Vice President,
Chief Information Officer at Life Technologies*

Aravinda Chakravarti, PhD

Director, Center for Complex Disease Genomics
McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins University
Baltimore, Maryland
Director, Institute of Molecular Medicine
Delhi, India

Evan Eichler, PhD

Investigator of the Howard Hughes Medical Institute
Professor of Genome Sciences
University of Washington
Seattle, Washington

William E. Evans, PharmD

Director and Chief Executive Officer
St. Jude Children's Research Hospital
Professor of Pediatrics and Pharmacy
University of Tennessee Colleges of Medicine and
Pharmacy
Memphis, Tennessee

James H. Fowler

Professor of Medical Genetics
And Political Science
University of California, San Diego
La Jolla, California

Fred H. Gage, PhD

Professor and Vi and John Adler Chair for Research on
Age-Related Neurodegenerative Diseases
Laboratory of Genetics
Salk Institute for Biological Studies
La Jolla, California

Bruce D. Gelb, MD

The Gogel Family Professor and Director, Child Health
and Development Institute
Professor of Pediatrics and Genetics & Genomic Sciences
Mount Sinai School of Medicine
New York, New York

Joseph G. Gleason, MD

Investigator, Howard Hughes Medical Institute
Professor Neurosciences and Pediatrics
Rady Children's Hospital
University of California, San Diego
La Jolla, California

Hakon Hakonarson, MD, PhD

Director, Center for Applied Genomics
Associate Professor of Pediatrics
Department of Pediatrics and Division of Genetics
The Children's Hospital of Philadelphia
University of Pennsylvania School of Medicine
Philadelphia, Pennsylvania

Michael R. Hayden MB, ChB, PhD

Canada Research Chair in Human
Genetics and Molecular Medicine
University Killam Professor
Department of Medical Genetics

Matthew Herper

Senior Editor, Forbes Magazine

Thomas J. Hudson, MD

President and Scientific Director
Ontario Institute for Cancer Research
MaRS Centre, South Tower
Toronto, Ontario, Canada

Howard J. Jacob, PhD

Director, Human and Molecular Genetics Center
Warren P. Knowles Chair of Genetics
Professor, Department of Physiology & Pediatrics
Medical College of Wisconsin
Vice Chair of Research, Department of Pediatrics
Children's Hospital of Wisconsin
Milwaukee, Wisconsin

Julie Johnson, PharmD

V. Ravi Chandran Professor of Pharmaceutical Sciences
Distinguished Professor of Pharmacy & Medicine
Director, UF&Shands Personalized Medicine Program
Director of the Center for Pharmacogenomics
University of Florida, Colleges of Pharmacy and Medicine
Gainesville, Florida

Lynn Jorde, PhD

Professor and Chair of Human Genetics
University of Utah
Salt Lake City, Utah

Isaac S. Kohane, MD, PhD

Henderson Professor of Health Sciences and Technology
Children's Hospital and Harvard Medical School
Director, Countway Library of Medicine
Director, i2b2 national Center for Biomedical Computing
Co-Director, HMS Center for Biomedical Informatics
Boston, Massachusetts

Daniel MacArthur, PhD

Assistant in Genetics
Analytic and Translational Genetics Unit
Massachusetts General Hospital
Boston, Massachusetts



Elaine R. Mardis, PhD

Professor, Genetics and Molecular Microbiology
 Co-Director of The Genome Institute
 Washington University
 Washington University School of Medicine at Washington
 University Medical Center
 St. Louis, Missouri

Joseph H. Nadeau, PhD

Director of Research and Academic Affairs
 Institute for Systems Biology (ISB)
 Seattle, Washington

Stanley F. Nelson, MD

Professor
 Departments of Human Genetics, Pathology and
 Laboratory Medicine, and Psychiatry
 David Geffen School of Medicine
 University of California, Los Angeles
 Los Angeles, California

Matthew J. Price, MD, FACC, FSCAI

Assistant Professor, Scripps Translational Science Institute
 Director, Cardiac Catheterization Laboratory
 Division of Cardiovascular Diseases
 Scripps Clinic
 La Jolla, California

Hakan Sakul, PhD

Executive Director and Head of Diagnostics
 Worldwide R&D, Pfizer
 San Diego, California

Leila Shepherd, PhD

Chief Technology Officer
 DNA Electronics Ltd.

John A. Todd, FRS, PhD

Director, Juvenile Diabetes Research
 Foundation/Wellcome Trust Diabetes &
 Inflammation Laboratory
 Department of Medical Genetics,
 Cambridge Institute for Medical Research
 University of Cambridge, Addenbrooke's Hospital
 Cambridge, United Kingdom

Jeffrey M. Trent, PhD

President and Research Director
 Translational Genomics Research Institute (TGen)
 Phoenix, Arizona

Reed V. Tuckson, MD, FACP

Executive Vice President and Chief of Medical Affairs
 UnitedHealth Group

Chris Van Gorder, FACHE

President and Chief Executive Officer
 Scripps Health
 San Diego, California

Daniel D. Von Hoff, MD

Physician-in-Chief, Distinguished Professor
 Translational Genomics Research Institute (TGen)
 Professor of Medicine, Mayo Clinic and
 Chief Scientific Officer
 Scottsdale Healthcare and US Oncology

Joseph C. Wu, MD, PhD

Associate Professor of Medicine & Radiology
 Institute of Stem Cell Biology
 Stanford University School of Medicine
 Stanford, California

Issam Zineh, PharmD, MPH

Associate Director of Genomics
 Office of Clinical Pharmacology
 United States Food and Drug Administration
 Co-Director, Biomarker Qualification Program
 Office of Translational Sciences
 Silver Spring, Maryland



Sarah S. Murray, PhD
Director, Genetics
Scripps Translational Science Institute
Scripps Health and The Scripps Research Institute
La Jolla, California

Biography

Sarah Shaw Murray, PhD is Director of Genetics at Scripps Genomic Medicine, and Associate Professor of Translational Genomics at The Scripps Research Institute. Dr. Murray directs a high-throughput genotyping laboratory aimed at discovering genetic components of both various diseases and health, and determining risk profiles based on combinations of specific risk alleles in large prospective studies. Before her position at Scripps Genomic Medicine, Dr. Murray was the staff geneticist at Illumina, a San Diego biotech, where she played a large role in developing the technology that has enabled large-scale genome-wide genetic studies. Dr. Murray has a long history in both theoretical and applied genetics research. She has published over 70 articles in the human genetics literature that focus on the discovery and analysis of DNA sequence polymorphism. Dr. Murray received her Ph.D. from the University of Pittsburgh's Department of Human Genetics, and was a postdoctoral fellow at Case Western Reserve's Department of Genetics.

Notes

Bradley A. Patay, MD
Assistant Professor
Scripps Translational Science Institute
Head, Section of Internal Medicine
Scripps Clinic Torrey Pines
La Jolla, California

Biography

Head of the internal medicine section at Scripps Torrey Pines since 2008, Bradley A. Patay, MD, is Assistant Professor at STSI, where he is the principal investigator of a novel clinical trial to evaluate whether intensive monitoring and genomic analysis improve blood pressure control. Brad, who also assists with the *Welllderly* study, is a Diplomat in Internal Medicine in the American College of Physicians and a Diplomat in Pediatrics in the American Academy of Pediatrics.

Dedicated to improving health by integrating genomic knowledge into primary care settings, he is a member of the board and vice president of the College of Genomic Medicine, established in 2010 to educate physicians and other health care professionals about genomic medicine.

His diverse clinical experience prior to joining Scripps Clinic in 2005 includes four years as an internist and pediatrician at both Neighborhood Healthcare, a private, nonprofit community healthcare practice, and Palomar Hospital. At these two institutions in Escondido, CA, he cared for a wide range of patients, from neonates to the elderly, in both intensive care and the general wards. Through his service on several committees, including patient safety, he helped improve the health care institutions' systems.

He graduated with honors in economics and molecular cell biology from the University of California at Berkeley, where he was elected to Phi Beta Kappa. After receiving his M.D. degree at UC San Diego in 1997, he completed a medical residency at Banner Health in Phoenix Arizona and earned board certifications in internal medicine and pediatrics.

Notes

Chris Van Gorder, FACHE
President and Chief Executive Officer
Scripps Health
San Diego, California

Biography

As president and CEO of Scripps Health since 2000, Chris Van Gorder has been instrumental in positioning Scripps among the nation's foremost health care institutions. Now he is leading the restructure of the \$2.3 billion, integrated health system to best prepare for the changes of health care reform.

Board certified in health care management and an American College of Healthcare Executives (ACHE) Fellow, he also serves as Immediate Past Chairman of ACHE, an international professional society of more than 40,000 health care executives who lead hospitals, health care systems and other health care organizations. As the 2010 chairman, he provided leadership for an established network of more than 80 ACHE chapters that provide local access to networking, education and career development. ACHE is also known for its prestigious FACHE credential, signifying board certification in health care management.

At Scripps Health, Van Gorder oversees all functions of the integrated health system, including its five acute-care hospital campuses, dozens of outpatient centers throughout San Diego County, and a regional home health service. More than 13,000 employees and 2,600 affiliated physicians provide care at Scripps, which has made the Fortune magazine "100 Best Hospitals to Work For" list for four consecutive years. Recently it was named by AARP as the No. 1 employer in the nation for workers 50 and older – the first California-based company to take top honors.

Named CEO shortly after arriving at Scripps in 1999, the organization was losing \$15 million a year, and employee and physician confidence had hit bottom. He responded with a transparent, co-management style, configured an award-winning executive team, streamlined business operations and focused on internal efficiencies, physician relations and workplace culture – leading to a landmark turnaround.

Scripps now has its sights on regional growth and expansion, more than doubling its outpatient neighborhood locations, and expanding on nearly all its hospital campuses. The organization's \$2 billion growth and expansion plan for the San Diego region includes two new critical care centers, a landmark cardiovascular institute, and a new radiation therapy center. The Scripps Proton Therapy Center is scheduled to open in 2013 – one of a handful available in the United States.

Van Gorder's rise to health care executive has been unconventional. His journey began as a hospital patient, when as a police officer he was critically injured during a family dispute call. After a lengthy recovery and starting a new career in hospital security, Van Gorder continued his education in health care management and rose to levels of increased responsibility. He continues to serve the public today as a reserve commander in the San Diego County Sheriff's Department Search & Rescue Unit, as a licensed emergency medical technician (EMT) and as an instructor for the American Red Cross.

In 2006, Van Gorder's volunteer work was honored with the Maltese Cross Award by the San Diego County Fire Chiefs, and again in 2007 with the MedAssets Outstanding Humanitarian Award. In 2007, he also received USC's prestigious Guardian Award for his significant achievements in health care and commitment to the community. In 2009, he received the distinguished B'nai B'rith National Healthcare Leader award. And this year, Van Gorder was recognized by the American College of Healthcare Executives with the ACHE Exemplary Service Award.

Van Gorder received his master's degree in public administration/health services administration at the University of Southern California, completed the Wharton CEO Program at the University of Pennsylvania and earned his bachelor's degree from California State University, Los Angeles.

In March 2006, California Gov. Arnold Schwarzenegger appointed Van Gorder to the California Commission on Emergency Medical Services and, in 2010 Van Gorder was reappointed. In January 2007, U.S. Secretary of State Condoleezza Rice reappointed Van Gorder to the U.S. Commission for the United Nations Educational, Scientific and Cultural Organization (UNESCO). He is a clinical professor in health administration at the University of Southern California, where he also serves on the Board of Councilors of the university's school of policy, planning and development. In addition, Van Gorder is a member of the editorial boards of HealthLeaders and the Governance Institute.

In 2008, Van Gorder and his executive team were named the Top Leadership Team in Health Care for large health systems by HealthLeaders magazine. Since 2007, Van Gorder has been included on the "100 Most Powerful in Health Care" list, compiled by Modern Healthcare magazine. In 2011, he was listed as No. 18.

Van Gorder was part of medical history following Hurricane Katrina in 2005, when U.S. Surgeon General Richard Carmona, M.D., asked a Scripps medical unit to staff a temporary clinic for hurricane survivors in the Houston Convention Center. The request for Scripps' assistance marked the first time the federal government asked a private health care organization for long-term support for a nationally organized disaster relief plan. The Scripps Medical Response Team later provided community medical support following the 2007 San Diego County wildfires, and again in January 2010, when Van Gorder and the Scripps Medical Response Team traveled to Port au Prince, Haiti to aid victims of the devastating earthquake.

In October 2010, Van Gorder announced a new direction for Scripps Health. In anticipation of dramatic change in health care, he "turned the organization on its side," creating a horizontally matrixed management structure to identify and significantly reduce unnecessary variation in patient care and health care operations. The new "One Scripps" approach led to more than \$70 million in performance improvements in the first year.

Notes

Joe Beery
Senior Vice President
Chief Information Officer
Life Technologies

Biography



Joe Beery is Chief Information Officer for Life Technologies, a role he has held since taking the position at Invitrogen beginning September 2008. He has spearheaded the implementation of IT programs to support the merger of Invitrogen and Applied Biosystems and the seamless integration of systems supporting a 50,000-product portfolio. Prior to Invitrogen, Mr. Beery held the executive position of Chief Information Officer at US Airways and America West Airlines. Previously, Mr. Beery spent ten years at Motorola Semiconductor, holding various positions in the computer integrated manufacturing group. Mr. Beery also served as a manufacturing and software engineer at

NV Philips in Albuquerque, N.M. Mr. Beery holds a B.S. in business administration and business computer systems from the University of New Mexico.

Notes

Bloomberg

Genome Proving Cure for Ailing Twins Paves Breakthrough to Doctor's Office

By John Lauerman - Jan 30, 2012

By the time his twins Noah and Alexis were 12 years old, Joe Beery and his wife Retta had spent a decade trying to figure out what made their children so ill. After Joe took a job at [Life Technologies Corp. \(LIFE\)](#), a California company that makes DNA sequencers, their luck turned.

The company's machines revealed that the twins had been misdiagnosed and incompletely treated for more than a decade. New medication put an end to an illness that had caused vomiting, muscle weakness and seizures. Their daughter, who had spells of breathing difficulties that turned her skin blue, was healthy again.

"Genome sequencing literally saved her life," Retta Beery said.

Nine years after scientists sequenced the first [complete human genome](#) -- the instruction manual for making all the body's cells -- the industry is poised for a series of takeovers and technological breakthroughs that will bring the technology into doctor's offices and patient hospital rooms. Equipment made by Life Technologies and [Illumina Inc. \(ILMN\)](#) is spewing out human genome sequences faster than ever and prices will soon drop to \$1,000, below that of many widely used diagnostic procedures, such as colonoscopies.

Roche Holding AG's hostile \$5.7 billion bid for Illumina may spark additional deals as pharmaceutical and diagnostic companies race to bring DNA scanning into routine medical use, analysts said.

Changes in Oncology

The explosion of genomic data is creating a revolution in the treatment of some genetically driven diseases, especially cancer, said [Harold Varmus](#), director of the U.S. National Cancer Institute. With an exact understanding of the genetic alterations causing individual tumors to grow uncontrollably, doctors can target therapies for better effectiveness, he said.

"It's the biggest change I've ever seen in oncology," he said in an interview. "People are taking genetic

information they see in patients' tumors and changing therapy in dramatic ways."

The surge in clinical use of genomics will increase the global market for sequencing products to about \$10 billion from the current \$1.5 billion over the "next few years," according to Life Technologies Chief Executive Officer Greg Lucier. That would exceed the \$3.8 billion global market for magnetic resonance imaging machines. Ross Muken, an analyst with Deutsche Bank AG in [New York](#), said it may take 15 years for the sequencing market to reach Lucier's estimate.

Life Technologies rose 1 cent to \$48.56 at the close of trading in New York. The stock has gained 25 percent this year.

Inaccessible to Most

Huge obstacles loom. While the Beerys benefited from Joe's job at Carlsbad, California-based Life Technologies, whole genome sequencing isn't routinely covered by insurers. The procedure currently remains inaccessible to most patients unless they have the persistence to be chosen for clinical research studies, or can pay out of pocket.

Just as important, the volume of data that needs to be organized and analyzed is smothering efforts to make the genome applicable to day-to-day medicine. Each person's DNA code contains 6 billion chemical letters, called bases, and differs from what's considered "normal" at more than 3 million of those points. Doctors are still learning the medical significance of the millions of variations. Most of the gigabytes of genetic data churned out by sequencers around the world is, for now, incomprehensible to scientists, let alone people who see patients daily.

'Tidal Wave'

"We are standing on the beach, and a tidal wave of information is coming toward us," says [David King](#), CEO of [Laboratory Corp. of America Holdings](#) in Burlington, [North Carolina](#), which does medical testing.

Health insurers are also reluctant to underwrite testing such as a full genome sequence that may encourage plan participants to use and demand services that might be needless in many cases, King said at a conference at [Harvard Medical School](#) in [Boston](#) last year.

"If you want to see their hair stand on end, ask them about a test that's going to be offered to everyone to see if they have some genetic predisposition to obesity or diabetes or something else," King said. "They will run from the room screaming."

Even if it costs just \$1,000 to sequence a genome, there are questions that must be answered before

insurers will pay for it, said Susan Pisano, a spokeswoman for America's Health Insurance Plans, a Washington-based industry group.

How It's Done

"It's whether there's evidence that it makes a difference to the health of individuals," Pisano said. "Are we confident that there's an adequate level of accuracy? And what are we going to do with the results?"

(Bloomberg News is examining the personal, medical and business impact of genome sequencing. Reporter John Lauerman gave a blood sample in September for sequencing, and the results will be detailed in a future article.)

Most sequencers are high-powered cameras costing from \$50,000 to \$700,000 that can read and arrange the four chemical bases of DNA -- called A, C, G and T -- quickly, accurately and in order. DNA from a person's tissue is chopped into pieces, assembled on a slide, and the four bases are labeled fluorescently or with some other marker, so they can be detected by a camera. Finally those millions of pieces are reassembled into one single DNA sequence.

Machines with newer technology, such as Life Technologies' Personal Genome Machine and Ion Proton sequencers, sense the different chemical bases by measuring the release of hydrogen ions from each "letter" of the DNA alphabet. [QuantuMDx](#), a closely held U.K. company, has designed a mini-sequencer the size of an iPod that could be used to test patients for infections and drug responses.

Venture Backing

Sequencing efficiency is racing ahead. Clifford Reid, CEO of [Complete Genomics Inc. \(GNOM\)](#) in [Mountain View](#), California, said his capacity will increase 100 times in the coming years.

The next challenge is developing software and gadgets to help put sequencing information to use in hospitals and physician offices. Google Inc.'s Google Ventures fund, which invests as much as \$200 million annually in startups, has put a "significant" chunk of its money into companies aiming to speed the use of the genome, said Krishna Yeshwant one of its partners.

Closely held companies such as GenomeQuest, based in Westborough, [Massachusetts](#), Emeryville, California-based Omicia Inc. and [Softgenetics LLC](#) in [State College, Pennsylvania](#), are making software to search through gene variations and find those that are "actionable," for doctors trying to diagnose and treat disease.

Months to Minutes

Knome Inc., co-founded by Harvard Medical School geneticist [George Church](#) to sequence genomes for the public, has reorganized as a DNA analytics company, and will begin licensing software to hospitals and clinics this year, said Jason Lee, chief marketing officer.

Martin Reese, CEO of Omicia, hired a [software developer](#) with experience both at [Apple Inc. \(AAPL\)](#) and [Wal-Mart Stores Inc. \(WMT\)](#) to develop a system that would sort huge numbers of genetic variations, select those that were most likely to be associated with disease, and make it easy for users to process the results.

“We’re taking a process, hand annotation of the genome, that can take up to six months to do, and automating it so that it takes just a half hour or an hour,” Reese said.

Like the computer industry 20 years ago, gene-sequencing companies are also aiming to make smaller, friendlier products.

During an interview in his office, Jay Flatley, the CEO of San Diego-based Illumina, waved an Apple iPad loaded with an app that displayed his genome. He skipped from page to page with his fingertips. Patients could take this information with them to a doctor’s appointment, he said, and discuss treatments.

Genetic Variants

“What people want to know is, what genetic variants do I have that are different for a disease or drug?” he says, quickly paging through screens of diagrams, text and color icons. He stopped on one describing a gene that would affect his response to the blood-thinner warfarin, should he ever take it. “I’m in the high-sensitivity category, so I should get a lower dose.”

Roche said Jan. 25 that it had offered \$5.7 billion, or \$44.50 a share, for Illumina. The sequencer manufacturer countered with a so-called “poison pill” measure that would give shareholders the option of buying preferred stock to match each of their common shares, which may make the company too expensive for Roche to buy.

Hospital Investment

The power of sequencing is spurring hospitals and clinics to invest. [Dana-Farber Cancer Institute](#) in Boston plans to buy at least two sequencers from Illumina this year, said Barrett Rollins, chief scientific officer of the Harvard-affiliated oncology center. Doctors at the University of [Iowa](#) use GenomeQuest’s software to sift through deaf children’s DNA to determine whether they’re likely to

benefit from hearing aids or cochlear implants.

The Beerys' introduction to sequencing began at a 2008 job interview. Joe, 49, was chief information officer at US Airways Group Inc. and Greg Lucier, chief executive of Invitrogen -- which later became Life Technologies -- was trying to recruit him.

At a dinner overlooking the Pacific shore in Del Mar, [California](#), the Beerys described to Lucier and his wife how their children had been misdiagnosed with cerebral palsy at the age of 2. Alexis had muscle weakness and general fatigue that worsened each day from the morning on.

"We went through so many invasive tests through the years," Retta said. "Between the two of them, I'm sure it's been more than a million dollars from insurance and what we've paid."

Good Night's Sleep

Retta described scouring the literature for conditions that fit the strange symptoms. When the twins were 5 1/2 years old, she read about a rare disorder called a "dystonia," caused by a deficiency of a nervous system chemical. It sounded similar to what the children were suffering from and was caused by a shortage of dopamine, a nervous system messenger molecule. Doctors treated the condition with a dose of a drug called Sinemet.

"Alexis took a quarter of a pill, and she slept through the night for the first time in her life," Retta said.

Lucier told the Beerys that Invitrogen was about to purchase a sequencing company called Applied Biosystems, a move that would later result in the company being renamed Life Technologies. Applied's sequencers had the potential to find faulty genes in children in just weeks or months, rather than the years-long odyssey the Beery twins, then in their early teens, had endured, Lucier said.

Symptoms Reappeared

"When you help me merge these companies, there's a possibility we'll have the technology to sequence kids with problems like these at birth," Joe Beery recalled Lucier saying.

"On the way home, I already knew that this was a company that I wanted to be a part of," said Joe, who signed with Life Technologies within about eight weeks.

A few months after that, the Beery twins' treatment started to lose effectiveness, and some of the children's symptoms reappeared. Alexis's were particularly frightening and severe. During a two-month period, she went to the emergency room seven times because she was turning blue from lack

of oxygen, Retta recalled.

In 2009, Joe and Retta were listening to a presentation by Eric Topol, a Scripps Research Institute scientist who has organized several large genetic studies. As Topol talked about the power of sequencing to solve medical mysteries, a thought began to take shape in Retta's mind: why not sequence the children's genomes?

DNA Decoded

Through Life Technologies' sequencing division, Joe and Retta got in touch with a team of doctors and scientists at Baylor College of Medicine's [Human Genome Sequencing Center](#) in Houston. One of the Baylor researchers, James Lupski, had sequenced his own genome to identify the mutation behind his case of an inherited nerve disorder, called [Charcot-Marie-Tooth](#) disease.

The Beery children's DNA was decoded in two months by Baylor's Richard Gibbs on Life Technologies' machines, the sequencers the Beerys learned about in the dinner at Del Mar. Life helped pay for the procedure, along with Baylor research funds. Baylor's doctors and scientists then performed the analysis. As the process went forward, Joe would take a few minutes during his regular work meetings with Lucier to update him on progress.

"I wanted to stay close to it," recalled Lucier, who later had his own genome sequenced. "It was one of the first full human genome sequencing efforts to study a disease, and I wanted to make sure it did not disappoint."

Alive and Healthy

The procedure revealed that the twins had been incompletely diagnosed. In addition to the dopamine deficiency, the dystonia was being caused by a second genetic mutation that interfered with a separate nervous system chemical, called serotonin. The twins' doctors found that the dystonia could be fully treated with a serotonin replacement that was readily available through pharmacies. The change was dramatic: soon Alexis was back to school, track and basketball. A small dose also helped Noah, who had been less severely affected by the deficiency.

News spread quickly through Life Technologies' offices that the sequencing had helped save the Beery twins. Last summer, Joe told the story to a crowd of hundreds of Life Technologies employees, with the twins present. He explained how he saw the work of the company's employees playing a key role in keeping his children alive and healthy.

'Prayers Answered'

“People were crying,” Lucier recalled. “People came up to the Beerys afterwards to hug them and thank them for their courage in doing this. It allowed all our employees to draw a line between their work and making life better for patients.”

Earlier this month, the company introduced the Ion Proton sequencer, a \$150,000 machine Lucier says will be able to sequence an entire human genome in one day for a cost of about \$1,000. Until now, most sequencers have been sold to the research community. While that’s an important market, the potential for sales to hospitals, clinics and testing laboratories is far larger, Lucier said.

The Beerys said they hope more parents and children will be helped as they have been. Joe Beery said that while he and his family always prayed for guidance, they never dreamed that the answer to their children’s health mystery would arrive because he landed at a company that took a gamble on genome sequencing.

“The fact that I ended up where I ended up, you have to believe those prayers were answered,” he said.

To contact the reporter on this story: John Lauerman in Boston at jlauerman@bloomberg.net

To contact the editor responsible for this story: [Jonathan Kaufman](mailto:jkaufman17@bloomberg.net) at jkaufman17@bloomberg.net.

©2012 BLOOMBERG L.P. ALL RIGHTS RESERVED.

Matthew J. Price, MD, FACC, FSCAI
Assistant Professor, Scripps Translational Science Institute
Director, Cardiac Catheterization Laboratory
Division of Cardiovascular Diseases
Scripps Clinic
La Jolla, California

Biography

Matthew J. Price, MD, FACC, FSCAI, is Director of the Cardiac Catheterization Laboratory and Attending Physician, Department of Interventional Cardiology, at the Scripps Clinic Green Hospital in La Jolla, California.

Dr Price earned his MD at the University of California, Los Angeles (UCLA) School of Medicine and completed his internship and residency in Internal Medicine at the UCLA Center for Health Sciences. He subsequently completed a fellowship in Cardiovascular Diseases at Cedars-Sinai Medical Center in Los Angeles, where he served as Chief Fellow and was awarded the Eliot Corday, MD, Outstanding Cardiology Fellow of the Year for his “outstanding clinical performance and commitment to patient care.” Dr Price completed his training in Interventional Cardiology at Scripps Clinic, where he currently acts as an Attending Interventional Cardiologist and Director of the Cardiac Catheterization Laboratory.

A fellow of numerous professional organizations, including the American College of Cardiology and the Society of Cardiovascular Angiography and Interventions, Dr Price is also involved with several studies and clinical investigations that are currently recruiting patients, and serves on the leadership of several randomized clinical trials of antiplatelet therapy in PCI. He was the Principal Investigator of the GRAVITAS Trial, the first randomized clinical trial examining the efficacy and safety of individualized antiplatelet therapy after PCI. He also specializes in transcatheter approaches to structural and congenital heart disease. He serves as an ad hoc reviewer for several professional journals, including *the New England Journal of Medicine*, *JAMA*, and *Circulation*, is a contributing editor for *Reviews in Cardiovascular Medicine*, and is a member of the Editorial Board of the *Journal of the American College of Cardiology*.

Notes

Samuel Levy, PhD
Director of Genomic Sciences
Scripps Translational Science Institute
Scripps Health and The Scripps Research Institute
La Jolla, California

Biography

Samuel Levy, PhD, Director of Genomic Sciences, has an ongoing interest in characterizing and refining the structure and function of the genome and epigenome in human populations. His work at the Scripps Translational Sciences Institute (STSI) involves characterizing genomic changes in tumors to enable the development of specific treatment regimens and determining how genomic modifications impact our response to different drug therapies. Further, he is investigating how individual genomic modifications mitigate patient risk to adverse drug effects. Lastly, his group is using induced pluripotent stem cells to employ “disease in a dish” approaches to associating specific genomic profiles to cellular behavior in human disease tissue.

Prior to joining STSI in 2009 Dr Levy held the role of Director and Professor in Human Genomics at the J. Craig Venter Institute (JCVI). He was the leading scientist on the first published diploid genome sequence of a human, J. Craig Venter, additionally he led studies involving the functional characterization of DNA variants in protein coding and cis-regulatory regions. Dr. Levy also provided a clinical sequencing pipeline for the National Heart, Lung and Blood Institute and has worked with over 15 different research groups over a five-year period, providing high throughput discovery of DNA variants on clinical samples for disease association studies.

Prior to joining JCVI in 2002, Dr Levy worked at Celera Genomics for a three-year period and was involved in developing computational tools for gene discovery and functional annotation in the *Drosophila* and human genome sequences. Prior to his work at Celera Genomics, from 1989-1999, Dr. Levy was a postdoctoral worker and then independent investigator in the Molecular, Cellular and Developmental Biology (MCDB) Department at the University of Colorado at Boulder. At MCDB, he worked on a variety of projects ranging from extracellular Golgi-based trafficking in plants, plant cell wall development and structure-function relationship in complex carbohydrates using a range of approaches involving biochemistry, cell biology and computational chemistry.

In 1982, Dr. Levy received a B.Sc. in Molecular Biophysics at the University of Leeds, UK, in 1986 was awarded the Ph.D., in cell and computational biology at the University of Bristol, UK. He was in receipt of a NATO/SERC postdoctoral fellowship for a two-year period studying structural and biochemical changes in plant cell walls at the École Normale Supérieure, France.

Notes

Julie Johnson, PharmD

V. Ravi Chandran Professor of Pharmaceutical Sciences
Distinguished Professor of Pharmacy & Medicine
Director, UF&Shands Personalized Medicine Program
Director of the Center for Pharmacogenomics
University of Florida, Colleges of Pharmacy and Medicine
Gainesville, Florida

Biography

Julie A. Johnson, Pharm.D., BCPS, FCCP, FAHA is the V. Ravi Chandran Professor of Pharmaceutical Sciences, Distinguished Professor of Pharmacy (Departments of Pharmacotherapy and Translational Research (primary), and Pharmaceutics) and Medicine; (Department of Medicine, Division of Cardiovascular Medicine) at the University of Florida Colleges of Pharmacy and Medicine. She is also Director, University of Florida Center for Pharmacogenomics and Director of the UF&Shands Personalized Medicine Program. She joined the faculty at the University of Florida in May 1998, and before that spent 9 years on the University of Tennessee College of Pharmacy faculty. She received her B.S. in Pharmacy from the Ohio State University (1985) and her Pharm.D. from the University of Texas at Austin and the University of Texas Health Science Center at San Antonio (1987). She completed a post-doctoral fellowship in cardiovascular pharmacology/ pharmacokinetics at the Ohio State University (1989).

Dr. Johnson's research focuses on cardiovascular pharmacogenomics. She leads a research group in the NIH-supported Pharmacogenomics Research Network, with a project focused on pharmacogenomics of antihypertensive drugs. She is a leader in the field of cardiovascular pharmacogenomics, with over 150 peer reviewed publications. She is also leading the University of Florida's effort in clinical translation of pharmacogenomics.

Dr. Johnson served on the Nonprescription Drugs Advisory Committee of the FDA (2000-2004), is currently serving on the XNDA Study Section at NIH (2007-2012), and is a member of a Protocol Review Committee and Data Safety Monitoring Board, for two NHLBI (NIH) networks. She was also a member of the American Heart Association Committee on Scientific Sessions Programming (2009-2011) and was co-chair for the 2009 AHA International Congress on Genetics and Genomics of Cardiovascular Disease. She served as a Regent of the American College of Clinical Pharmacy (2000-2003), and has had numerous other leadership roles in a variety of national organizations. She is on the editorial boards of the journals *Clinical Pharmacology and Therapeutics*, *Pharmacogenetics and Genomics*, *Journal of the American Heart Association*, *Psychosomatic Medicine*, and *Pharmacogenomics*, is a scientific editor for *Pharmacotherapy*.

Dr. Johnson has received numerous awards including teaching awards from both the University of Tennessee (1996) and the University of Florida (2001); the Ohio State University Alumni Association William Oxley Thompson Award for early career achievement (1997), the Leon I Goldberg Young Investigator Award from the American Society for Clinical Pharmacology and Therapeutics (2004), the Distinguished Alumnus Award from the Ohio State University College of Pharmacy (2005), Paul Dawson Biotechnology Research Award, American Association of Colleges of Pharmacy (2007), the Therapeutic Frontiers Award (2009) and the Russell R Miller Award for Contributions to the Literature (2010), both from the American College of Clinical Pharmacy; among others. She also has numerous lectureship awards.

Notes

Michael R. Hayden MB, ChB, PhD
Canada Research Chair in Human
Genetics and Molecular Medicine
University Killam Professor
Department of Medical Genetics

Biography

Dr. Michael Hayden is the Killam Professor of Medical Genetics at the UBC and Canada Research Chair in Human Genetics and Molecular Medicine. He is the Director of the Center for Molecular Medicine and Therapeutics (CMMT) and founder of three biotechnology companies: NeuroVir Therapeutics, Inc, Xenon Pharmaceuticals, Inc. and Aspreva Pharmaceuticals Corp. He is also co-leader of the Canadian Pharmacogenomics Network for Drug Safety project, a BC-led Genome Canada-funded, national strategy to prevent adverse drug reactions.

Author of over 700 peer-reviewed publications and invited submissions, Michael focuses his research primarily on genetic diseases, including genetics of lipoprotein disorders, Huntington disease, predictive and personalized medicine. Michael and his group have been researching the identification of genetic variation influencing drug response (toxicity, efficacy) in patients, and been working on the development of personalized medicine to incorporate these findings to optimize drug therapy for patients to ensure maximum efficacy with minimal adverse effects. Michael is also the most cited author in the world for ABCA1 and Huntington Disease.

Notes

William E. Evans, PharmD
Director and Chief Executive Officer
St. Jude Children's Research Hospital
Professor of Pediatrics and Pharmacy
University of Tennessee Colleges of Medicine and Pharmacy
Memphis, Tennessee

Biography



Dr. William E. Evans is Director and Chief Executive Officer of St. Jude Children's Research Hospital (SJCRH), and holds the St. Jude Professorship and Endowed chair at the University of Tennessee Colleges of Medicine and Pharmacy.

For the past 30 years, his research at St. Jude has focused on the pharmacogenomics of anticancer agents in children, for which he has received three consecutive NIH MERIT Awards from the National Cancer Institute (1987-2015). The major disease focus of his research is acute lymphoblastic leukemia in children.

Dr. Evans has authored over 300 articles and numerous book chapters. He has received several national awards for his research, most recently the 2009 Pediatric Cancer Award from the ASCO (shared with Mary V. Relling of SJCRH) and the 2009 Team Science Prize from AACR (shared with colleagues at St. Jude). He is recognized by ISI as a "Highly Cited Scientist" in pharmacology, based on citations of his research publications. He was elected to the Institute of Medicine of the National Academy of Sciences in 2002.

Dr. Evans has served as CEO of SJCRH since 2004, during which time St. Jude has been ranked annually in the top 10 best places to work in academia by *The Scientist* magazine.

Notes

Pharmacogenomics to optimize treatment of childhood leukemia

William E. Evans, PharmD

There is a growing body of evidence that inherited germline genome variation can influence the effects of cancer chemotherapy, such as the highly penetrant effects of genetic polymorphisms in the thiopurine methyl-transferase gene (*TPMT*) and its relation to dose-limiting hematological toxicity of mercaptopurine (reviewed in Evans and Relling, *Nature* 2004; Relling et al *CPT* 2011). Other examples include methotrexate and *SLCO1B1* polymorphisms, irinotecan and *UGT1A1* polymorphisms, and tamoxifen and *CYP2D6* polymorphisms (reviewed in Paugh et al, *CPT* 2011). However, in each cancer patient, there are at least two genomes of potential importance in determining the efficacy of cancer chemotherapy, the host genome and the cancer cell genome(s). There are examples where cancer treatment has been targeted toward genomic abnormalities in cancer cells (e.g., tyrosine kinase inhibitors targeted against aberrant kinase activity due to chromosomal translocations or somatic mutations). However, most cancers are treated with chemotherapy that is not targeted toward specific genomic abnormalities in cancer cells, and little is currently known about somatic genome variation that alters the effects of anticancer agents. This lecture will focus on examples of somatic and inherited genome variations that can influence the effects of cancer chemotherapy for acute lymphoblastic leukemia, including copy number alterations, focal deletions, and epigenetic changes that alter the sensitivity of leukemia cells to specific anticancer agents (e.g., Qing et al, *Nature Gen.* 2005, Diouf et al, *Nature Med.* 2011). This presentation will also address ongoing strategies for identifying somatic changes that alter drug response, including GWAS and whole genome sequencing (e.g. The St. Jude-Wash. U. Pediatric Cancer Genome Project).

Notes

Reed V. Tuckson, MD, FACP

Executive Vice President and Chief of Medical Affairs UnitedHealth Group

Biography



A graduate of Howard University, Georgetown University School of Medicine, and the Hospital of the University of Pennsylvania's General Internal Medicine Residency and Fellowship Programs, Dr. Tuckson is currently Executive Vice President and Chief of Medical Affairs at UnitedHealth Group, a Fortune 25 diversified health and well-being company. As the most senior clinician of UnitedHealth Group, Dr. Tuckson is responsible for working with all the company's diverse and comprehensive business units to improve the quality and efficiency of the health services we provide to the 75 million members we are privileged to serve worldwide.

Formerly, Dr. Tuckson served as Senior Vice President, Professional Standards, for the American Medical Association (AMA). He is former President of the Charles R. Drew University of Medicine and Science in Los Angeles; has served as Senior Vice President for Programs of the March of Dimes Birth Defects Foundation; and is a former Commissioner of Public Health for the District of Columbia.

Dr. Tuckson is an active member of the prestigious Institute of Medicine of the National Academy of Sciences. He was appointed to the National Institutes of Health's Advisory Committee to the Director and the Department of Health and Human Services' Health Information Technology (HIT) Policy Committee -- Enrollment Workgroup. He is past Chair of the Secretary of Health and Human Services' Advisory Committee on Genetics, Health and Society.

Dr. Tuckson currently serves on the Board of Directors for several national organizations including the Alliance for Health Reform; the American Telemedicine Association; the National Patient Advocate Foundation; the Arnold P. Gold Foundation; Project Sunshine; Cell Therapeutics, Inc. and Howard University. Additionally, he serves on several Boards within his local community of Minneapolis, including Big Brothers Big Sisters of the Greater Twin Cities and Minnesota Public Radio.

Dr. Tuckson has also held other federal appointments, including cabinet level advisory committees on health reform, infant mortality, children's health, violence, and radiation testing.

Dr. Tuckson is the author of *The Doctor in the Mirror*, a patient empowerment book and media presentation, which was released in November, 2011. He was also selected as one of *Modern Healthcare/Modern Physician* "50 Most Powerful Physician Executives" in Healthcare in 2009 and 2010. Dr. Tuckson was named one of *Modern Healthcare*'s "Top 25 Minority Executives" in Healthcare in 2008 and 2010 and to *Ebony* magazine's "2008 Power 150: The Most Influential Blacks in America" list. In May 2010, Project Sunshine honored Dr. Tuckson for his work with children, volunteerism, and social responsibility. In 2011, he was ranked #8 on *Uptown Professional* magazine's list of Top 100 Executives.

Leila Shepherd, PhD
Chief Technology Officer
DNA Electronics Ltd.

Biography

Leila Shepherd, PhD is Chief Technology Officer at DNA Electronics Ltd, a company developing rapid, lab-free tests for DNA analysis based on semiconductor microchip technology. She has a background in electronic engineering, specifically in CMOS-based chemical sensors and circuits. She holds a MEng degree and a PhD, both from the Department of Electrical and Electronic Engineering, Imperial College London, and spent a year at ENSTA, part of the Paris Institute of Technology in France studying engineering with management.

Notes

James H. Fowler
Professor of Medical Genetics
And Political Science
University of California, San Diego
La Jolla, California

Biography



[James H. Fowler](#) earned a PhD from Harvard in 2003 and is currently Professor of Medical Genetics and Political Science at the University of California, San Diego. His work lies at the intersection of the natural and social sciences, with a focus on social networks, behavioral economics, evolutionary game theory, political participation, cooperation, and genopolitics. James was recently named a Fellow of the John Simon Guggenheim Foundation, one of *Foreign Policy's* Top 100 Global Thinkers, and Most Original Thinker of the year by *The McLaughlin Group*. His research has been featured in numerous best-of lists including *New York Times Magazine's* Year in Ideas, *Time's* Year in Medicine, *Discover Magazine's* Year in Science, and *Harvard Business Review's* Breakthrough Business Ideas. Together with Nicholas Christakis, James wrote a book on

social networks for a general audience called *Connected*. Winner of a Books for a Better Life Award, it has been translated into twenty languages, named an Editor's Choice by the *New York Times Book Review*, and featured in *Wired*, Oprah's Reading Guide, *Business Week's* Best Books of the Year, and a cover story in *New York Times Magazine*.

Notes

Evan Eichler, PhD
Investigator of the Howard Hughes Medical Institute
Professor of Genome Sciences
University of Washington
Seattle, Washington

Biography

Evan Eichler, PhD, is a Professor and Howard Hughes Medical Institute Investigator in the Department of Genome Sciences, University of Washington School of Medicine. He graduated with a B.Sc. Honours degree in Biology from the University of Saskatchewan, Canada, in 1990. He received his Ph.D. in 1995 from the Department of Molecular and Human Genetics at Baylor College of Medicine, Houston. After a Hollaender postdoctoral fellowship at Lawrence Livermore National Laboratory, he joined the faculty of Case Western Reserve University in 1997 and later the University of Washington in 2004. He was a March of Dimes Basil O'Connor Scholar (1998-2001), appointed as an HHMI Investigator (2005), and awarded an AAAS Fellowship (2006) and the American Society of Human Genetics Curt Stern Award (2008). He is an editor of *Genome Research* and has served on various scientific advisory boards for both NIH and NSF. His research group provided the first genome-wide view of segmental duplications within human and other primate genomes and he is a leader in an effort to identify and sequence normal and disease-causing structural variation in the human genome. The long-term goal of his research is to understand the evolution and mechanisms of recent gene duplication and its relationship to copy number variation and human disease.

Notes

Hakon Hakonarson, MD, PhD

Director, Center for Applied Genomics

Associate Professor of Pediatrics

Department of Pediatrics and Division of Genetics

The Children's Hospital of Philadelphia

University of Pennsylvania School of Medicine

Philadelphia, Pennsylvania

Biography

Hakon Hakonarson, MD, PhD, is an associate professor of Pediatrics at The University of Pennsylvania School of Medicine. He is a physician-scientist and director of The Children's Hospital of Philadelphia's Center for Applied



Genomics (CAG), a high-throughput highly automated genotyping facility founded to identify the genetic causes of complex medical disorders in children, such as autism and cancer, with the objective of developing new therapies. The Center represents a \$40 million commitment from CHOP to genotype approximately 100,000 children a research undertaking that has gained nationwide attention, including news features in the *Wall Street Journal*, *New York Times*, *Time Magazine*, *Nature* and *Science*. Dr. Hakonarson has an extensive track record in human genetics and has developed an international reputation amongst his peers. He has served previously in several senior posts in the biopharmaceutical industry, including as the director of Inflammatory and Pharmacogenomics Research and the vice president of Clinical Sciences and Development and CSO. Dr. Hakonarson has also been the principal and co-principal investigator on several NIH-sponsored grants, and he has published numerous high-

impact papers on genomic discoveries and their translations in some of the most prestigious scientific medical journals, including *Nature*, *Nature Genetics* and *The New England Journal of Medicine*. *Time Magazine* listed Dr. Hakonarson's autism gene discovery reported in *Nature*, 2009, among the top 10 medical breakthroughs of that year. With over ten years of experience in pioneering genomics research and genome-wide mapping and association studies, Dr. Hakonarson has intimate knowledge of the complexities of large-scale genomics projects and has put together the necessary infrastructure and workflow processes to unravel these complexities.

Notes

Isaac S. Kohane, MD, PhD

Henderson Professor of Health Sciences and Technology
Children's Hospital and Harvard Medical School
Director, Countway Library of Medicine
Director, i2b2 national Center for Biomedical Computing
Co-Director, HMS Center for Biomedical Informatics
Boston, Massachusetts

Biography

Isaac (Zak) Kohane is the director of the Children's Hospital Informatics Program and is the Henderson Professor of Pediatrics and Health Sciences and Technology at Harvard Medical School (HMS). He is also the co-Director of the HMS Center for Biomedical Informatics and Director of the HMS Countway Library of Medicine. Dr. Kohane leads multiple collaborations at Harvard Medical School and its hospital affiliates in the use of genomics and computer science to study diseases (particularly cancer and autism) through the perspective of biological development. He also has developed several computer systems to allow multiple hospital systems to be used as "living laboratories" to study the genetic basis of disease while preserving patient privacy. Among these, the i2b2 (Informatics for Integrating Biology and the Bedside) National Computing Center has been deployed at over 52 academic health centers internationally.

Dr. Kohane has published over 200 papers in the medical literature and authored a widely used book on Microarrays for an Integrative Genomics. He has been elected to multiple honor societies including the American Society for Clinical Investigation, the American College of Medical Informatics, and the Institute of Medicine. He leads a doctoral program in genomics and bioinformatics at the Division of Health Sciences and Technology at Harvard and MIT. He is also a practicing pediatric endocrinologist and father of three energetic children.

Notes

John A. Todd, FRS, PhD

Director, Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes & Inflammation Laboratory
Department of Medical Genetics,
Cambridge Institute for Medical Research
University of Cambridge, Addenbrooke's Hospital
Cambridge, United Kingdom

Biography

John Todd FRS, FMedSci, PhD is Professor of Medical Genetics at Cambridge University and Director of the Juvenile Diabetes Research Foundation (JDRF)/Wellcome Trust Diabetes and Inflammation Laboratory in the University's Cambridge Institute for Medical Research. Todd researches type 1 diabetes genetics and disease mechanisms, in collaboration with Linda Wicker and David Clayton. Previously, Todd was a JDRF Career Development Fellow, Professor of Human Genetics at Oxford University and a Wellcome Trust Principal Research Fellow. He received his B.Sc. from Edinburgh University in Biological Sciences, and his Ph.D. from Cambridge University in Biochemistry. Post-doctoral training was undertaken at Cambridge and Stanford in genetics, molecular biology and immunology. He has over 400 publications, almost 300 of which were peer-reviewed with over 29,000 citations and H-index = 85. He has received several awards and honours for his research. Todd has trained so far 30 PhD students and many more research assistants and postdoctoral fellows. His main goal is to provide knowledge based on genetic aetiological findings that can inform in the prevention of autoimmunity in type 1 diabetes and in other immune-mediated diseases.

Notes

Acknowledgements

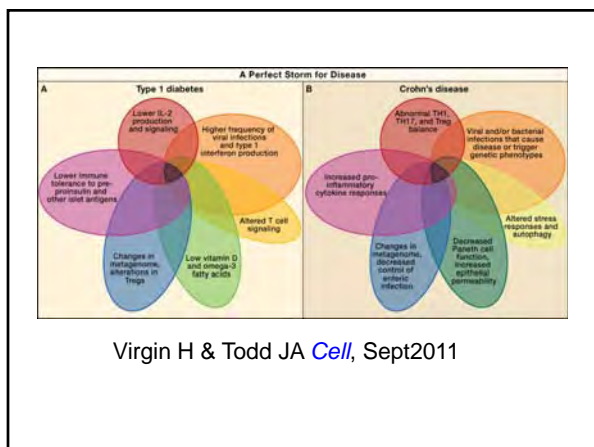
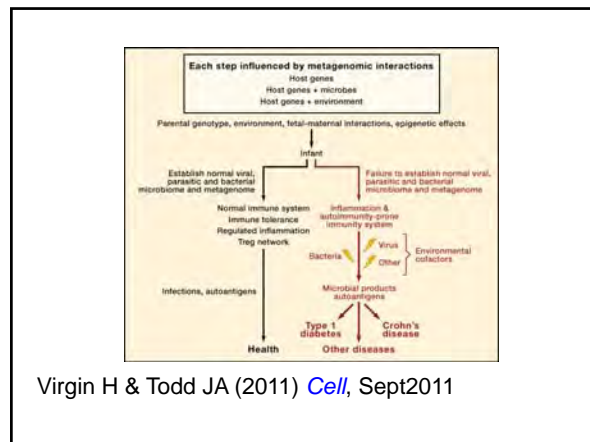
JDRF/WT Diabetes & Inflammation Laboratory (DIL)
JDRF Diabetes—Genes, Autoimmunity and Prevention (D-GAP)
David Clayton & Linda Wicker, co-PIs

Sarah Nutland & team: NIHR BRC Cambridge BioResource
 Helen Stevens & team: samples; Matt Woodburn; Sample IT
 Neil Walker & team: data
 Olly Burren & team: informatics – T1DBase & ImmunoBase
 Jason Cooper, Jo Howson, Vincent Plagnol, Chris Wallace: stats
 Debbie Smyth: genotyping
 Jennie Yang, Kate Downes, Marcin Pekalski, Xaq Castro, Calli Dendrow, Laura Esposito, Lucy Davison, Sarah Howlett, Jan Clark, Dan Rainbow
 D-GAP: Mark Peakman, Tim Tree, David Dunger, Polly Bingley
 NAIMIT: Chantal Mathieu

Juvenile Diabetes Research Foundation & Wellcome Trust; NIDDK; EU FP7; NIHR Cambridge Biomedical Research Centre







UNIVERSITY OF CAMBRIDGE

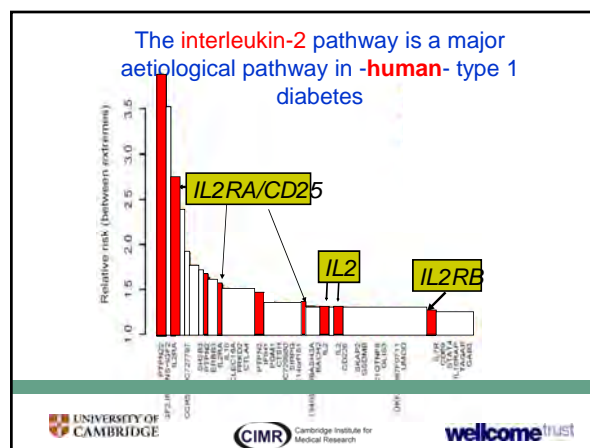
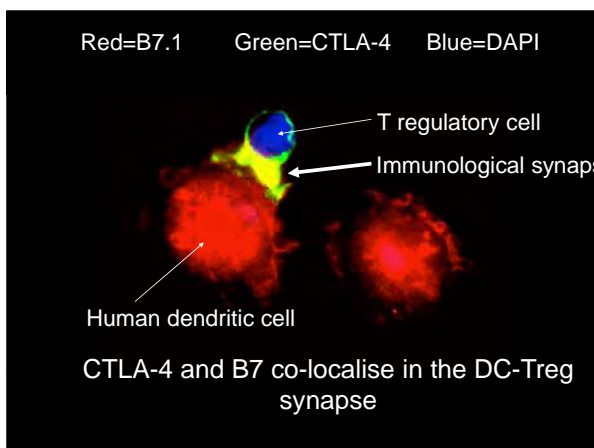
Over 50 regions of the human genome control diagnosis of type 1 diabetes

www.t1dbase.org

Genes to biology & mechanism

Four major predisposing pathways (so far):

- (1) Cytokine production and signalling *IL2, IL2RA (CD25), PTPN2, IL10, IL27*
- (2) Decreased T cell signalling and activation
- (3) Increased type 1 interferon production and anti-viral responses
- (4) Antigen presentation and T cell repertoire formation



Genes to biology

The Cambridge BioResource: a resource of >9,000 local volunteers including T1D, SLE, RA, and MS patients willing to be invited to a wide range of medical research studies, SELECTED by genotype & RECALLABLE

The Cambridge BioResource

UNIVERSITY OF CAMBRIDGE
MRC
NHS
National Institute for Health Research
Cambridge Biomedical Research Centre
Blood and Transplant

The interleukin-2 pathway is a major aetiological pathway in **-human-** type 1 diabetes

Dendrou, Plagnol, Nutland, Todd, Wicker, et al. *Nature Genetics* 41:1011-1015 (2009)
Gregersen. News & Views. *Nature Genetics* 41:1011-1015 (2009)

CD4 T cell phenotypes associated with *IL2RA* (CD25)

Red = protected donor
Blue = susceptible donor

Lower CD25 surface expression is associated with T1D-predisposing CD25 genotype

T1D most associated SNP homozygous genotype only 1% of the population of 2,000 volunteers in Cambridge BioResource

$P = 2.468 \times 10^{-8}$

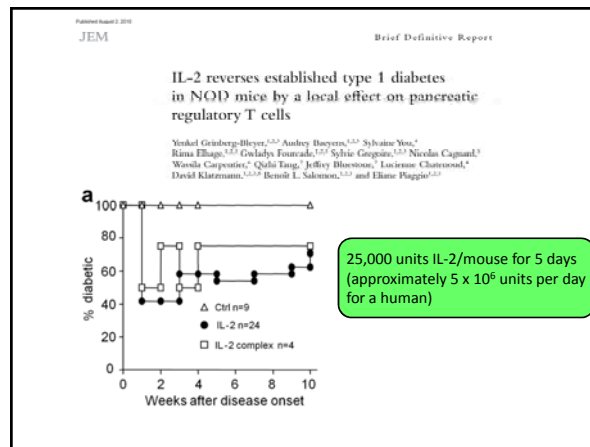
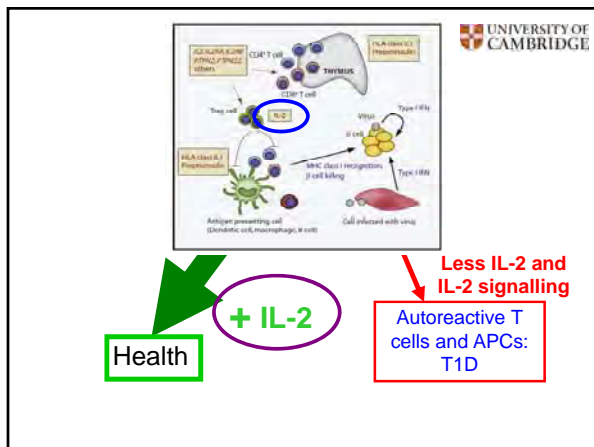
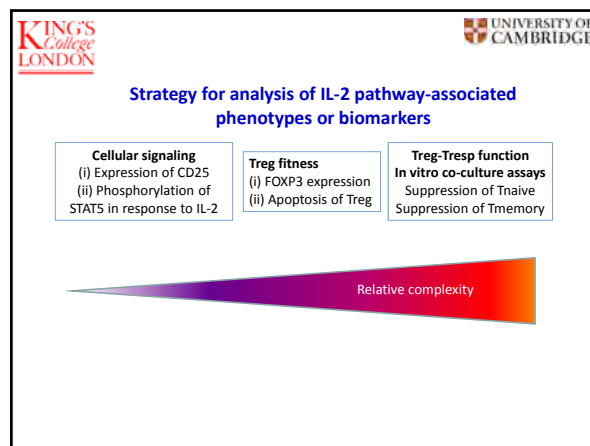
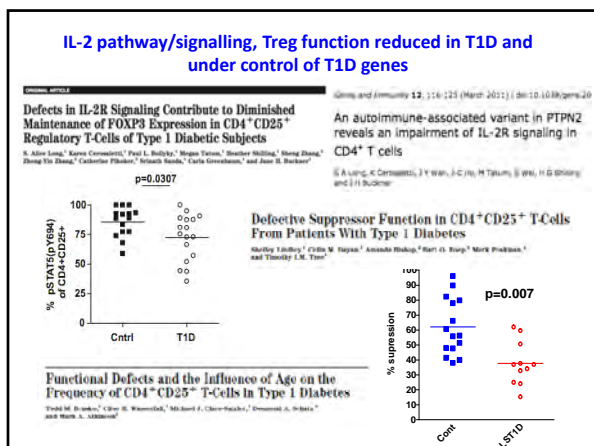
CD25 surface expression on CD4 memory T cells

Dendrou, Plagnol, Nutland, Todd, Wicker, et al. (2009) *Nature Genetics* 41, 1011-1015.

Lower CD25 surface expression is associated with T1D-predisposing CD25 genotype and lower IL-2 production

CD25 surface expression on CD4 memory T cells

Dendrou, Plagnol, Nutland, Todd, Wicker, et al. (2009) *Nature Genetics* 41, 1011-1015.



Alice Long, Jane Buckner, TRIALNET
Immunophenotyping of the (discontinued) IL-2+rapamycin trial: Data presented at FOCIS, Washington DC, June 2011

- N = 9 T1D patients, >18<40 yrs old, within 4 yr of diagnosis
- One month of IL-2, and then rapamycin for 3 months
- Used IL-2 (Proleukin) at a dose of **4.5 x 10⁶ IU s.c., 3x/wk x 4 weeks**
 This dose is in the range (1.5 - 7.5 x 10⁶ IU) of low dose IL-2 therapy for HIV, and also is the IL-2 dose and duration of therapy that was found effective (together with rapamycin) in preventing diabetes and reducing hyperglycemia after diabetes onset in NOD mice.
- At day 28, Helios positive and Helios negative Tregs increased in number significantly and then decreased
- Helios-pos Tregs secreted IFN-g
- Diminished pSTAT5 signalling in response to IL-2 was reversed with IL-2 injections.
- IL-2 increased eosinophils, **NK cells, Teff**
- Administration of rapamycin caused very rapid loss of C-peptide....think that rapamycin killed beta cells directly

ClinicalTrials.gov

Study ID: NCT00170534
 Study Title: **Dose-Effect Relationship of Low-dose IL-2 in Type 1 Diabetes (DF-IL2)**

This study is currently recruiting participants.
Updated on May 22, 2011 by assistance@clinicaltrials.gov

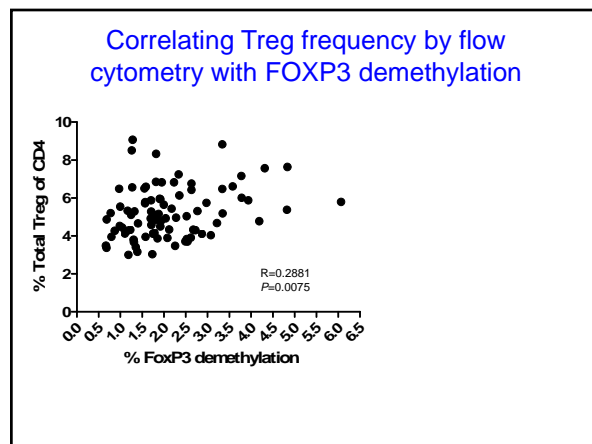
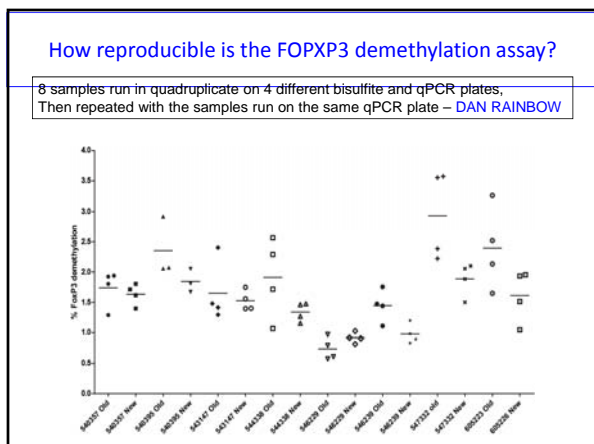
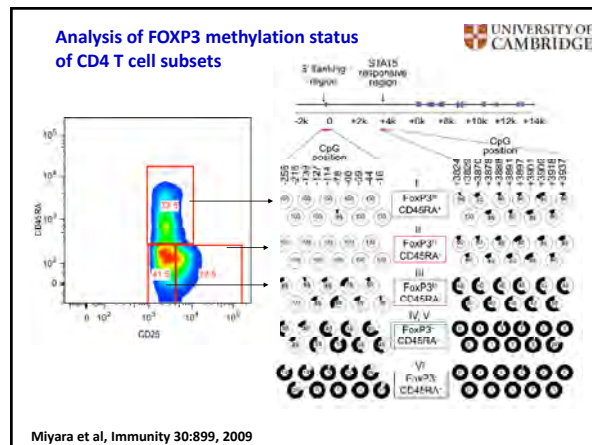
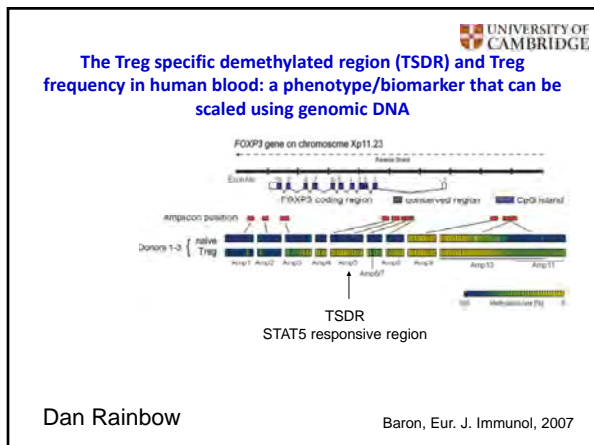
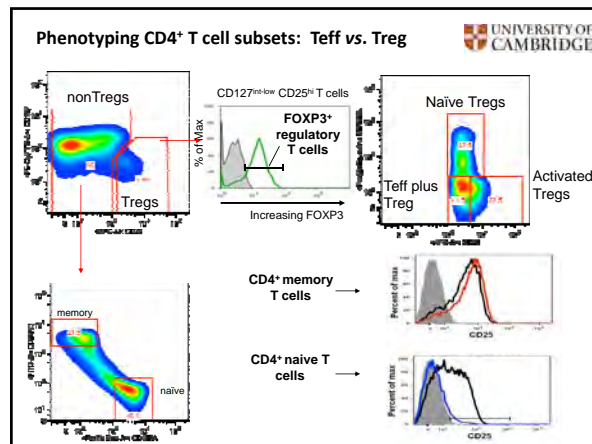
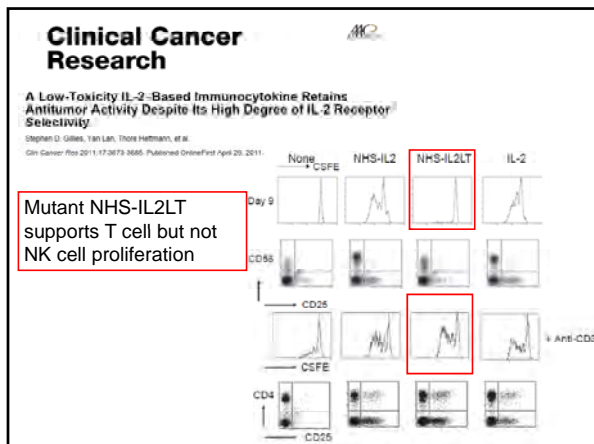
Primary Outcome Measures:

- Kinetic parameters of Treg proportions variation within CD4⁺ T cells in peripheral blood
 - [Time Frame: from Day+0 to Day+60]
 - [Designated as safety issue: No]

Secondary Outcome Measures:

- Improvement of residual secretion of insulin assessed by the AUC of peptide C during a standardized test meal in IL-2 vs placebo treated patients
 - [Time Frame: at Day+0 and Day+60]
 - [Designated as safety issue: No]
- Doses 3, 1 and 0.33 million units s.c. for 5 days

Future: David Klatzmann, Paris, applying for European funding (4 Oct, LOI) for a Phase 2 trial of low dose proleukin in 200 newly-diagnosed T1D: Linda Wicker and John Todd to participate and recruit to the trial and analyse samples



ORIGINAL ARTICLE

Vitamin D Supplementation and Regulatory T Cells in Apparently Healthy Subjects: Vitamin D Treatment for Autoimmune Diseases?

Vitamin D controls T cell antigen receptor signaling and activation of human T cells

Vitamin D is Required for IFN- γ -Mediated Antimicrobial Activity of Human Macrophages

VOLUME 11 NUMBER 4 APRIL 2010 NATURE IMMUNOLOGY

Available online at www.sciencedirect.com

ScienceDirect

Current Contents in Pharmacology

Vitamin D: modulator of the immune system
 Fermeke Baeke, Tatiana Takishi, Hannelie Korf, Conny Gysemans and Chantal Mathieu
 2010

THE JOURNAL OF IMMUNOLOGY

1,25-Dihydroxyvitamin D₃ and IL-2 Combine to Inhibit T Cell Production of Inflammatory Cytokines and Promote Development of Regulatory T Cells Expressing CTLA-4 and FoxP3

This information is current as of November 7, 2011

Louisa E. Jeffery, Fiona Burke, Mameela Muna, Yong Zheng, Omar S. Qureshi, Martin Hewison, Lucy S. K. Walker, David A. Lammens, Karim Raza and David M. Sissons

J Immunol 2009;183:5458-5467
 doi:10.4049/jimmunol.0803217
<http://www.jimmunol.com/content/183/10/5458>


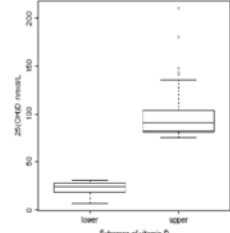
ORIGINAL ARTICLE

Inherited Variation in Vitamin D Genes Is Associated With Predisposition to Autoimmune Disease Type 1 Diabetes

Jason D. Cooper,¹ Deborah J. Smyth,¹ Neil M. Walker,¹ Helen Stevens,¹ Oliver S. Barron,¹ Chris Wallace,¹ Christopher Greissl,² Elizabeth Ramos-Lopez,² Elinu Hyppönen,³ David B. Dunger,⁴ Timothy D. Spector,⁵ Willem H. Ouwens,^{6,7} Thomas J. Wang,^{8,9,10} Klaus Borch-Johnsen,⁷ and John A. Todd¹

Multiple sclerosis GWAS:
 CYP27B1 and CYP24A1
 Nature, Aug2011

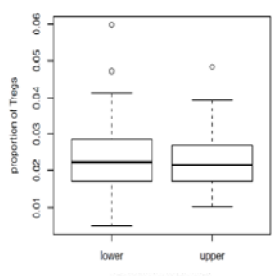
Extremes of vitamin D status and immunophenotype

- 80 pairs containing a sample from the top and bottom quintiles of circulating levels of vitamin D [25(OH)D]
- Matched on sex, age at bleed, month of bleed, BMI

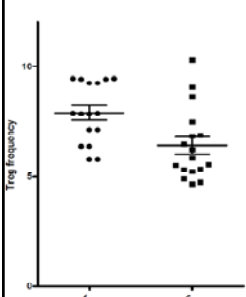
Median = 25.07 nmol/L 90.95 nmol/L

No correlation between levels of Vitamin D and demethylation of FOXP3/Treg number

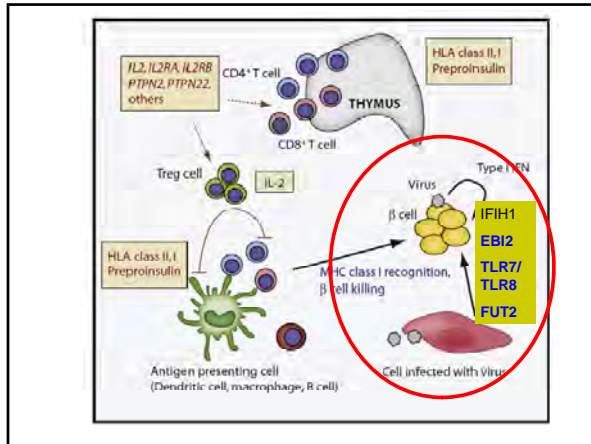


BUT we are doing gene-SNP association analysis of FOXP3 demeth trait and we have our first (replicated) hit

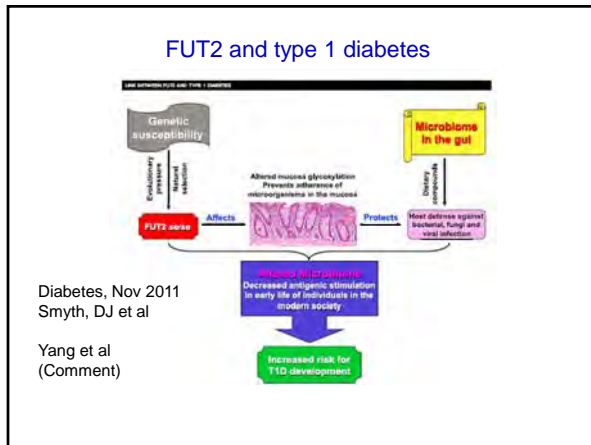
Treg frequency differs by PTPN22 genotype discovered using the PCR-based Treg specific demethylation assay using whole blood DNA



Confirmed by flow cytometry using matched pair volunteers from the Cambridge BioResource



ORIGINAL ARTICLE
FUT2 Nonsecretor Status Links Type 1 Diabetes Susceptibility and Resistance to Infection
 Deborah J. Smyth, Jason D. Cooper, Joanna M.M. Hovson, Pamela Clarke, Kate Dunwars, Trupti Mistry, Helen Stevens, Neil M. Walker, and John A. Todd
 DIABETES, VOL. 60, NOVEMBER 2011
 May 2011 | Volume 6 | Issue 5 | e2013
Secretor Genotype (*FUT2* gene) Is Strongly Associated with the Composition of *Bifidobacteria* in the Human Intestine
 Peijo Huokari, Aari Mäkelä, Teemu Alakotila, Jarmo Nikkila, Heikki Teiskonen, Jarmo Rintala, Jukka Partanen, Kari Aitakko, Juana Mäkitö



The Cambridge BioResource

UNIVERSITY OF CAMBRIDGE
 MRC Medical Research Council
 National Institute for Health Research
 Cambridge Biomedical Research Centre
 Blood and Transplant

What is the microbiome content in FUT2 wild type versus mutant donors?

Acknowledgements

Each patient, control, donor, volunteer and family

Howard J. Jacob, PhD

Director, Human and Molecular Genetics Center
Warren P. Knowles Chair of Genetics
Professor, Department of Physiology & Pediatrics
Medical College of Wisconsin
Vice Chair of Research, Department of Pediatrics
Children's Hospital of Wisconsin
Milwaukee, Wisconsin

Biography



Professor Jacob received his Ph.D. in Pharmacology from University of Iowa in 1989. He completed two parallel post-doctoral fellowships in functional genomics and molecular genetics/genomics at Harvard, Stanford and MIT with Victor J. Dzau, M.D. and Eric S. Lander, Ph.D. In 1992 he was appointed an Assistant Professor at Harvard Medical School at Massachusetts General Hospital. He joined the Medical College of Wisconsin in 1996 as an Associate Professor, Department of Physiology with full Professorship and Tenure in 2001. He was appointed the Founding Director of the Human and Molecular Genetics Center (HMGC) and awarded the Warren P. Knowles Chair of Genetics in 1999. Dr. Jacob along with Drs. Reza Shaker (PI) and Ellis Avner developed the Clinical Translational Science Institute at MCW. In 2009, he became the Vice Chair of Research in Pediatrics. Dr. Jacob is also an entrepreneur having co-founded the company,

PhysioGenix Inc. in 1997 and Primary Genetics in 2010.

Under his leadership the Human and Molecular Genetics Center (HMGC) has grown from two faculty members to 30 and is one of the top funded genetic programs based on NIH funding. In 2009, the grant support for the HMGC grew by 46%, including a Center of Excellence in Genomic Sciences, and winning a Grand Opportunities (ARRA) grant. The HMGC has faculty from 9 different departments and has initiated numerous novel research programs. The HMGC also provides clinical services with faculty providing clinical genetics, pre-implementation genetic diagnosis, molecular diagnostics and recently genomic sequencing.

His research interests center on physiological genetics, genetic dissection of and analysis of common, complex disease, especially target organ damage associated with diabetes mellitus and hypertension. His laboratory is known for applying genomics, and high through-put phenotyping, genomic sequencing as well as comparative genomics and bioinformatics. Dr. Jacob previously chaired the Coordinating Committee for the Program for Genomic Applications (PGA), which consisted of 11 institutions and \$33M annual budget funded by the National Heart, Lung and Blood Institute. Dr. Jacob currently participates in 6 NIH grants and 2 European Union grants. He has published 204 scientific papers, and 11 book chapters.

In 2010 Dr. Jacob led a team of researchers at the Medical College who used an innovative DNA sequencing technique to unravel the medical mystery of Nicholas Volker, a young boy whose life-threatening disease had baffled his doctors and tested his family's faith. Working with Medical College scientists and physicians at the Children's Hospital of Wisconsin, Dr. Jacob's team used Nicholas' DNA to diagnose his disease and recommend a course of treatment. This treatment has so far been successful.

Because of this success the Medical College of Wisconsin and Children's Hospital and Health System have developed and implemented a clinic for whole genome sequencing.

Joseph G. Gleeson, MD

Investigator, Howard Hughes Medical Institute
Professor Neurosciences and Pediatrics
Rady Children's Hospital
University of California, San Diego
La Jolla, California

Biography

Dr. Gleeson received a BA degree from UCSD in Chemistry and an MD from the University of Chicago Pritzker School of Medicine. He completed residency in Pediatrics and Neurology followed by a research fellowship at Harvard Medical School and Children's Hospital, Boston. His lab at UCSD studies the genetic basis of pediatric brain disorders. He has received many awards including a Searle Scholarship, Klingenstein Award, Burroughs Wellcome Fund Clinical Scientist Award in Translational Research and Howard Hughes Medical Institute award for Translational Research.

Dr. Gleeson has focused on identifying causes of a host of devastating neurological conditions affecting children. His research has identified over a dozen genes associated with altered brain development in humans. When mutated, these lead to a host of clinical manifestations, such as autism, epilepsy and mental retardation. Gleeson's strategy, which continues to evolve, is to use a combination of careful assessments, brain MRI, and electrophysiology, in consultation with a team of international experts, to identify new types of neurological conditions. He then develops strategies to identify the molecular basis of these newly defined conditions.

He focused recruitment in areas of the world with high rates of consanguinity, including the Middle East and North Africa, then utilizes next-generation sequencing to identify causative genetic variants. He has found that early incorporation of genetic sequencing in the clinic can improve diagnostic accuracy and in many cases suggest new approaches to treatment. Once new genes are identified, they are subject to functional analysis and validation using animal models and induced pluripotent stem cell models. The approach to identify nervous system specific phenotypes from iPSCs is a new focus for the lab, and offers the possibility to more rapidly identify treatment approaches for previously untreatable neurodevelopmental diseases.

Notes

Stanley F. Nelson, MD

Professor

Departments of Human Genetics, Pathology and Laboratory Medicine, and Psychiatry

David Geffen School of Medicine

University of California, Los Angeles

Los Angeles, California

Biography

Stanley F. Nelson, M.D. is Professor of Human Genetics, Pathology and Laboratory Medicine, and Psychiatry. His research work includes the identification of gene variants involved in neuropsychiatric disorders and cancer through the use of novel genomic technologies and information tools. He is leading the effort at UCLA to implement Clinical Exome Sequencing for the identification of disease causing mutations/rare alleles for Mendelian disorders for clinical diagnostic and for the identification of novel disease genes.

Notes

Efficient detection of causative mutations for rare diseases: Rethinking clinical practice

Stanley F. Nelson, MD

Clinical genetic diagnostics are changing rapidly through the implementation of efficient and accurate next generation sequencing tools. It is now possible to perform targeted capture of about 95% of all exons within known disease causing genes through 'Exome Sequencing'. These tools are being used for the identification of novel disease causing variants, especially for rare mendelian disorders, which demonstrates that exome sequencing is an efficient and cost effective means of performing detailed analyses through the vast majority of protein coding genes. This process including DNA extraction, library preparation, capture, sequencing, alignment to a reference genome, detection of approximately 22,000 DNA variants/ exome and determination of potential causality can be completed in on a number of independent genomes in just a few weeks for a cost that is comparable to single gene testing or small disease centric disease panels. Thus, the logic of clinical approach to molecular diagnosis is poised to undergo a dramatic reconfiguration. At UCLA, we have implemented Clinical Exome Sequencing for the purposed of mendelian disease diagnosis through the Clinical Genomics Center. This process is fully CLIA/CAP compliant and the processes for mutation/variant detection reliable and sensitive. This service is now offered to outpatients and inpatients through physician ordering in order to simplify molecular diagnosis. Our approach recognizes the large number of unknowns and complexity of the interpretation of genome-scale data. We have implemented a novel strategy employing a Genomic Data Board consisting of Pathologist, Medical Geneticist, Genetic Counselor, Informaticist, Genomics Expert, and relevant subspecialty clinicians in order to thorough evaluate the complex landscape of an individual genome for the purpose of diagnosis difficult cases. While not always successful, the enterprise is informing us as to the best means to identify likely causal mutations, highlighting inefficiencies in patient phenotypic descriptions critical for the diagnostic process, and pointing the way forward for more efficient diagnosis. In a number of instances, Clinical Exome Sequencing, has established a new diagnosis that led physicians to perform more focused phenotyping which was supportive of the genetic diagnosis and which optimized patient management. Genetic diagnosis for human disorders is frequently difficult with a large fraction of Medical Genetics patients referred for diagnosis to Genetics Clinics not receiving clear molecular diagnosis. The observed clinical heterogeneity of many rare mendelian disorders makes the selection of individual genes for sequencing difficult. We intend to implement Clinical Exome Sequencing as a first line genetic diagnostic tool, but recognize that better patient phenotyping or better tools that permit integration with patient data will be necessary for optimal implementation. In my talk, I discuss the practical implementation of Clinical Exome Sequencing and discuss a few informative patient vignettes.

Notes

Efficient detection of causative mutations for rare diseases: Rethinking clinical practice

Stanley F. Nelson, MD
 Professor
 Departments of Human Genetics, Pathology and
 Laboratory Medicine and Psychiatry
 snelson@ucla.edu

Human Genome

3 billion letters (GATC), ~20,000 genes, ~240,000 exons (85% of disease causing mutation here!)

There are ~3,000 Different Genetic Diseases for which the exact genetic mutation is known

Genetic heterogeneity in Human Disease

Genetic Heterogeneity in Human Disease

48 different known genes for hereditary deafness

Confusing Landscape for Clinicians

Pace of sequencing is accelerating

Human Genome Project:
 10 years, millions in technology
 \$1 billion = One genome

2007, second genome, \$30m, one Year
 2007, third genome, \$3million, 6 months
 2008, 4th-5th genomes, \$0.5million, 6m
 2009, 6th-7th genomes, \$200,000, 4months
 2010, 8th genome, \$30,000, 6 weeks

Now: ~\$4,000 to generate ~100Gb sequence

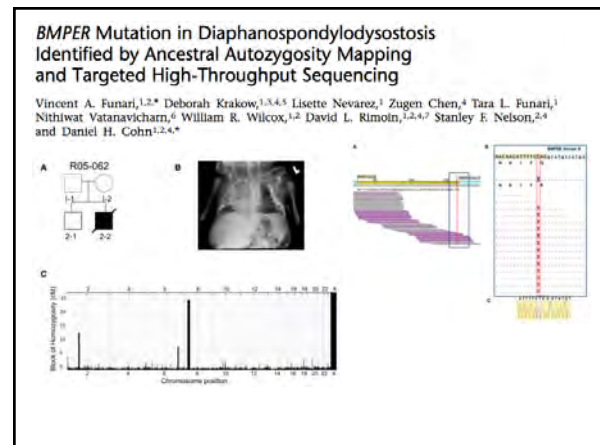
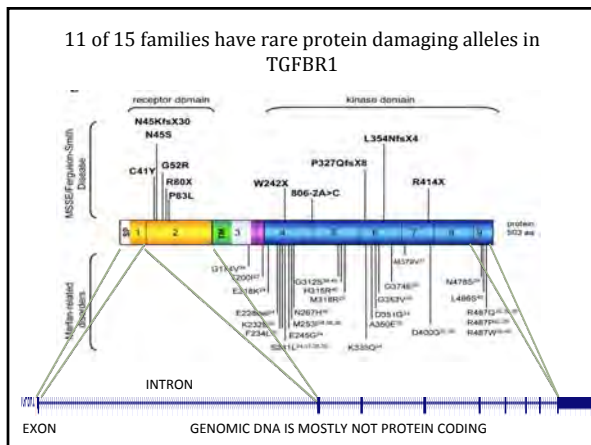
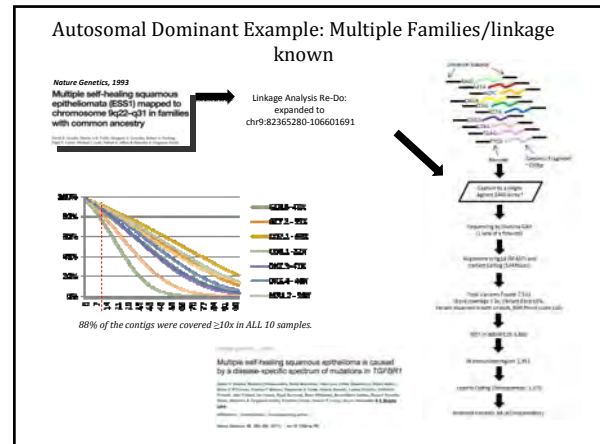
Novel Mendelian Disease Gene: Research Findings Summary

| Condition | Findings | Causative Gene | Report (Publication) |
|-------------------------|------------------|----------------|--------------------------|
| Acrolystosis | New Gene Finding | PDE4D | AJHG Accepted |
| Acrolystosis | New Gene Finding | PRKAR1A | AJHG Accepted |
| Walker Warburg Syndrome | New Gene Finding | ISPD | Nature Genetics Accepted |
| (confidential) | New Gene Finding | | In preparation |
| (confidential) | New Gene Finding | | In preparation |
| MSSE | New Gene Finding | TGFBR1 | Nat Genet 2011 |
| Cuts Laxa | New Gene Finding | PYCR1 | Nat Genet 2009 |
| Brachydactyly | New Gene Finding | CHSY2 | Am J Hum Genet 2010 |
| (confidential) | New Gene Finding | | In revision |
| (confidential) | New Gene Finding | | In preparation |
| (confidential) | New Gene Finding | | In review |

| Primary Diagnosis / Conditions | Findings | Causative Genes |
|--------------------------------|----------------------|-----------------|
| SMDS | Change of Diagnosis? | COL2A1 |
| SMDS | Change of Diagnosis? | FGFR3 |
| Opsismodysplasia | Change of Diagnosis? | TRIP21 |

32 Research Disorders in Process

| Category | Conditions |
|--------------------------|--|
| Complex Disorders | Bipolar, OCD/Tics, Autism |
| Cancers | GBM, Throat Cancer, Acral Melanoma, Uveal Melanoma, Congenital ALL, Schwannomatosis, Stomach Cancer |
| Rare Mendelian Disorders | Familial Hemophagocytic Lymphohistiocytosis, Walker Warburg Syndrome, Asphyxiating Thoracic Dystrophy, SMCS, Opsismodyplasia, SED, Pseudochondroplasia, Spontylthoracic / VATER, Nonprogressive cerebellar ataxia, Ponto cerebellar hypoplasia, Sick sinus syndrome, Horizontal gaze palsy with progressive scoliosis, Ataxia paraplegia, Intestinal atresia, Chronic idiopathic intestinal pseudo-obstruction, Fanconi syndrome, Glucose galactose malabsorption, Congenital rickets, Secretory diarrhea, Bile malabsorption, Generalized malabsorptive diarrhea, Coarse hair syndrome, Tufting enteropathy |



Clinical Case Sampling: Findings Summary

| Primary Diagnosis/ Conditions | Findings | Causative Genes | Final Diagnosis |
|--|-----------------------------|-----------------|-------------------------|
| Juvenile ALS | Change of Diagnosis | AAAS | Allogrove Syndrome |
| Unknown | Diagnosis of a known allele | MT-CO3 | Leber Optic Atrophy |
| Fatal Familial Insomnia / Presumed Prion Disease | Diagnosis of a known allele | PRNP | Fatal Familial Insomnia |
| Congenital malabsorptive diarrhea | Change of Diagnosis | PCK1 | Diabetes Insipidus |
| Mycoclonic Epilepsy | Diagnosis of a known gene | TBC1D24 | Mycoclonic Epilepsy |
| Hypoplastic right heart | No significant findings | - | - |
| Neuropathy | No significant findings | - | - |
| Joubert-like | In Process | - | - |
| Mild Course DMG | In Process | - | - |
| Autism / Developmental Delay / Seizures | In Process | - | - |
| Developmental Delay | In Queue | - | - |
| Muscular Dystrophy | In Queue | - | - |

- ### Current Medical Genetics Practice
- Presentation with rare or common phenotype
 - Family history
 - Clinician search for additional phenotypic information
 - Imaging
 - Biopsy
 - Functional assessments
 - Longitudinal assessment
 - GENETIC SEQUENCING TO CONFIRM DIAGNOSIS
 - 3,000 mendelian disorders, 1,700 gene tests offered!
 - Low success rate except in specific diseases with modest or no genetic heterogeneity (i.e. Duchenne Muscular Dystrophy, Cystic Fibrosis)
 - Ultimate cost to establish these rare diagnoses are high and MOST patients never get a precise genetic diagnosis

Upcoming Medical Genetics Practice

- Presentation with rare or common phenotype
- Family history
- **GENETIC SEQUENCING TO IDENTIFY CAUSATIVE MUTATION AND GUIDE DIAGNOSTIC PROCESS**
- Clinician search for additional phenotypic information
 - Imaging
 - Biopsy
 - Functional assessments
 - Longitudinal assessment

UCLA Clinical Genomics Center

Leading the Way

Toward a Definitive Diagnosis

85% OF DISEASE-CAUSING DNA VARIANTS

1% OF THE ENTIRE HUMAN GENOME

Pre- and Post-Test Counseling
Certified genetic counselors available in person and on a remote basis for patient convenience.

Clinical Exome Sequencing
Detects DNA changes throughout the entire protein-coding regions of patient's genome.

Integrated Bioinformatics and Data Storage
Independently for future interpretation.

Result Interpretation by Our Expert Team
Specialist physicians and board-certified laboratory directors interpret results.

Fast Report Turnaround Time
Results available within 8 weeks.

Human Exome

30million bases (about 1% of the genome),
~20,000 genes, ~240,000 exons

ONLY CAPTURE THE REGIONS OF THE GENOME OF INTEREST

Clinical Exome Sequencing Analysis Workflow

```

    graph TD
      A[Library Preparation  
Exon Capture] --> B[Illumina HiSeq  
250 million short reads]
      B --> C[Alignment  
Find the correct location for each read in the human genome]
      C --> D[Post-Alignment Process  
Realignment  
Indel/realignment  
FINETUNING]
      D --> E[Variant Calling  
Use coverage of 100 reads over flanking of a given base to call variants]
      E --> F[Variant Filtration  
Remove variants with low confidence]
      F --> G[Genomic Annotator  
Determine effect on protein sequence]
      F --> H[Crosslink Data Based Prediction  
Determine likelihood of causing given disease]
    
```

Typical exome output: Highly consistent across samples

| | Mean | Standard Deviation |
|--------------------------------------|--------|--------------------|
| Total Number of SNVs | 22,594 | 1899 |
| dbSNP132 rate (%) | 96.1 | 0.88 |
| Concordance rate (%) | 99.9 | 0.02 |
| Known Variants (dbSNP132) | | |
| Total Number | 21,568 | 1,466 |
| Het/Homo ratio | 1.57 | 0.19 |
| T _i /T _v ratio | 3.16 | 0.03 |
| Novel Variants (dbSNP132) | | |
| Total Number | 899 | 249 |
| Het/Homo ratio | 19.6 | 6.87 |
| T _i /T _v ratio | 2.7 | 0.22 |
| Small Indels | | |
| Total Number | 1,165 | 60 |
| dbSNP132 rate (%) | 50.2 | 0.86 |
| Concordance rate (%) | 92.3 | 0.29 |

New Diagnosis...
Change of Diagnosis...
New Gene Discovery...

CLINICAL CASE EXAMPLES

Congenital Malabsorptive Diarrhea

CASE 1

Patient GMD45:


A perplexing child with congenital malabsorptive diarrhea

- Assessed at three weeks of age for recurrent diarrhea and associated metabolic acidosis
- Known congenital disorders were NOT identified** by metabolic assays or genotyping
- Small bowel follow-through, abdominal ultrasounds, upper and lower GI endoscopies, small and large bowel biopsies including electron microscopy were normal
- Five weeks after discharge, he presented with hypovolemic shock and profound metabolic acidosis
 - central venous catheter was placed, and he started on TPN
- Multiple prolonged hospitalizations followed, and he was treated for
 - central venous catheter occlusions
 - heparin-induced thrombocytopenia
 - plasminogen inhibitor deficiency
 - multiple deep venous thrombi
 - pneumonia and respiratory distress
 - hyperglycemia
 - hypokalemia
 - acidosis
 - left ventricular dysfunction.

Yourshaw M, et al., 2012 in submission

Patient GMD45: Whole-exome sequencing

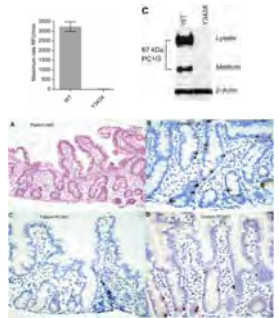
- We sequenced the protein coding regions of the patient's DNA.
- Four genes had homozygous or compound heterozygous mutations that were rare in the population, not present in 74 unrelated control exomes, and predicted to have an adverse effect on protein structure or function
- A **Tyr343X mutation in PCSK1** was highlighted by the fact that the protein would be truncated by a premature stop codon within its catalytic domain
- Defects in PCSK1 are the cause of proprotein convertase 1 deficiency**, which had been identified thus far in three subjects and is characterized by obesity, hypogonadism, hypoadrenalism, and reactive hypoglycemia as well as significant small-intestinal absorptive dysfunction, and DI.



Yourshaw M, et al., 2012 in submission

Patient GMD45: Validation

- The Y343X mutation rendered the gene product undetectable in either cells or media and resulted in a total lack of enzyme activity
- Small and large bowel mucosa were histologically normal except for the loss of PC1/3 positive enteroendocrine cells
 - The architecture, immune cell complement, and epithelium were indistinguishable from normal mucosa (2A)
 - The appearance and number of chromogranin A positive enteroendocrine cells were normal (2B)
 - PC1 expressing enteroendocrine cells were absent in the small (2C, D) and large bowel (data not shown)



Yourshaw M, et al., 2012 in submission

Patient GMD45: Follow-up & treatment

- A follow up evaluation was prompted by identification of the genetic mutation to determine if known clinical manifestations of PC1/3 deficiency were observed in this child
- The family reported evidence suggestive of polydipsia, polyuria, enuresis and polyphagia, but he had not been evaluated for diabetes insipidus
- The subject was found to have an undetectable serum vasopressin level
- Intranasal desmopressin (DDAVP) improved significantly the severity of the polydipsia and enuretic episodes.
- This case confirms the **value and efficiency of exome sequencing as a primary diagnostic mode** to identify mutations of genes associated with rare clinical conditions
- Multiple hospitalizations, and a barrage of various indirect, redundant, and expensive tests failed to establish an actionable diagnosis
- Exome sequencing** provided a diagnosis that resulted in **immediate changes in patient care and an improved ability to predict clinical progression**

Yourshaw M, et al., 2012 in submission

Juvenile ALS

CASE 2

Patient B83: Change of Diagnosis based on Genetic Findings

- Clinical Courses
 - 18 year old female of Persian descent
 - Walked normally as toddler
 - Never ran well
 - Clumsy since age 5
 - Noticeable leg weakness and atrophy age 10
 - Right arm weakness, slurred speech and swallowing difficulties for 5 years
 - Unable to participate in PE throughout school
 - Fatigue
 - Muscle twitches of UE and LE
- Family History
 - None
 - Parents are first cousins once removed
 - 23 year old sister is unaffected
- Exam
 - UMN: Bulbar pathological reflexes / Arms: hyperreflexia / Legs: hyperreflexia
 - LMN: Tongue fascic, atrophy / Arms: atrophy, weakness / Legs: atrophy, weakness

Primary Diagnosis:
Juvenile ALS (Amyotrophic Lateral Sclerosis, Lou Gehrig's Disease) 18 year old girl with a very early age of onset of over the past 8 years (so also a slow progression)

Dr. Martina Wiedow-Piess

Patient B83: Sequence Data

- Exome capture was performed using Agilent SureSelect 50Mb
- Sequencing was performed
 - on a HiSeq2000
 - using V2 chemistry
 - on one flowcell lane

| | | |
|--|------------------|-------------|
| Paired Reads (50bp + 50bp) | 125,749,775 | ~2x ~80% |
| Total 50bp Reads | 251,499,550 | |
| Uniquely Aligned 50 bp Reads | 221,968,363 | |
| | | |
| Total Sequence Aligned to the Human Reference Genome | 9,144,552,729 bp | ~37% |
| Total Sequence Aligned within the RefSeq Gene Coding Exons | 3,425,592,770 bp | |
| Average Coverage Achieved Across Exome | 100.9X | |
| % of Exome Covered with 10 or more reads | 92.4% | |

Patient B83: Keywords & Primary Gene List

- **Keywords:**
 - Amyotrophic lateral sclerosis, ALS
- **Requested Gene List:**
 - UBQLN2, SOD1, SETX, FUS, VAPB, ANG, TARDBP, FIG4, OPTN, DAO, VCP
- **Primary Gene List:**
 - Generated by searching HGMD (Human Gene Mutation Database) Professional Version 2011.3 with the keywords.
 - ALAD, ALS2, ANG, APEX1, ATXN2, CDH13, CHGB, CHMP2B, CRYM, DAO, DCTN1, FIG4, FUS, GRN, HSPB1, KIFAP3, LUM, NAIP, NEFH, OPTN, PARK7, PON1, PON2, PON3, PRPH, SETX, SOD1, SPG11, TAF15, TARDBP, TARDBP, VAPB, VCP, VEGFA, ZNF512B

Patient B83: Results I

A variant was identified in the primary gene list and which is predicted to be altering the function of the protein. This variant is heterozygous and in SETX. Heterozygous variants have been described in SETX that lead to 1) autosomal dominant distal hereditary motor neuropathy; 2) ataxia-ocular apraxia 2 and 3) amyotrophic lateral sclerosis type 4 (ALS4) [MIM: 602433]. Defects in SETX are also known to cause spinocerebellar ataxia autosomal recessive type 1 (SCAR1) [MIM: 606002]; This exact mutation (K992R) was reported to cause ataxia-ocular apraxia 2. However, the allele frequency of this variant in the public database is C=116/T=6904. **Polymorphism**

| Gene Name | Genomic Position (hg19) | Observed Alleles | Zygosity | Amino Acid Change |
|-----------|-------------------------|------------------|--------------|-------------------|
| SETX | 9:135204010 | T/C | heterozygous | K992R |

By using the keywords as search terms, the following three heterozygous, protein changing variants were identified. However, none of these observed variants are of known significance.

| Gene Name | Genomic Position (hg19) | Observed Alleles | Zygosity | Amino Acid Change |
|-----------|-------------------------|------------------|--------------|-------------------|
| ALS2CR11 | 2:202400845 | T/G | heterozygous | S469R |
| RGNF | 5:73072501 | G/T | heterozygous | D274Y |
| BHD | 22:18222869-18222881 | GCCAGCCTCAACT/- | heterozygous | . |

Patient B83: Pedigree

~3% of the genome is homozygous

Patient B83: Results II

Protein changing variants within the homozygous intervals and observed to associate with a human disease: One of these is a non-synonymous novel variant in GRM6. Defects in GRM6 are the cause of recessive congenital stationary night blindness type 1B (CSNB1B) [MIM:257270]. Another homozygous novel missense mutation was identified in AAAS, which is predicted to cause protein truncation in AAAS. Defects in AAAS are reported to cause recessive achalasia-addisonianism-alacrima syndrome (AAAS or 'Triple A syndrome') which is phenotypically heterogeneous and patients with AAAS mutations have been described with juvenile onset progressive spinobulbar amyotrophy.

| Gene Name | Genomic Position (hg19) | Observed Alleles | Zygosity | Amino Acid Change |
|-----------|-------------------------|------------------|------------|-------------------|
| GRM6 | 5:178417746 | G/A | homozygous | R287W |
| AAAS | 12:53702226 | C/A | homozygous | E359X |
| IRAK3 | 12:66622062 | C/T | homozygous | R267X |
| RAFI | 17:17698436 | C/T | homozygous | A725V |

Change of diagnosis to AAAS?

Patient B83: Results II

AAAS 12:53702226 C>A Homozygous

Patient B83: Re-examination

- No tears (alacrima)
- Achalasia age 15
- Peripheral neuropathy
- Ataxia gait/stance
- Laboratory Tests

| Test | Level | Reference Range |
|----------------------------|------------|---|
| Cortisol | 15 mcg/dL | Morning draw: 8-25 Evening draw: <6 |
| DHEA-Sulfate | 140 ng/mL | Adult female: 400-3600 Postmenopausal: 100-1600 Prepubertal Children: 100-600 |
| Folate, Serum | 23.4 ng/mL | 5.2-26.1 |
| Homocysteine, Total | 9 mcmol/L | 5-14 |
| Magnesium | 1.8 mEq/L | 1.3-1.9 |
| Methylmalonic Acid | 125 nmol/L | 87-318 |
| PEP & Total Protein, Serum | 7.2 g/dL | 6.2-8.3 |

AAAS: *A*chalasia, *A*drenocortical insufficiency, *A*lacrimia (aka Allgrove, triple-A)
Change of Diagnosis based on Genetic Findings

Dr. Martina Wislau-Pazos

Unknown

CASE 3

Patient LE: Diagnosis based on genetic findings

- Key issues and terms:
Opsoclonus-Myoclonus, Cataracts, Vertical nystagmus, Dyskinesia/Chorea, Lipid accumulation in muscle
Primary Diagnosis: ?
- Genes tested:
 - PLP1 (Pelizaeus-Merzbacher disease, spastic paraplegia type 2)
 - GJC2 (leukodystrophy hypomyelinating type 2 (HLD2), spastic paraplegia autosomal recessive type 44 (SPG44), lymphedema hereditary type 1C (LMPH1C))
- Other genetic disorders tested & found to be *negative*: GM1 Gangliosidosis, Beta Mannosidosis, Fucosidosis, Tay-Sachs, Sandhoff, Krabbe Disease, Schindler Disease, Metachromatic Leukodystrophy
- Many other biochemical tests were done
- Microdeletion shared with Dad on 1q25.1

Dr. Derek Wong, Dr. Harley Kornblum & Dr. Lekha Rao

Patient LE: Sequence Data

- Exome capture was performed using Agilent SureSelect 50Mb
- Sequencing was performed
 - on a HiSeq2000
 - using V2 chemistry
 - on one flowcell lane

| | | |
|--|------------------|-------|
| Paired Reads (50bp x 50bp) | 180,440,509 | } 72% |
| Total >80bp Reads | 290,881,018 | |
| Uniquely Aligned >80bp Reads | 172,061,237 | |
| Total Sequence Aligned to the Human Reference Genome | 7,189,992,276 bp | } 43% |
| Total Sequence Aligned within the RefSeq Gene Coding Exons | 3,095,991,423 bp | |
| Average Coverage Achieved Across Exome | 91.2X | |
| % of Exome Covered with 10 or more reads | 91.2% | |

Patient LE: Keywords & Primary Gene List

- Keywords:**
 - myoclonus, cataract, nystagmus, dyskinesia, chorea, Krabbe Disease, Metachromatic Leukodystrophy, Sandhoff, Tay-Sachs, Fucosidosis, Gangliosidosis GM1
- Requested Gene List:**
 - PLP1, GJC2
- Primary Gene List:**
 - Generated by searching HGMD (Human Gene Mutation Database) Public version with the keywords.
 - CSTB, DRD2, EPM2A, GOSR2, PRICKLE1, PRICKLE2, SCARB2, SCN2A, SERPINI1, SGCE, CHMP4B, DNMT2, EPHA2, EYA1, CTDP1, JAM3, FOXE3, CACNA1A, CASK, FRMD7, GPR143, PAX6, C14orf104, CCDC39, CCDC40, DNAH11, DNAH5, DNAAF1, DNAAF2, DNAAF3, DNAAF4, DNAAF5, DNAAF6, DNAAF7, DNAAF8, DNAAF9, DNAAF10, DNAAF11, DNAAF12, DNAAF13, DNAAF14, DNAAF15, DNAAF16, DNAAF17, DNAAF18, DNAAF19, DNAAF20, DNAAF21, DNAAF22, DNAAF23, DNAAF24, DNAAF25, DNAAF26, DNAAF27, DNAAF28, DNAAF29, DNAAF30, DNAAF31, DNAAF32, DNAAF33, DNAAF34, DNAAF35, DNAAF36, DNAAF37, DNAAF38, DNAAF39, DNAAF40, DNAAF41, DNAAF42, DNAAF43, DNAAF44, DNAAF45, DNAAF46, DNAAF47, DNAAF48, DNAAF49, DNAAF50, DNAAF51, DNAAF52, DNAAF53, DNAAF54, DNAAF55, DNAAF56, DNAAF57, DNAAF58, DNAAF59, DNAAF60, DNAAF61, DNAAF62, DNAAF63, DNAAF64, DNAAF65, DNAAF66, DNAAF67, DNAAF68, DNAAF69, DNAAF70, DNAAF71, DNAAF72, DNAAF73, DNAAF74, DNAAF75, DNAAF76, DNAAF77, DNAAF78, DNAAF79, DNAAF80, DNAAF81, DNAAF82, DNAAF83, DNAAF84, DNAAF85, DNAAF86, DNAAF87, DNAAF88, DNAAF89, DNAAF90, DNAAF91, DNAAF92, DNAAF93, DNAAF94, DNAAF95, DNAAF96, DNAAF97, DNAAF98, DNAAF99, DNAAF100

Patient LE: Results

- No significant findings from the Keyword search and within the Primary Gene List
- However, a homoplasmic MITOCHONDRIAL mutation known to cause Leber Hereditary Optic Neuropathy (LHON) was identified:

| Gene Name | Genomic Position (hg19) | Observed Alleles | Zygosity | Amino Acid Change |
|-----------|-------------------------|------------------|------------|-------------------|
| MT-CO3 | MT:9804 | G/A | homozygous | A200T |

- OMIM
 - .0002 LEBER OPTIC ATROPHY
 - MTCO3, LHON9804A
 - This allele changes the highly conserved alanine at amino acid 200 to a threonine (A200T). It was present in 3 prior independent patients, was homoplasmic, and was not found in 400 disease and normal controls (Johns and Neufeld, 1993).

Patient LE: Results

MT-CO3 MT:9804 G>A Homozygous

Patient LE: Clinical Follow up

- Mutation validated by UCLA ODTG (Orphan Disease Testing Center)

- As the patients course developed, unusual presentation of LHON considered likely clinical diagnosis
- Recommended management:
 - Co Q 20 mg/kg/day – a very high dose but little toxicity
 - Riboflavin 100 mg/day
 - Carnitine 100 mg/kg/day
- First child: important diagnostic considerations for mother

Dr. Derek Wong, Dr. Harley Kornblum & Dr. Leifka Rao

CHALLENGES

Technical Challenges

Misalignment causes false positive somatic indel

False positive SNVs caused by misalignment

Misalignment corrected by local indel realignment

Technical Challenges

Regions of overalignment

Regions of low coverage (consistently across samples)

Extremely Repetitive Regions get over-covered by incorrect alignments

Main reason we cannot achieve 100% coverage with current technology

Tandem repeats are not well covered and it's hard to detect expansion

Interpretation Limitation

- In B83 and LE cases, the causal mutations were not found by the keyword search or within the primary gene lists built upon clinician's descriptions or the primary diagnosis
 - Without the **Genomic Data Board** discussions with involvement of the subspecialty expert physicians, the causal mutations could have been missed.
 - **Primary Physician's Input** Will Change interpretation!
 - Thorough Description of the Phenotype is Very Important
 - Providing Family History, Pedigree, Previous Clinical Reports, etc will Greatly Improve Interpretation
- Over 200 heterozygous protein changing rare variants per patient:
- Possible disease causing, but if NOVEL, we will not establish clear causality
- These data are stored with the intent of integrating with other additional patients with similar phenotype to identify novel **Discovery of new gene variants. This requires good phenotyping/storage and query capabilities and will be best performed across disease 'names' and across institutions.**

UCLA CLINICAL EXOME SEQUENCING

UCLA Clinical Exome Sequencing

- Suspected genetic disorder (MUST specify diagnosis)
- Blood draw (under 1ml) into lavender tube
- Turnaround time:
 - 6-8 weeks to the Genomic Data Board
 - 12 weeks to the patients
- Protein coding DNA variants matched to known disease causing genes.
 - If single causal gene identified, report generated
 - If multiple potential causal genes identified, Genomic Data Board--- data stored for future interpretations
- Data considered by UCLA Genomic Data Board: **Genomics, Pathology, Medical Genetics, Genetics Counseling**
- Variants of unknown significance (Many) not reported but available without interpretation
- Incidental findings of variants of medically actionable results are reported with patient consent

Conclusions

- Exome Sequencing
 - Routine and reliable
 - High likelihood of success in small patient groups
 - Improved quality of sequencing is not critical
 - Further improvements in sampling of genome in progress
 - About 93% of exome is targetable, about 97% of all known disease causing bases observed (No simple sequence repeats detected)
- Clinical Exome Sequencing is becoming a standard and powerful aspect of our clinical approach to individuals with rare diseases--**Outstanding first line screening tool to identify causative mutations**
- This will drive the generation of a vast knowledge of human genetic variation as it is systematically applied to patients.
- Future: We will move to more routine implementation generally (disease screening)

Acknowledgments

Clinical Genomics Center

Hane Lee, PhD
Traci Toy
Bobby Chin
Naghme Dorrani, MS, CGC
Josh Deignan, PhD
Kingshuk Das, MD
Sharon Webb
Bret Harry
Wayne Grody, MD
Eric Vilain, MD
Genomic Data Board

Nelson Lab

Michael Yourshaw
Barry Merriman, PhD(Life)
Vivian Chang, MD
Paige Taylor
Valerie Arboleda
Kevin Squire, PhD
Ascia Eskin, MS
Aliz Raksi, MS
Zugen Chen, PhD
Nils Homer, PhD (Life)
Brian O'Connor, PhD
Fah Sathirapongsasuti (Harvard)
Sam Strom, PhD

Collaborators

Martin Martin, MD
Martina Wiedau-Pazos, MD
Derek Wong, MD
Harley Kornblum, MD
Dan Cohn, MD
Deb Krakow, MD
Katrina Dipple, MD
Julian Martinez, MD
Bruno Reversade, PhD
Birgitte Lane, FRSE, FMedSci
David Goudie, PhD
Kevin Campbell, PhD
Tobias Willer, PhD
Perry Shieh, MD
and many more...

Lynn Jorde, PhD
Professor and Chair of Human Genetics
University of Utah
Salt Lake City, Utah

Biography

Dr. Lynn Jorde has been on the faculty of the University of Utah School of Medicine since 1979 and holds an H.A. and Edna Benning Presidential Endowed Chair in the Department of Human Genetics. He was appointed Chair of the Department of Human Genetics in September 2009. Dr. Jorde's laboratory has been involved in studies of human genetic variation, mobile element evolution, the genetic basis of human limb malformations, and the genetics of common diseases such as hypertension, juvenile idiopathic arthritis, and inflammatory bowel disease. He has published more than 200 scientific articles on these and related subjects.

Dr. Jorde is the lead author of *Medical Genetics*, a textbook that is now in its 4th edition. He has received 12 teaching awards at the University of Utah School of Medicine, including the Outstanding Pre-Clinical Professor Awards from the graduating medical classes of 2002, 2003, 2004, and 2005; the Leonard W. Jarcho Distinguished Teaching Award in 2003; and the University of Utah Distinguished Teaching Award in 2006. He is the co-recipient (with Dr. Louisa Stark and Dr. John Carey) of the 2008 Award for Excellence in Education from the American Society of Human Genetics. He was president of the American Society of Human Genetics in 2011.

Dr. Jorde has served on several advisory panels for the National Science Foundation and the National Institutes of Health. He completed a 4-year term as a member of the Mammalian Genetics review panel at the National Institutes of Health and a 3-year term on the Board of Directors of the American Society of Human Genetics. He was a member of the Center for Inherited Disease Research Access Committee and currently serves on the 1000 Genomes Advisory Committee for the National Human Genome Research Institute. He is on the scientific advisory boards of the Burroughs Wellcome Foundation and the Institute for Systems Biology. He has served on the editorial boards of *Human Biology*, the *American Journal of Human Biology*, *Gene*, and the *American Journal of Human Genetics*. He has been an expert witness in a number of court cases involving DNA evidence.

Notes

Whole-genome sequencing and the detection of disease-causing mutations

Lynn Jorde, PhD

Whole-genome sequences of related individuals in large pedigrees provide new opportunities and challenges for disease-gene discovery. They also permit direct estimates of sex-specific human mutation rates. The VAAST software package incorporates pedigree information and the observed inheritance pattern with information about genetic variation in a control population and amino acid substitution severity under a unified likelihood analysis framework. We applied VAAST to a four-person pedigree in which the two offspring have Miller syndrome and primary ciliary dyskinesia. By incorporating whole-genome sequence information from all four family members, VAAST unambiguously identified the two genes responsible for these conditions with high statistical significance. VAAST has also been used to identify a gene responsible for a new lethal X-linked condition, Ogden syndrome, using exome data from a single family. We have applied VAAST to whole-genome data from 21 individuals in a 5-generation pedigree to identify a highly-penetrant congenital heart defect mutation. Of the 21 individuals in the pedigree, 11 are affected with cardiac septal defects. The variant responsible for septal defects in the family is a known missense mutation in the *GATA4* gene that had been previously identified through traditional linkage analysis. In our whole-genome analysis of the septal defect phenotype, the *GATA4* gene was highly significant (Bonferroni corrected p-value = 8.9×10^{-5}), and no other gene reached statistical significance. To estimate the male-specific intergenerational mutation rate, we identified novel single nucleotide variants (SNVs) that were absent in a father but were present on the paternal chromosomes of one of the father's offspring. We identified 12 de novo mutations in approximately 600 Mb of sequence data, with estimated false-positive and false-negative rates of less than 1×10^{-3} , for a male-specific, intergenerational mutation rate that is approximately five times greater than the female-specific mutation rate. This result agrees well with estimates based on phylogenetic comparisons.

Notes



A probabilistic disease-gene finder for personal genomes

Mark Yandell, Chad Huff, Hao Hu, et al.

Genome Res. published online June 23, 2011

Access the most recent version at doi:[10.1101/gr.123158.111](https://doi.org/10.1101/gr.123158.111)

**Supplemental
Material**

<http://genome.cshlp.org/content/suppl/2011/06/13/gr.123158.111.DC1.html>

P<P

Published online June 23, 2011 in advance of the print journal.

Open Access

Freely available online through the Genome Research Open Access option.

**Email alerting
service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *Genome Research* go to:
<http://genome.cshlp.org/subscriptions>

Research

A probabilistic disease-gene finder for personal genomes

Mark Yandell,^{1,3,4} Chad Huff,^{1,3} Hao Hu,^{1,3} Marc Singleton,¹ Barry Moore,¹ Jinchuan Xing,¹ Lynn B. Jorde,¹ and Martin G. Reese²

¹Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah and School of Medicine, Salt Lake City, Utah 84112, USA; ²Omicia, Inc., Emeryville, California 94608, USA

VAAST (the Variant Annotation, Analysis & Search Tool) is a probabilistic search tool for identifying damaged genes and their disease-causing variants in personal genome sequences. VAAST builds on existing amino acid substitution (AAS) and aggregative approaches to variant prioritization, combining elements of both into a single unified likelihood framework that allows users to identify damaged genes and deleterious variants with greater accuracy, and in an easy-to-use fashion. VAAST can score both coding and noncoding variants, evaluating the cumulative impact of both types of variants simultaneously. VAAST can identify rare variants causing rare genetic diseases, and it can also use both rare and common variants to identify genes responsible for common diseases. VAAST thus has a much greater scope of use than any existing methodology. Here we demonstrate its ability to identify damaged genes using small cohorts ($n = 3$) of unrelated individuals, wherein no two share the same deleterious variants, and for common, multigenic diseases using as few as 150 cases.

[Supplemental material is available for this article.]

The past three decades have witnessed major advances in technologies for identifying disease-causing genes. As genome-wide panels of polymorphic marker loci were developed, linkage analysis of human pedigrees identified the locations of many Mendelian disease-causing genes (Altshuler et al. 2008; Lausch et al. 2008). With the advent of SNP microarrays, the principle of linkage disequilibrium was used to identify hundreds of SNPs associated with susceptibility to common diseases (Wellcome Trust Case Control Consortium 2007; Manolio 2009). However, the causes of many genetic disorders remain unidentified because of a lack of multiplex families, and most of the heritability that underlies common, complex diseases remains unexplained (Manolio et al. 2009).

Recent developments in whole-genome sequencing technology should overcome these problems. Whole-genome (or exome) sequence data have indeed yielded some successes (Choi et al. 2009; Lupski et al. 2010; Ng et al. 2010; Roach et al. 2010), but these data present significant new analytic challenges as well. As the volume of genomic data grows, the goals of genome analysis itself are changing. Broadly speaking, discovery of sequence dissimilarity (in the form of sequence variants) rather than similarity has become the goal of most human genome analyses. In addition, the human genome is no longer a frontier; sequence variants must be evaluated in the context of preexisting gene annotations. This is not merely a matter of annotating nonsynonymous variants, nor is it a matter of predicting the severity of individual variants in isolation. Rather, the challenge is to determine their aggregative impact on a gene's function, a challenge unmet by existing tools for genome-wide association studies (GWAS) and linkage analysis.

Much work is currently being done in this area. Recently, several heuristic search tools have been published for personal

genome data (Pelak et al. 2010; Wang et al. 2010). Useful as these tools are, the need for users to specify search criteria places hard-to-quantify limitations on their performance. More broadly, applicable probabilistic approaches are thus desirable. Indeed, the development of such methods is currently an active area of research. Several aggregative approaches such as CAST (Morgenthaler and Thilly 2007), CMC (Li and Leal 2008), WSS (Madsen and Browning 2009), and KBAC (Liu and Leal 2010) have recently been published, and all demonstrate greater statistical power than existing GWAS approaches. But as promising as these approaches are, to date they have remained largely theoretical. And understandably so: creating a tool that can use these methods on the very large and complex data sets associated with personal genome data is a separate software engineering challenge. Nevertheless, it is a significant one. To be truly practical, a disease-gene finder must be able to rapidly and simultaneously search hundreds of genomes and their annotations.

Also missing from published aggregative approaches is a general implementation that can make use of Amino Acid Substitution (AAS) data. The utility of AAS approaches for variant prioritization is well established (Ng and Henikoff 2006); combining AAS approaches with aggregative scoring methods thus seems a logical next step. This is the approach we have taken with the Variant Annotation, Analysis & Search Tool (VAAST), combining elements of AAS and aggregative approaches into a single, unified likelihood framework. The result is greater statistical power and accuracy compared to either method alone. It also significantly widens the scope of potential applications. As our results demonstrate, VAAST can assay the impact of rare variants to identify rare diseases, and it can use both common and rare variants to identify genes involved in common diseases. No other published tool or statistical methodology has all of these capabilities.

To be truly effective, a disease-gene finder also needs many other practical features. Since many disease-associated variants are located in noncoding regions (Hindorf et al. 2009), a disease-gene finder must be able to assess the cumulative impact of variants in

³These authors contributed equally to this work.

⁴Corresponding author.

E-mail myandell@genetics.utah.edu.

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.123158.111>. Freely available online through the *Genome Research* Open Access option.

both coding and noncoding regions of the genome. A disease-gene finder must also be capable of dealing with low-complexity and repetitive genome sequences. These regions complicate searches of personal genomes for damaged genes, as they can result in false-positive predictions. The tool should also be capable of using pedigree and phased genome data, as these provide powerful additional sources of information. Finally, a disease-gene finder should have the same general utility that has made genomic search tools such as BLAST (Altschul et al. 1990; Korf et al. 2003), GENSCAN (Burge and Karlin 1997), and GENIE (Reese et al. 2000) so successful: It must be portable, easily trained, and easy to use; and, ideally, it should be an ab initio tool, requiring only very limited user-specified search criteria. Here we show that VAAST is such a tool.

We demonstrate VAAST's ability to identify both common and rare disease-causing variants using several recently published personal genome data sets, benchmarking its performance on more than 100 Mendelian conditions including congenital chloride diarrhea (Choi et al. 2009) and Miller syndrome (Ng et al. 2010; Roach et al. 2010). We also show that VAAST can identify genes responsible for two common, complex diseases, Crohn disease (Lesage et al. 2002) and hypertriglyceridemia (Johansen et al. 2010).

Collectively, our results demonstrate that VAAST provides a highly accurate, statistically robust means to rapidly search personal genome data for damaged genes and disease-causing variants in an easy-to-use fashion.

Results

VAAST scores

VAAST combines variant frequency data with AAS effect information on a feature-by-feature basis (Fig. 1) using the likelihood ratio (λ) shown in Equations 1 and 2 in Methods. Importantly, VAAST can make use of both coding and noncoding variants when doing so (see Methods). The numerator and denominator in Equation 1 give the composite likelihoods of the observed genotypes for each feature under a healthy and disease model, respectively. For the healthy model, variant frequencies are drawn from the combined control (background) and case (target) genomes (p_i in Eq. 1); for the disease model, variant frequencies are taken

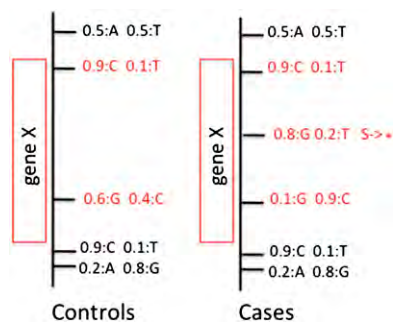


Figure 1. VAAST uses a feature-based approach to prioritization. Variants along with frequency information, e.g., 0.5:A 0.5:T, are grouped into user-defined features (red boxes). These features can be genes, sliding windows, conserved sequence regions, etc. Variants within the bounds of a given feature (shown in red) are then scored to give a composite likelihood for the observed genotypes at that feature under a healthy and disease model by comparing variant frequencies in the cases (target) compared to control (background) genomes. Variants producing non-synonymous amino acid changes are simultaneously scored under a healthy and disease model.

separately from the control genomes (p_i^U in Eq. 2) and the case genomes file (p_i^A in Eq. 1), respectively. Similarly, genome-wide Amino Acid Substitution (AAS) frequencies are derived using the control (background) genome sets for the healthy model; for the disease model, these are based either on the frequencies of different AAS observed for OMIM (Yandell et al. 2008) alleles or from the BLOSUM (Henikoff and Henikoff 1992) matrix, depending on user preference. Figure 2 shows the degree to which AAS frequencies among known disease-causing alleles in OMIM and AAS frequencies in healthy personal genomes differ from the BLOSUM model of amino acid substitution frequencies. As can be seen, the AAS frequency spectra of these data sets differ markedly from one another. The differences are most notable for stop codons, in part because stop gains and losses are never observed in the multiple protein alignments used by AAS methods and LOD-based scoring schemes such as BLOSUM (Henikoff and Henikoff 1992).

VAAST aggregately scores variants within genomic features. In principle, a feature is simply one or more user-defined regions of the genome. The analyses reported here use protein-coding human gene models as features. Each feature's significance level is the one-tailed probability of observing λ , which is estimated from a randomization test that permutes the case/control status of each individual. For the analyses reported below, the genome-wide statistical significance level (assuming 21,000 protein-coding human genes) is $0.05/21,000 = 2.4 \times 10^{-6}$.

Comparison to AAS approaches

Our approach to determining a variant's impact on gene function allows VAAST to score a wider spectrum of variants than existing AAS methods (Lausch et al. 2008) (for more details, see Eq. 2. in Methods). SIFT (Kumar et al. 2009), for example, examines non-synonymous changes in human proteins in the context of multiple alignments of homologous proteins from other organisms. Because not every human gene is conserved and because conserved genes often contain unconserved coding regions, an appreciable fraction of nonsynonymous variants cannot be scored by this approach. For example, for the genomes shown in Table 2, ~10% of nonsynonymous variants are not scored by SIFT due to a lack of conservation. VAAST, on the other hand, can score all non-synonymous variants. VAAST can also score synonymous variants and variants in noncoding regions of genes, which typically account for the great majority of SNVs (single nucleotide variants) genome-wide. Because AAS approaches such as SIFT cannot score these variants, researchers typically either exclude them from the search entirely or else impose a threshold on the variants' frequencies as observed in dbSNP or in the 1000 Genomes Project data set (The 1000 Genomes Project Consortium 2010). VAAST takes a more rigorous, computationally tractable approach: The VAAST score assigned to a noncoding variant is not merely the reciprocal of the variant's frequency; rather, the noncoding variant's score is a log-likelihood ratio that incorporates an estimate of the severity of the substitution as well as the allele frequencies in the control and case genomes (for details, see Scoring Noncoding Variants section in Methods).

To illustrate the consequences of VAAST's novel approach to nonsynonymous variant scoring, we compared it to two widely used tools for variant prioritization, SIFT (Kumar et al. 2009) and ANNOVAR (Wang et al. 2010). Using a previously published data set of 1454 high-confidence known disease-causing and predisposing coding variants from OMIM (Yandell et al. 2008), we asked what fraction were identified as deleterious by each tool. SIFT correctly identified 69% of the disease-causing variants ($P < 0.05$),

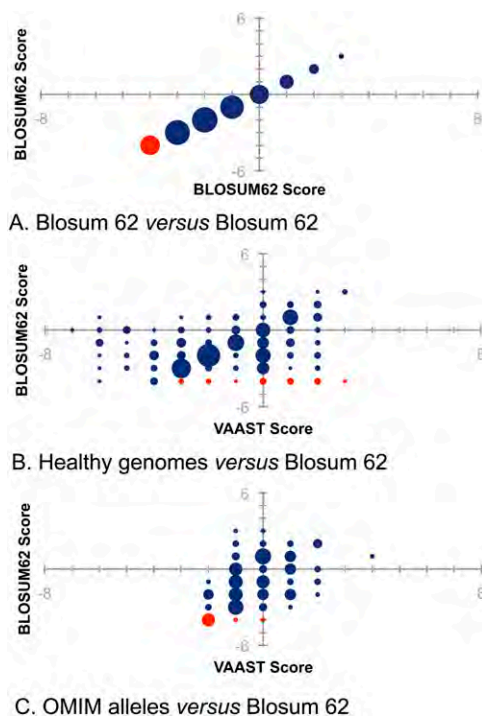


Figure 2. Observed amino acid substitution frequencies compared to BLOSUM62. Amino acid substitution frequencies observed in healthy and reported for OMIM disease alleles were converted to LOD-based scores for purposes of comparison to BLOSUM62. The BLOSUM62 scores are plotted on the y-axis throughout. (Red circles) stops; (blue circles) all other amino acid changes. The diameter of the circles is proportional to the number of changes with that score in BLOSUM62. (A) BLOSUM62 scoring compared to itself. Perfect correspondence would produce the diagonally arranged circles shown. (B) Frequencies of amino acid substitutions in 10 healthy genomes compared to BLOSUM62. (C) OMIM nonsynonymous variant frequencies compared to BLOSUM62.

ANNOVAR (Wang et al. 2010) identified 71%, and VAAST identified 98.0% (Table 1). We then carried out the same analysis using 1454 nonsynonymous variants, randomly drawn from five different European-American (CEU) genome sequences by the 1000 Genomes Project (The 1000 Genomes Project Consortium 2010). These variants are unlikely to be disease-causing given that the individuals are healthy adults. SIFT incorrectly identified 18% of the “healthy” variants as deleterious ($P < 0.05$), ANNOVAR (Wang et al. 2010) identified 1%, and VAAST identified 8%. Under the assumption that there are 1454 true positives and an equal number of true negatives, these two analyses indicate that overall the accuracy [(Sensitivity + Specificity)/2] of SIFT was 75%, ANNOVAR 85%, and VAAST 95% (Table 1). Figure 5C below provides a comparison of the same three tools in the context of genome-wide disease-gene hunts.

We also used these data to investigate the relative contribution of AAS and variant frequency information to VAAST’s allele prioritization accuracy. Running VAAST without using any AAS information, its accuracy decreased from 95% to 80%, demonstrating that the AAS information contributes significantly to VAAST’s accuracy in identifying deleterious alleles.

Population stratification

The impact of population stratification on VAAST’s false-positive rate is shown in Figure 3A (red line). In this test we used 30 European-American genomes as a background file and various mixtures

of 30 European-American and Yoruban (African) genomes as targets. We then ran VAAST on these mixed data sets and observed the number of genes with VAAST scores that reached genome-wide significance, repeating the process after replacing one of the target or background genomes with a Yoruban genome from the 1000 Genomes data set (The 1000 Genomes Project Consortium 2010), until the target contained 30 Yoruban genomes and the background set contained 30 European-American genomes. The resulting curve shown in red in Figure 3A thus reports the impact of differences in population stratification in cases and controls on VAAST’s false-positive prediction rate. With complete stratification (e.g., all genomes in the target are Yoruban and all background genomes are CEU), 1087 genes have LD-corrected genome-wide statistically significant scores ($\alpha = 2.4 \times 10^{-6}$).

Platform errors

We also investigated the impact of bias in sequencing platform and variant-calling procedures on false-positive rates, using a similar approach to the one we used to investigate population stratification effects. Here we varied the number of case genomes drawn from different sequencing platforms and alignment/variant-calling pipelines. We began with 30 background genomes drawn from the CEU subset of the 1000 Genomes Project (The 1000 Genomes Project Consortium 2010) initial release. All of the selected genomes were sequenced to $\sim 6\times$ and called using the 1000 Genomes Project variant-calling pipeline. The target file in this case consisted of 30 similar 1000 Genomes Project CEU genomes that were not included in the background file. This was the starting point for these analyses. We then ran VAAST and recorded the number of genes with LD-corrected genome-wide statistically significant scores ($\alpha = 2.4 \times 10^{-6}$), repeating the process after substituting one of the target genomes with a non-1000 Genomes Project European-American (CEU) genome (Reese et al. 2000; Li et al. 2010). We repeated this process 30 times. These results are shown in Figure 3B (red line). Taken together, these results (Fig. 3) quantify the impact of population stratification and the cumulative effects of platform differences, coverage, and variant-calling procedures on false-positive rates and allow comparisons of the relative magnitude of platform-related biases to population stratification effects. With all background genomes from the subset of the 1000 Genomes Project data (The 1000 Genomes Project Consortium

Table 1. Variant prioritization accuracy comparisons

| | Percent judged deleterious | | |
|----------|----------------------------|---------|-------|
| | SIFT | ANNOVAR | VAAST |
| Diseased | 69% | 71% | 98% |
| Healthy | 18% | 1% | 8% |
| Accuracy | 75% | 85% | 95% |

SIFT, ANNOVAR, and VAAST were run on a collection of 1454 known disease-causing variants (Diseased) and 1454 presumably healthy variants randomly chosen from five different CEU genomes (Healthy). The top portion of the table reports the percentage of variants in both sets judged deleterious by the three tools. The bottom row reports the accuracy of each tool. The filtering criteria used in ANNOVAR excluded all variants present in the 1000 Genomes Project data and dbSNP130 as well as any variant residing in a segmentally duplicated region of the genome. For the “Diseased” category, the VAAST control data set contained 196 personal genomes drawn from the 1000 Genomes Project and 10Gen data sets and dbSNP130. For the “Healthy” category, the VAAST control data set contained 55 other European-American genomes drawn from the 1000 Genomes Project data set (to match the ethnicity of the 1454 CEU alleles).

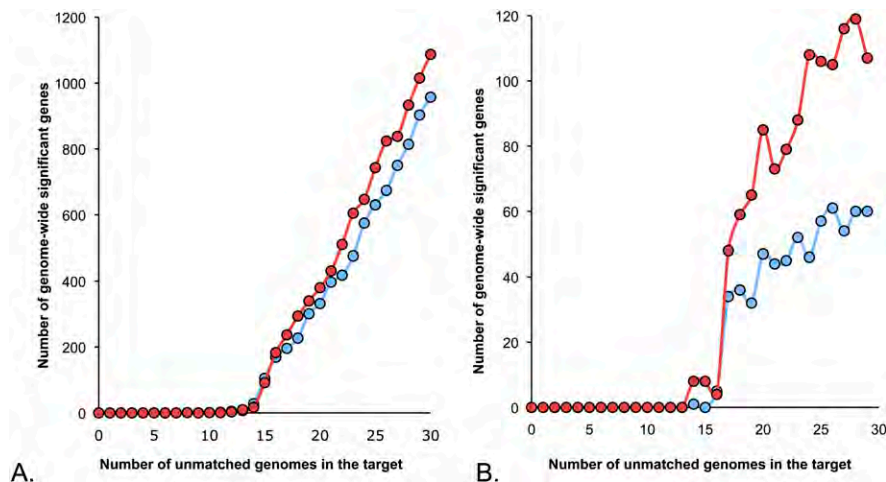


Figure 3. Impact of population stratification and platform bias. Numbers of false positives with and without masking. (A) Effect of population stratification. (B) Effect of heterogeneous platform and variant calling procedures. (Red line) Number of false positives without masking; (blue line) after masking. Note that although masking has little effect on population stratification, it has a much larger impact on platform bias. This is an important behavior: Population stratification introduces real, but confounding signals into disease gene searches; these signals are unaffected by masking (A); in contrast, VAAST's masking option removes false positives due to noise introduced by systematic errors in platform and variant calling procedures (B).

2010) described above and all target genomes from data sets other than the 1000 Genomes data set (Reese et al. 2000; Li et al. 2010), 107 genes have genome-wide LD-corrected statistically significant scores ($\alpha = 2.4 \times 10^{-6}$), compared to the 1087 observed in our population stratification experiments ($\alpha = 2.4 \times 10^{-6}$).

Variant masking

The limited number of personal genomes available today necessitates comparisons of genomes sequenced on different platforms, to different depths of coverage, and subjected to different variant-calling procedures. As shown in Figure 3B, these factors can be a major source of false positives in disease-gene searches. Based on an analysis of these data, we found variant-calling errors to be over-represented in low-complexity and repetitive regions of the genome, which is not unexpected. We therefore developed a VAAST runtime option for masking variants within these regions of the genome. VAAST users specify a read length and paired or unpaired reads. VAAST then identifies all variants in non-unique regions of the genome meeting these criteria and excludes them from its calculations. The blue lines in Figure 3 plot the number of genes attaining LD-corrected genome-wide significance after masking. As can be seen, whereas masking has negligible impact on false positives due to population stratification, it has a much larger impact on sequencing platform and variant-calling bias. This is a desirable behavior. Population stratification introduces real, but confounding, signals into disease-gene searches, and these real signals are unaffected by masking (Fig. 3A). In contrast, masking eliminates many false positives due to noise introduced by systematic errors in sequencing platform and variant-calling procedures (Fig. 3B).

Identification of genes and variants that cause rare diseases

Miller syndrome

Our targets in these analyses were the exome sequences of two siblings affected with Miller syndrome (Ng et al. 2010; Roach et al.

2010). Previous work (Ng et al. 2010; Roach et al. 2010) has shown that the phenotypes of these individuals result from variants in two different genes. The affected siblings' craniofacial and limb malformations arise from compromised copies of *DHODH*, a gene involved in pyrimidine metabolism. Both affected siblings also suffer from primary ciliary dyskinesia as a result of mutations in another gene, *DNAH5*, that encodes a ciliary dynein motor protein. Both affected individuals are compound heterozygotes at both of these loci. Thus, this data set allows us to test VAAST's ability to identify disease-causing loci when more than one locus is involved and the mutations at each locus are not identical by position or descent.

Accuracy on the Miller syndrome data

We carried out a genome-wide search of 21,000 protein-coding genes using the two affected Miller syndrome exomes as targets and using two different healthy-genome background files. The first back-

ground file consists of 65 European-American (CEU) genomes selected from the 1000 Genomes Project data (The 1000 Genomes Project Consortium 2010) and the 10Gen data set (Reese et al. 2010). The second, larger background file consists of 189 genomes selected from the same data sources, but, in distinction to the first, is ethnically heterogeneous and contains a mixture of sequencing platforms, allowing us to assay the impact of these factors on VAAST's performance in disease-gene searches. In these experiments, we ran VAAST using its recessive disease model option (for a description of VAAST disease models, see Methods), and with and without its variant-masking option. Depending on whether or not its variant-masking option was used, VAAST identified a maximum of 32, and a minimum of nine, candidate genes. Variant masking, on average, halved the number of candidates (Table 2). The best accuracy was obtained using the larger background file together with the masking option. *DHODH* ranked fourth and *DNAH5* fifth among the 21,000 human genes searched. This result demonstrates that VAAST can identify both disease genes with great specificity using a cohort of only two related individuals, both compound heterozygotes for a rare recessive disease. Overall, accuracy was better using the second, larger background file, demonstrating that, for rare diseases, larger background data sets constructed from a diverse set of populations and sequencing platforms improve VAAST's accuracy, despite the stratification issues these data sets introduce.

We also took advantage of family quartet information (Ng et al. 2010; Roach et al. 2010) to demonstrate the utility of pedigree information for VAAST searches. When run with its pedigree and variant-masking options, only two genes are identified as candidates: *DNAH5* is ranked first, and *DHODH* is ranked second, demonstrating that VAAST can achieve perfect accuracy using only a single family quartet of exomes (Fig. 4). Our previously published analysis (Roach et al. 2010) identified four candidate genes, and further, expert post hoc analyses were required to identify the two actual disease-causing genes. The results shown in Figure 4 thus demonstrate that VAAST can use pedigree data to improve its accuracy, even in the face of confounding signals due to relatedness

Table 2. Effect of background file size and stratification on accuracy

| | Genome-wide significant genes | DHODH | | DNAH5 | |
|---------------------------------|-------------------------------|-------|-----------------------|-------|-----------------------|
| | | Rank | P-value | Rank | P-value |
| Caucasian only (65 genomes) | | | | | |
| UMSK | 32 | 25 | 7.92×10^{-7} | 32 | 1.98×10^{-6} |
| MSK | 17 | 14 | 9.93×10^{-7} | 19 | 5.79×10^{-5} |
| Mixed ethnicities (189 genomes) | | | | | |
| UMSK | 16 | 9 | 6.78×10^{-9} | 5 | 2.00×10^{-9} |
| MSK | 9 | 4 | 7.60×10^{-9} | 5 | 1.18×10^{-8} |

Results of searching the intersection of two Utah Miller Syndrome affected genomes against two different background files, with and without masking. (Caucasians only) 65 Caucasian genomes drawn from six different sequencing/alignment/variant calling platforms; (mixed ethnicities) 189 genomes (62 YRI, 65 CAUC, 62 ASIAN), from the 1000 Genomes Project and 10Gen data set; (UMSK) unmasked; (MSK): masked; (genome-wide significant genes) number of genes genome-wide attaining a significant non-LD corrected *P*-value; (rank) gene rank of *DHODH* and *DNAH5* among all scored genes; (*P*-value) non-LD corrected *P*-value; genome-wide significant alpha is 2.4×10^{-6} . Data were generated using a fully penetrant, monogenic recessive model. The causative allele incidence was set to 0.00035.

of target exomes, significant population stratification, and platform-specific noise.

Impact of noncoding SNVs

We used these same data sets to investigate the impact of using both coding and noncoding variants in our searches. To do so, we extended our search to include all SNVs at synonymous codon positions and in conserved DNase hypersensitive sites and transcription factor-binding sites (for details, see Methods). Doing so added an additional 36,883 synonymous and regulatory variants to the 19,249 nonsynonymous changes we screened in the analyses reported above. Using only the two Utah siblings, 189 candidate genes are identified. *DHODH* is ranked 15th and *DNAH5* is sixth among them. Repeating the analysis using family quartet information, 23 candidate genes are identified; *DHODH* is ranked fourth and *DNAH5* is ranked first. Thus, increasing the search space to include almost 37,000 additional noncoding variants had little negative impact on accuracy.

Impact of cohort size

We also used the Miller syndrome data to assess the ability of VAAST to identify disease-causing genes in very small case cohorts wherein no two individuals in the target data set are related or share the same disease-causing variants. We also wished to determine the extent to which the relatedness of the two siblings introduced spurious signals into the analyses reported in Table 2. We used information from additional Miller syndrome kindreds (Ng et al. 2010; Roach et al. 2010) to test this scenario. To do so, we used a publicly available set of Danish exome sequences (Li et al. 2010). We added two different disease-causing variants in *DHODH* reported in individuals with Miller syndrome (Ng et al. 2010; Roach et al. 2010) to six different Danish exomes to produce six unrelated Danish exomes, each carrying two different Miller syndrome causative alleles. The background file consisted of the same 189 genome equivalents of mixed ethnicities and sequencing platforms used in Table 2. We then used VAAST to carry out a genome-wide screen using these six exomes as targets. We first used one exome as a target, then the union of two exomes as a target, and so on, in order to investigate VAAST's performance in a series of case cohorts containing pools of one to six exomes. The results are shown in Table 3.

DHODH is the highest ranked of two candidates for a cohort of three unrelated individuals and the only candidate to achieve LD-corrected genome-wide statistical significance (Table 3). In this data set no two individuals share the same variants, nor are any homozygous for a variant. This data set thus demonstrates VAAST's ability to identify a disease-causing gene in situations in which the gene is enriched for rare variants, but no two individuals in the case data set share the same variant, and the cohort size is as small as three unrelated individuals. VAAST's probabilistic framework also makes it possible to assess the relative contribution of each variant to the overall VAAST score for that gene, allowing users to identify and prioritize for follow-up studies those variants predicted to have the greatest functional impact on a gene. Table 4 shows these scored variants for the Miller syndrome alleles of all six affected individuals.

Congenital Chloride Diarrhea (CCD) data set

We tested VAAST's ability to identify the genetic variant responsible for a rare recessive disease using the whole-exome sequence of a patient diagnosed with congenital chloride diarrhea (CCD) due to a homozygous D652N mutation in the *SLC26A3* gene (Choi et al. 2009). In this analysis the background data set consisted of 189 European-American genomes (Table 5). Using the single affected exome as a target, *SLC26A3* is ranked 21st genome-wide. We also evaluated the impact of bias in platform and variant-calling procedures on this result. To do so, we added the CCD causative allele as a homozygote to an ethnically matched genome drawn from the 1000 Genomes data set (Table 5; The 1000 Genomes Project Consortium 2010), in the same manner that was used to generate the data in Table 3. Under the assumption that this rare recessive disease is due to variants at the same location in each affected genome (intersection by position), only a single pair of unrelated exomes is required to identify CCD with perfect specificity. Adding a third affected exome is sufficient to obtain LD-corrected genome-wide statistical significance, even when the selection criteria are relaxed to include the possibility of different disease-causing alleles at different positions in different individuals (union of variants by position).

Impact of recessive modeling on accuracy

We also investigated the impact of VAAST's recessive inheritance model on our rare disease analyses (Supplemental Tables 2, 3). In general, running VAAST with this option yielded improved specificity but had little impact on gene ranks. For a cohort of three unrelated Miller syndrome individuals, the recessive inheritance model had no impact on rank or specificity (Supplemental Table 2). For CCD, using a cohort of three unrelated individuals, *SLC26A3* was ranked first in both cases, but the recessive model decreased the number of candidate genes from seven to two (Supplemental Table 3). These results demonstrate VAAST's ab initio capabilities: It is capable of identifying disease-causing alleles with great accuracy, even without making assumptions regarding mode of inheritance. Our large-scale performance analyses, described below, support and clarify these conclusions.

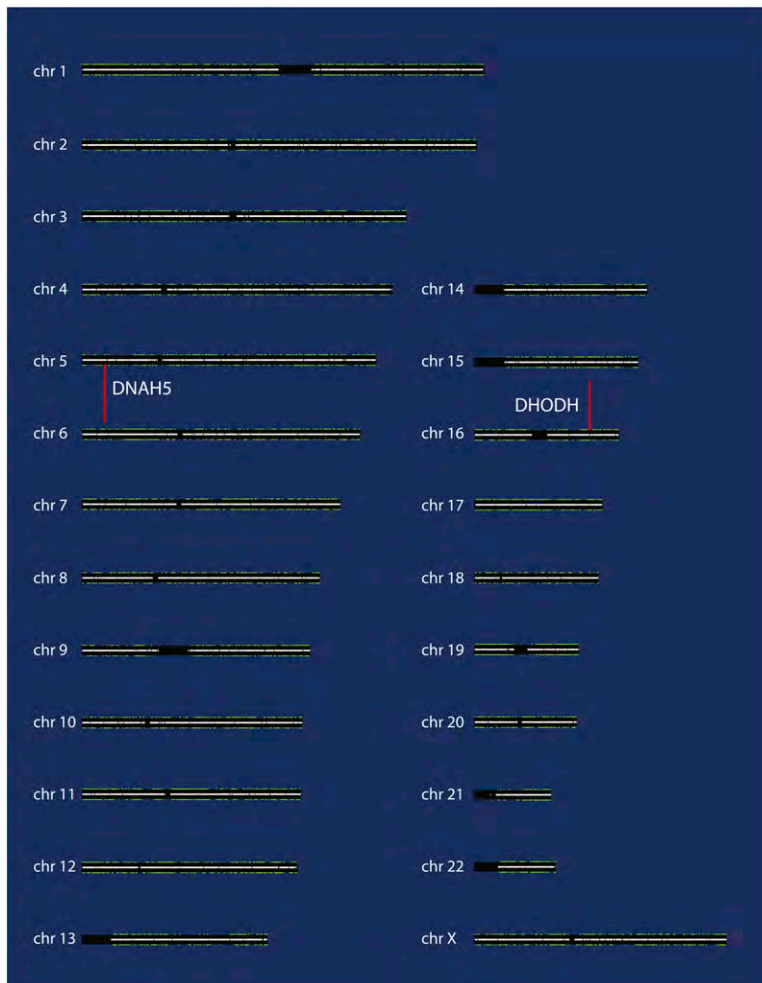


Figure 4. Genome-wide VAAST analysis of Utah Miller Syndrome Quartet. VAAST was run in its quartet mode, using the genomes of the two parents to improve specificity when scoring the two affected siblings. Gray bars along the center of each chromosome show the proportion of unique sequence along the chromosome arms, with white denoting completely unique sequence; black regions thus outline centromeric regions. Colored bars above and below the chromosomes (mostly green) represent each annotated gene; plus strand genes are shown above and minus strand genes below; their width is proportional to their length; height of bar is proportional to their VAAST score. Genes colored red are candidates identified by VAAST. Only two genes are identified in this case: *DNAH5* and *DHODH*. Causative allele incidence was set to 0.00035, and amino acid substitution frequency was used along with variant-masking. This view was generated using the VAAST report viewer. This software tool allows the visualization of a genome-wide search in easily interpretable form, graphically displaying chromosomes, genes, and their VAAST scores. For comparison, the corresponding figure, without pedigree information, is provided as Supplemental Figure 1.

Benchmark on 100 different known disease genes

To gain a better understanding of VAAST's performance characteristics, we also evaluated its ability to identify 100 different known disease-causing genes in genome-wide searches. For these analyses, we first randomly selected (without replacement) a known disease-causing gene from OMIM for which there existed at least six different published nonsynonymous disease-causing alleles. See Supplemental File 2 for a complete listing of diseases, genes, and alleles. Next we randomly selected known disease-causing alleles at the selected locus (without replacement) and inserted them at their reported positions within the gene into different whole-genome sequences drawn for the Complete Genomics Diversity Panel (<http://www.completegenomics.com/>

[sequence-data/download-data/](#)). We then ran VAAST under a variety of scenarios (e.g., dominant, recessive, and various case cohort sizes) and recorded the rank of the disease gene, repeating the analyses for 100 different known disease genes. We also compared the performance of VAAST to SIFT and ANNOVAR using these same data sets. (Details of the experimental design can be found in the Methods section.) The results of these analyses are shown in Figure 5. In this figure the height of each box is proportional to the mean rank of the disease-causing gene for the 100 trials, and the number shown above each box is the mean rank from among 17,293 RefSeq genes. The error bars delimit the spread of the ranks, with 95% of the runs encompassed within the bars.

Figure 5A summarizes VAAST's performance on this data set under both dominant and recessive disease scenarios. For these experiments, we assayed the average rank for three different cohort sizes: two, four, and six individuals for the dominant scenario, and one, two, and three individuals for the recessive analyses. For both scenarios, the mean and variance rapidly decrease as the cohort size increases. **For the dominant scenario, using a case cohort of six unrelated individuals, each carrying a different disease-causing allele, VAAST ranked the disease-causing gene on average ninth out of 17,293 candidates with 95% of the runs having ranks between 5 and 40 in 100 different genome-wide searches. For the recessive scenario, using a case cohort of three unrelated individuals each carrying two different disease-causing variants at different positions (all compound heterozygotes), VAAST ranked the disease-causing gene on average third out of 17,293 candidates, with 95% of the runs having ranks between 2 and 10. None of the individuals had any disease-causing alleles in common.**

Figure 5B summarizes VAAST's performance when only a subset of the case cohort contains a disease-causing allele, which could result from (1) no calls at the disease-causing allele during variant calling; (2) the presence of phenocopies in the case cohort; and (3) locus heterogeneity. As can be seen in Figure 5B, averages and variances decrease monotonically as increasing numbers of individuals in the case cohort bear disease-causing alleles in the gene of interest. Moreover, for dominant diseases, even when one-third of the cases lack disease-causing alleles in the selected OMIM disease gene, VAAST achieves an average rank of 61 with 95% of the runs having ranks between 5 and 446. For recessive diseases the average was 21, with 95% of the disease genes ranking between 7 and 136 out of 17,293 genes, genome-wide.

Figure 5C compares VAAST's accuracy to that of ANNOVAR and SIFT. For these analyses, we used the same data used to produce

Table 3. Impact of cohort size on VAAST's ability to identify a rare disease caused by compound heterozygous alleles

| Target genome(s) | Genome-wide | | | DHODH rank | | |
|--------------------------|-------------------|------------------|--------------|------------|------------------------|------------------------|
| | Significant genes | | | P-value | | |
| | Genes scored | Non-LD-corrected | LD-corrected | Rank | Non-LD-corrected | LD-corrected |
| 1 Compound heterozygote | 92 | 67 | 0 | 86 | 2.36×10^{-4} | 5.26×10^{-3} |
| 2 Compound heterozygotes | 4 | 3 | 0 | 2 | 2.81×10^{-8} | 5.51×10^{-5} |
| 3 Compound heterozygotes | 2 | 2 | 1 | 1 | 2.61×10^{-11} | 8.61×10^{-7} |
| 4 Compound heterozygotes | 1 | 1 | 1 | 1 | 1.99×10^{-15} | 1.78×10^{-8} |
| 5 Compound heterozygotes | 1 | 1 | 1 | 1 | 6.95×10^{-15} | 4.60×10^{-10} |
| 6 Compound heterozygotes | 1 | 1 | 1 | 1 | 5.79×10^{-17} | 1.42×10^{-11} |

The background file consisted of 189 genomes of mixed ethnicity from the 1000 Genomes Project combined with nine additional genomes of mixed ethnicity and sequencing platforms drawn from the 10Gen genome set (Reese et al. 2010). Causative alleles reported in the six individuals described in Ng et al. (2010) were added to unrelated exomes from re-sequenced individuals from Denmark reported in Li et al. (2010). Data were generated using a fully penetrant monogenic recessive model (see Supplemental Table 2). Causative allele incidence was set to 0.00035 (for details, see Supplemental Table 2), and amino acid substitution frequency was used along with masking of repeats. (Genes scored) Number of genes in the genome with variant distributions consistent with VAAST's fully penetrant monogenic recessive model and causative allele incidence threshold. Scoring was evaluated by permutation by gene and permutation by genome.

Figure 5A, running all three tools on a case cohort of six and three individuals for the dominant and recessive comparisons, respectively (for details, see Methods). In these analyses, all members of the case cohort contain disease-causing alleles. For ANNOVAR, we set the expected combined disease-allele frequency at <5% (see Methods) as this improved ANNOVAR's performance (data not shown), but for VAAST no prior assumptions were made regarding the disease-causing alleles' frequencies in the population. VAAST outperforms both SIFT and ANNOVAR—both as regards to mean ranks and variances. VAAST, for example, achieves a mean rank of 3 for recessive diseases using three compound heterozygous individuals as a case cohort. SIFT achieves an average rank of 2317, and ANNOVAR an average rank of 529. There is also much less variance in the VAAST ranks than in those of the other tools. For example, in the recessive scenario, using three compound heterozygous individuals as a case cohort, in 95% of the VAAST runs the rank of the disease-causing gene was between ranks 2 and 10. By comparison, ANNOVAR's ranks varied between 67 and 8762 on the same data sets, and SIFT's varied between 66 and 9107. See Supplemental Figures 2 and 3 for the complete distributions. We also investigated the possibility that taking the intersection of ANNOVAR and SIFT calls might improve accuracy compared to either of these tools alone. It did not; see Supplemental Figure 4. Closer inspection of these data revealed the reasons for the high variances characteristic of SIFT and ANNOVAR. In SIFT's case, the variance is due to failure to identify one or more of the disease-causing alleles as deleterious, a finding consistent with our accuracy analysis presented in Table

1. This, coupled with its inability to make use of variant frequencies, means that SIFT also identifies many very frequent alleles genome-wide as deleterious, increasing the rank of the actual disease-causing gene. ANNOVAR's performance, because it can filter candidate variants based on their allele frequencies, is thus better than SIFT's (average rank of 529 vs. 2317). However, its variance from search to search remains high compared to VAAST, as the OMIM alleles in the analysis are distributed across a range of frequencies, and unlike VAAST, ANNOVAR is unable to leverage this information for greater accuracy.

Identification of genes and variants causing common multigenic diseases

Power analyses

Our goal in these analyses was twofold: first, to benchmark the statistical power of VAAST compared to the standard single nucleotide variation (SNV) GWAS approach; and second, to determine the relative contributions of variant frequencies and amino acid substitution frequencies to VAAST's statistical power. We also compared the statistical power of VAAST's default scoring algorithm to that of WSS (Madsen and Browning 2009), one of the most accurate aggregative methods to date for identifying common disease genes using rare variants. Figure 6A shows the results for the *NOD2* gene, implicated in Crohn's disease (CD) (Lesage et al. 2002). This data set contains both rare (minor allele frequency [MAF] <5%) and common variants. Figure 6B shows the same power analysis

Table 4. Relative impacts of observed variants in DHODH

| Sequence Information | | | | VAAST Scoring | SIFT Scoring | |
|----------------------|--------------------|------------------|-------------------------|---------------|--------------|-------------------------|
| Genomic Position | Reference Sequence | Variant Genotype | Amino Acid Substitution | Score | Score | Impact |
| chr16:70599943 | T | C,T | Promoter | 0.00 | N/A | UNABLE TO SCORE |
| chr16:70600183 | A | C,C | K->Q | 0.00 | 0.19 | TOLERATED (rs3213472:C) |
| chr16:70603484 | G | G,A | G->E | 4.87 | 0.05 | DAMAGING (novel) |
| chr16:70606041 | C | C,T | R->C | 6.21 | 0.00 | DAMAGING (novel) |
| chr16:70608443 | G | G,A | G->R | 19.08 | 0.00 | DAMAGING (novel) |
| chr16:70612601 | C | C,T | R->C | 6.21 | 0.00 | DAMAGING (novel) |
| chr16:70612611 | G | G,C | G->A | 25.17 | 0.16 | TOLERATED (novel) |
| chr16:70612617 | T | T,C | L->P | 5.19 | 0.02 | DAMAGING (novel) |
| chr16:70613786 | C | C,T | R->W | 6.66 | 0.02 | DAMAGING (novel) |
| chr16:70614596 | C | C,T | T->I | 3.52 | 0.00 | DAMAGING (novel) |
| chr16:70614936 | C | C,T | R->W | 13.27 | 0.00 | DAMAGING (novel) |
| chr16:70615586 | A | A,G | D->G | 5.16 | 0.05 | TOLERATED (novel) |

The "score contribution" column shows the magnitude of impact of each observed variant in *DHODH* to its final score. (Red) Most severe; (green) least severe. For comparison, SIFT values are also shown. Note that SIFT judges two of the known disease-causing alleles as tolerated and is unable to score the noncoding SNV. The target file contains six unrelated individuals with the compound heterozygous variants described in Table 3. The background file consisted of 189 genomes of mixed ethnicity from the 1000 Genomes Project combined with nine additional genomes of mixed ethnicity and sequencing platforms drawn from the 10Gen set (Reese et al. 2010). Data were generated using VAAST's fully penetrant monogenic recessive model and masking. Causative allele incidence was set to 0.00035.

Table 5. Impact of cohort size on VAAST's ability to identify a rare recessive disease

| Target genome(s) | Genome-wide | | | Rank | SLC26A3 | |
|----------------------------|-------------------|------------------|--------------|------|------------------------|-----------------------|
| | Significant genes | | | | P-value | |
| | Genes scored | Non-LD-corrected | LD-corrected | | Non-LD-corrected | LD-corrected |
| 1 Homozygote | 127 | 69 | 0 | 21 | 1.22×10^{-5} | 5.26×10^{-3} |
| Union 2 homozygotes | 7 | 7 | 0 | 3 | 4.74×10^{-10} | 5.51×10^{-5} |
| Intersection 2 homozygotes | 3 | 3 | 0 | 1 | 7.47×10^{-10} | 5.51×10^{-5} |
| Union 3 homozygotes | 2 | 2 | 2 | 1 | 2.83×10^{-13} | 8.61×10^{-7} |
| Intersection 3 homozygotes | 1 | 1 | 1 | 1 | 1.29×10^{-13} | 8.61×10^{-7} |

The background file consists of 189 genomes of mixed ethnicity from the 1000 Genomes Project combined with nine additional genomes of mixed ethnicity and sequencing platforms drawn from the 10Gen set (Reese et al. 2010). (Targets) The first homozygote affected is the single CCD affected exome reported in Choi et al. (2009); (remaining target genomes) unrelated exomes from re-sequenced individuals from Denmark reported in Li et al. (2010) with the causative allele added. Data were generated on either the union or intersection of affecteds using VAAST's fully penetrant monogenic recessive model. Causative allele incidence was set to 0.013; masking was also used. Scoring was evaluated by non-LD and LD-corrected permutation. (Genes scored) The number of genes in the genome receiving a score >0.

using *LPL*, a gene implicated in hypertriglyceridemia (HTG) (Johansen et al. 2010). This analysis uses a data set of 438 re-sequenced subjects (Johansen et al. 2010). For the *LPL* gene, only rare variants (MAF < 5%) were available; therefore, this analysis tests VAAST's ability to detect disease genes for common diseases in which only rare variants contribute to disease risk. To control for Type I error in this analysis, we applied a Bonferroni correction, with the number of tests approximately equal to the number of genes that would be included in a genome-wide analysis ($\alpha = 0.05/21,000 = 2.4 \times 10^{-6}$).

VAAST rapidly obtains good statistical power even with modest sample sizes; its estimated power is 89% for *NOD2* using as few as 150 individuals ($\alpha = 2.4 \times 10^{-6}$). By comparison, the power of GWAS is <4% at the same sample size. Notably, for *NOD2*, nearly 100% power is obtained with VAAST when a GWAS would still have <10% power. Also shown is VAAST's power as a function of sample size without the use of amino acid substitution data. The red and blue lines in Figure 6A show the power curves for VAAST using OMIM and BLOSUM, respectively, for its AAS disease models. As can be seen, power is improved when AAS information is used.

In general, the *LPL* results mirror those of *NOD2*. Although VAAST obtained less power using the *LPL* data set compared to *NOD2*, this was true for every approach. Interestingly, for *NOD2*, BLOSUM attains higher power using smaller sample sizes compared to OMIM. The fact that the trend is reversed for *LPL*, however, suggests that the two AAS models are roughly equivalent. We also compared VAAST's performance to that of WSS (Madsen and Browning 2009), another aggregative prioritization method. VAAST achieves greater statistical power than WSS on both data sets, even when VAAST is run without use of AAS information.

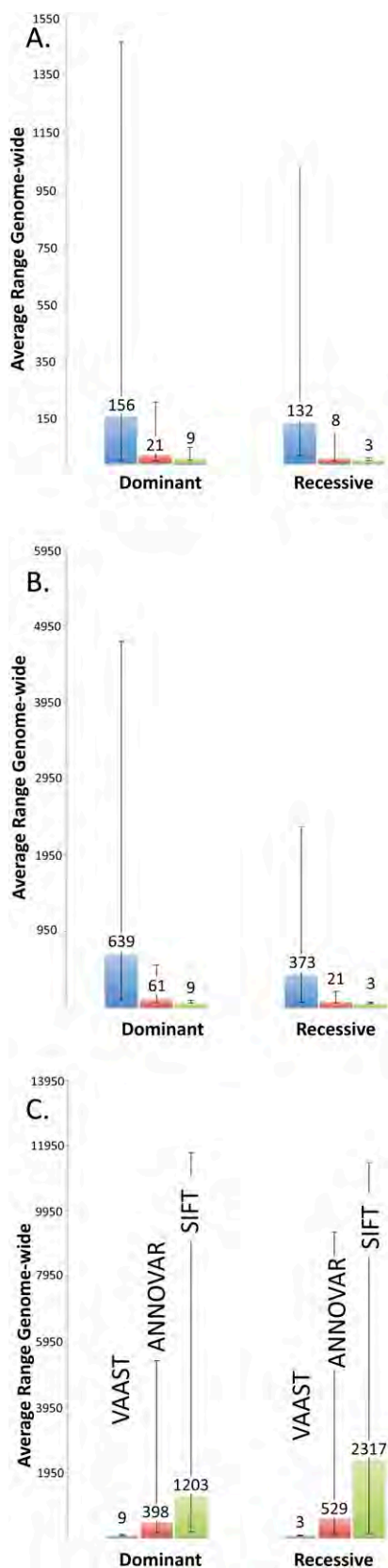
Discussion

VAAST uses a generalized feature-based prioritization approach, aggregating variants to achieve greater statistical search power. VAAST can score both coding and noncoding variants, evaluating the aggregative impact of both types of SNVs simultaneously. In this first study, we have focused on genes, but in principle, the tool can be used to search for disease-causing variants in other classes of features as well; for example, regulatory elements, sets of genes belonging to a particular genetic pathway, or genes belonging to a common functional category, e.g., transcription factors.

In contrast to GWAS approaches, which evaluate the statistical significance of frequency differences for individual variants in cases versus controls, VAAST evaluates the likelihood of observing the aggregate genotype of a feature given a background data set of control genomes. As our results demonstrate, this approach greatly improves statistical power, in part because it bypasses the need for large statistical corrections for multiple tests. In this sense, VAAST resembles several other methods that aggregate variants: CAST (Morgenthaler and Thilly 2007), CMC (Li and Leal 2008), WSS (Madsen and Browning 2009), and KBAC (Liu and Leal 2010). However, in contrast to these methods, VAAST also uses AAS information. Moreover, it uses a new approach to do so, one that allows it to score more SNVs than existing AAS methods such as SIFT (Kumar et al. 2009) and Polyphen (Sunyaev et al. 2001).

Much additional statistical power and accuracy are also gained from other components of the VAAST architecture, such as its ability to use pedigrees, phased data sets, and disease inheritance models. No existing AAS (Ng and Henikoff 2006) or aggregating method (Morgenthaler and Thilly 2007; Li and Leal 2008; Madsen and Browning 2009; Liu and Leal 2010) has these capabilities. The power of VAAST's pedigree approach is made clear in the quartet-based Miller syndrome analysis shown in Figure 4, where genome-wide only the two disease-causing genes are identified in a genome-wide screen of 19,249 nonsynonymous variants. Another important feature of VAAST is its ability to identify and mask variants in repetitive regions of the genome. As our results show, this provides a valuable method for mitigating platform-specific sequencing errors in situations in which it is cost-prohibitive to obtain a sufficiently large control set of genomes matched with regard to sequencing and variant calling pipeline. VAAST also differs in important ways from published heuristic search tools such as ANNOVAR (Wang et al. 2010). Unlike these tools, VAAST is not designed specifically to identify rare variants responsible for rare diseases. Instead, VAAST can search any collection of variants, regardless of their frequency distributions, to identify genes involved in both rare and common diseases.

Collectively, our results make clear the synergy that exists between these various components of the VAAST architecture. For example, they grant VAAST several unique features that distinguish it from commonly used AAS methods such as SIFT. Unlike AAS approaches, VAAST can score all variants, coding and non-coding, and in nonconserved regions of the genome. In addition, VAAST can obtain greater accuracy in judging which variants are deleterious. Comparison of the two Utah Miller syndrome exomes serves to highlight these differences. The two Miller syndrome exomes (Ng et al. 2010; Roach et al. 2010), for example, share 337 SNVs that are judged deleterious by SIFT; these 337 shared SNVs are distributed among 277 different genes. Thus, although AAS tools such as SIFT are useful for prioritizing the variants within a single known disease gene for follow-up studies, they are of limited use when carrying out genome-wide disease-gene searches, especially when the affected individuals are compound heterozygotes, as in the Miller syndrome examples.



In comparison to SIFT, VAAST scores 10% more nonsynonymous SNVs but identifies only nine candidate genes (Table 2), with the two disease-causing genes ranked fourth and fifth. When run in its pedigree mode, only the four disease-causing variants in the two disease genes are judged deleterious by VAAST genome-wide. The original analysis (Roach et al. 2010) of the family of four required 3 mo and identified eight potential disease-causing variants in four genes. An exome analysis required four affected individuals in three families to identify *DHODH* as the sole candidate for Miller syndrome (Ng et al. 2010). In contrast, using only the data from the family of four, VAAST identified the two disease genes in ~11 min using a 24-CPU compute server, and with perfect accuracy. Even when an additional 36,883 synonymous and noncoding regulatory variants are included in this genome-wide screen, only 23 candidate genes are identified, with *DHODH* still ranked fourth and *DNAH5* ranked first.

Our benchmark analyses using 100 different known diseases and 600 different known disease-causing alleles make it clear that our Miller syndrome and CCD analyses are representative results, and that VAAST is both a very accurate and a very reliable tool. VAAST consistently ranked the disease gene in the top three candidates genome-wide for recessive diseases and in the top nine gene candidates for dominant diseases. Equally important is reliability. VAAST has a much lower variance than either SIFT or ANNOVAR. In the recessive scenario, using three compound heterozygous individuals as a case cohort, for 95% of the VAAST runs, the disease-causing gene was ranked between second and 10th genome-wide; in comparison, ANNOVAR's ranks varied between 67 and 8762 on the same data sets, and SIFT's varied between 66 and 9107. Thus, VAAST is not only more accurate, it is also a more reliable tool. These same analyses also demonstrate that VAAST remains a reliable tool even when confronted with missing data due to phenomena such as missed variants, locus heterogeneity, and phenocopies in the case cohorts. Even when one-third of the cohort lacked disease-causing alleles at the locus, the average rank was still 61 for dominant diseases and 21 for recessive diseases (Fig. 5B).

VAAST can also be used to search for genes that cause common diseases and to estimate the impact of common alleles on gene function, something tools like ANNOVAR are not able to do. For example, when run over a published collection of 1454

Figure 5. Benchmark analyses using 100 different known disease genes. In each panel the y -axis denotes the average rank of the disease gene among 100 different disease genes. Heights of boxes are proportional to the mean rank, with the number above each box denoting the mean rank of the disease gene among all RefSeq annotated human genes. Error bars encompass the maximum and minimum observed ranks for 95% of the trials. (A) Average ranks for 100 different VAAST searches. (Left half of panel) The results for genome-wide searches for 100 different disease genes assuming dominance using a case cohort of two (blue box), four (red box), and six (green box) unrelated individuals. (Right half of panel) The results for genome-wide searches for 100 different recessive disease genes using a case cohort of 1 (blue box), 2 (red box), and 3 (green box). (B) Impact of missing data on VAAST performance. (Left and right half of panel) Results for dominant and recessive gene searches as in panel A, except in this panel the case cohorts contain differing percentages of individuals with no disease-causing variants in the disease gene. (Blue box) Two-thirds of the individuals lack a disease-causing allele; (red box) one-third lack a disease-causing allele; (green box) all members of the case cohort contain disease-causing alleles. (C) Comparison of VAAST performance to that of ANNOVAR and SIFT. (Left half of panel) The results for genome-wide searches using VAAST, ANNOVAR, and SIFT to search for 100 different dominant disease genes using a case cohort of six unrelated individuals. (Right half of panel) The results for genome-wide searches using VAAST, ANNOVAR, and SIFT to search for 100 different recessive disease genes using a case cohort of three unrelated individuals.

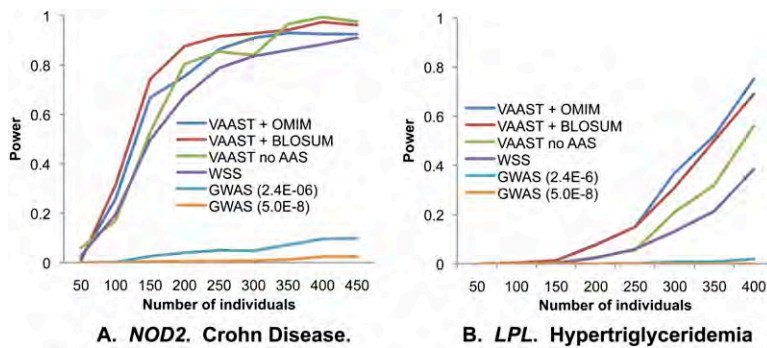


Figure 6. Statistical power as a function of number of target genomes for two common disease genes. (A) *NOD2*, using a data set containing rare and common nonsynonymous variants. (B) *LPL*, using a data set containing only rare nonsynonymous variants. For each data point, power is estimated from 500 bootstrapped resamples of the original data sets, with $\alpha = 2.4 \times 10^{-6}$ except where specified. y-axis: probability of identifying gene as implicated in disease in a genome-wide search; x-axis: number of cases. The number of controls is equal to the number of cases up to a maximum of 327 for *LPL* (original data set) and 163 for *NOD2* (original data set + 60 Europeans from 1000 Genomes). (VAAST + OMIM) VAAST using AAS data from OMIM as its disease model; (VAAST + BLOSUM) VAAST using BLOSUM62 as its disease model; (VAAST no AAS) VAAST running on allele frequencies alone; (WSS) weighted sum score of Madsen and Browning (2009); (GWAS) single variant GWAS analysis. *NOD2* and *LPL* data sets were taken from Lesage et al. (2002) and Johansen et al. (2010), respectively.

high-confidence disease-causing and predisposing SNVs from OMIM (Yandell et al. 2008), VAAST identifies all but 29 (2%) of these SNVs as damaging. ANNOVAR (Wang et al. 2010), in comparison, excludes 427 (29%) of these SNVs from further analysis because they are present in the 1000 Genomes Project data (The 1000 Genomes Project Consortium 2010), dbSNP130, or in segmentally duplicated regions. These results underscore the advantages of VAAST's probabilistic approach. VAAST can assay the impact of rare variants to identify rare diseases and both common and rare variants to identify the alleles responsible for common diseases, and it operates equally well on data sets (e.g., *NOD2*) wherein both rare and common variants are contributing to disease. Our common-disease analyses serve to illustrate these points. These results demonstrate that VAAST can achieve close to 100% statistical power on common-disease data sets, where a traditional GWAS test has almost no power. We also demonstrate that VAAST's own feature-based scoring significantly outperforms WSS (Madsen and Browning 2009), which, like all published aggregative scoring methods, does not use AAS information. These analyses also demonstrate another key feature of VAAST: **While the controls in the Crohn's disease data set were fully sequenced at *NOD2*, only a small subset of the cases was sequenced, and the rest were genotyped at sites that were polymorphic in the sample. VAAST does well with this mixed data set.** It is likely that VAAST would do even better using a data set of the same size consisting only of sequence data, as such a cohort would likely contain additional rare variants not detectable with chip-based technologies. Consistent with this hypothesis, VAAST also attains high statistical power compared to traditional GWAS methods on the *LPL* data set, which only contains alleles with a frequency of <5%. This demonstrates that VAAST can also identify common-disease genes even when they contain no common variants that contribute to disease risk.

These results suggest that VAAST will prove useful for re-analyses of existing GWAS and linkage studies. Targeted VAAST analyses combined with region-specific resequencing around GWAS hits will allow smaller Bonferroni corrections (Nicodemus et al. 2005) than the genome-wide analyses presented here, resulting in still greater statistical power, especially in light of VAAST's feature-based ap-

proach. The same is true for linkage studies. **In addition, because much of the power of VAAST is derived from rare variants and amino acid substitutions, the likelihood of false positives due to linkage disequilibrium with causal variants is low.** Thus, it is likely that VAAST will allow identification of disease genes and causative variants in GWAS data sets in which the relationships of hits to actual disease genes and the causative variants are unclear, and for linkage studies, where only broad spans of statistically significant linkage peaks have been detected to date.

VAAST is compatible with current genomic data standards. Given the size and complexity of personal genome data, this is not a trivial hurdle for software applications. VAAST uses GFF3 (<http://www.sequenceontology.org/resources/gff3.html>), and GVF (Reese et al. 2010) and VCF (<http://www.1000genomes.org/wiki/Analysis/vcf4.0>), standardized file formats for genome annotations and personal

genomes data. The size and heterogeneity of the data sets used in our analyses make clear VAAST's ability to mine hundreds of genomes and their annotations at a time. We also point out that VAAST has a modular software architecture that makes it easy to add additional scoring methods. Indeed, we have already done so for WSS (Madsen and Browning 2009). This is an important point, as aggregative scoring methods are a rapidly developing area of personal genomics (Morgenthaler and Thilly 2007; Li and Leal 2008; Madsen and Browning 2009; Liu and Leal 2010). VAAST thus provides an easy means to incorporate and compare new scoring methods, lending them its many other functionalities.

Although there exist other tools with some of its features, to our knowledge, VAAST is the first generalized, probabilistic *ab initio* tool for identifying both rare and common disease-causing variants using personal exomes and genomes. VAAST is a practical, portable, self-contained piece of software that substantially improves on existing methods with regard to statistical power, flexibility, and scope of use. It is resistant to no calls, automated, and fast; works across all variant frequencies; and deals with platform-specific noise.

Methods

Inputs and outputs

The VAAST search procedure is shown in Figure 7. VAAST operates using two input files: a background and a target file. The background and target files contain the variants observed in control and case genomes, respectively. Importantly, the same background file can be used again and again, obviating the need—and expense—of producing a new set of control data for each analysis. Background files prepared from whole-genome data can be used for whole-genome analyses, exome analyses and for individual gene analyses. **These files can be in either VCF (<http://www.1000genomes.org/wiki/Analysis/vcf4.0>) or GVF (Reese et al. 2010) format.** VAAST also comes with a series of premade and annotated background condenser files for the 1000 genomes data (The 1000 Genomes Project Consortium 2010) and the 10Gen data set (Reese et al. 2010). Also needed is

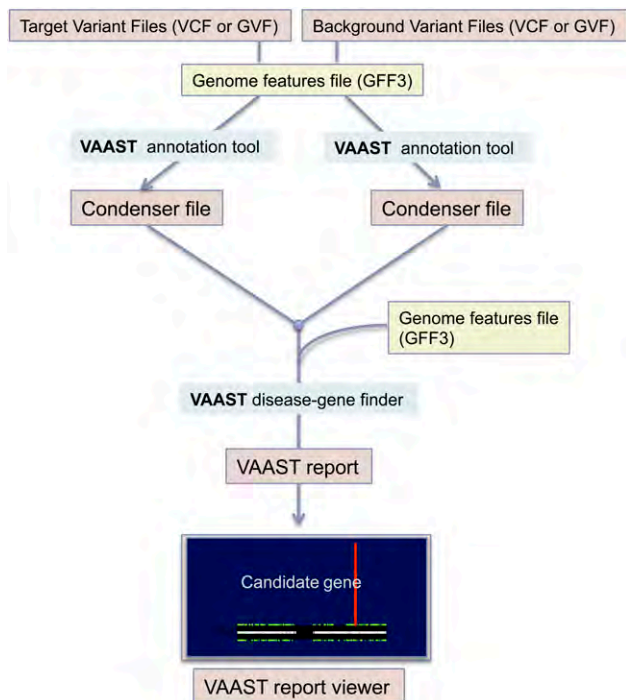


Figure 7. VAAST search procedure. One or more variant files (in VCF or GVF format) are first annotated using the VAAST annotation tool and a GFF3 file of genome annotations. Multiple target and background variant files are then combined by the VAAST annotation tool into a single condenser file; these two files, one for the background and one for the target genomes, together with a GFF3 file containing the genomic features to be searched are then passed to VAAST. VAAST outputs a simple text file, which can also be viewed in the VAAST viewer.

a third file in GFF3 (<http://www.sequenceontology.org/resources/gff3.html>) containing genome features to be searched.

Basic CLR method

The composite likelihood ratio (CLR) test is designed to evaluate whether a gene or other genomic feature contributes to disease risk. We first calculate the likelihood of the null and alternative models assuming independence between nucleotide sites and then evaluate the significance of the likelihood ratio by permutation to control for LD. The basic method is a nested CLR test that depends only on differences in allele frequencies between affected and unaffected individuals. In a manner similar to the CMC method (Li and Leal 2008), we collapse sites with rare minor alleles into one or more categories, but we count the total number of minor allele copies among all affected and unaffected individuals rather than just the presence or absence of minor alleles within an individual. For our analyses, we set the collapsing threshold at fewer than five copies of the minor allele among all affected individuals, but this parameter is adjustable. Let k equal the number of uncollapsed variant sites among n_i^U unaffected and n_i^A affected individuals, with n_i equal to $n_i^U + n_i^A$. Let $l_{k+1} \dots l_{k+m}$ equal the number of collapsed variant sites within m collapsing categories labeled $k+1$ to m , and let $l_1 \dots l_k$ equal 1. Let X_i, X_i^U , and X_i^A equal the number of copies of the minor allele(s) at variant site i or collapsing category i among all individuals, unaffected individuals, and affected individuals, respectively. Then the log-likelihood ratio is equal to:

$$\lambda = \ln \left(\frac{L_{Null}}{L_{Alt}} \right) = \sum_{i=1}^{k+m} \ln \left[\frac{\binom{n_i}{X_i} (p_i)^{X_i} (1-p_i)^{2l_i n_i - X_i}}{\binom{n_i^U}{X_i^U} (p_i^U)^{X_i^U} (1-p_i^U)^{2l_i n_i^U - X_i^U} \binom{n_i^A}{X_i^A} (p_i^A)^{X_i^A} (1-p_i^A)^{2l_i n_i^A - X_i^A}} \right] \quad (1)$$

where p_i, p_i^U , and p_i^A equal the maximum-likelihood estimates for the frequency of minor allele(s) at variant site i or collapsing category i among all individuals, unaffected individuals, and affected individuals, respectively. When no constraints are placed on the frequency of disease-causing variants, the maximum-likelihood estimates are equal to the observed frequencies of the minor allele(s). Assuming that variant sites are unlinked, -2λ approximately follows a χ^2 distribution with $k+m$ degrees of freedom. We report the non-LD-corrected χ^2 P -value as the VAAST score to provide a statistic for rapid prioritization of disease-gene candidates. To evaluate the statistical significance of a genomic feature, we perform a randomization test by permuting the affected/unaffected status of each individual (or each individual chromosome, when phased data are available). Because the degrees of freedom can vary between iterations of the permutation test, we use the χ^2 P -value as the test statistic for the randomization test.

Extensions to the basic CLR method

In the basic CLR method, the null model is fully nested within the alternative model. Extensions to this method result in models that are no longer nested. Because the χ^2 approximation is only appropriate for likelihood ratio tests of nested models, we apply Vuong's closeness test in extended CLR tests using the Akaike Information Criterion correction factor. Thus, the test statistic used in the permutation tests for these methods is $-2\lambda - 2(k+m)$. To efficiently calculate the non-LD-corrected P -value for non-nested models, we use an importance sampling technique in a randomization test that assumes independence between sites by permuting the affected/unaffected status of each allele at each site. To evaluate the LD-corrected statistical significance of genomic features for these models, we permute the affected/unaffected status of each individual (or each individual chromosome).

For rare diseases, we constrain the allele frequency of putative disease-causing alleles in the population background such that p_i^U cannot exceed a specified threshold, t , based on available information about the penetrance, inheritance mode, and prevalence of the disease. With this constraint, the maximum-likelihood estimate for p_i^U is equal to the minimum of t and $X_i/l_i n_i$.

The framework can incorporate various categories of indels, splice-site variants, synonymous variants, and noncoding variants. Methods incorporating amino acid severity and constraints on allele frequency can result in situations in which the alternative model is less likely than the null model for a given variant. In these situations, we exclude the variant from the likelihood calculation, accounting for the bias introduced from this exclusion in the permutation test. For variants sufficiently rare to meet the collapsing criteria, we exclude the variant from the collapsing category if the alternative model is less likely than the null model prior to variant collapse.

Severity of amino acid changes

To incorporate information about the potential severity of amino acid changes, we include one additional parameter in the null and alternative models for each variant site or collapsing category. The

parameter h_i in the null model is the likelihood that the amino acid change does not contribute to disease risk. We estimate h_i by setting it equal to the proportion of this type of amino acid change in the population background. **The parameter a_i in the alternative model is the likelihood that the amino acid change contributes to disease risk. We estimate a_i by setting it equal to the proportion of this type of amino acid change among all disease-causing mutations in OMIM** (Yandell et al. 2008). Incorporating information about amino acid severity, λ is equal to:

$$\lambda = \ln \left(\frac{L_{Null}}{L_{Alt}} \right) = \sum_{i=1}^{k+m} \ln \left[\frac{h_i (\hat{p}_i)^{X_i} (1 - \hat{p}_i)^{2L_i m - X_i}}{a_i (\hat{p}_i^U)^{X_i^U} (1 - \hat{p}_i^U)^{2L_i m^U - X_i^U} (\hat{p}_i^A)^{X_i^A} (1 - \hat{p}_i^A)^{2L_i m^A - X_i^A}} \right] \quad (2)$$

To include the severity of amino acid changes for collapsed rare variants, we create m collapsing categories that are divided according to the severity of potential amino acid changes. To create the collapsing categories, we first rank all possible amino acid changes according to their severity. We then assign an equal number of potential changes to each category, with the first category receiving the least severe changes and each subsequent category receiving progressively more severe changes. Each rare variant is then included in the category with its corresponding amino acid change (Tavtigian 2009). For each collapsing category i , we set the parameters h_i and a_i equal to their average values among all variants present in the category. We first calculate the likelihood of the null and alternative models assuming independence between nucleotide sites and then evaluate the significance of the likelihood ratio by permutation to control for LD.

Scoring noncoding variants

The VAAST CLR framework can also score noncoding variants and synonymous variants within coding regions. Because ascertainment bias in OMIM can cause a bias against such variants, we took an evolutionary approach to estimate the relative impacts of noncoding and synonymous variants using the vertebrate-to-human genome multiple alignments downloaded from the UCSC Genome Browser (<http://hgdownload.cse.ucsc.edu/goldenPath/hg18/multiz44way/maf/>). For each codon in the human genome, we calculated the frequency in which it aligns to other codons in primate genomes (wherever an open reading frame [ORF] in the corresponding genomes is available). Then for every codon alignment pair involving one or fewer nucleotide changes, we calculated its Normalized Mutational Proportion (NMP), which is defined as the proportion of occurrences of each such codon pair among all codon pairs with the identical human codon and with one or fewer nucleotide changes. For example, suppose the human codon GCC aligned to codons in primate genomes with the following frequencies: GCC → GCC: 1000 times; GCC → GCT: 200 times; GCC → GCG: 250 times; GCC → GGG: 50 times. The NMP value of GCC → GCT would be 0.134 [i.e., 200/(1000 + 200 + 250)]. For every codon pair that involves a nonsynonymous change, we then calculated its severity parameter from the OMIM database and 180 healthy genomes from the 1000 Genomes Project (a_i/h_i in Eq. 2). Linear regression analysis indicates that $\log(a_i/h_i)$ is significantly correlated with $\log(\text{NMP})$ ($R^2 = 0.23$, $p < 0.001$). This model allows us to estimate the severity parameter of synonymous variants (again by linear regression), which by this approach is 0.01 (100 times less severe than a typical nonsynonymous variant). We used a similar approach to derive an equivalent value for SNVs in noncoding regions. To do so, we again used the primate alignments from UCSC, but here we restricted our analysis to primate

clustered DNase hypersensitive sites and transcription factor binding regions as defined by ENCODE regulation tracks, calculating NMP for every conserved trinucleotide. The resulting severity parameter for these regions of the genome is 0.03.

Inheritance and penetrance patterns

VAAST includes several options to aid in the identification of disease-causing genes matching specific inheritance and penetrance patterns. These models enforce a particular disease model within a single gene or other genomic feature. Because the disease models introduce interdependence between sites, VAAST does not provide a site-based non-LD-corrected P -value for these models.

For recessive diseases, VAAST includes three models: recessive, recessive with complete penetrance, and recessive with no locus heterogeneity. In the basic recessive model, the likelihood calculation is constrained such that no more than two minor alleles in each feature of each affected individual will be scored. The two alleles that receive a score are the alleles that maximize the likelihood of the alternative model. The complete penetrance model assumes that all of the individuals in the control data set are unaffected. As the genotypes of each affected individual are evaluated within a genomic feature, if any individual in the control data set has a genotype exactly matching an affected individual, the affected individual will be excluded from the likelihood calculation for that genomic feature. This process will frequently remove all affected individuals from the calculation, resulting in a genomic feature that receives no score. In the recessive with no locus heterogeneity model, genomic features are only scored if all affected individuals possess two or more minor alleles at sites where the alternative (disease) model is more likely than the null (healthy) model. The two alleles can be present at different nucleotide sites in each affected individual (i.e., allelic heterogeneity is permitted), but locus heterogeneity is excluded. The models can be combined, for example, in the case of a completely penetrant disease with no locus heterogeneity.

The three dominant disease models parallel the recessive models: dominant, dominant with complete penetrance, and dominant with no locus heterogeneity. For the basic dominant model, only one minor allele in each feature of each affected individual will be scored (the allele that maximizes the likelihood of the alternative model). **For the complete penetrance dominant model, alleles will only be scored if they are absent among all individuals in the control data set. For the dominant with no locus heterogeneity model, genomic features are only scored if all affected individuals possess at least one minor allele at variant sites where the alternative model is more likely than the null model.**

Protective alleles

For non-nested models, the default behavior is to only score variants in which the minor allele is at higher frequency in cases than in controls, under the assumption that the disease-causing alleles are relatively rare. This assumption is problematic if protective alleles are also contributing to the difference between cases and controls. By enabling the “protective” option, VAAST will also score variants in which the minor allele frequency is higher in controls than in cases. This option also adds one additional collapsing category for rare protective alleles. Because we have no available AAS model for protective alleles, we set h_i and a_i equal to 1 for these variants.

Variant masking

The variant-masking option allows the user to exclude a list of nucleotide sites from the likelihood calculations based on information obtained prior to the genome analysis. The masking

files used in these analyses exclude sites where short reads would map to more than one position in the reference genome. This procedure mitigates the effects introduced by cross-platform biases by excluding sites that are likely to produce spurious variant calls due to improper alignment of short reads to the reference sequence. The three masking schemes we used were (1) 60-bp single-end reads, (2) 35-bp single-end reads, and (3) 35-bp paired-end reads separated by 400 bp. These three masking files are included with the VAAST distribution, although VAAST can mask any user-specified list of sites. Because variant masking depends only on information provided prior to the genome analysis, it is compatible with both nested and non-nested models CLR models.

Trio option

By providing the genomes of the parents of one or more affected individuals, VAAST can identify and exclude Mendelian inheritance errors for variants that are present in the affected individual but absent in both parents. Although this procedure will exclude both de novo mutations and sequencing errors, for genomes with an error rate of ~ 1 in 100,000, $\sim 99.9\%$ of all Mendelian inheritance errors are genotyping errors (Roach et al. 2010). This option is compatible with both nested and non-nested models.

Minor reference alleles

Most publicly available human genome and exome sequences do not distinguish between no calls and reference alleles at any particular nucleotide site. For this reason, VAAST excludes reference alleles with frequencies of $<50\%$ from the likelihood calculation by default. This exclusion can be overridden with a command-line parameter.

VAAST options, including command lines used to generate each table and figure, are provided in the Supplemental Material.

Benchmark analyses

We assayed the ability of VAAST, SIFT, and ANNOVAR to identify mutated genes and their disease-causing variants in genome-wide searches. To do so, we randomly selected a set of 100 genes, each having at least six SNVs that are annotated as deleterious by OMIM. For each run, the OMIM variants from one of the 100 genes were inserted into the genomes of healthy individuals sampled from the Complete Genomics Diversity Panel (<http://www.completegenomics.com/sequence-data/download-data/>). For the partial representation panel (Fig. 5B), we inserted the OMIM variants into only a partial set of the case genomes. For example, in the panel of 66% partial representation and dominant model, we inserted four OMIM variants into four of the six case genomes for each gene, so that 66% of the case genomes have deleterious variants; for 66% representation under the recessive model, we inserted four OMIM variants into two of the three case genomes.

We ran VAAST using 443 background genomes (including 180 genomes from the 1000 Genomes Project pilot phase, 63 Complete Genomics Diversity panel genomes, nine published genomes, and 191 Danish exomes) and with the inheritance model option (-iht). We ran SIFT using its web service (http://sift.jcvi.org/www/SIFT_chr_coords_submit.html, as of 5/3/2011). For ANNOVAR, we used version: 2011-02-11 00:07:48 with the 1000 Genomes Project 2010 July release as the variant database. We used its automatic annotation pipeline (auto_annoar.pl) and default parameters for annotation, setting its -maf option to the upper 99% confidence interval of the expected minor allele frequency (MAF), such that the combined MAF for inserted alleles did not exceed 5%. The dbSNP database was not used in this analysis because ANNOVAR's dbSNP130 database does not provide MAF information, and

a portion of the disease-causing OMIM alleles are collected by dbSNP130. We found that setting -maf and excluding dbSNP130 for this analysis greatly improved the accuracy of ANNOVAR in comparison to its default parameters (data not shown); thus we used these more favorable parameters for our comparisons.

To compare the performance of the three algorithms with a sample size of six under a dominant model, for each of the 100 genes, we inserted the six different OMIM variants located in this gene into six different healthy genomes, making all of them heterozygous for a different disease-causing SNV at that locus. Under the recessive model, with a sample size of two, for example, we inserted four different OMIM variants located in each gene into two healthy genomes, so that each case genome carries two different OMIM variants in this gene, i.e., the individuals are compound heterozygotes.

Scalability

VAAST computes scale linearly with the number of features (genes) being evaluated and the number of variants in the targets. The maximum number of permutations needed is bounded by $O(n^k)$, where n equals the number of background and target genomes, and k equals the number of target genomes. VAAST is a multi-threaded, parallelized application designed to scale to cohorts of thousands of genomes.

Data access

VAAST is available for download at <http://www.yandell-lab.org> with an academic user license.

Acknowledgments

This work was supported by NIH SBIR grant 1R4HG003667 to M.G.R. and M.Y., NIH ARRA GO grant 1RC2HG005619 to M.Y. and M.G.R., NIH GM59290 to L.B.J., and NIH K99HG005846 to J.X., all administered by the National Human Genome Research Institute (NHGRI). C.H. was supported by 1RC2HG005619, the University of Luxembourg-Institute for Systems Biology Program, and NIH 1T32HL105321-01. An allocation of computer time from the Center for High Performance Computing at the University of Utah is gratefully acknowledged.

Authors' contributions: M.Y. and M.G.R. conceived of the project. M.Y. oversaw and coordinated the research. M.Y. designed the software architecture. M.Y., H.H., C.H., and B.M. wrote the software package. C.H., H.H., and M.Y. contributed to the statistical method development. M.S., H.H., C.H., and J.X. performed experiments. M.Y., C.H., L.B.J., and M.G.R. wrote the paper.

References

- The 1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. *Nature* **467**: 1061–1073.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403–410.
- Altshuler D, Daly MJ, Lander ES. 2008. Genetic mapping in human disease. *Science* **322**: 881–888.
- Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* **268**: 78–94.
- Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Özen S, Sanjad S, et al. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci* **106**: 19096–19101.
- Henikoff S, Henikoff J. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci* **89**: 10915–10919.
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. 2009. Potential etiologic and functional implications of

- genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci* **106**: 9362–9367.
- Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, et al. 2010. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* **42**: 684–687.
- Korf I, Bedell J, Yandell M. 2003. *BLAST: An essential guide to the Basic Local Alignment Search Tool*. O'Reilly, Beijing.
- Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**: 1073–1081.
- Lausch E, Hermanns P, Farin HF, Alanay Y, Unger S, Nikkel S, Steinwender C, Scherer G, Spranger J, Zabel B, et al. 2008. TBX15 mutations cause craniofacial dysmorphism, hypoplasia of scapula and pelvis, and short stature in Cousin syndrome. *Am J Hum Genet* **83**: 649–655.
- Lesage S, Zouali H, Cézard JP, Colombel JF, Belaiche J, Almer S, Tysk C, O'Morain C, Gassull M, Binder V, et al. 2002. CARD15/NOD2 mutational analysis and genotype–phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* **70**: 845–857.
- Li B, Leal SM. 2008. Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. *Am J Hum Genet* **83**: 311–321.
- Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersen G, Cao H, Korneliusen T, et al. 2010. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet* **42**: 969–972.
- Liu DJ, Leal SM. 2010. A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. *PLoS Genet* **6**: e1001156. doi: 10.1371/journal.pgen.1001156.
- Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, et al. 2010. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* **362**: 1181–1191.
- Madsen BE, Browning SR. 2009. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet* **5**: e1000384. doi: 10.1371/journal.pgen.1000384.
- Manolio TA. 2009. Cohort studies and the genetics of complex disease. *Nat Genet* **41**: 5–6.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461**: 747–753.
- Morgenthaler S, Thilly WG. 2007. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res* **615**: 28–56.
- Ng PC, Henikoff S. 2006. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* **7**: 61–80.
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, et al. 2010. Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet* **42**: 30–35.
- Nicodemus KK, Liu W, Chase GA, Tsai YY, Fallin MD. 2005. Comparison of type I error for multiple test corrections in large single-nucleotide polymorphism studies using principal components versus haplotype blocking algorithms. *BMC Genet (Suppl 1)* **6**: S78. doi: 10.1186/1471-2156-6-S1-S78.
- Pelak K, Shianna KV, Ge D, Maia JM, Zhu M, Smith JP, Cirulli ET, Fellay J, Dickson SP, Gumbs CE, et al. 2010. The characterization of twenty sequenced human genomes. *PLoS Genet* **6**: e1001111. doi: 10.1371/journal.pgen.1001111.
- Reese MG, Kulp D, Tammanna H, Haussler D. 2000. Gene–gene finding in *Drosophila melanogaster*. *Genome Res* **10**: 529–538.
- Reese MG, Moore B, Batchelor C, Salas F, Cunningham F, Marth GT, Stein L, Flicek P, Yandell M, Eilbeck K. 2010. A standard variation file format for human genome sequences. *Genome Biol* **11**: R88. doi: 10.1186/gb-2010-11-8-r88.
- Roach J, Glusman G, Smit A, Huff C, Hubley R, Shannon P, Rowen L, Pant K, Goodman N, Bamshad M, et al. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**: 636–639.
- Sunyaev S, Ramensky V, Koch I, Lathe W III, Kondrashov AS, Bork P. 2001. Prediction of deleterious human alleles. *Hum Mol Genet* **10**: 591–597.
- Tavtigian SV, Oefner PJ, Babikyan D, Hartmann A, Healey S, Le Calvez-Kelm F, Lesueur F, Byrnes GB, Chuang SC, Forey N, et al. 2009. Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *Am J Hum Genet* **85**: 427–446.
- Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**: e164. doi: 10.1093/nar/gkq603.
- Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**: 661–678.
- Yandell M, Moore B, Salas F, Mungall C, MacBride A, White C, Reese MG. 2008. Genome-wide analysis of human disease alleles reveals that their locations are correlated in paralogous proteins. *PLoS Comput Biol* **4**: e1000218. doi: 10.1371/journal.pcbi.1000218.

Received March 9, 2011; accepted in revised form June 8, 2011.

Using VAAST to Identify an X-Linked Disorder Resulting in Lethality in Male Infants Due to N-Terminal Acetyltransferase Deficiency

Alan F. Rope,¹ Kai Wang,^{2,19} Rune Evjenth,³ Jinchuan Xing,⁴ Jennifer J. Johnston,⁵ Jeffrey J. Swensen,^{6,7} W. Evan Johnson,⁸ Barry Moore,⁴ Chad D. Huff,⁴ Lynne M. Bird,⁹ John C. Carey,¹ John M. Opitz,^{1,4,6,10,11} Cathy A. Stevens,¹² Tao Jiang,^{13,14} Christa Schank,⁸ Heidi Deborah Fain,¹⁵ Reid Robison,¹⁵ Brian Dalley,¹⁶ Steven Chin,⁶ Sarah T. South,^{1,7} Theodore J. Pysher,⁶ Lynn B. Jorde,⁴ Hakon Hakonarson,² Johan R. Lillehaug,³ Leslie G. Biesecker,⁵ Mark Yandell,⁴ Thomas Arnesen,^{3,17} and Gholson J. Lyon^{15,18,20,*}

We have identified two families with a previously undescribed lethal X-linked disorder of infancy; the disorder comprises a distinct combination of an aged appearance, craniofacial anomalies, hypotonia, global developmental delays, cryptorchidism, and cardiac arrhythmias. Using X chromosome exon sequencing and a recently developed probabilistic algorithm aimed at discovering disease-causing variants, we identified in one family a c.109T>C (p.Ser37Pro) variant in *NAA10*, a gene encoding the catalytic subunit of the major human N-terminal acetyltransferase (NAT). A parallel effort on a second unrelated family converged on the same variant. The absence of this variant in controls, the amino acid conservation of this region of the protein, the predicted disruptive change, and the co-occurrence in two unrelated families with the same rare disorder suggest that this is the pathogenic mutation. We confirmed this by demonstrating a significantly impaired biochemical activity of the mutant hNaa10p, and from this we conclude that a reduction in acetylation by hNaa10p causes this disease. Here we provide evidence of a human genetic disorder resulting from direct impairment of N-terminal acetylation, one of the most common protein modifications in humans.

Introduction

Researchers have used exon capture and high-throughput sequencing (exome sequencing) to identify the molecular etiology of several Mendelian disorders.^{1–5} It has been used in studies of a multigenerational pedigree⁶ and a de novo disorder,¹ and it has been applied in molecular diagnostics.⁷ Most of these efforts have focused on characterizing previously described and recognizable genetic syndromes, such as Kabuki (MIM 147920),¹ Miller (MIM 263750),^{2,8} and TARP (MIM 311900)⁹ syndromes. These efforts to search for the molecular etiology of known syndromes benefited from years of clinical evaluations, allowing researchers to combine unrelated individuals in the same cohort for analysis.

In contrast, exon capture and sequencing can help to identify previously unrecognized syndromes. We have characterized such a syndrome, in which the afflicted boys have an aged appearance, craniofacial anomalies,

hypotonia, global developmental delays, cryptorchidism, and cardiac arrhythmias. To determine the genetic basis of this syndrome, we used X chromosome exon capture and sequencing and a recently developed probabilistic algorithm aimed at discovering disease-causing variants. We also demonstrate that this phenotype results from a decrease in the function of an enzyme involved in N-terminal acetylation of proteins.

Subjects and Methods

We describe two parallel genetic research efforts that converged on the same gene variant. The two groups working on this project became aware of each other's work after the genetic analyses had been completed, and therefore the methodologies used varied somewhat. When these methods differed, they are described separately.

The sample collection and analyses of families 1 and 2 were approved by the institutional review boards at the University of Utah and the National Human Genome Research Institute,

¹Department of Pediatrics (Medical Genetics), University of Utah School of Medicine, Salt Lake City, UT 84112, USA; ²Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; ³Department of Molecular Biology, University of Bergen, N-5020 Bergen, Norway; ⁴Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA; ⁵Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health (NIH), Bethesda, MD 20892, USA; ⁶Department of Pathology, University of Utah, Salt Lake City, UT 84112, USA; ⁷ARUP Laboratories, Salt Lake City, UT 84112, USA; ⁸Department of Statistics, Brigham Young University, Provo, UT 84602, USA; ⁹Rady Children's Hospital and University of California, San Diego, Department of Pediatrics, San Diego, CA 92123, USA; ¹⁰Department of Obstetrics and Gynecology, University of Utah, Salt Lake City, UT 84112, USA; ¹¹Department of Neurology, University of Utah, Salt Lake City, UT 84112, USA; ¹²Department of Pediatrics, University of Tennessee College of Medicine, Chattanooga, TN 38163, USA; ¹³BGI-Shenzhen, Shenzhen 518083, China; ¹⁴Genome Research Institute, Shenzhen University Medical School, Shenzhen 518060, China; ¹⁵Department of Psychiatry, University of Utah, Salt Lake City, UT 84112, USA; ¹⁶Huntsman Cancer Institute, Salt Lake City, UT 84112, USA; ¹⁷Department of Surgery, Haukeland University Hospital, N-5021 Bergen, Norway; ¹⁸New York University Child Study Center, New York, NY 10016, USA

¹⁹Present Address: Zilkha Neurogenetic Institute, Department of Psychiatry and Preventive Medicine, University of Southern California, Los Angeles, CA 90089, USA

²⁰Present Address: Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

*Correspondence: lyong1@email.chop.edu

DOI 10.1016/j.ajhg.2011.05.017. ©2011 by The American Society of Human Genetics. All rights reserved.

respectively. Written informed consent was obtained from the parents of affected children for collected blood samples and extracting DNA from stored samples. Blood samples were collected and genomic DNA extracted with alkaline lysis and ethanol precipitation (Gentra Puregene, QIAGEN, USA). Two DNA samples from family 1 were extracted from stored formalin-fixed-paraffin-embedded (FFPE) tissues from the prior autopsies of two of the deceased boys. Slices from FFPE tissue blocks were digested overnight in proteinase K and then boiled for 10 min to inactivate the enzyme. The crude lysates were diluted 1:10 for PCR.

Genomic Microarrays and Copy-Number Variation Analysis

Copy-number variation in individual III-4 in family 1 was evaluated with two oligonucleotide-based genomic microarray platforms, the U-array Cyto6000 (manufactured by Agilent Technologies, Santa Clara, CA) and the SignatureChip Oligo Solution microarray (manufactured by Roche NimbleGen). Both platforms use a custom design with approximately 10 kb coverage for many clinically relevant syndromes and telomere and pericentromeric regions; the genome-wide coverage is 75 kb for the U-array and 35 kb for the SignatureChipOS. The U-array data were analyzed with DNA Analytics 4.0.76 software from Agilent Technologies with the ADM-1 algorithm. The SignatureChipOS data were analyzed with Genoglyphix software (Signature Genomics, Spokane, WA).

X-Chromosome Exon Capture and Sequencing

Exon capture for family 1 was carried out with a commercially available Agilent in-solution method (SureSelect Human X chromosome kit, Agilent) as per the manufacturer guidelines with minor modifications. Randomly fragmenting the pure and high-molecular-weight genomic DNA samples with a Covaris S series Adaptive Focused Acoustics machine resulted in DNA fragments with a base pair peak of 150 to 200 bps. Adaptors were then ligated to both ends of the resulting fragments. The adaptor-ligated templates were purified by the Agencourt AMPure SPRI beads. Adaptor-ligated products were amplified by PCR with Phusion polymerase (six cycles) and subsequently purified with a QIAGEN QIAquick PCR purification kit. The quality and size distribution of the amplified product was assessed on an Agilent Bioanalyzer DNA 1000 chip. Amplified library (500 ng) with a size range of 200–400 bp was hybridized to SureSelect Biotinylated RNA Library (BAITS) for enrichment. Hybridized fragments were bound to the streptavidin beads whereas nonhybridized fragments were washed out after a 24 hr hybridization. The captured library was enriched by amplification by PCR (12 cycles), and the amplified product was subsequently qualified on an Agilent High Sensitivity DNA Bioanalyzer chip to verify the size range and quantity of the enriched library. Each captured library was sequenced with 76 bp single-end reads on one lane each of the Illumina GAIIX platform. Raw image files were processed by Illumina Pipeline v1.6 for base-calling with default parameters. All coordinates are based on the human genome build hg18. The pseudoautosomal regions were defined based on annotations within the UCSC genome browser. Specifically, these are the regions chrX:1–2709520, chrX:154584238–154913754, chrY:1–2709520, and chrY:57443438–57772954. For family 2, solution hybridization selection of the X exome (Sure Select, Agilent, Santa Clara, CA) was used to produce a paired-end sequencing library (Illumina, San Diego, CA) as previously described.⁹ Paired-end

75 bp reads were generated from the target-selected DNA library in one lane of an Illumina GAIIX. Coding changes were predicted with custom-designed software.⁹

Sanger Sequence-Confirmation Analysis

Amplification by PCR and Sanger sequencing were performed as described so that mutations and the cosegregation could be confirmed.¹⁰ Mutation numbering was performed according to Human Gene Variation Society nomenclature with reference NM_003491.2.

Sequence Alignment

For family 1, sequence reads were converted from Illumina fastq format to fastq files that conform to the Sanger specification for base-quality encoding with perl scripts. Burroughs Wheeler alignment (BWA)¹¹ version 0.5.8 was used to align the sequencing reads, and the default parameters were used for fragment reads, to the human genome sequence build 36 downloaded from the UCSC Genome Browser or the 1000 Genomes Project websites.¹² Alignments were converted from SAM format to sorted, indexed BAM files with SamTools.¹³ The picard tool was used to remove invalid alignments and remove duplicate reads from the BAM files. Regions surrounding potential indels were realigned with the GATK IndelRealigner tool.¹⁴ Variants were called with SamTools pileup command with the default parameters except that no upper limit was placed on the depth of coverage for calling variants. In addition, haploid chromosomes were called with the -N 1 options SamTools pileup. Variants were annotated with respect to their impact on coding features with perl scripts and the knownGenes and RefGenes tracks from the UCSC Genome Browser.¹⁵

Genotype Calling

Regions surrounding potential indels were realigned with the GATK IndelRealigner tool.¹⁴ Genotypes were called with both SamTools (pileup command) and the GATK UnifiedGenotypeCaller and IndelCaller.¹⁴ We then analyzed the union of single nucleotide variant (SNV) and indel variant calls from GATK and SamTools with the ANNOVAR program¹⁶ to identify exonic variants and to identify variants not previously reported in the 1000 Genomes Project and the dbSNP version 130. All variant calls not present on the X chromosome were removed given the X-linked nature of the disease. We analyzed the location and genotype of variants for each individual to locate the subset of variants on the X chromosome that were heterozygous in female carriers, hemizygous in individual III-4, and a hemizygous reference in the unaffected males. Each candidate variant was also screened for presence or absence in dbSNP 130,¹⁷ the 10Gen Dataset¹⁸ and variant data from the 1000 Genomes Project.¹²

We also used the read-mapping and SNV-calling algorithm GNUMAP¹⁹ independently to align the reads from the Illumina.qseq files to the X chromosome (human sequence build 36) and to simultaneously call SNVs. GNUMAP utilizes a probabilistic pair-hidden Markov model (PHMM) for base calling and SNV detection that incorporates base uncertainty on the basis of the quality scores from the sequencing run, as well as mapping uncertainty from multiple optimal and suboptimal alignments of the read to a given location to the genome. In addition, this approach applies a likelihood ratio test that provides researchers with straightforward SNV-calling cutoffs based on a p value cutoff or a false discovery control. Reads were aligned and SNVs called for the five samples. SNV calls for individual III-4, his brother, and his

uncle were made assuming a haploid genome (because the calls are on the X chromosome), whereas heterozygous calls were allowed for the mother and grandmother. SNVs were selected based on a p value cutoff of 0.001. Because of the X-linked nature of the disease, candidate SNVs were selected that are heterozygous in the mother and grandmother and different between the uncle and brother and individual III-4.

Variant Annotation, Analysis, and Selection Tool Analysis

The SNV filtering was performed with the simple selection tool (SST) module in variant annotation, analysis, and selection tool (VAAST) on the basis of on the SNV position. Our analysis applied a disease model that did not require complete penetrance or locus homogeneity. We restricted the expected allele frequency of putative disease-causing variants within the control genomes to 0.1% or lower. The background file used in the analysis is composed of variants from dbSNP (version 130), 189 genomes from the 1000 Genomes Project,¹² the 10Gen dataset,¹⁸ 184 Danish exomes,²⁰ and 40 whole genomes from the Complete Genomics Diversity Panel. VAAST candidate-gene prioritization analysis was performed with the likelihood ratio test under the dominant-inheritance model. An expected allele frequency of 0.1% or lower was assumed for the causal variant in the general population. After masking out loci of potentially low variant quality, SNVs in each gene were scored as a group. The significance level was assessed with individual permutation tests (the following VAAST analysis parameters were used “-m lrt -c X -g 4 -d 1.E8 -r 0.001 -x 35bp_se.wig -less_ram -inheritance d”).

Short-Tandem-Repeat Genotyping

Genotyping on family 1 was performed on DNA extracted from peripheral blood or FFPE tissues with a panel of 16 polymorphic short-tandem-repeat (STR) genotyping markers spanning the long arm of chromosome X (Applied Biosystems Linkage Mapping Set v2.5- MD10 and custom primers DXS8020, DXS1275, DXS6799, DXS1203, DXS8076, and DXS8037). Fluorescently labeled PCR products were separated with an Applied Biosystems 3130XL Genetic Analyzer and analyzed with GeneMapper Software version 3.7. Genotyping and haplotype analysis for family 2 was performed for a subset of markers on the X chromosome (DXS9900, DXS9896, DXS8016, DXS1003, DXS2132, DXS990, DXS6797, DXS8057, GATA31E08, DXS7127, and DXS1073) as described.²¹

Analysis of Haplotype Sharing

Using GNUMAP software and the X chromosome exon sequence data from family 1, we derived 1322 polymorphic markers on the mother's X chromosomes. We then determined the regions shared or not shared by individual III-4 and his unaffected brother.

Screening for Frequency of Causal Variants

We manually screened for the presence or absence of c.109T>C in *NAA10* in dbSNP (version 130), 401 participants in the ClinSeq project,²² 180 genomes from the 1000 Genomes Project,¹² the 10Gen dataset,¹⁸ 184 Danish exomes,²⁰ and 40 whole genomes from the Complete Genomics Diversity Panel.

Sorting Intolerant From Tolerant Analysis

The sorting intolerant from tolerant (SIFT) sequence was used with the sequence of hNaa10p (NP_003482) as input. The following

parameters were used: the database searched was the UniProt-SwissProt + TrEMBL 2010_09 and the median conservation of sequences was 3.00 (between 0 and 4.32, between 2.75 and 3.25 is recommended). The substitution at position 37 from S to P was predicted to affect protein function with a score of 0.00. The median sequence conservation was 3.02; 119 sequences were represented at this position. The threshold for intolerance is 0.05 (less is predicted not tolerated).

Plasmid Construction, Mutagenesis, Protein Production, Purification, and In Vitro Acetyltransferase Assays

The cDNA encoding hNaa10p WT was cloned into the pETM-41 vector (Maltose Binding Protein [MBP]/His-fusion) (from G. Stier, EMBL, Heidelberg, Germany) for expression in *Escherichia coli*. The plasmid encoding hNaa10p p.Ser37Pro was made by site-directed mutagenesis with the primers hNAA10 T109C F: 5'-C TTCTACCATGGCCTTCCTGGCCCCAGCTC-3' and hNAA10 T109C R: 5'-GAGCTGGGGCCAGGGAAGGCCATGGTAGAAG-3' according to the instruction manual (QuikChange Site-Directed Mutagenesis Kit, Stratagene). Correct cloning and mutagenesis were verified by DNA sequencing and the plasmids were transformed into *E. coli* BL21 Star (DE3) cells (Invitrogen) by heat shock. The 200 ml cell cultures were grown in Luria Bertani (LB) medium to an OD_{600 nm} of 0.6 at 37°C and subsequently transferred to 20°C. After 30 min of incubation, protein production was induced by adding IPTG to a final concentration of 0.5 mM. After 17 hr of incubation, the cells were harvested by centrifugation, and the bacterial pellets were stored at -20°C. *E. coli* pellets containing recombinant proteins were thawed at 4°C and lysed by sonication and French press in lysis buffer (1 mM DTT, 50 mM Tris-HCl [pH 7.4], 300 mM NaCl, and one tablet EDTA-free protease inhibitor cocktail [Roche] per 50 ml). The cell extracts were applied on a metal affinity fast protein liquid chromatography column (HisTrap HP, GE Healthcare, Sweden). Fractions containing recombinant protein were pooled and subjected to size-exclusion chromatography (Superdex 75 16/60, GE Healthcare). Fractions containing the monomeric recombinant protein were pooled, and the protein purity was analyzed by SDS-PAGE gel electrophoresis. The protein concentrations were determined by OD_{280 nm} measurements.

In the assays investigating the N-terminal acetyltransferase activity toward selected substrate peptides, purified MBP-hNaa10p WT or p.Ser37Pro were mixed with oligopeptides, acetyl-CoA and acetylation buffer. After incubation at 37°C, the acetylation reaction products were quantified with RP-HPLC as described previously.^{11,13} In the assays investigating the N-terminal acetyltransferase activity toward the acidic N termini of actins (DDIA or EEEIA), 15 nM of purified MBP-hNaa10p WT or S37P were mixed with 250 μM of oligopeptides. In the assays where the AVFAD and SESSS based oligopeptides were tested, 250 nM of purified enzyme were mixed with 200 μM of mentioned oligopeptides. All samples were further mixed with 400 μM acetyl-CoA and acetylation buffer (1 mM DTT, 50 mM Tris-HCl [pH 8.5], 800 μM EDTA, 10% glycerol) in a total volume of 100 μl. The samples with actin-based oligopeptides as substrates were incubated at 37°C for 15 min, whereas the SESSS and AVFAD reactions were incubated at 37°C for 20 min. In the time-course acetylation experiments, the same conditions as mentioned above were used, and samples were collected at indicated times. For all aliquots collected, the enzyme activities were quenched by adding 5 μl of 10% TEA.

The acetylation reaction products were quantified with RP-HPLC as described previously.¹¹ All peptides were custom-made (Biogenes) to a purity of 80%–95%. All peptides (24-mers) used as substrates contain seven unique N-terminal amino acids because these are the major determinants influencing N-terminal acetylation. The next 17 amino acids are essentially identical to the ACTH peptide sequence (RWGRPVGRRRRPVRVYP); however, lysines were replaced by arginines so that any potential interference by N^ε-acetylation would be minimized. The following oligopeptides were used, and proteins from which the seven N-terminal amino acids are derived are indicated: [H]AVFADLRWGRPVGRRRRPVRVYP[OH], RNaseP protein p30 (P78346); [H]SESSKSRWGRPVGRRRRPVRVYP[OH], high-mobility group protein A1 (P17096); [H]DDIAALRWGRPVGRRRRPVRVYP[OH], β-actin (NP 001092); and [H]EEEEALRWGRPVGRRRRPVRVYP[OH], γ-actin (NP 001092).

Clinical Reports

Family 1

Individual II-1 was born at 37½ weeks of gestation to a healthy 19-year-old G₁P₀₋₁. There had been concern for placental insufficiency. His birth weight was 2140 g (3rd–10th centile), and his length was 47 cm, (25th–50th centile); his orbitofrontal circumference (OFC) was 32 cm, (10th–25th centile), and his Apgar scores were 1¹ and 1.⁵ His perinatal course had been complicated by meconium aspiration leading to central depression. A number of distinctive features were noted at birth: large anterior and posterior fontanels, prominent eyes, large ears, flared nares, a narrow palate, a short neck, right cryptorchidism, fifth finger clinodactyly, relatively large great toes, metatarsus valgus, and very little subcutaneous fat.

At 6 months of age, his weight was 4.34 kg, (<< 5th centile), and his OFC was 39 cm, (<< 5th centile). He had severe global delays and had only achieved a social smile as a developmental milestone. His clinical course had been complicated by frequent apneic episodes, poor feeding, and eczema. Diagnostic studies had included karyotype analysis (46,XY), TORCH titers (which were normal), a computed tomography (CT) scan of the head demonstrating cerebral atrophy with enlarged ventricles and echocardiography suggesting peripheral pulmonary stenosis and possibly a septal defect.

At 11½ months old, he presented after 2–3 weeks of emesis and diarrhea. Upon admission he was noted to have an electrolyte imbalance, arrhythmias (premature ventricular complexes [PVCs], premature atrial contractions [PACs], supraventricular tachycardia [SVT], and ventricular tachycardia [Vtach]), and cardiomegaly, and went into cardiopulmonary arrest from which he could not be resuscitated. It should be noted that dysrhythmias occurred after the electrolyte imbalance had been corrected.

Individual II-6 was born at 38¹ weeks of gestation to a healthy 27-year-old G₆P₅₋₆. There had been concern for placental insufficiency, and his birth weight was 3065 g, (50th centile); his length was 47 cm, (25th–50th centile), and his OFC was 33.75 cm, (50th centile). His Apgar scores were 4¹ and 6.⁵ His perinatal course was complicated by respiratory distress and recognition of multiple minor anomalies: large anterior and posterior fontanels, prominent eyes, large ears, flared nares, a short neck, hypotonia, neurologic depression, and very little subcutaneous fat. Echocardiography was performed because of a murmur, which demonstrated persistence of the ductus arteriosus.

At 9 months of age, his weight was 6.38 kg, (<5th centile), his length was 68 cm, (5th centile), and his OFC was at 44.2 cm,

(10th–25th centile). He had moderate to severe global developmental delays, achieving only a social smile and the ability to raise his head and roll over. He was described as being very fussy and irritable. His clinical course had been complicated by being a poor feeder, experiencing frequent otitis media, iron deficiency anemia, and mild eczema. His physical examination was notable because the anterior fontanel was open and he had down-slanting palpebral fissures; prominent eyes; long lashes; large ears (75th–97th centile); flared nares; a short columella; a short philtrum with an overhanging upper lip; a thin vermilion border of the lips; a narrow palate; microretrognathia; an umbilical hernia; small, “high riding” testes; long fingers; fullness to the dorsum of the feet; metatarsus valgus; capillary malformation over the glabella, eyelids, philtrum, and nape of neck; and hypertonia.

Diagnostic studies had included karyotype analysis (46, XY), very long-chain fatty acids (normal), glycerine kinase (normal), total carnitine (slightly elevated), urine organic acids (normal), biotinidase activity (normal), an EKG that was somewhat irregular, and a CT scan of the head demonstrating cerebral atrophy versus dysgenesis. At 9½ months of age, he died after presenting to the hospital with multiple apneic episodes.

Individual III-7 was born at 33 weeks of gestation to a 21-year-old G₁P₀₋₁, reported as having gestational diabetes; he was delivered by elective C-section because of concern that he suffered from polycystic kidneys, oligohydramnios, and pulmonary hypoplasia. His birth weight was 1559 g, (10th–25th centile), and he had a length of 39 cm, (10th centile); his OFC was 27 cm, (3th–10th centile), and his Apgar scores were 4¹ and 8.⁵ He was noted to have distinctive facial features and a large anterior fontanel and right-sided cryptorchidism and inguinal hernia. His newborn course was complicated by polycythemia, jaundice, low cortisol, and mild pulmonary hypoplasia. He spent approximately 5 weeks in the neonatal intensive care unit (NICU); he required assisted ventilation for 3 days with the rest of the time dedicated to feeding and growing. Diagnostic studies conducted during the neonatal period included renal ultrasonography (which was normal) and echocardiography, demonstrating a small persistent ductus arteriosus, a mildly decreased left ventricular systolic function, an abnormal appearing aortic valve, and an enlargement of the right ventricle, decreased right ventricular systolic function, and persistence of the foramen ovale.

At 5½ months of age, he was in his usual state of health until 2 days prior to admission, when he developed rhinorrhea; one day prior to admission he developed a fever. He presented to the hospital with increased irritability. At first he did not appear very ill, but he quickly progressed to shock and death.

At the autopsy he was noted to have a congenital hypotonia-lymphedema sequence; hypertelorism; a high, broad forehead with a frontal furrow; micrognathia; cardiomegaly (the right half of the heart was greater than left) and the persistence of the foramen ovale; adrenomegaly; and bilateral enlarged kidneys; and abnormal development, including fetal lobulations, glomerulocystic change, sclerotic glomeruli, micronodular distal tubule proliferation, hypoplastic or small testes, a right-sided inguinal hernia, and right-sided cryptorchidism.

Individual III-4 (Figures 1 and 2B) was born at 37³ weeks of gestation to a 28-year-old G₄P₃₋₄, reported as having gestational diabetes. She had experienced preterm labor at 35 weeks; there had been decreased fetal movements, and concern for intrauterine growth retardation. His birth weight was 2410 g, (10th–25th centile), and he had a length of 44 cm (<10th centile), an OFC of 32 cm (10th–25th centile), and Apgar scores of 6,¹ 7,⁵ and 9.¹⁰



Figure 1. Triptych of Individual III-4 from Family 1

These pictures demonstrate the prominence of eyes, down-slanted palpebral fissures, thickened lids, large ears, flared nares, hypoplastic alae, short columella, protruding upper lip, and microretrognathia.

He was noted to have large anterior fontanels, prominent eyes, large ears, flared nares, a short neck, and very little subcutaneous fat. His newborn course had been complicated by hyperbilirubinemia, thrombocytopenia, and polycythemia.

At 3½ months of age his weight was 3.18 kg (<5th centile), and he had a length of 48 cm (<5th centile), and an OFC of 35 cm (2nd centile). He was able to smile. His clinical course was complicated by feeding difficulties and growth failure, but he had had no significant illness. He was noted to have a number of distinctive features, including prominent eyes, downslanted palpebral fissures, ocular hypertelorism, prominence of the cheeks, relatively large ears, a short nose, a longer philtrum, a narrow palate, microretrognathia, hypotonia, redundancy of the nuchal skin, skin laxity, and little subcutaneous fat.

At 15 months of age, his weight was 9.06 kg (<3rd centile), and he had a length of 78 cm (25th–50th centile) and an OFC of 43 cm (<3rd centile). He was able to smile. His clinical course had been complicated by severe scoliosis leading to restrictive lung disease, dysphagia, and eczema. He was unable to sit up, and he could not take food by mouth. He had been hospitalized on four separate occasions for apnea, a viral respiratory tract infection, an aspiration pneumonia, and a complication of surgery infection. He was noted to have a number of distinctive features including an anterior fontanel closed with fibrous tissue, a palpable metopic ridge, relative ocular hypertelorism (75th centile), deep-set eyes, downslanting palpebral fissures, long eyelashes, a bifid and depressed nasal tip, prominent nares, prominent ears, a short neck with excess nuchal skin, torticollis, pectus excavatum, scoliosis, and a small scrotal sack with left cryptorchidism. He had been able to build up subcutaneous fat stores, presumably because of G tube feeding. He developed a very thick, dark head of hair by his first birthday.

Diagnostic studies had included urine organic acids (normal); serum immunoglobulins (normal); comparative genomic hybridization U-array 44 k platform (normal); and an MRI of the brain with spectroscopy that demonstrated bilateral symmetric globus pallidus T2 prolongation without diffusion restriction, a nonspecific elevation of choline and prominence of the Sylvian fissures, immature myelination of the splenium, and moderate lateral and third ventricular dilation without identified cause. An MRI of the spine demonstrated convex-right curvature reflecting compensation for moderately severe convex right C-shaped thoracic neuromuscular scoliosis without any noted congenital vertebral segmentation anomalies.

radiography demonstrated a small sella turcica. Holter monitoring and the EKG had been entirely normal, but he eventually developed a nonspecific T-wave abnormality.

He died at 15 months of age after a protracted hospitalization during which he originally presented with multiple hypoxic episodes. He had a surgical correction of his severe scoliosis to optimize his pulmonary function. He developed a bradyarrhythmia and hypoxia and eventual pulseless electrical activity. After then suffering a full arrest and resuscitation, he was supported by mechanical ventilation until his parents chose to withdraw support.

An extensive autopsy of individual III-4 revealed multiple minor external anomalies, severe scoliosis, an enlarged heart with a form fruste of a perimembranous VSD, and several findings that could be attributed to hypoxia and ischemia during individual III-4's terminal course, including slight waviness of myocardial fibers (early ischemic change), pulmonary congestion and intra-alveolar edema, serous effusions in all body cavities, mid- to central zone hepatocytic necrosis, and acute neuronal ischemia in the brain. Focal segmental and global glomerulosclerosis in the outer renal cortex suggested a more chronic or remote injury that might or might not be related to the underlying genetic disorder, but no other histologic or ultrastructural lesion was identified.

Individual III-6 was recently born at 35⁴ weeks of gestation to a 28-year-old G₂P₁₋₂, reported as having gestational diabetes. His birth weight was 2604 g, (50th–75th centile), and he had a length of 48 cm, (75th centile), an OFC of 32.5 cm, (50th–75th centile), and an Apgar scores of 4¹ and 9.⁵ He was admitted directly to the NICU because of respiratory distress and remained there because of feeding difficulties and mild hyperbilirubinemia. He was noted to have a number of distinctive features including a relatively large anterior fontanel, prominent eyes, downslanted palpebral fissures, ocular hypertelorism, prominent cheeks, relatively large ears, a short nose, a longer philtrum, a narrow palate, microretrognathia, left cryptorchidism, a left inguinal hernia and a large right hydrocele, skin laxity, little subcutaneous fat, broad great toes, and hypotonia. Echocardiography demonstrated a thickened bicuspid aortic valve and mild pulmonary hypertension. A CT scan of the head was remarkable only for a cephalohematoma. His clinical course has been complicated thus far by continued feeding difficulties and growth failure.

Family 2

Family 2 presented originally in 1977 when individual II-1 (Figure 2D) was born at 43 weeks of gestation to his 25-year-old primigravid mother by Cesarean section because of cephalopelvic

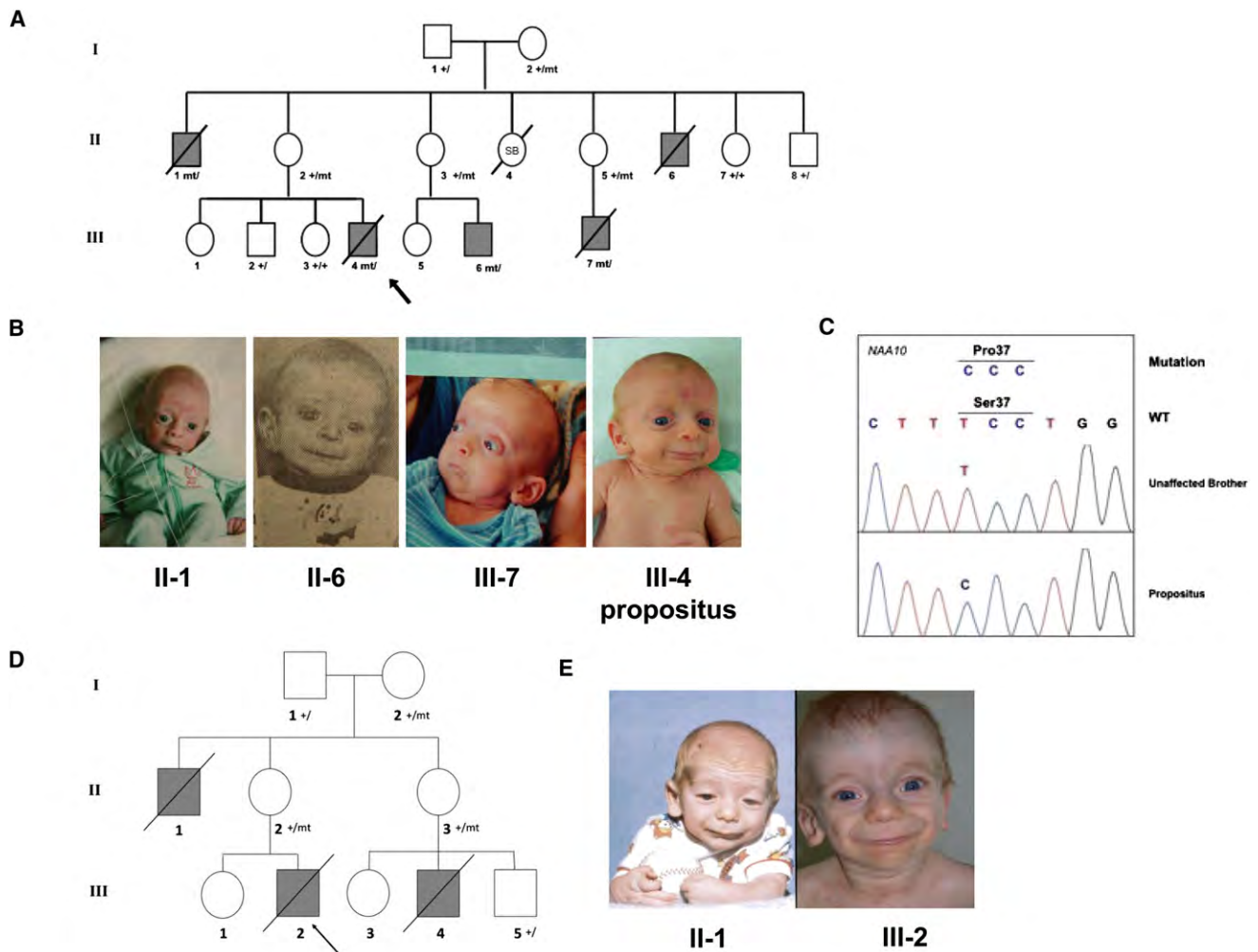


Figure 2. Pedigree Drawing and Pictures of Families 1 and 2

(A) Pedigree drawing for family 1. The most recent deceased individual, III-4, is the most well-studied subject in the family and is indicated by an arrow. Genotypes are marked for those in which DNA was available and tested. The following abbreviations are used: SB, stillborn; +, normal variant; mt, rare mutant variant.

(B) Pictures of four affected and deceased boys in this family, showing the aged appearance.

(C) Sanger sequencing results of *NAA10* in individual III-4 from family 1.

(D) Pedigree for family 2. Individual III-2 is the most well-studied subject in the family and is indicated by an arrow.

(E) Picture of individuals II-1 and III-2 in family 2 at ~1 year of age.

disproportion. He weighed 3.3 kg (25th–50th centile), measured 51 cm in length (50th–75th centile), and had a head circumference of 32 cm (<5th centile); he was assigned Apgar scores of 6¹ and 8.⁵ His testicles were undescended. He had poor feeding, jaundice with a peak bilirubin of 15 mg/dl, and vasomotor instability. He was transferred to a tertiary hospital at 8 days of age. At 7 months of age, he weighed 7.33 kg, measured 59 cm, and had a head circumference of 40.75 cm (all <5th centile). His ears and palpebral fissures were large for his chronological age, and he appeared to have bilateral ptosis. His had coarse facial features, horizontal wrinkles on his forehead, lax skin with irregular fat deposits beneath, and increased facial and body hair. His phallus was large at 4 cm, and his testes had descended. A skeletal survey showed delayed osseous development. His psychomotor development was delayed. A diagnosis of Donohue syndrome (MIM 246200) was suggested. At 8 months of age, he was admitted for traction of a congenitally dislocated left hip. Traction was insufficient to reduce the dislocation so a closed reduction under anesthesia

was performed, and he was placed in a half-body cast. On the evening after the procedure, he apparently aspirated and had a respiratory arrest, and then had generalized seizures and an apparent posthypoxia encephalopathy. He recovered from this and was discharged but died at home several days later.

In 2005, individual III-2 (Figure 2D) was born at 39.7 weeks of gestation to a 25 year-old primigravida. The pregnancy was complicated by fetal supraventricular tachycardia, which occurred at 20 weeks of gestation and was controlled with maternal administration of flecainide. At birth he weighed 2.66 kg (5th centile), measured 46 cm (5th–10th centile), had a head circumference of 32 cm (5th–10th centile), and was assigned Apgar scores of 7¹ and 9.⁵ He had infraorbital creases, a horizontal crease across his chin, and prominent nasolabial folds. He had a glabellar vascular stain; his forehead was wrinkled, giving him a worried look, and he appeared aged. He had prominent areolae and nipples, which were normally spaced, but no palpable breast tissue. He had a generalized increase in fine body hair. He had bilateral inguinal

hernias. The first and second toes appeared widely spaced. His karyotype was normal (500 bands). His copper and ceruloplasmin levels were low, but his plasma catecholamine levels were normal, ruling out Menkes syndrome (MIM 309400). His carbohydrate-deficient transferrin levels were normal. An echocardiogram revealed no structural or functional abnormalities. Inguinal herniorrhaphy was performed uneventfully at 6 weeks of age. At 5 months of age, he weighed 4.35 kg, measured 56.2 cm, and had a head circumference of 39 cm (all <5th centile). Despite hypertonia and hyperreflexia, his development was proceeding normally. He had eczema that was responding to topical corticosteroids. At 7.5 months, he weighed 5 kg, measured 61.3 cm and had a head circumference of 40 cm (all <5th centile), and his weight-for-length was 5%. He had little subcutaneous fat, which gave him a wizened look. His ears measured 4.6 cm (50th centile), appeared large for the size of his body, and had fleshy lobules. His philtrum measured 1.5 cm and appeared long and prominent. He was socially engaging, vocalized mostly with vowels, and could roll over, but could not sit independently. Subtelomeric fluorescence in situ hybridization analysis was normal. Eye evaluation showed only mild lagophthalmos. By 11 months of age, two incisors had erupted; a pincer grasp was emerging, and he could maintain a sitting position if placed into one. He vocalized with screeching and squealing but did not babble. At 11.5 months of age, he presented with symptoms of acute gastroenteritis and was admitted for rehydration. *Clostridium difficile* toxin testing was positive. He experienced three tonic-clonic seizures, and there was decerebrate posturing during the last. After loading with phenobarbital was performed, no further seizures occurred, and he returned to baseline. He continued to have excessive stool output, but electrolyte levels remained normal. He experienced increased difficulty in breathing and poor perfusion and was transferred to the prenatal intensive care unit, where he had a wide complex rhythm at 120–140 beats per minute. Ultimately, the rhythm became irregular, pulses were not palpable, and attempts to resuscitate him were unsuccessful. The autopsy showed severe micro- and macrovesicular steatosis. Because of the prenatal SVT and the rhythm disturbance without any causal electrolyte disturbance at the end of his life, his heart was evaluated postmortem for abnormalities of the conduction system. This analysis showed focal subendocardial scarring, but no cardiomyopathy or conduction system abnormalities were found. Rare inflammatory infiltrates suggested the possibility of a resolved infection.

In 2007, individual III-4 (Figure 2D) was born at 34 weeks of gestation to his 24-year-old mother, who had a 3-year-old daughter at the time. At 6.5 months of age, he had an episode of aspiration, after which gastrostomy and a tracheostomy tubes were placed. At the age of 15 months, he was hospitalized for 5 weeks for an infection. The family was told a defibrillator was needed and surgery was done for this. Two days after his discharge, he died at home at the age of 16 months.

Table 1. Features of the Syndrome in Family 1

| Category | Features |
|-------------------------|---|
| Growth | postnatal growth failure |
| Development | global, severe delays |
| Facial | wrinkled foreheads; prominence of eyes, down-sloping palpebral fissures, thickened lids; large ears; flared nares, hypoplastic alae, short columella; protruding upper lip; microretrognathia |
| Skeletal | delayed closure of fontanel; broad great toes |
| Integument | redundancy/laxity of skin, minimal subcutaneous fat, cutaneous capillary malformations, very fine hair and eyebrows |
| Cardiac ^a | structural anomalies (ventricular septal defect, atrial level defect, pulmonary artery stenoses), arrhythmias (Torsade de points, PVCs, PACs, SVtach, Vtach), death usually associated with cardiogenic shock preceded by arrhythmia. |
| Genital ^a | inguinal hernia, hypo- or cryptorchidism |
| Neurologic ^a | hypotonia progressing to hypertonia, cerebral atrophy neurogenic scoliosis |

^a Features of the syndrome demonstrating more variability. Though variable findings of the cardiac, genital and neurologic systems were observed, all affected individuals manifested some pathologic finding of each.

Results

Family 1

Individuals II-1 and II-6 from family 1 (Figure 1 and Figure 2) presented in the mid-1980s to the University of Utah Medical Center. These boys had striking similarity to each other with an array of shared manifestations (Table 1). Both subsequently died in infancy. At that time no specific diagnosis could be made, and the inheritance pattern was uncertain, though autosomal-recessive and X-linked inheritance modes were considered. X-linked inheritance was confirmed in the next generation, when individuals III-4 (Figure 1) and III-7 (Figure 2A) presented. The aged appearance was the most striking part of the disease.

Family 1 X-Chromosome Exon Capture and Sequencing

Two genomic microarray analyses in individual III-4 did not show any likely causal copy-number variants. Accordingly, X chromosome exon capture and sequencing were used to screen for variants within coding regions. We

Table 2. Coverage Statistics in Family 1 Based on GNUMAP

| Region | RefSeq Transcripts | Unique Exons | Percent Exon Coverage ≥ 1× | Percent Exon Coverage ≥ 10× | Unique Genes | Average Base Coverage | VAAST Candidate SNVs |
|---------------------------|--------------------|--------------|----------------------------|-----------------------------|--------------|-----------------------|----------------------|
| X chromosome ^a | 1959 | 7486 | 97.8 | 95.6 | 913 | 214.6 | 1 (<i>NAA10</i>) |
| chrX:10054434–40666673 | 262 | 1259 | 98.1 | 95.9 | 134 | 213.5 | 0 |
| chrX:138927365–153331900 | 263 | 860 | 97.1 | 94.9 | 132 | 177.1 | 1 (<i>NAA10</i>) |

^a On the X chromosome, there are 8222 unique RefSeq exons. Of these exons, 736 were excluded from the SureSelect X-Chromosome Capture Kit because they were designated as pseudoautosomal or repetitive sequences (UCSC Genome Browser).

Table 3. The SNV Count, Nonsynonymous Coding SNV Count, and Ti/Tv Ratio for Each Individual for Each Variant Analysis Pipeline in Family 1

| Sample | Pipeline | SNV Count | Nonsynonymous | Ti/Tv |
|--------|----------|-------------------------|------------------------------------|-------|
| III-4 | samtools | 1499 | 114 | 2.0 |
| | GATK | 1546 + 236 ^a | 146 + 6 (nonsynonymous + frame) | 2.0 |
| | GNUMAP | 2168 | 155 | 2.0 |
| II-2 | samtools | 2512 | 219 | 1.6 |
| | GATK | 1999 + 270 ^a | 168 + 8 | 2.1 |
| | GNUMAP | 2893 | 183 | 2.0 |
| III-2 | samtools | 1491 | 106 | 2.0 |
| | GATK | 1509 + 252 ^a | 134 + 10 | 2.0 |
| | GNUMAP | 2062 | 131 | 2.0 |
| I-2 | samtools | 2637 | 229 | 1.5 |
| | GATK | 2032 + 278 ^a | 160 + 10 | 2.0 |
| | GNUMAP | 2920 | 183 | 1.9 |
| II-8 | samtools | 1513 | 108 | 1.9 |
| | GATK | 1572 + 243 ^a | 136 + 8 | 1.9 |
| | GNUMAP | 1924 | 139 | 2.0 |

^a Microindels ascertained with GATK pipeline.

established the following filtering criteria for five samples to determine the final set of variant calls: variants must be present on the X chromosome in individual III-4 (hemizygous), heterozygous in individual II-2 (the mother), heterozygous in individual I-2 (the grandmother), and absent in individuals III-2 and II-8 (the unaffected brother and unaffected uncle). Table 2 shows the coverage statistics of the X chromosome exon capture. The exon sequencing reads were processed by three independent variant-calling pipelines to increase the accuracy for SNVs and/or microindels (Table 3). These three variant-calling pipelines converged on a small list of candidate variants, which were annotated by ANNOVAR¹⁶ for functional importance. One mutation had not been seen previously in dbSNP or the 1000 Genomes Project database. It is missense mutation c.109T>C in *NAA10*, which predicts p.Ser37Pro (MIM 300013) and was confirmed by Sanger sequencing (Figure 2C). We confirmed that this mutation was not present in 401 participants (90% white and mixed European descent and 96% nonlatino) in the ClinSeq project²² nor was it seen in a combination of ~6000 genomes or exomes (the majority of white and mixed European descent) collected in ongoing projects at Children's Hospital of Philadelphia, University of Utah, and/or BGI.

VAAST Analysis

We also used a recently developed tool (VAAST),²³ which identifies disease-causing variants, to analyze the exon

Table 4. Summary of the Filtering Procedure and Candidate Genes with VAAST

| SNV-Calling Pipeline | GATK | Samtools | GNUMAP |
|--|-------|----------|--------|
| III-4 (total SNVs) | 1546 | 1499 | 2168 |
| III-4 (nsSNVs) | 146 | 114 | 155 |
| VAAST candidate genes (NAA10 ranking) | 4 (3) | 3 (2) | 5 (2) |
| Present in III-4 and mother II-2 (nsSNVs) | 122 | 107 | 116 |
| VAAST candidate genes (NAA10 ranking) | 3 (2) | 2 (1) | 2 (2) |
| Present in III-4, mother II-2, and grandmother I-2 (nsSNVs) | 115 | 95 | 104 |
| VAAST candidate genes (NAA10 ranking) | 2 (1) | 2 (1) | 1 (1) |
| Present in III-4, II-2, and I-2, absent in brother III-2 and uncle II-8 (nsSNVs) | 8 | 6 | 8 |
| VAAST candidate genes (NAA10 ranking) | 1 (1) | 1 (1) | 2 (1) |

capture data from family 1. VAAST annotates the SNVs on the basis of their effect on coding sequences, selects the variants compatible with the pedigree, and performs a statistical analysis on the X chromosome genes to identify the variant(s) most likely to be disease-causing. In the candidate-gene identification step, VAAST uses a likelihood ratio test that incorporates both amino acid substitution frequencies and allele frequencies to prioritize candidate genes on the basis of SNVs present in those genes. The analyses by VAAST of the variant sets generated from the three variant-calling pipelines yielded similar results and identified the same causal mutation in *NAA10*. With SNVs from the affected child (individual III-4) alone, VAAST was able to narrow the candidate-gene list to fewer than five genes (Table 4). We then filtered the data by only selecting SNVs shared by the mother (individual II-2) and the affected child (individual III-4). This subset resulted in three, two, and two candidate genes in the GATK, Samtools, and GNUMAP datasets, respectively (Table 4). Next, we filtered the data by selecting SNVs shared by the mother (individual II-2), the maternal grandmother (individual I-2), and the affected child (individual III-4). This subset resulted in two, two, and one hits in the three datasets, respectively, and *NAA10* ranked first in all three lists. When we further excluded the SNVs that were present in the unaffected brother and uncle, VAAST identified a single candidate disease-causing variant in *NAA10* in the GATK and the Samtools datasets and two candidate disease-causing variants in *NAA10* and *RENBP* in the GNUMAP dataset (Table 4).

Confirmation in Other Family Members

DNA samples from other members of family 1 were obtained from the medical examiner; the samples include

DNA isolated from FFPE tissues from two of the deceased boys. Sanger sequencing in 14 DNA samples derived from blood or FFPE tissue from the family confirmed that the mutation in *NAA10* cosegregated according to the deduced affection status or carrier status in all members of the family (Figure 2A and Figures S1–S5, available online). During the course of this work, another boy (individual III-6) was born to individual II-3 in family 1 (Figure S6). This boy was clinically diagnosed as having the syndrome and also was confirmed to have the mutation by Sanger sequencing.

Haplotype Analysis

Having identified a putative causative mutation from the X chromosome exon analysis, we performed haplotype analysis on family 1 to exclude regions of the X chromosome from which we could have possibly missed a mutation because of a lack of exon capture or a lack of sequencing coverage. We performed haplotype analysis by using two data sets: (1) the SNVs derived from X chromosome exon sequencing of individual III-4 and (2) STR genotyping of individual III-4, two other affected males (individuals II-6 and III-7) in the family, and carrier females I-2, II-2, and II-5. Haplotype analysis narrowed the possible candidate regions to X chromosome positions 10,054,434–40,666,673 (~30 Mb) and 138,927,365–153,331,900 (~14 Mb) (Table 5 and Table S1). The sequencing data showed that of the ~155 Mb on the X chromosome, the affected brother (III-4) and unaffected brother (III-2) were identical at ~71% and were recombinant at only 29%. The ~14 Mb region is located on the telomeric end of the long arm of the X chromosome and includes *NAA10*. Subsequent addition of STR genotyping from the newborn affected male (III-6) further narrowed this interval to chrX:143,836,276–154,913,754 (telomere) (~11 Mb), which still includes *NAA10* (Table 5).

Coverage Analysis

Exon capture techniques cannot capture all exonic regions because of design issues with replicative regions; this can therefore result in an incomplete set of variants. The X chromosome Agilent design consists of over 3 Mb of exon intervals. We evaluated the capture and sequencing coverage for all the currently known 1959 chrX transcripts in RefSeq (hg18). On the X chromosome, there are 8222 unique RefSeq exons. Of these exons, 736 were excluded from the SureSelect X-Chromosome Capture Kit because they were designated as pseudoautosomal or repetitive sequences (UCSC Genome Browser). The remaining 7486 exons have an average length of 358.9 bases and are covered on average by 5.7 bait probes per exon (bait probes were 120 bases long). The average read coverage in these regions was 214.6 reads per base; in these regions an average of 97.8% of the bases were covered by one or more reads, and 95.6% were covered at 10× or better.

We also analyzed the average coverage for each gene (calculated as the total read bases within exons divided

by the exon length) more specifically in the affected haplotype in the 14 Mb critical region derived from analysis of family 1 (Table S2). The average coverage was 185× among the 167 nonpseudoautosomal or duplicative RefSeq transcripts (110 unique genes) in the 14 Mb region. These 167 transcripts consist of 864 exons, of which 860 had some read coverage and 809 had reads covering more than 95% of the bases in the exon. Most of the exon portions in the 14 Mb region that were not covered by reads were from the 89 exons not included in the SureSelect X-Chromosome Capture Kit because they were in the pseudoautosomal or duplicative regions. There was some coverage on a handful of these exons as well as intronic and intergenic regions flanking covered exons as a consequence of the experimental protocol. For example, if an excluded exon or intronic sequence is close (~500 bases) to an exon that is captured, the long fragment size (and the fact that we only sequence the end of the fragment) can lead to reads for the exon or intron. In the *NAA10* haplotype region, we had greater than 10× coverage for 86% of the exons in the region for family 1; the corollary of this means that 97 exons do not have 10× coverage in the *NAA10* region. However, of the 856 exons in the 14 Mb region that did have >10× coverage, the GNUMAP and VAAST approach found only two candidate SNVs that met our criteria (present in proband, not present in his brother and uncle, polymorphic in his mother and grandmother, nonsynonymous, not in the 1000 genomes or dbSNP), and in the exons in the 30 Mb region, no SNVs met our criteria. Theoretically, we could have nonetheless missed a crucially important variant, so we needed to use parallel approaches to prove causality of the variant in *NAA10*.

A Second Family

During preparation of a manuscript reporting the above results, another group (L.M.B., J.J.J., L.G.B.) communicated to us that they had also identified the *NAA10* c.109T>C mutation in an apparently unrelated family (herein designated as family 2) with an indistinguishable phenotype (see Figure 2D and 2E). Haplotype analysis had been performed for family 2, and it identified a shared region of chrX:140,061,918–154,913,754 (~14 Mb) in three carrier females (I-2, II-2, and II-3) that was not present in an unaffected male (III-5). The borders of the region were defined by recombinant marker *GATA31E08* and the q terminus.

Massively parallel sequencing of X chromosome exons was done on DNA samples from an obligate carrier and her unaffected child as previously described.⁹ The entire X chromosome exon region target sequence was 2,784,426 bp, and the X chromosome exon region oligonucleotide library was designed to target 2,264,175 bp of this (81.3%). The target-selected DNA libraries from one female heterozygote (M87_4) and one unaffected male child (M87_5) were sequenced on one lane each of an Illumina GAIIx in paired-end 75 bp configuration, which yielded, respectively, 75,382,114 and 65,015,016 reads (separate

Table 5. Haplotype Analysis in Affected Males from Family 1: FineMapChX_STR Results

| Sample | Allele | Marker ^a | hg18 Position | Chromosome |
|---------------------|--------|----------------------|---------------|------------|
| III-4 | 219 | DXS1275 (AFM261ZH5) | 68,431,124 | Xq13.1 |
| III-6 | 225 | | | |
| II-6 (FFPE tissue) | ? | | | |
| III-7 (FFPE tissue) | 225 | DXS8037 (AFMA285XG5) | 74,040,314 | Xq13.3 |
| III-4 | 253 | | | |
| III-6 | 257 | | | |
| II-6 (FFPE tissue) | 242 | DXS986 (AFM116XG1) | 79,267,784 | Xq21.1 |
| III-7 (FFPE tissue) | 257 | | | |
| III-4 | 162 | | | |
| III-6 | 164 | DXS8076 (AFMB357XE5) | 82,666,040 | Xq21.1 |
| II-6 (FFPE tissue) | 168 | | | |
| III-7 (FFPE tissue) | 164 | | | |
| III-4 | 106 | DXS1203 (AFM262VG1) | 92,654,766 | Xq21.32 |
| III-6 | 98 | | | |
| II-6 (FFPE tissue) | 98 | | | |
| III-7 (FFPE tissue) | 98 | DXS990 (AFM136YC7) | 92,887,320 | Xq21.32 |
| III-4 | 217 | | | |
| III-6 | 217 | | | |
| II-6 (FFPE tissue) | 217 | DXS6799 (GATA29G07) | 97,265,664 | Xq21.33 |
| III-7 (FFPE tissue) | 217 | | | |
| III-4 | 131 | | | |
| III-6 | 127 | DXS8020 (AFMA162TC1) | 99,455,440 | Xq22.1 |
| II-6 (FFPE tissue) | 127 | | | |
| III-7 (FFPE tissue) | 127 | | | |
| III-4 | 253 | DXS1106 (AFM263WE1) | 102,618,641 | Xq22.2 |
| III-6 | 257 | | | |
| II-6 (FFPE tissue) | 257 | | | |
| III-7 (FFPE tissue) | 257 | DXS8055 (AFMB291YE5) | 114,561,258 | Xq23 |
| III-4 | 201 | | | |
| III-6 | 212 | | | |
| II-6 (FFPE tissue) | 203 | DXS1106 (AFM263WE1) | 102,618,641 | Xq22.2 |
| III-7 (FFPE tissue) | 212 | | | |
| III-4 | 128 | | | |
| III-6 | 130 | DXS8055 (AFMB291YE5) | 114,561,258 | Xq23 |
| II-6 (FFPE tissue) | 126 | | | |
| III-7 (FFPE tissue) | 130 | | | |
| III-4 | 317 | DXS8055 (AFMB291YE5) | 114,561,258 | Xq23 |
| III-6 | 315 | | | |
| II-6 (FFPE tissue) | 317 | | | |
| III-7 (FFPE tissue) | 315 | | | |

Table 5. Continued

| Sample | Allele | Marker ^a | hg18 Position | Chromosome |
|---------------------|--------|--------------------------|---------------------------|------------|
| III-4 | 206 | DXS1001 (AFM248WE5) | 119,720,593 | Xq24 |
| III-6 | 199 | | | |
| II-6 (FFPE tissue) | 206 | | | |
| III-7 (FFPE tissue) | 199 | | | |
| III-4 | 165 | DXS1047 (AFM150XF10) | 128,902,983 | Xq25 |
| III-6 | 165 | | | |
| II-6 (FFPE tissue) | 165 | | | |
| III-7 (FFPE tissue) | 160 | | | |
| III-4 | 90 | DXS1227 (AFM317YE9) | 140,630,173 | Xq27.2 |
| III-6 | 82 | | | |
| II-6 (FFPE tissue) | 90 | | | |
| III-7 (FFPE tissue) | 82 | | | |
| III-4 | 170 | DXS8043 (AFMB018WD9) | 143,836,276 | Xq27.3 |
| III-6 | 156 | | | |
| II-6 (FFPE tissue) | 170 | | | |
| III-7 (FFPE tissue) | 170 | | | |
| III-4 | 83 | DXS8091 (AFM345WG9) | 147,410,588 | Xq28 |
| III-6 | 83 | | | |
| II-6 (FFPE tissue) | 83 | | | |
| III-7 (FFPE tissue) | 83 | | | |
| | | c.109T>C in <i>NAA10</i> | ~152,850,921 ^b | Xq28 |
| III-4 | 310 | DXS1073 (AFM276XH9) | 153,482,054 | Xq28 |
| III-6 | 310 | | | |
| II-6 (FFPE tissue) | 310 | | | |
| III-7 (FFPE tissue) | 310 | | | |

^a Genomic position represents a single nucleotide within the STR.

^b Maximum Shared Region (hg18): chrX:143,836,276–154,913,754 (telomere) (11 Mb).

results are given for each sample) or 5,729,040,664 and 4,941,141,216 bp of total sequence. Of the filtered aligning sequence, 20.9% and 11.9% could be uniquely aligned to the entire X chromosome exon region target (2,784,426 bases). This aligned sequence yielded a gross overall coverage of 2058× and 1775× of the entire X chromosome exon region target. The capture efficiency varied across the targets; 2,538,791 and 2,464,367 bp (89.8%) of the entire X chromosome exon region yielded ≥1× coverage, 2,305,571 and 2,251,868 bp (81.8%) yielded ≥10× coverage, and 2,254,446 and 2,183,075 bp (79.7%) yielded ≥20× coverage. The Most Probable Genotype (MPG) variant-calling software²⁴ was able to make base calls on 2,295,223 and 2,378,985 bp of this sequence (82.4% and 85.4%).

The genotypes were filtered on the basis of several attributes that were felt to be appropriate for this disorder (Table 6). Heterozygosity was used because the test subject

was an unaffected female carrier for an X-linked trait. We performed further filtering to include nonsynonymous, splice-site, frame-shifting, and nonsense variants. Filtering to exclude variants present in dbSNP or in the ClinSeq cohort (401 control individuals)²² was also performed. Finally, a filter that bounded the variants genomically within the shared 14 Mb haplotype region was applied. This left one variant, a single missense mutation, c.109T>C in *NAA10*. This variant was confirmed in the other individuals, and mutation status segregated with affection status and carrier status (Figure 2D).

Genome-wide Significance

We next assessed whether adding members of family 2 would improve the power of the VAAST analysis. We combined variants from individual III-4 in family 1 with the obligate carrier mother (II-2) in family 2 and performed VAAST analysis on the 216 coding SNVs present in either of

Table 6. Filtering Analysis of Family 2

| | X chromosome exons | Exons in Shared Haplotype |
|------------------------------------|---------------------------|----------------------------------|
| Total SNVs | 3441 | 585 |
| Heterozygous | 2381 | 418 |
| Stop/NonSyn/FS/Splice ^a | 136 | 35 |
| Not in dbSNP | 40 | 10 |
| Not in ClinSeq | 20 | 1 (<i>NAA10</i>) |

^a The following abbreviations are used: Stop, nonsense variants; NonSyn, non-synonymous variants; FS, frame-shifting variants; Splice, splice-site variants.

the two individuals. After incorporating the mother from family 2, *NAA10* was the only candidate gene, and the result was statistically significant (p value = 3.8×10^{-5} ; Bonferroni corrected p value = $3.8 \times 10^{-5} \times 729 = 0.028$).

Lack of Relatedness of the Two Families

Both families are of mixed European ancestry. To evaluate whether the two families inherited the *NAA10* variant from a recent shared founder, we examined SNVs surrounding this variant to test for evidence of identity-by-descent (IBD). Assuming no sequencing errors, the largest region consistent with IBD between the two probands is only 30 kb (Table 7). By relaxing this assumption to allow for multiple sequencing errors, we find the

largest region consistent with IBD is 704 kb, which is approximately 1.5 centiMorgans (cM)²⁵ (Figure 3). If the two probands were as distantly related as fourth cousins, the expected IBD segment size would be 20 cM (conditioned on ascertainment from the presence of a shared variant).²⁶ Therefore, the variation between the two probands in this region is inconsistent with recent shared ancestry. If the two probands inherited the mutation from a more distant founder and if they share the entire 700 kb region, the estimated time to the founder would be around 3300 years ago. Because this variant is an X-linked recessive lethal mutation, the probability of its being eliminated in each transmission is 1/3, and the variant is therefore highly unlikely to have occurred 3300 years ago. We conclude that the disease-causing variant in the two families resulted from two independent mutational events.

Functional Analysis

N-terminal acetylation of proteins is catalyzed by N-terminal acetyltransferases (NATs). The primary NAT in terms of targeted substrates is the evolutionarily conserved NatA.^{27–29} The functional impact of N-terminal acetylation remains quite elusive, but recent data suggest a role as a destabilization signal for proteins.³⁰ The human NatA complex is composed of the catalytic subunit hNaa10p (hARD1) and the auxiliary subunit hNaa15p (NATH/hNAT1), both essential for its activity.³¹ Increased NatA levels have been linked to tumor progression, and

Table 7. Relationship Inference of the Two Families

| Position | First Family | | Second Family | | |
|-----------|-----------------------|---------|-----------------------------|----------------|--------|
| | Inconsistent with IBD | Proband | Imputed Genotype of Proband | Unaffected Son | Mother |
| 152802312 | X | G | A | G | AG |
| 152804123 | X | A | G | A | GA |
| 152821601 | X | A | G | A | GA |
| 152821643 | X | C | T | C | TC |
| 152825187 | X | G | A | G | AG |
| 152829448 | X | G | A | G | AG |
| 152853035 | | G | G | A | GA |
| 152860231 | X | T | G | G | GG |
| 152932023 | | A | A | A | AA |
| 152937386 | | G | G | G | GG |
| 152945374 | | C | C | C | CC |
| 153317692 | X | G | A | A | AA |
| 153321366 | | G | G | G | GG |
| 153534024 | X | C | A | C | AC |
| 153534719 | X | C | G | C | GC |
| 153556589 | X | A | G | A | GA |
| 153557591 | X | G | A | G | AG |
| 153557667 | X | T | C | T | CT |

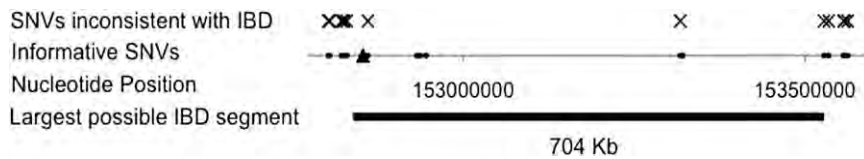


Figure 3. IBD Analysis of the Two Families in the Genomic Region Surrounding c.109T>C

Informative SNVs consist of two classes: rare SNVs present in the proband from family 1 in genomic regions with high coverage in the second family, and

common SNVs present in the first proband and at least one member of the second family. Variant c.109T>C is indicated by the triangle. We imputed the genotype of the proband from family 2 from the genotypes of the mother and unaffected sibling of the second family (see Table 7). SNVs inconsistent with IBD, in which the imputed genotype of the second proband does not match the first proband, are indicated with an X. After allowing for multiple sequencing errors, the largest genomic segment consistent with IBD is around 700 kb in length.

depletion of NatA subunits from cancer cells induces cell-cycle arrest and apoptosis.³² Human Naa10p is a protein of 235 amino acid residues, of which the first 178 residues compose a globular region, whereas the latter 57 residues are predicted to form an unstructured and flexible C-terminal tail.³³ Thus, Ser37 is located in a structured part of hNaa10p. For many soluble globular proteins, Pro is known to be potentially disruptive for secondary structure elements such as the alpha-helix and beta-sheets.³⁴ Thus, although the structure of hNaa10p is undetermined, the p.Ser37Pro mutation could indeed affect the structure of hNaa10p and thereby the catalytic activity. Ser37 and its surrounding residues are highly conserved among eukaryotes,³¹ suggesting an essential function. A SIFT analysis predicting whether an amino acid substitution affects protein function³⁵ strongly suggested that a substitution from Ser to Pro at position 37 would affect protein function with a score of 0.00.

In order to directly assess the functional consequences of the hNaa10p p.Ser37Pro mutation, we analyzed the wild-type (WT) and the mutant proteins by a quantitative in vitro N-terminal acetylation assay. The enzyme activities of hNaa10p WT and hNaa10p p.Ser37Pro were determined with four unique peptides as substrate; these peptides having been previously shown to be acetylated by hNaa10p and NatA (Figure 4). Compared to hNaa10p WT, hNaa10p p.Ser37Pro displayed a 60%–80% reduction in NAT activity toward the in vivo substrate RNaseP protein p30 (AVFAD-) and toward β -actin (DDDIA-) and γ -actin (EEEIA-). In contrast, the activity toward the NatA substrate high-mobility group protein A1 (SESSS-) was only reduced by 20% (Figure 4). The oligopeptides AVFAD- and SESSS- represent classical cotranslational NatA substrates, being the N termini of proteins that are partially (AVFAD-) or fully (SESSS-) acetylated by NatA in HeLa cells.²⁷ On the other hand, β -actin (DDDIA-) and γ -actin (EEEIA-) are nonclassical substrates recently shown to be acetylated more efficiently by hNaa10p/NatA than the classical substrates were, representing a posttranslational NAT activity.³⁶

Discussion

The data presented here show that a mutation in an enzyme involved in N-terminal acetylation of proteins

leads to a distinct, previously undescribed X-linked phenotype in humans and that males who carry the hypomorphic hNaa10p p.Ser37Pro allele die in infancy with cardiac arrhythmias. N-terminal acetylation is one of the most common protein modifications in humans, occurring on approximately 80% of all human proteins²⁷ It is catalyzed by several distinct NAT enzymes, of which the major one is NatA.²⁹ The catalytic subunit of the NatA complex, hNaa10p, is essential for survival in the organisms *Drosophila melanogaster*,³⁷ *Trypanosoma brucei*,³⁸ and *Caenorhabditis elegans*.³⁹ It is presumed that an amorphic *NAA10* mutation would lead to embryonal lethality in humans, although one can only prove this by analyzing tissues from pregnancies that did not survive to term. The strong conservation of Naa10p Ser37, the predicted functional effect of p.Ser37Pro due to structural distortion, and the demonstrated disruption of catalytic activity by p.Ser37Pro (Figure 4) strongly imply that the hemizygous males with hNaa10p p.Ser37Pro have impaired NatA function. Thus, a variety of protein N termini, both those that are cotranslationally and those that are posttranslationally acetylated (e.g., actins), are likely to be insufficiently acetylated. Most likely, the serious consequences of the p.Ser37Pro mutation are caused by the lack of N-terminal acetylation for one or several proteins strictly requiring this modification for function or for maintenance of adequate amounts in the cell. Because hNaa10p has also been suggested to perform N-lysine (N-epsilon) acetylation of proteins such as beta-catenin,⁴⁰ a lack of N-lysine acetylation of selected substrates could also cause the observed effects. Finally, proposed noncatalytic functions of hNaa10p^{41,42} could be affected in the p.Ser37Pro mutant and thereby also play a role in the observed phenotypes.

We have demonstrated herein that a probabilistic disease-causing variant discovery algorithm can readily identify and characterize the genetic basis of a previously unrecognized X-linked syndrome. We have also shown that this algorithm, when used in parallel with high-throughput sequencing, can identify variants with high prioritization for causing disease with as few as two individuals. In this instance, we screened ~150 variants distributed among ~2000 transcripts on the X chromosome in one sample from individual III-4 in family 1. With no prior filtering, we prioritized three to five possible candidate genes, and the mutation in *NAA10* ranked second overall (Table 4). Including exon capture data from relatives

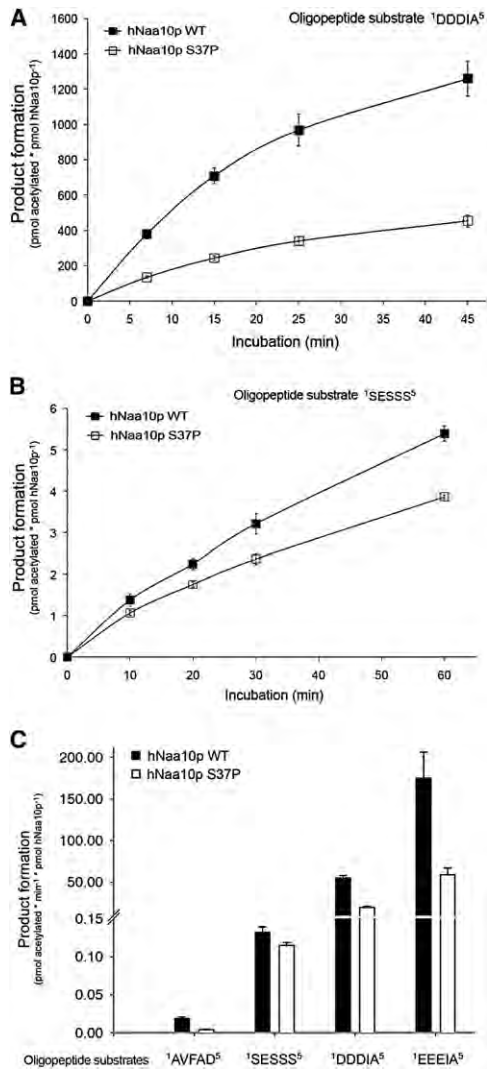


Figure 4. NAT Activity of Recombinant hNaa10p WT or p.Ser37Pro toward Synthetic N-Terminal Peptides

(A and B) Purified MBP-hNaa10p WT or p.Ser37Pro were mixed with the indicated oligopeptide substrates (200 μ M for SESSS and 250 μ M for DDDIA) and saturated levels of acetyl-CoA (400 μ M). Aliquots were collected at indicated time points and the acetylation reactions were quantified with reverse phase HPLC peptide separation. Error bars indicate the standard deviation based on three independent experiments. The five first amino acids in the peptides are indicated, for further details see [Subjects and Methods](#). Time-dependent acetylation reactions were performed to determine initial velocity conditions when comparing the WT and Ser37Pro NAT activities toward different oligopeptides. (C) Purified MBP-hNaa10p WT or p.Ser37Pro were mixed with the indicated oligopeptide substrates (200 μ M for SESSS and AVFAD and 250 μ M for DDDIA and EEEIA) and saturated levels of acetyl-CoA (400 μ M) and incubated for 15 min (DDDIA and EEEIA) or 20 min (SESSS and AVFAD) at 37°C in acetylation buffer. The acetylation activity was determined as above. Error bars indicate the standard deviation based on three independent experiments. Black bars indicate the acetylation capacity of the MBP-hNaa10p WT, whereas white bars indicate the acetylation capacity of the MBP-hNaa10p mutant p.Ser37Pro. The five first amino acids in the peptides are indicated.

increased *NAA10*'s ranking to first overall in just one family. Furthermore, after combining variants from the proband in family 1 with the obligate carrier mother in family 2, VAAST identified *NAA10* as the only statistically significant candidate.

Although we have noted that the affected infants have an aged appearance, we have not established any direct link with progeria or other progeroid syndromes. The autopsies did not reveal any premature arteriosclerosis or degeneration of vascular smooth muscle cells, as is seen in Hutchinson-Gilford progeria syndrome (MIM 176670).^{43,44} Cell lines now being derived from family 1 and possibly future animal models will provide important insights about the pathophysiology underlying this previously unrecognized syndrome.

Supplemental Data

Supplemental Data include six figures and two tables and can be found with this article online at <http://www.cell.com/AJHG/>.

Acknowledgments

We express our gratitude to the families for their extraordinary cooperation and assistance. We also thank David Nix, Nina Glomnes, and Whitney Fitts for advice and/or technical assistance with family 1. Exon capture and sequencing for family 1 was paid for by the Department of Psychiatry, University of Utah (to G.J.L.). Collection of DNA and phenotyping for family 1 was supported by the Clinical Genetics Research Program: Phenotyping Core, under CCTS grant UL1RR025764 at the University of Utah. The University of Utah Microarray and Genomic Analysis core facility was supported by award number P30CA042014 from the National Cancer Institute. Functional analyses were supported by the Research Council of Norway (grant 197136 to T.A.) and the Norwegian Cancer Society (to J.R.L. and T.A.). B.M. and M.Y. were supported by National Human Genome Research Institute (NHGRI) 1RC2HG005619, K.W. by a pilot/methodological study award from NIH/National Center for Research Resources grant UL1 RR025774, J.X. by NIH/NHGRI K99HG005846, and W.E.J. by NHGRI 5R01HG5692. The research on family 2 was supported by Intramural Funds of the NHGRI, NIH (L.G.B.). These authors thank Danielle Brinkman for the initial consenting and records gathering for family 2 and Caitlin Krause and Jamie Teer for technical support.

Received: April 30, 2011

Revised: May 18, 2011

Accepted: May 19, 2011

Published online: June 23, 2011

Web Resources

The URLs for data presented herein are as follows:

- ANNOVAR Software, <http://www.openbioinformatics.org/annovar/>
- Complete Genomics Diversity Panel, <http://www.completegenomics.com/sequence-data/download-data/>
- GATK Software, http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit
- GNUMAP, <http://dna.cs.byu.edu/gnumap/>

Online Mendelian Inheritance in Man(OMIM), <http://www.omim.org>
 Picard, <http://sourceforge.net/projects/picard/>
 SIFT (Sorting Intolerant From Tolerant) Analysis, http://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html
 UCSC Genome Browser, <http://genome.ucsc.edu/>

References

- Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C., et al. (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.* **42**, 790–793.
- Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., et al. (2010). Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **42**, 30–35.
- Choi, M., Scholl, U.I., Ji, W., Liu, T., Tikhonova, I.R., Zumbo, P., Nayir, A., Bakkaloğlu, A., Ozen, S., Sanjad, S., et al. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. USA* **106**, 19096–19101.
- Pierce, S.B., Walsh, T., Chisholm, K.M., Lee, M.K., Thornton, A.M., Fiumara, A., Opitz, J.M., Levy-Lahad, E., Kleivit, R.E., and King, M.C. (2010). Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault Syndrome. *Am. J. Hum. Genet.* **87**, 282–288.
- Bilgüvar, K., Oztürk, A.K., Louvi, A., Kwan, K.Y., Choi, M., Tatli, B., Yalnizoglu, D., Tüysüz, B., Çağlayan, A.O., Gökben, S., et al. (2010). Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* **467**, 207–210.
- Hedges, D.J., Burges, D., Powell, E., Almonte, C., Huang, J., Young, S., Boese, B., Schmidt, M., Pericak-Vance, M.A., Martin, E., et al. (2009). Exome sequencing of a multigenerational human pedigree. *PLoS ONE* **4**, e8232.
- Bonnefond, A., Durand, E., Sand, O., De Graeve, F., Gallina, S., Busiah, K., Lobbens, S., Simon, A., Bellanné-Chantelot, C., Létourneau, L., et al. (2010). Molecular diagnosis of neonatal diabetes mellitus using next-generation sequencing of the whole exome. *PLoS ONE* **5**, e13630.
- Roach, J.C., Glusman, G., Smit, A.F., Huff, C.D., Hubley, R., Shannon, P.T., Rowen, L., Pant, K.P., Goodman, N., Bamshad, M., et al. (2010). Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636–639.
- Johnston, J.J., Teer, J.K., Cherukuri, P.F., Hansen, N.F., Loftus, S.K., Chong, K., Mullikin, J.C., and Biesecker, L.G.; NIH Intramural Sequencing Center (NISC). (2010). Massively parallel sequencing of exons on the X chromosome identifies RBM10 as the gene that causes a syndromic form of cleft palate. *Am. J. Hum. Genet.* **86**, 743–748.
- Johnston, J.J., Olivos-Glander, I., Killoran, C., Elson, E., Turner, J.T., Peters, K.F., Abbott, M.H., Aughton, D.J., Aylsworth, A.S., Bamshad, M.J., et al. (2005). Molecular and clinical analyses of Greig cephalopolysyndactyly and Pallister-Hall syndromes: Robust phenotype prediction from the type and position of GLI3 mutations. *Am. J. Hum. Genet.* **76**, 609–622.
- Evjenth, R., Hole, K., Ziegler, M., and Lillehaug, J.R. (2009). Application of reverse-phase HPLC to quantify oligopeptide acetylation eliminates interference from nonspecific acetyl CoA hydrolysis. *BMC Proc* **3** (Suppl 6), S5.
- Durbin, R.M., Abecasis, G.R., Altshuler, D.L., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurler, M.E., and McVean, G.A.; 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073.
- Evjenth, R., Hole, K., Karlsen, O.A., Ziegler, M., Arnesen, T., and Lillehaug, J.R. (2009). Human Naa50p (Nat5/San) displays both protein N alpha- and N epsilon-acetyltransferase activity. *J. Biol. Chem.* **284**, 31122–31129.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.
- Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A., et al. (2011). The UCSC Genome Browser database: Update 2011. *Nucleic Acids Res.* **39** (Database issue), D876–D882.
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164.
- Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetverin, V., Church, D.M., Dicuccio, M., Federhen, S., et al. (2010). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **38** (Database issue), D5–D16.
- Reese, M.G., Moore, B., Batchelor, C., Salas, F., Cunningham, F., Marth, G.T., Stein, L., Flicek, P., Yandell, M., and Eilbeck, K. (2010). A standard variation file format for human genome sequences. *Genome Biol.* **11**, R88.
- Clement, N.L., Snell, Q., Clement, M.J., Hollenhorst, P.C., Purwar, J., Graves, B.J., Cairns, B.R., and Johnson, W.E. (2010). The GNUMAP algorithm: Unbiased probabilistic mapping of oligonucleotides from next-generation sequencing. *Bioinformatics* **26**, 38–45.
- Li, Y., Vinckenbosch, N., Tian, G., Huerta-Sanchez, E., Jiang, T., Jiang, H., Albrechtsen, A., Andersen, G., Cao, H., Korneliusson, T., et al. (2010). Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat. Genet.* **42**, 969–972.
- Kurpinski, K.T., Magyari, P.A., Gorlin, R.J., Ng, D., and Biesecker, L.G. (2003). Designation of the TARP syndrome and linkage to Xp11.23-q13.3 without samples from affected patients. *Am. J. Med. Genet. A.* **120A**, 1–4.
- Biesecker, L.G., Mullikin, J.C., Facio, F.M., Turner, C., Cherukuri, P.F., Blakesley, R.W., Bouffard, G.G., Chines, P.S., Cruz, P., Hansen, N.F., et al; NISC Comparative Sequencing Program. (2009). The ClinSeq Project: Piloting large-scale genome sequencing for research in genomic medicine. *Genome Res.* **19**, 1665–1674.
- Yandell, M., Huff, C.D., Hu, H., Singleton, M., Moore, B., Xing, J., Jorde, L.B., and Reese, M.G. (2011). A probabilistic disease-gene finder for personal genomes. *Genome Res.* **21** 10.1101/gr.123158.111.
- Teer, J.K., Bonnycastle, L.L., Chines, P.S., Hansen, N.F., Aoyama, N., Swift, A.J., Abaan, H.O., Albert, T.J., Margulies, E.H., Green, E.D., et al; NISC Comparative Sequencing Program. (2010). Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. *Genome Res.* **20**, 1420–1431.
- Kong, A., Gudbjartsson, D.F., Sainz, J., Jonsdottir, G.M., Gudjonsson, S.A., Richardsson, B., Sigurdardottir, S., Barnard, J.,

- Hallbeck, B., Masson, G., et al. (2002). A high-resolution recombination map of the human genome. *Nat. Genet.* *31*, 241–247.
26. Huff, C.D., Witherspoon, D.J., Simonson, T.S., Xing, J., Watkins, W.S., Zhang, Y., Tuohy, T.M., Neklason, D.W., Burt, R.W., Guthery, S.L., et al. (2011). Maximum-likelihood estimation of recent shared ancestry (ERSA). *Genome Res.* *21*, 768–774.
27. Arnesen, T., Van Damme, P., Polevoda, B., Helsens, K., Evjenth, R., Colaert, N., Varhaug, J.E., Vandekerckhove, J., Lillehaug, J.R., Sherman, F., and Gevaert, K. (2009). Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc. Natl. Acad. Sci. USA* *106*, 8157–8162.
28. Mullen, J.R., Kayne, P.S., Moerschell, R.P., Tsunasawa, S., Gribskov, M., Colavito-Shepanski, M., Grunstein, M., Sherman, F., and Sternglanz, R. (1989). Identification and characterization of genes and mutants for an N-terminal acetyltransferase from yeast. *EMBO J.* *8*, 2067–2075.
29. Polevoda, B., Arnesen, T., and Sherman, F. (2009). A synopsis of eukaryotic N-terminal acetyltransferases: Nomenclature, subunits and substrates. *BMC Proc* *3* (Suppl 6), S2.
30. Hwang, C.S., Shemorry, A., and Varshavsky, A. (2010). N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* *327*, 973–977.
31. Arnesen, T., Anderson, D., Baldersheim, C., Lanotte, M., Varhaug, J.E., and Lillehaug, J.R. (2005). Identification and characterization of the human ARD1-NATH protein acetyltransferase complex. *Biochem. J.* *386*, 433–443.
32. Gromyko, D., Arnesen, T., Rynningen, A., Varhaug, J.E., and Lillehaug, J.R. (2010). Depletion of the human N-terminal acetyltransferase A induces p53-dependent apoptosis and p53-independent growth inhibition. *Int. J. Cancer* *127*, 2777–2789.
33. Sánchez-Puig, N., and Fersht, A.R. (2006). Characterization of the native and fibrillar conformation of the human N-terminal acetyltransferase ARD1. *Protein Sci.* *15*, 1968–1976.
34. Chou, P.Y., and Fasman, G.D. (1974). Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. *Biochemistry* *13*, 211–222.
35. Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* *31*, 3812–3814.
36. Van Damme, P., Evjenth, R., Foyn, H., Demeyer, K., De Bock, P.J., Lillehaug, J.R., Vandekerckhove, J., Arnesen, T., and Gevaert, K. (2011). Proteome-derived peptide libraries allow detailed analysis of the substrate specificities of N-terminal acetyltransferases and point to hNaa10p as the post-translational actin N-terminal acetyltransferase. *Mol. Cell. Proteomics* *10*, M110, 004580.
37. Wang, Y., Mijares, M., Gall, M.D., Turan, T., Javier, A., Bornemann, D.J., Manage, K., and Warrior, R. (2010). *Drosophila* variable nurse cells encodes arrest defective 1 (ARD1), the catalytic subunit of the major N-terminal acetyltransferase complex. *Dev. Dyn.* *239*, 2813–2827.
38. Ingram, A.K., Cross, G.A., and Horn, D. (2000). Genetic manipulation indicates that ARD1 is an essential N-terminal acetyltransferase in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* *111*, 309–317.
39. Sönnichsen, B., Koski, L.B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., Alleaume, A.M., Artelt, J., Betten-court, P., Cassin, E., et al. (2005). Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* *434*, 462–469.
40. Lim, J.H., Park, J.W., and Chun, Y.S. (2006). Human arrest defective 1 acetylates and activates beta-catenin, promoting lung cancer cell proliferation. *Cancer Res.* *66*, 10677–10682.
41. Hua, K.T., Tan, C.T., Johansson, G., Lee, J.M., Yang, P.W., Lu, H.Y., Chen, C.K., Su, J.L., Chen, P.B., Wu, Y.L., et al. (2011). N-terminal acetyltransferase 10 protein suppresses cancer cell metastasis by binding PIX proteins and inhibiting Cdc42/Rac1 activity. *Cancer Cell* *19*, 218–231.
42. Lee, C.F., Ou, D.S., Lee, S.B., Chang, L.H., Lin, R.K., Li, Y.S., Upadhyay, A.K., Cheng, X., Wang, Y.C., Hsu, H.S., et al. (2010). hNaa10p contributes to tumorigenesis by facilitating DNMT1-mediated tumor suppressor gene silencing. *J. Clin. Invest.* *120*, 2920–2930.
43. Capell, B.C., Tloutan, B.E., and Orlow, S.J. (2009). From the rarest to the most common: Insights from progeroid syndromes into skin cancer and aging. *J. Invest. Dermatol.* *129*, 2340–2350.
44. Liu, G.H., Barkho, B.Z., Ruiz, S., Diep, D., Qu, J., Yang, S.L., Panopoulos, A.D., Suzuki, K., Kurian, L., Walsh, C., et al. (2011). Recapitulation of premature ageing with iPSCs from Hutchinson-Gilford progeria syndrome. *Nature* *472*, 221–225.

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of May 4, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/328/5978/636>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/science.1186802/DC1>

This article **cites 24 articles**, 4 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/328/5978/636#otherarticles>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

and neuropeptides (e.g., thyrotropin-releasing hormone) (24) are secreted by granular glands, and the first group represents an important defense against pathogens (25). Antimicrobial peptides are clustered in at least seven transcription units >350 kbp on scaffold 811, with no intervening genes.

X. tropicalis occupies a key phylogenetic position among previously sequenced vertebrate genomes, namely amniotes and teleost fish. Given the utility of the frog as a genetic and developmental biology system and the large and increasing amounts of cDNA sequence from the pseudo-tetraploid *X. laevis*, the *X. tropicalis* reference sequence is well poised to advance our understanding of genome and proteome evolution in general, and vertebrate evolution in particular.

References and Notes

1. L. Hogben, C. Gordon, *J. Exp. Biol.* **7**, 286 (1930).
2. D. D. Brown, *J. Biol. Chem.* **279**, 45291 (2004).
3. J. Tymowska, *Cytogenet. Cell Genet.* **12**, 297 (1973).
4. Supporting material is available on Science Online.
5. International Chicken Genome Sequencing Consortium, *Nature* **432**, 695 (2004).
6. International Human Genome Sequencing Consortium, *Nature* **409**, 860 (2001).
7. Mouse Genome Sequencing Consortium, *Nature* **420**, 520 (2002).
8. V. V. Kapitonov, J. Jurka, *Genetica* **107**, 27 (1999).
9. V. V. Kapitonov, J. Jurka, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6569 (2003).
10. International Rice Genome Sequencing Project, *Nature* **436**, 793 (2005).
11. N. L. Craig, R. Craigie, M. Gellert, A. M. Lambowitz, Eds., *Mobile DNA II* (American Society for Microbiology, Washington, DC, 2002).
12. V. V. Kapitonov, J. Jurka, *DNA Cell Biol.* **23**, 311 (2004).
13. N. H. Putnam *et al.*, *Science* **317**, 86 (2007).
14. Dataset S1 is available on Science Online.
15. I. G. Woods *et al.*, *Genome Res.* **15**, 1307 (2005).
16. A. Abu-Daya, A. K. Sater, D. E. Wells, T. J. Mohun, L. B. Zimmerman, *Dev. Biol.* **336**, 20 (2009).
17. M. A. Nobrega, I. Ovcharenko, V. Afzal, E. M. Rubin, *Science* **302**, 413 (2003).
18. R. J. Garriock, A. S. Warkman, S. M. Meadows, S. D'Agostino, P. A. Krieg, *Dev. Dyn.* **236**, 1249 (2007).
19. M. Kofron *et al.*, *Development* **126**, 5759 (1999).
20. L. Du Pasquier, J. Schwager, M. F. Flajnik, *Annu. Rev. Immunol.* **7**, 251 (1989).
21. J. Robert, Y. Ohta, *Dev. Dyn.* **238**, 1249 (2009).
22. Y. Ohta, M. Flajnik, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10723 (2006).
23. Y. Zhao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12087 (2006).
24. G. Kreil, in *The Biology of Xenopus*, R. C. Tinsley, H. R. Kobels, Eds. (The Zoological Society of London, Oxford, 1996), pp. 263–277.
25. L. A. Rollins-Smith, *Integr. Comp. Biol.* **45**, 137 (2005).
26. This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory, under contract DE-AC02-05CH11231, Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344, and Los Alamos National Laboratory under contract DE-AC02-06NA25396. This research was supported in part by the Intramural Research Program of the NIH, National Library of Medicine, and by a grant to R.K.W. from the National Human Genome Research Institute (NHGRI U01 HG02155) with supplemental funds provided by the National Institute of Child Health and Human Development. We thank R. Gibbs and S. Scherer of the Human Genome Sequencing Center, Baylor College of Medicine, for their contributions to identification and mapping of simple sequence length polymorphisms.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5978/633/DC1

Materials and Methods

SOM Text

Figs. S1 to S9

Tables S1 to S17

References

Dataset S1

22 October 2009; accepted 16 February 2010

10.1126/science.1183670

Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing

Jared C. Roach,^{1*} Gustavo Glusman,^{1*} Arian F. A. Smit,^{1*} Chad D. Huff,^{1,2*} Robert Hubley,¹ Paul T. Shannon,¹ Lee Rowen,¹ Krishna P. Pant,³ Nathan Goodman,¹ Michael Bamshad,⁴ Jay Shendure,⁵ Radoje Drmanac,³ Lynn B. Jorde,² Leroy Hood,^{1,†} David J. Galas^{1†}

We analyzed the whole-genome sequences of a family of four, consisting of two siblings and their parents. Family-based sequencing allowed us to delineate recombination sites precisely, identify 70% of the sequencing errors (resulting in >99.999% accuracy), and identify very rare single-nucleotide polymorphisms. We also directly estimated a human intergeneration mutation rate of $\sim 1.1 \times 10^{-8}$ per position per haploid genome. Both offspring in this family have two recessive disorders: Miller syndrome, for which the gene was concurrently identified, and primary ciliary dyskinesia, for which causative genes have been previously identified. Family-based genome analysis enabled us to narrow the candidate genes for both of these Mendelian disorders to only four. Our results demonstrate the value of complete genome sequencing in families.

Whole-genome sequences from four members of a family represent a qualitatively different type of genetic data than whole-genome sequences from individu-

al or sets of unrelated genomes. They enable inheritance analyses that detect errors and permit the identification of precise locations of recombination events. This leads in turn to near-complete knowledge of inheritance states through the precise determination of the parental chromosomal origins of sequence blocks in offspring. Confident predictions of inheritance states and haplotypes power analyses that include the identification of genomic features with nonclassical inheritance patterns, such as hemizygous deletions or copy number variants (CNVs). Identification of inheritance patterns in the pedigree permits the detection of $\sim 70\%$ of sequencing errors and sharply reduces the search space for

disease-causing variants. These analyses would be far less powerful in studies that had fewer markers (such as standard genotype or exome data sets) or that had sequences from fewer family members.

DNA from each family member was extracted from peripheral blood cells and sequenced at CGI (Mountain View, California) with a nanoarray-based short-read sequencing-by-ligation technology (1), including an adaptation of the pairwise end-sequencing strategy (2). Reads were mapped to the National Center for Biotechnology Information (NCBI) reference genome (fig. S1 and tables S1 and S2). Polymorphic markers used for this analysis were single-nucleotide polymorphisms (SNPs) with at least two variants among the four genotypes of the family, averaging 802 base pairs (bp) between markers. We observed 4,471,510 positions at which at least one family member had an allele that varied from the reference genome. This corresponds to a Watterson's theta (θ_w) of 9.5×10^{-4} per site for the two parents and the reference sequence (3), given the fraction of the genome successfully genotyped in each parent (fig. S1). This is a close match to the estimate of $\theta_w = 9.3 \times 10^{-4}$ that we obtained by combining two previously published European genomes and the reference sequence (4). Of the 4.5 million variant positions, 3,665,772 were variable within the family; the rest were homozygous and identical in all four members. Comparisons to known SNPs show that 323,255 of these 3.7 million SNPs are novel.

For each meiosis in a pedigree, each base position in a resulting gamete will have inherited one of two parental alleles. The number of inheritance patterns of the segregation of alleles in

¹Institute for Systems Biology, Seattle, WA 98103, USA.

²Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT 84109, USA.

³Complete Genomics, Inc. (CGI), Mountain View, CA 94043, USA.

⁴Department of Pediatrics, University of Washington, Seattle, WA 98195, USA.

⁵Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail:

dgalas@systemsbiology.org (D.J.G.); lhood@systemsbiology.org (L.H.)

gametes is therefore 2^n , where n is the number of meioses in a pedigree. In a nuclear family of four, the Mendelian inheritance patterns can be grouped into four inheritance states for each variant position, with children receiving (i) the same allele from both the mother and the father (identical), (ii) the same allele from the mother but opposites from the father (haploidentical maternal), (iii) the same allele from the father, but opposites from the mother (haploidentical paternal), or (iv) opposites from both parents (nonidentical) (fig. S2). Adjacent variant base pairs in alignments of the family genomes have the same inheritance state unless a recombination has occurred between these bases in one of the meioses. This delineates inheritance blocks.

Many algorithms can identify the boundaries of blocks, and theory-driven implementations are in wide use (5–7). For our complete genome sequence data, we developed an algorithm to identify all states, including non-Mendelian states. One non-Mendelian state will occur in regions where highly similar sequences are inadvertently compressed computationally (for example, during sequence assembly of CNVs). In such a “compression block,” many positions will appear to be heterozygous in all individuals, regardless of the inheritance patterns of the positions contributing to the compression. Other non-Mendelian patterns are seen in regions prone to errors in sequence calling or assembly or that have inherited hemizygous deletions. For both of these patterns, many positions will be observed as Mendelian inheritance errors (MIEs). Our algorithm identified six states: one for each of the four Mendelian inheritance states, one for a compression state, and one for a MIE-prone state (4). We identified 1.5% of the genome in this pedigree as 409 compression blocks and 1.7% as 126 error-prone blocks. Because these blocks are a source of false positives for recombination predictions, SNPs, and disease candidate alleles, their identification is important (Fig. 1). The power to precisely determine inheritance-state boundaries is striking in families of at least four and would be reduced had we sequenced fewer individuals (Fig. 2). Meiotic gene conversions could in principle be recognized in the same way as inheritance blocks; they would be indistinguishable from a short region flanked by meiotic recombinations in the same meiosis. We found that the great majority of candidate gene-conversion regions were caused by reads mismapped to repetitive DNA, such as CNVs or satellites, and did not conclusively identify gene-conversion regions.

Recombination in maternal meioses is thought to occur 1.7 times more frequently than in paternal meioses (8). We inferred 98 crossovers in maternal and 57 in paternal meioses (count includes both offspring), which is consistent with this estimate. The median resolution of the 155 crossover sites was 2.6 kb, with a few sites localized within a 30-bp window (Fig. 1). Crossover sites were significantly correlated with hotspots of recombination as inferred from HapMap data, in

which a hotspot is defined as a region with ≥ 10 centimorgan (cM)/Mb; 92 of the 155 recombinations took place in a hotspot.

By identifying inconsistencies across the 22% of the genomes of the two children in “identical” blocks, for which they are effectively twins, we computed an error rate of 1.0×10^{-5} . We also determined error rate through other methods, including resequencing, which gave similar estimates, ranging from 8.1×10^{-6} to 1.1×10^{-5} (4). Furthermore, $\sim 70\%$ of the errors in a four-person pedigree can be detected as apparent MIEs and inconsistencies in inheritance state blocks, so the effective base-pair error rate in the context of a pedigree is $\sim 3 \times 10^{-6}$.

Analysis of the mutation rate, including germline and early embryonic somatic mutations, requires highly accurate sequence data. Even with such data, however, most apparent aberrations in allele inheritance will be due to errors in the data and not to mutation. Our data had thousands of such false-positive candidates for each true de novo mutation. Our initial data encompassed 2.3 billion bases and contained 49,720 candidate MIEs that were consistent with the presence of a single-nucleotide mutation. After excluding sites in MIE-prone and compression

states as well as sites that were unsuitable for probe design, 33,937 potential mutations among 1.83 billion bases remained. We resequenced each of these candidates and applied a stringent base-calling algorithm to confirm 28 candidates as de novo mutations. In a final confirmation step, we verified all 28 mutations with mass spectrometry (table S3) (4), corresponding to a mutation rate of 3.8×10^{-9} per position per generation per haploid genome.

Because the raw estimate of 3.8×10^{-9} does not account for the true mutations that were not conclusively identified through resequencing, we estimated a false-negative rate by applying the base-calling algorithm to 5 Mb of independent resequencing data, divided into 25 randomly selected regions of the genome. A comparison of the resequencing data with the complete genome sequence for the same regions provided a de novo mutation false negative rate of 0.662 [95% confidence interval (CI) 0.644 to 0.680]. Adjusting for the false-negative rate produced an unbiased mutation rate estimate of 1.1×10^{-8} per position per haploid genome, corresponding to approximately 70 new mutations in each diploid human genome (95% CI of 6.8×10^{-9} to 1.7×10^{-8}) (4). In great apes, CpG sites are

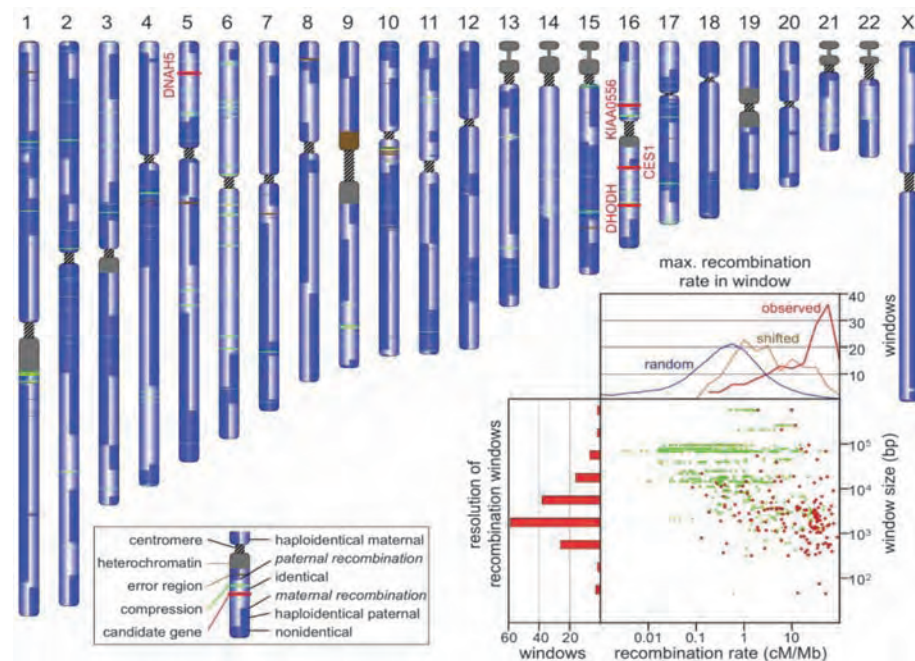


Fig. 1. The landscape of recombination. Each chromosome in this schematic karyotype is used to represent information abstracted from the four corresponding chromosomes of the two children in the pedigree. It is vertically split to indicate the inheritance state from the father (left half) and mother (right half), as shown in the key. The three compound heterozygous (*DHODH*, *DNAH5*, and *KIAA0556*) and one recessive (*CES1*) candidate gene, depicted by red bands, lie in “identical” blocks. (Inset) Scatterplot of HapMap recombination rates (in centimorgans per megabase) within the predicted crossover regions. The maximum value of centimorgans per megabase found in each window is shown in red. The left histogram shows the size distribution of recombination windows (\log_{10} value of -0.58 ± 0.92). The top graph shows the centimorgans per megabase distribution for the observed maximal values (red), for similarly sized windows shifted by 6 kb (orange), and for similarly sized windows randomly chosen from the entire genome (blue). A shift of 6 kb from the observed locations eliminates the correlation with hotspots. Of 155 recombination windows, 92 contained a HapMap site with >10 cM/Mb. Only five randomly picked windows are expected to contain such high recombination rates.

reported to mutate at a rate 11 times higher than other sites (9). We observed five CpG mutations, closely matching this estimate. Of the remaining 23 mutations, seven were transversions and 16 were transitions. This yields a transition-to-transversion ratio of 2.3 (table S3), which is once again similar to a previous estimate of 2.2 for non-CpG sites (10).

Although both the observed transition-to-transversion ratio and the proportion of CpG mutations in our data match predictions, our estimated human mutation rate is lower than previous estimates, the most widely cited of which is 2.5×10^{-8} per generation (10) based on three parameters: a human-chimpanzee nucleotide divergence per site (K_1) of 0.013, a species divergence time of 5 million years ago, and an ancestral effective population size of 10,000. More recent estimates indicate a nucleotide divergence of 0.012 (9), species divergence time between 6 and 7 million years ago (11–15), and ancestral effective population size between 40,000 and 148,000 (16–19). With these parameter ranges and a generation length of 15 to 25 years, the mutation rate estimate is between 7.6×10^{-9} and 2.2×10^{-8} per generation, which is consistent with our intergenerational estimate of 1.1×10^{-8} . Our estimate is within 1 SD of an earlier estimate of 1.7×10^{-8} (SD of 9×10^{-9}) based on 20 disease-causing loci (20). The rate we report is for autosomes and should be substantially lower than that of the Y chromosome because in the male germ line, more cell divi-

sions occur per generation. Although our rate differs approximately as expected from the recently reported estimate of 3.0×10^{-8} (95% CI, 8.9×10^{-9} to 7.0×10^{-8}) for the Y chromosome, this difference is not significant (21).

Genomic inheritance analysis facilitates the identification of alleles that cause genetic disorders. Because genome sequences from a family of four provide near-exact determination of inheritance-state boundaries, the number of false-positive disease-gene candidates is greatly reduced as compared with those of analyses lacking the context of a pedigree or complete genome sequence (Fig. 3 and tables S3 and S4). Two disorders in this family—Miller syndrome and primary ciliary dyskinesia, which affect both offspring but neither parent—provided an opportunity to test this application. A parsimonious explanation is that each phenotype arises from defects in a single gene or a site regulating a single gene. The inheritance mode is undetermined, but a recessive mode is more consistent with observed data. We therefore examined each candidate variant by testing each of three inheritance modes: dominant, simple recessive, or compound heterozygote (a subcategory of recessive).

The two recessive modes require that both offspring have identical dysfunctional variants for which the parents are heterozygous and which may come either from the same position (simple recessive) or occur at distinct positions within the same gene (compound heterozygote). Genes that are consistent with these two recessive

modes must lie in “identical” inheritance blocks because both offspring are affected, limiting the search space to the 22% of the genome in these blocks. Because the phenotypes are rare, they are likely to be encoded by rare variants, which further limits the possibilities. Only two missense SNPs in the *CESI* gene matched the simple recessive mode (table S4), whereas three genes fit the compound heterozygote mode: *DHODH*, *DNAH5*, and *KIAA0556* (Fig. 1). A small number of possibly detrimental variants outside exons also matched the simple recessive mode: two in highly conserved regions, one in an intronic sequence near a splice site, five in non-protein-coding transcripts, and one in an untranslated region (UTR). Concurrent with this study, the core exomes of the two affected offspring were sequenced along with those of two unrelated individuals with Miller syndrome (22). Compared with that study of only affected individuals, our analysis of just two affected

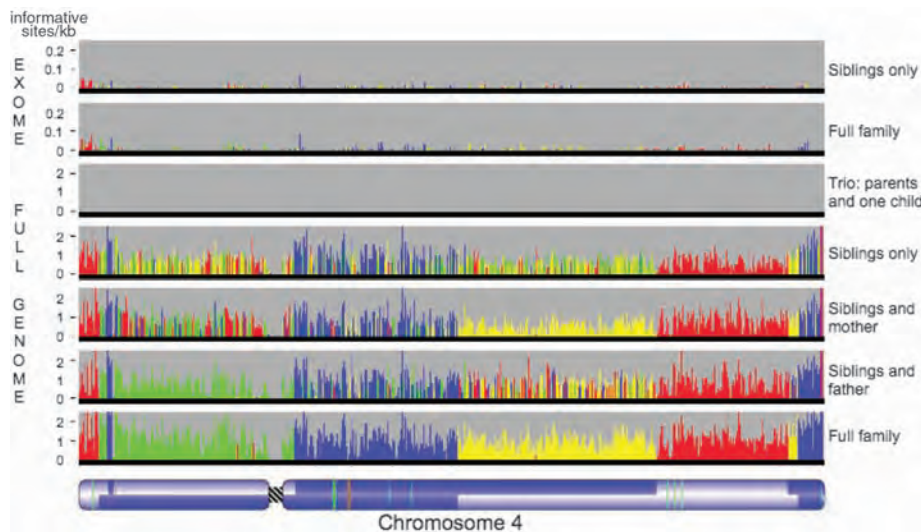


Fig. 2. Power of four. Shown are inheritance states for a single chromosome in six scenarios representing restrictions of the data set to the exome (for two siblings only or for the full family) or to subsets of the family (parents and one child, two siblings, or siblings and one parent), as compared with analysis with full data from all four family members. The most supported state for each bin is shown as a color; the height of each histogram bar is proportional to the number of informative markers supporting that state. The father has two regions of homozygosity (bottom, thin red lines) on the short arm of the chromosome, where it is not possible to distinguish the haploidentical maternal from identical states (fig. S2A, panel b). These regions are undetected when the mother’s genotypes are missing because all marker positions in the region are uninformative (second from bottom). A pedigree of two parents and one child has only one inheritance state and so provides no information on recombination. Red, identical; blue, nonidentical; green, haploidentical maternal; yellow, haploidentical paternal. Chromosome structure is annotated as in Fig. 1.

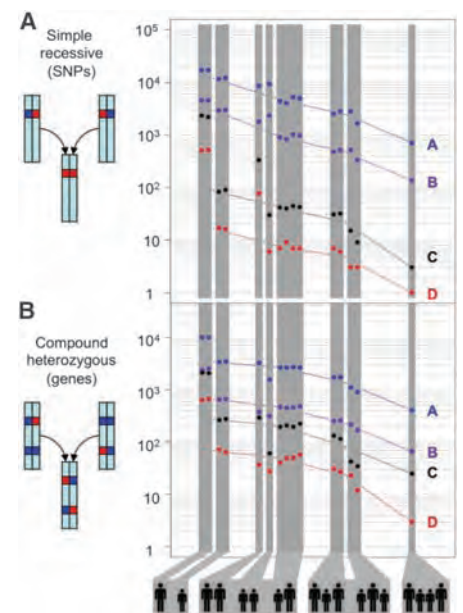


Fig. 3. The power of family genome inheritance analysis. The number of false-positive candidates drops exponentially as the number of family members increases. (A) Number of candidate SNPs that are consistent with a simple recessive inheritance mode. (B) Number of candidate genes that are consistent with a compound heterozygote model. The different groupings of parents (large silhouettes) and children (small silhouettes) are depicted below. Dashed lines join the average values of each grouping. For this figure, “probably detrimental” includes missense, nonsense, splice defect, and non-initiation; “possibly detrimental” also includes UTR, noncoding, and splice region. A block of SNPs so that all SNPs in the block are within 5 kb of another SNP in the block is counted only once because together these are likely to encode at most one phenotype. “A,” all probably detrimental SNPs; “B,” all possibly detrimental SNPs; “C,” rare possibly detrimental SNPs; “D,” rare probably detrimental SNPs.

offspring and their unaffected parents reduced the number of gene candidates in the core exome from nine to four; had we not sequenced the parents, we would have had 34 rather than four candidates (Fig. 3 and table S5). The exome study supported *DHODH* as the primary gene for Miller syndrome. *DNAH5* had been previously identified as a cause of primary ciliary dyskinesia, and so is probably the cause in these offspring as well (23).

Family genome analysis can clearly be effective for finding candidate genes that encode Mendelian traits because sequence accuracy is enhanced. In addition, delineation of recombination sites identifies inherited chromosome segments precisely and reduces the chromosomal search space for candidate genes (in this case to 22% of the genome). The ability to identify large effects of very rare alleles in small pedigrees can complement the power of genome-wide association studies in identifying weak effects of common alleles in large populations. An unknown fraction of important phenotypes in humans are encoded by nonexonic variants identified only by means of whole-genome sequencing. When the cost of recruiting additional families is expensive relative to sequencing costs, sequencing genomes of families will be an economical strategy for the identification of many disease-causing genes. Constraining searches to very rare variants can provide considerable power, as recently demonstrated for Freeman-Sheldon syndrome and congenital chloride diarrhea (24, 25). De novo mutations can be assayed, either as we have reported here

or through family sequencing of more than two generations. As our knowledge of gene function increases, we will be able to use the power of family genome analysis rapidly to identify disease-gene candidates. These data, along with relevant environmental and medical information, will characterize the integrated medical records of the future.

References and Notes

1. R. Drmanac *et al.*, *Science* **327**, 78 (2010).
2. J. C. Roach, C. Boysen, K. Wang, L. Hood, *Genomics* **26**, 345 (1995).
3. G. A. Watterson, *Theor. Popul. Biol.* **7**, 256 (1975).
4. Materials and methods are available as supporting material on Science Online.
5. K. P. Donnelly, *Theor. Popul. Biol.* **23**, 34 (1983).
6. L. Kruglyak, M. J. Daly, M. P. Reeve-Daly, E. S. Lander, *Am. J. Hum. Genet.* **58**, 1347 (1996).
7. G. R. Abecasis, S. S. Cherny, W. O. Cookson, L. R. Cardon, *Nat. Genet.* **30**, 97 (2002).
8. P. M. Petkov, K. W. Broman, J. P. Szatkiewicz, K. Paigen, *Trends Genet.* **23**, 539 (2007).
9. Chimpanzee Sequencing and Analysis Consortium, *Nature* **437**, 69 (2005).
10. M. W. Nachman, S. L. Crowell, *Genetics* **156**, 297 (2000).
11. Y. Haile-Selassie, *Nature* **412**, 178 (2001).
12. Y. Haile-Selassie, B. Asfaw, T. D. White, *Am. J. Phys. Anthropol.* **123**, 1 (2004).
13. Y. Haile-Selassie, G. Suwa, T. D. White, *Science* **303**, 1503 (2004).
14. A. L. Deino, L. Tauxe, M. Monaghan, A. Hill, *J. Hum. Evol.* **42**, 117 (2002).
15. M. Brunet *et al.*, *Nature* **418**, 145 (2002).
16. F. C. Chen, W. H. Li, *Am. J. Hum. Genet.* **68**, 444 (2001).
17. R. Burgess, Z. Yang, *Mol. Biol. Evol.* **25**, 1979 (2008).
18. N. Takahata, *Jpn. J. Genet.* **68**, 539 (1993).
19. J. D. Wall, *Genetics* **163**, 395 (2003).
20. A. S. Kondrashov, *Hum. Mutat.* **21**, 12 (2003).
21. Y. Xue *et al.*, *Curr. Biol.* **19**, 1453 (2009).
22. S. B. Ng *et al.*, *Nat. Genet.* **42**, 30 (2010).
23. H. Olbrich *et al.*, *Nat. Genet.* **30**, 143 (2002).
24. S. B. Ng *et al.*, *Nature* **461**, 272 (2009).
25. M. Choi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 190961 (2009).
26. This study was supported by the University of Luxembourg–Institute for Systems Biology Program and by these NIH grants: Center for Systems Biology GM076547 (L.H. and L.R.), R01GM081083 (A.F.S. and G.G.), R01HL094976 and RZ1HG004749 (J.S.), RC2HG005608 (M.D. and J.S.), and R01HD048895 (M.B.). H. Tabor assisted with ethical review. J. Xing performed the principal components analysis. H. Mefford performed CNV analysis. A. Bigham and K. Buckingham evaluated candidate genes in unrelated individuals. D. Ballinger, A. Sparks, A. Halpern, and G. Nilsen assisted with sequencing and analysis. R. Bressler, S. Dee, and D. Mauldin assisted with bioinformatics. S. Ng and R. Qiu performed the capture array. S. Bloom obtained the resequencing data on the Illumina Genome Analyzer. M. Janer and S. Li performed Sequenom analysis. D. Cox commented on an early version of the manuscript. R. Durbin and D. Altshuler granted permission for our use of 1000 genomes SNP data. CGI employees (R.D. and K.P.) have stock options in the company. J.S. has consulted for CGI. L.H. is a scientific advisor to CGI and holds stock in the company. The dbGAP accessions can be found at www.ncbi.nlm.nih.gov/projects/gap/cgibin/study.cgi?study_id=phs000244.v1.p1 (accession phs000244.v1.p1).

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1186802/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S5
References

7 January 2010; accepted 5 March 2010
Published online 11 March 2010;
10.1126/science.1186802
Include this information when citing this paper.

Eric J. Topol, MD
Director, Scripps Translational Science Institute
Chief Academic Officer, Scripps Health
Professor of Genomics, The Scripps Research Institute
The Gary and Mary West Chair of Innovative Medicine
La Jolla, California

Biography



Dr. Topol is a pioneer of the genomic and wireless digital innovative technologies to reshape the future of medicine.

He is a practicing cardiologist at Scripps in La Jolla, California and well known for leading the Cleveland Clinic to become the #1 center for heart care. While there he also started a new medical school, led many worldwide clinical trials to advance care for patients with heart disease, and spearheaded the discovery of multiple genes that increase susceptibility for heart attacks.

Since 2006, in La Jolla, he leads the flagship NIH supported Scripps Translational Science Institute and is a co-Founder and the Vice-Chairman of the West Wireless Health Institute. He also serves as Professor of Genomics at The Scripps Research

Institute and Chief Academic Officer of Scripps Health.

Topol pioneered the development of many medications that are routinely used in medical practice including t-PA, Plavix, Angiomax, and ReoPro and was the first physician to raise safety concerns on Vioxx. He has published 1100 peer-reviewed articles and over 30 medical textbooks. In 2009, along with Francis Collins and Harold Varmus, Topol was selected to be one of the country's 12 "Rock Stars of Science" in GQ Magazine. In 2011, the University of Michigan, where he had served on the faculty, initiated the Eric Topol Professor of Cardiovascular Medicine to recognize his contributions. The University of Rochester, his alma mater medical school, awarded him the Hutchinson Medal, the University's highest honor.

He was named as one of the *Top 100 Most Influential People in Healthcare* in 2011.

He was elected to the Institute of Medicine of the National Academy of Sciences and is one of the top 10 most cited researchers in medicine. His book *The Creative Destruction of Medicine* (Basic Books) was published in 2012.

Notes

Aravinda Chakravarti, PhD
Director, Center for Complex Disease Genomics
McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins University
Baltimore, Maryland
Director, Institute of Molecular Medicine
Delhi, India

Notes

Joseph H. Nadeau, PhD
Director of Research and Academic Affairs
Institute for Systems Biology (ISB)
Seattle, Washington

Biography

Joseph Nadeau is an internationally-recognized expert in genetic, genomic, metabolic, bioinformatics, computational and systems studies of mouse models of birth defects, cancer and metabolic disease as well as translating results to these studies in humans. He has been a pioneer in comparative genomics (comparative gene mapping), genetics and systems studies of mouse models of human disease (chromosome substitution strains) as well as transgenerational epigenetic effects on cancer, metabolism, embryogenesis and behavior based on discoveries that change the ways that we understand inheritance of phenotypic variation and disease susceptibility.

He is formerly James H. Jewel Professor and Chair of Genetics Department at Case Western Reserve University School of Medicine. He was a founding member of the International Mammalian Genome Society and a founding editor of both Mammalian Genome and WIREs Systems Biology and Medicine; the latter won the RR Hawkins Award from the American Publishers Awards for Professional & Scholarly Excellence (PROSE) – this is the top award for outstanding scholarly work in all disciplines of the arts and sciences. He was founder and director of the Mouse Genome Informatics Project and the Mouse Genome Database. He has served on numerous review panels and advisory groups at the National Institutes of Health, the National Science Foundation, the Wellcome Trust, and the Human Genome Database; he has also consulted for several biotech and major pharmaceutical companies. His work has received several awards and he is an Elected Fellow of the American Association for the Advancement of Science. His work was recently recognized with an NIH Pioneer Award. Finally, his students and fellows have won numerous local, national and international awards for their work.

Notes

Revisiting Lamarck and Mendel: Transgenerational genetic effects on health and disease

Joseph Nadeau, PhD


Both humans and animal models provide many examples of highly heritable traits where the disease-causing genetic variants elude discovery. In these cases, heritability is often high but the explained genetic variance is low. The usual explanation involves undetected genetic variants with weak and heterogeneous actions in affected individuals. This explanation is reasonable given the limited statistical power in most genetic studies. However, we recently found several examples of heritable epigenetic changes, through the germline, suggesting that modes of inheritance other than DNA can also be transmitted across generations to affect phenotypic variation and disease risk. In each case, genetic variants in ancestral generations lead to phenotypic variation in subsequent generations. These effects can be as common and strong as those resulting from conventional inheritance, and they can persist for multiple generations. Transgenerational inheritance affects cancer, metabolic diseases, behaviors and many other traits. The effects are usually specific to one germ-lineage, with examples of transmission through the female germ-lineage and other through the paternal germ-lineage. The identity of the genes that mediate these effects implicate aspects of RNA biology including RNA editing, demethylation, translation control, and regulation of miRNA biology.

References

- Youngren, K.K., D. Coveney, X. Peng, C. Battacharya, L.S. Schmidt, M.L. Nickerson, B.T. Lamb, J.M. Deng, R.R. Behringer, B. Capel, E.M. Rubin, J.H. Nadeau, and A. Matin 2005. The Ter mutation in the Dead-end gene causes germ cell loss and testicular germ cell cancer. *Nature* 435:360-365.
- Lam, M.-Y.J., J. Heaney, K.K. Youngren, J.H. Kawasoe, and J.H. Nadeau (2007) Transgenerational epistasis between *Dnd1*Ter* and other modifier genes controls susceptibility to testicular germ cell tumors. *Hum. Mol. Genet.* 16:2233-2240.
- Heaney, J.D., M.V. Michelson, K.K. Youngren, M.Y.J. Lam and J.H. Nadeau (2009) Deletion of translation initiation factor eIF2beta suppresses testicular cancer incidence and causes recessive lethality in agouti-yellow mice. *Hum. Mol. Genet.* 18:5193-5197.
- Nadeau, J.H. (2009) Transgenerational genetic effects on phenotypic variation and disease risk. *Hum. Mol. Genet.* 18: R202-10.
- Nelson V.R., S.H. Spiezio, and J.H. Nadeau (2010) Transgenerational genetic effects of the paternal Y chromosome on daughters' phenotypes. *Epigenomics* 2: 513-521.
- Nelson, V.R. and J.H. Nadeau (2010) Transgenerational genetic effects. *Epigenomics* 2: 797-806.

Yazbek, S.N., S.H. Spiezio, J.H. Nadeau, and D.A. Buchner (2010) Ancestral paternal genotype determines diet-induced obesity and food intake for multiple generations. *Hum. Mol. Genet.* 19: 4134-4144. PMID: 20696673

Notes



Mendel's laws of inheritance

genotype - phenotype association within individuals is the foundation of most genetic studies

Transgenerational genetic effects

Phenotypes and disease risk result from genetic variants in previous rather than the present generation


Genetic origins, heritable and familial, but genetic variants are not in affected individuals

Testicular cancer - PGC stem cells
principles and mechanisms

Daughters of fathers with different Y chromosomes

Diet-induced obesity and feeding habits

Mouse model of TGCTs
(testicular cancer)



- Spontaneous TGCTs occur at a measurable frequency only in the 129 family of inbred strains
- What do we know?
Genetically controlled, but complex
Stem cells are primordial germ cells (PGCs)
Age of onset (E11.5-E12.5)
Evident at 3-4 weeks of age

Reduced Kit ligand signaling increases susceptibility

Maternal *Kit^{Sl-gb}*

Kit^{Sl-gb} / +** **+ / +

Male progeny

modified (15 %) ***Kit^{Sl-gb} / +***

baseline (5 %) **+ / +**

Conventional

J. Heaney et al. Canc Res 2008

Reduced Kit ligand signaling increases susceptibility

Maternal *Kit^{Sl-gb}* Paternal *Kit^{Sl-gb}*

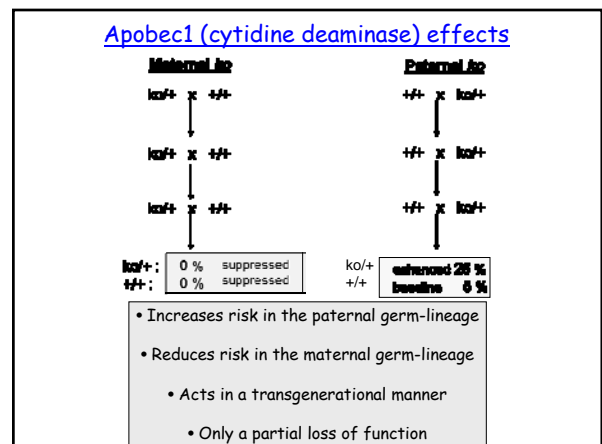
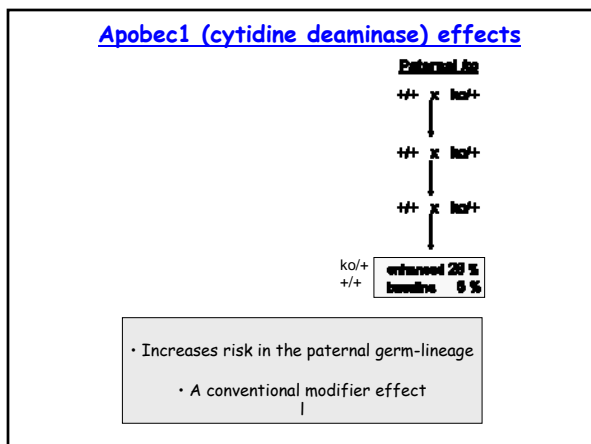
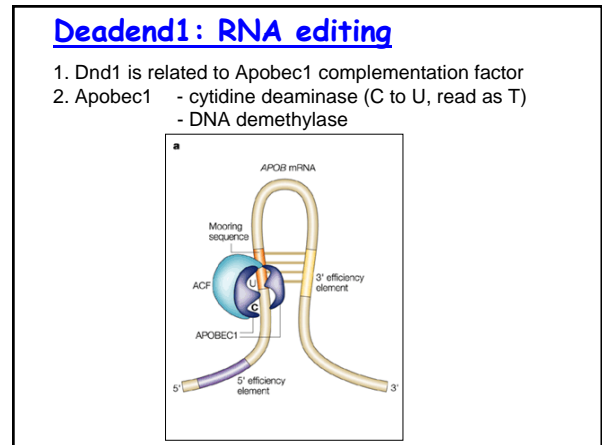
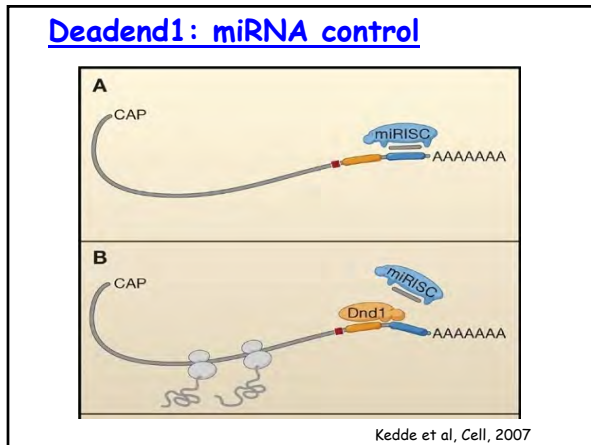
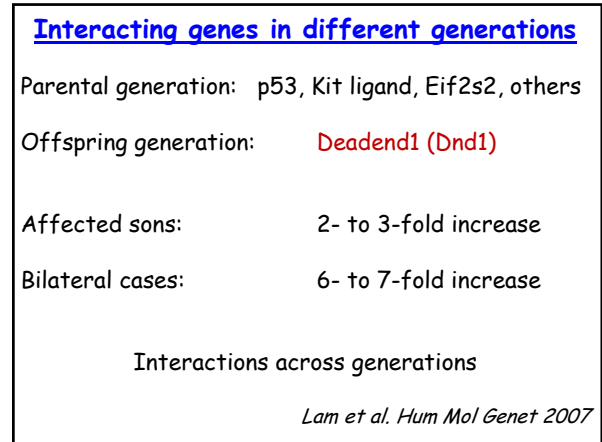
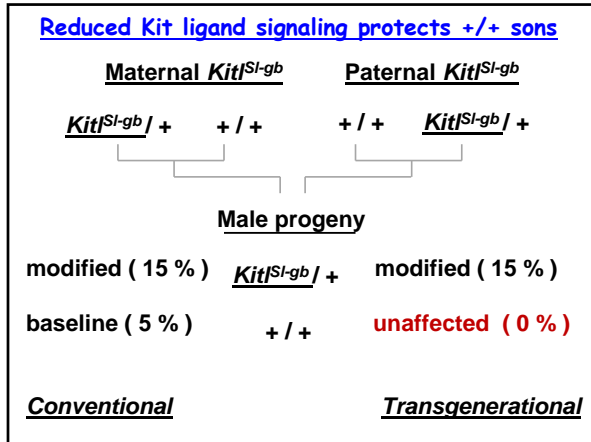
Kit^{Sl-gb} / +** **+ / +** **+ / +** **Kit^{Sl-gb} / +

Male progeny

modified (15 %) ***Kit^{Sl-gb} / +*** modified (15 %)

baseline (5 %) **+ / +**

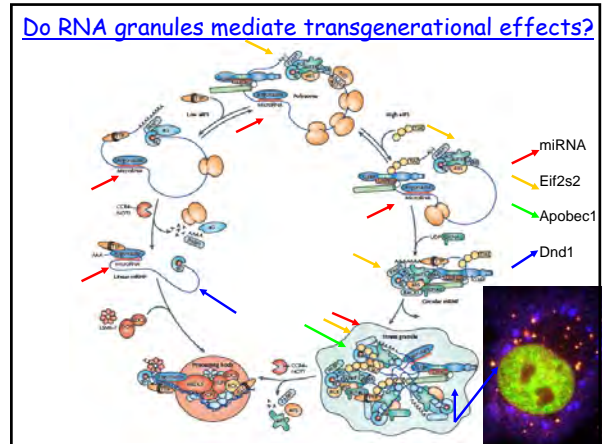
Conventional



Features of Apobec1 effects on heritable epigenetic changes

Wild-type sons of *Apobec1* *ko/+* females show reduced TGCT risk for at least the next three generations

Transgenerational effects can be reversed with two consecutive crosses through the alternative germ-lineage

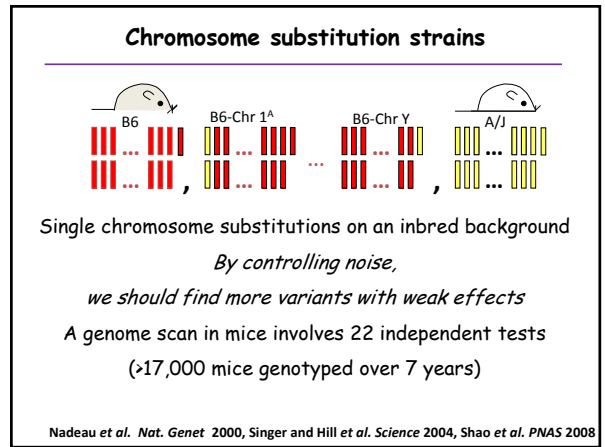


Second

Testicular cancer

Daughters of fathers with different Y chromosomes strong or weak, rare or general?

Diet-induced obesity and feeding habits



Daughters of fathers, different Y chromosomes

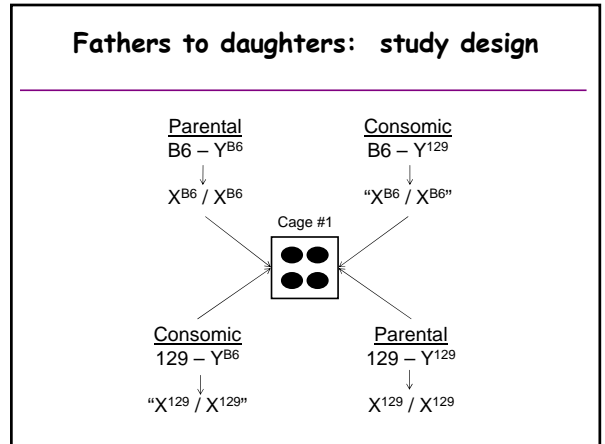
Can we arrange a test for an un-inherited chromosome, a test where daughters differ *only* in the genetics of the fathers' Y chromosome?

Assay: B6.Chr^{A/J} CSSs - 120 traits (and 41 traits in B6.Chr^{PWD})

Females: 22 behavioral, 7 cardiac, 39 hematological

Males: 9 behavioral, 4 cardiac, 4 metabolic, 32 hematological

Singer and Hill et al. Science (2004), Shao et al. PNAS 2008



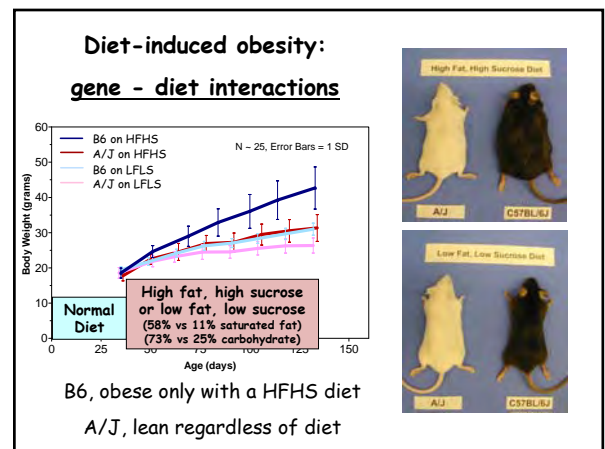
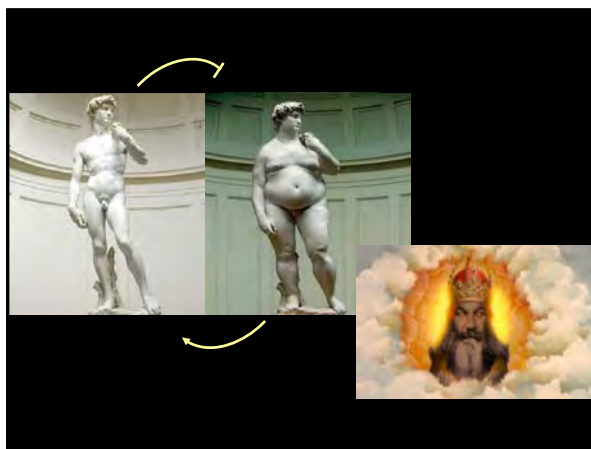
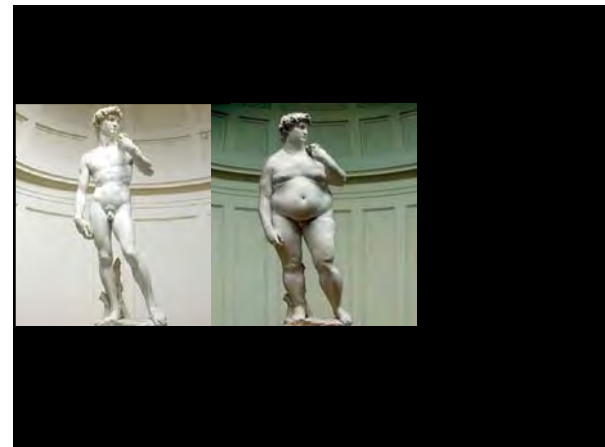
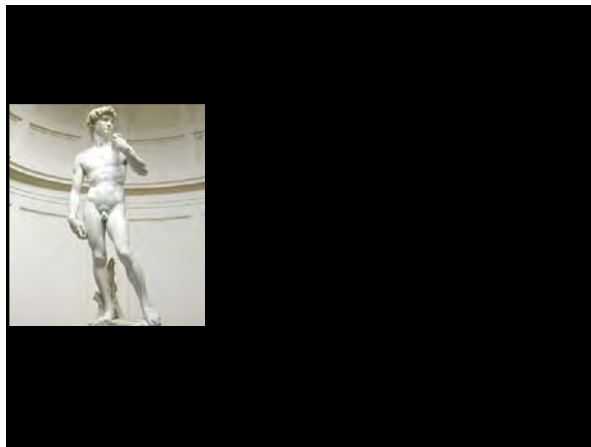
Fathers to Daughters:
 Frequency of significant phenotypic effects and their average effect sizes

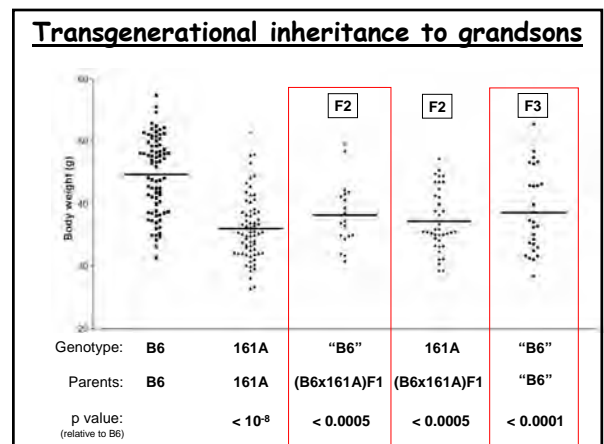
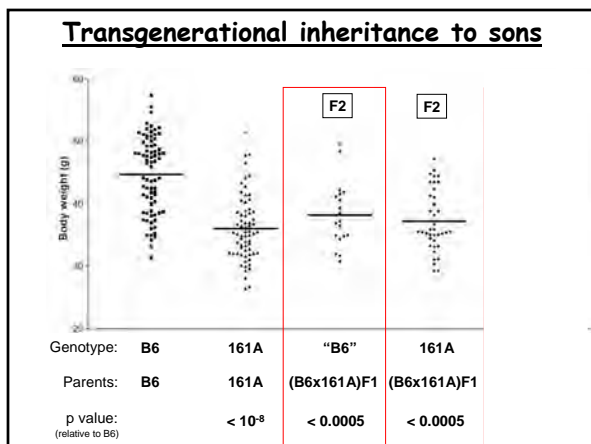
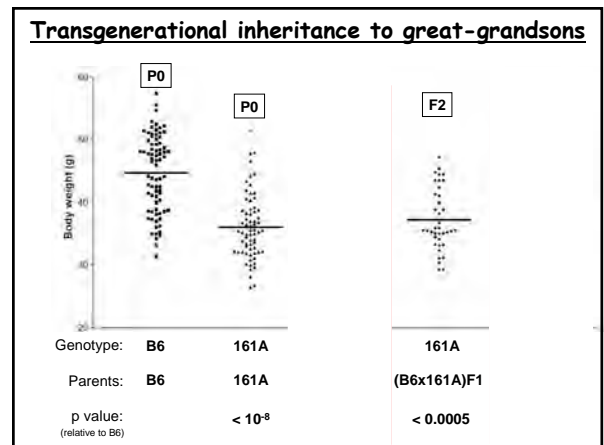
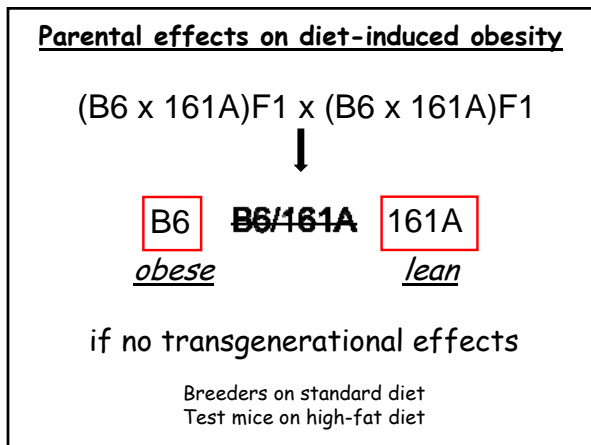
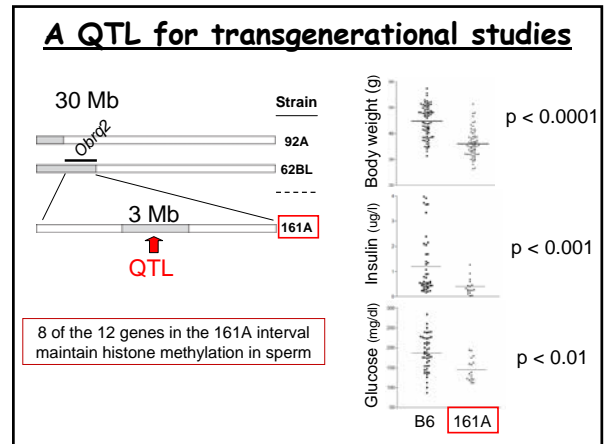
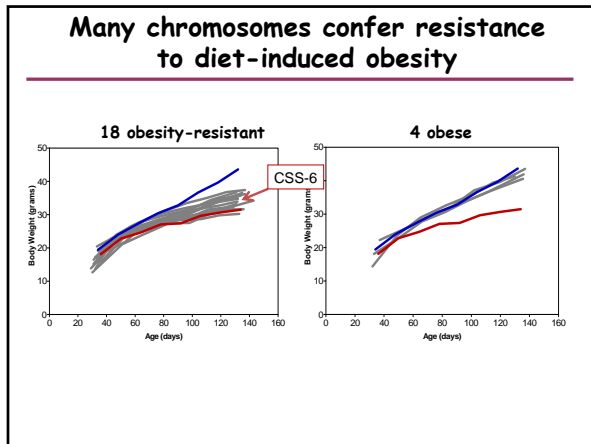
| | Chromosome | Females | | Males | |
|----------------------|-------------|----------|---------------------|----------|---------------------|
| | | % traits | Average effect size | % traits | Average effect size |
| Unconventional | Y (father) | 23 % | 91 % | | |
| | Y (brother) | | | 25 % | 79 % |
| Conventional Effects | X | 22 % | 80 % | 25 % | 93 % |
| | Autosomes | 32 % | 94 % | 25 % | 80 % |

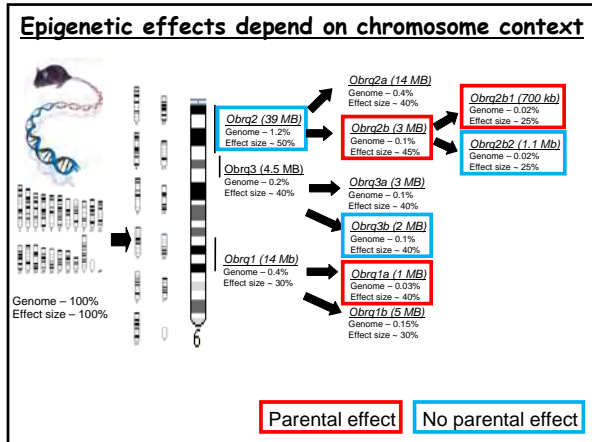
many traits, strong effects

Nelson and Nadeau, *Epigenomics* 2010

Third
Testicular cancer
 Daughters of fathers with different Y chromosomes
 Diet-induced obesity and feeding habits long lasting?
 fractal genetics and gene discovery





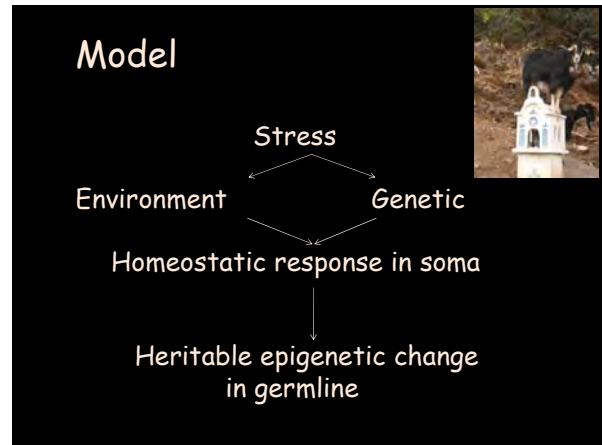


Testicular cancer
 strong, persist at least 3 generations, females

Daughters of fathers with different Y chromosomes
 strong, common, males

Diet-induced obesity and feeding habits
 strong, persists at least 3 generations, males

- ### Essential features
- Ancestral generation**
 - p53, Eif2s2, Kitlg*
 - Heterogeneous genetic functions*
 - Dosage effects, often partial loss of function*
 - Genes in both parents and offspring necessary*
 - Germline**
 - Female, male, or both*
 - Deadend1 and Apobec1*
 - Next generation**
 - At or soon after fertilization*
 - But before embryonic day 3.5*
 - Deadend1 and Apobec1*



- ### Key questions
- Germline, or soma to germline
 - Molecule that is not DNA
 - Mechanisms in the next generation
 - Mechanisms for reversing effects
 - Humans

Jason Heaney
 Vicki Nelson
 Jennifer Zechel
 Steph Doerner
 John Giesinger
 Soha Yazbek
 David Buchner
 Ghunwa Nakouzi
 Paola Raska
 Philip Anderson
 Annie Hill
 Sabrina Spiezio
 Elaine Leung

Josephine Lam, Cleveland Clinic
 Nick Davidson, Washington Univ.

NCI, NCCR and NIH Pioneer Award



Nicholas J. Schork, PhD
Director of Research, Scripps Genomic Medicine
Scripps Health
Director of Biostatistics and Bioinformatics
The Scripps Translational Science Institute
Professor, Department of Molecular and Experimental Medicine
The Scripps Research Institute
La Jolla, California

Biography

Dr. Schork is Professor, Department of Molecular and Experimental Medicine at The Scripps Research Institute; Director of Bioinformatics and Biostatistics at the Scripps Translational Science Institute and Director of Research, Scripps Genomic Medicine.

Dr. Schork's interest and expertise are quantitative human genetics and genomics, especially the design and implementation of methodologies to dissect the genetic basis of complex traits and diseases. He has published over 350 scientific articles and book chapters on the analysis of complex, multifactorial traits and diseases.

Prior to joining Scripps in 2007, Dr. Schork served for seven years as Professor of Biostatistics and Psychiatry, and Co-Director of the Center for Human Genetics and Genomics at the University of California, San Diego. From 1994 to 2000, he was an Associate Professor of Epidemiology and Biostatistics at Case Western Reserve University in Cleveland, Ohio, and an Associate Professor of Biostatistics at Harvard University. During 1999 and 2000, Dr. Schork took a leave of absence to conduct research as the Vice President of Statistical Genomics at the French biotechnology company, Genset, where he helped guide efforts to construct the first high-density map of the human genome.

A member of several scientific journal editorial boards, Dr. Schork is a frequent participant in U.S. National Institutes of Health-related steering committees and review boards, and has served on the advisory board of five companies. In addition, he is currently Director of the Bioinformatics and Biostatistics Core of the National Institute of Aging-sponsored Longevity Consortium as well as the Director of Statistical Genetics for the National Institute of Mental Health-sponsored Bipolar Genetics Study (BiGS) consortium.

Dr. Schork earned the B.A. in Philosophy, M.A. in Philosophy, M.A. in Statistics, and Ph.D. in Epidemiology, all from the University of Michigan in Ann Arbor.

Notes

Healthy vs. Unhealthy Genomes: Individual and Population Perspectives

Nicholas Schork, PhD

Advances in DNA sequencing technologies have made it possible to sequence entire human genomes efficiently and in a cost-effective manner. However, merely generating the sequence data is only a small fraction of what it will actually take to make whole genome sequencing (WGS) useful in clinical and public health settings. Making sense of DNA sequence variations across individuals in terms of their likelihood of contributing to disease susceptibility and drug responsiveness is absolutely crucial. In this talk, methods for 'annotating' or making sense of human DNA sequence variations are discussed and applied to a number of actual individual genomes. In addition, ways of defining individuals at risk for disease and how DNA sequence variation across different human populations impacts the interpretation of disease risk and disease diagnosis are also emphasized.

Notes

Clinical Development Out of Academia Based on Genetic Stratification
Hakon Hakonarson, MD, PhD

Notes

Matthew Herper

Senior Editor, Forbes Magazine

Biography

Matthew Herper has covered science and medicine for Forbes since June 2000, beginning with the tumult and hype surrounding the human genome project, continuing through Vioxx and the other drug safety crises of the past decade. He was the first reporter at a major outlet to focus on the efficacy questions surrounding Vytorin. Lately, the Moore's law-like ascent of DNA sequencing technology has brought his coverage full circle.

Notes

The Collision: What Happens When Medicine Gets Genetic?

Matthew Herper

The advent of cheap DNA sequencing could lead to dramatic changes in medicine. But it also sets the stage for a collision between two cultures and ways of doing business: the population-based approach to evidence-based medicine that has prevailed both in the pharmaceutical and device industries and in medicine at large for the past decade, and a new, more personalized approach that has much of its basis in academic groups and basic science. These two cultures have very different ideas about not only diagnosis but pricing, privacy, and ethics. Comparing them gives some sense of what about medicine is likely to change, and what is not.

Notes

Daniel MacArthur, PhD

Assistant in Genetics
Analytic and Translational Genetics Unit
Massachusetts General Hospital
Boston, Massachusetts

Biography

Daniel MacArthur is a newly appointed Assistant in Genetics at the Analytical and Translational Genetics Unit, Massachusetts General Hospital. He leads a research team developing methods for improved functional interpretation of genetic variation found in large-scale genome sequencing studies, with a particular focus on variants predicted to cause complete loss-of-function (LoF) of human protein-coding genes.

Daniel completed his PhD at the University of Sydney, Australia, on the functional effects and recent evolutionary history of a common LoF variant in the human *ACTN3* gene associated with muscle strength and athletic performance. During a postdoctoral position at the Wellcome Trust Sanger Institute in Hinxton, UK, he built on this analysis by leading an international consortium performing a systematic survey of all predicted LoF variants in 185 individuals, using complete genome sequences generated as part of the 1000 Genomes Project. This work was recently published in the journal *Science*.

Daniel has co-authored 30 peer-reviewed papers, including publications in *Nature Genetics*, *Nature* and *Science*, and three book chapters. He has also written extensively online about personal genomics and direct-to-consumer genetic testing on his blog *Genetic Future* (www.wired.com/wiredscience/author/danielmacarthur), and is co-founder of *Genomes Unzipped* (www.genomesunzipped.org), a collaborative online project seeking to improve public understanding of genomics.

Notes

All Genomes Are Dysfunctional: Gene-Disrupting Variants in Healthy Individuals

Daniel MacArthur

In order to take full advantage of the recent advantages in genome sequencing technology we must be able to accurately interpret the functional effects of the variation we observe in human genomes. In this presentation I will discuss our work on a specific class of human genetic variation: variants predicted to cause complete loss-of-function (LoF) of protein-coding genes. LoF variants play a role in many severe human diseases, and can also provide powerful information about gene function and potential therapeutic targets for common diseases. Surprisingly, recent studies have reported that sequenced genomes, even from apparently healthy individuals, carry hundreds of predicted LoF variants. These variants - such as nonsense SNPs and frameshift indels - are skewed towards low frequencies, suggesting an excess of deleterious variants, consistent with an impact on gene function and human disease risk. However, they are also enriched for sequencing and annotation artefacts, making careful filtering essential for interpretation.

I will present a pipeline for annotating LoF polymorphisms from large-scale human sequence data, which we have applied to 185 whole genomes sequenced by the 1000 Genomes pilot projects. The results of this analysis reveals the major challenges associated with the accurate detection of these variants, emphasising the need for great caution in interpreting the results of clinical sequencing studies.

After stringent filtering, we find that the average European still carries ~100 true LoF variants, ~20 in a homozygous state. As expected, validated LoF variants are enriched at low frequencies, and often alter RNA expression of affected exons. LoF-tolerant genes are less conserved, less tissue-specific and have fewer protein-protein interactions than other genes, and show striking differences in properties to known recessive disease genes. Understanding these differences is potentially extremely valuable for clinical studies: I will describe a machine-learning approach using both LoF-tolerant and known disease genes to help predict whether a novel mutation seen in a patient's genome is likely to be disease-causing.

This presentation will illustrate the profound challenges ahead for the accurate functional interpretation of sequenced genomes. Complete functional annotation will require integrating information from diverse factors including gene expression and alternative splicing, evolutionary conservation, interaction networks and the linkage structure of genetic variants. While difficult, this improved annotation of genetic variation will be essential for the success of large-scale clinical sequencing projects.

Useful references

MacArthur DG, Balasubramanian S, Frankish A, Morris J, Huang N, Jostins L, Albers CA, Pickrell JK, Montgomery SB, Walter K, Zhang Z, Rosenfeld J, Ayub Q, Conrad DF, Zheng H, DePristo MA, Banks E, Handsaker R, Habegger L, Mu X, Khurana E, Hu M, Xue Y, 1000 Genomes Project Consortium, Li Y, Gibbs RA, McCarroll SA, van Heel DA, Dermitzakis ET, Pritchard JK, Barrett JC, Harrow J, Hurles ME, Gerstein M, Tyler-Smith C. (2011) A systematic survey of loss-of-function polymorphisms in human protein-coding genes. *Science*. In press (expected to be published by time of meeting).

Balasubramanian S, Habegger L, Frankish A, MacArthur DG, Harte R, Tyler-Smith C, Harrow J, Gerstein M. (2011) Gene inactivation and its implications for annotation in the era of personal genomics. *Genes Dev.* 25(1):1-10.

MacArthur DG, Tyler-Smith C. (2010) Loss-of-function variants in the genomes of healthy humans. *Hum Mol Genet.* 19(R2):R125-130.

1000 Genomes Project Consortium. (2010) A map of human genome variation from population-scale sequencing. *Nature.* 467(7319):1061-1073.

Notes

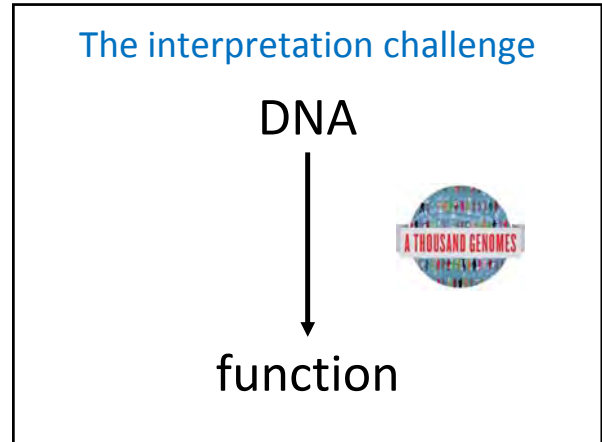


Dysfunction is the normal state:
 loss-of-function variants in
 "healthy" human genomes

Daniel MacArthur
 Wellcome Trust Sanger Institute
 1000 Genomes Project Consortium








- ### Coding loss-of-function variants
- SNPs introducing a premature **stop** codon
 - SNPs disrupting a **splice** site
 - Small indels creating a **frameshift**
 - Large **deletions** disrupting gene structure

- ### Why study LoF variants?
- (Relatively) easy to annotate
 - Enriched for functionality
 - Enriched for artefacts
 - More of them than we expected
- same error / fewer true variants = higher false positive rate


All genomes are dysfunctional



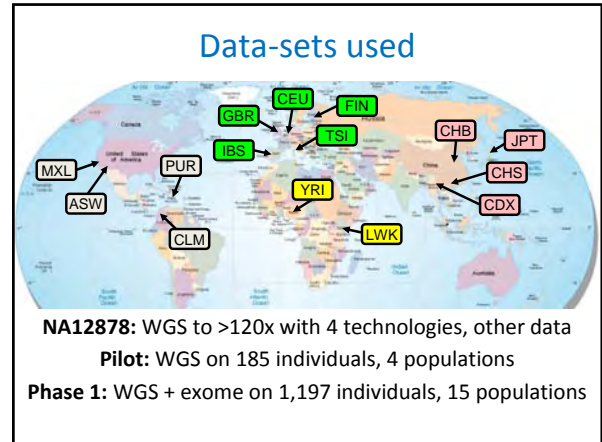
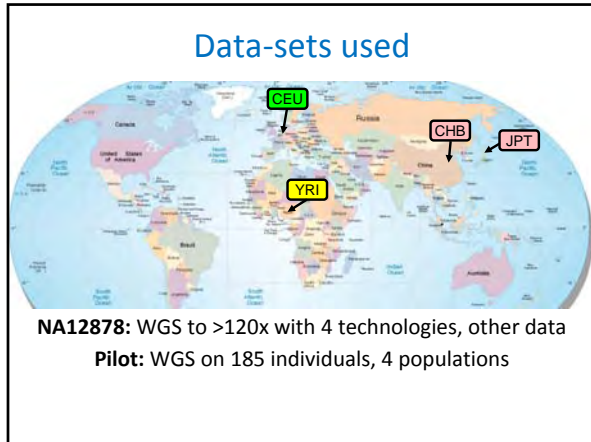
| | | | | |
|---------------|--------------|---------------|---------------|--------------|
| James Watson | Craig Venter | Seong-Jin Kim | Stephen Quake | Desmond Tutu |
| LoF SNPs: 196 | 141 | 130 | 182 | 149 |
| 454 | Sanger | Illumina | Helicos | SOLiD |

No large-scale experimental validation of LoF variants

Data-sets used



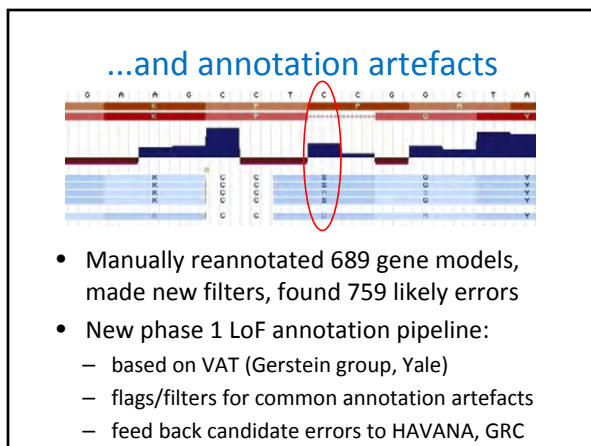
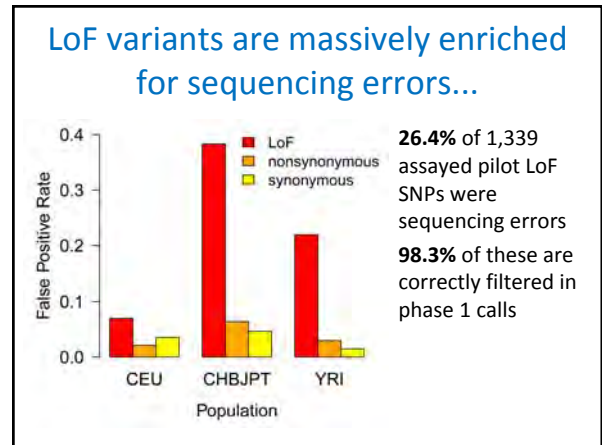
NA12878: WGS to >120x with 4 technologies, other data



Candidate LoF variants in Phase 1

| category | NA12878 | pilot | phase 1 |
|--------------------|------------|--------------|---------------|
| nonsense SNV | 115 | 1,088 | 5,864 |
| splice SNV | 95 | 639 | 3,092 |
| frameshift indel | 348 | 955 | 7,160 |
| large LoF deletion | 31 | 142 | 1,222 |
| total | 654 | 2,824 | 17,338 |

with Suganthi Balasubramanian and Ni Huang




LoF variants cluster in bad regions

| gene | candidate LoF | high-conf LoF |
|-------------------|---------------|---------------|
| <i>AC131157.4</i> | 16 | 0 |
| <i>SSPO</i> | 15 | 0 |
| <i>AC092143.1</i> | 12 | 0 |
| <i>MAN1B1</i> | 12 | 0 |
| <i>CDC27</i> | 10 | 0 |
| <i>MUC19</i> | 10 | 0 |
| <i>C17orf57</i> | 7 | 0 |
| <i>AC009063.1</i> | 7 | 0 |
| <i>AC073957.1</i> | 7 | 0 |
| <i>C11orf40</i> | 6 | 0 |

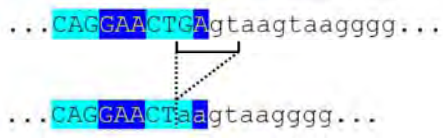
- Genes containing many LoF variants are nearly all artefactual:
 - duplicates not present in reference
 - other mapping errors
 - unannotated pseudogenes
- Can mostly be fixed with additional filters

Importance of haplotype effects: frame-restoring indels

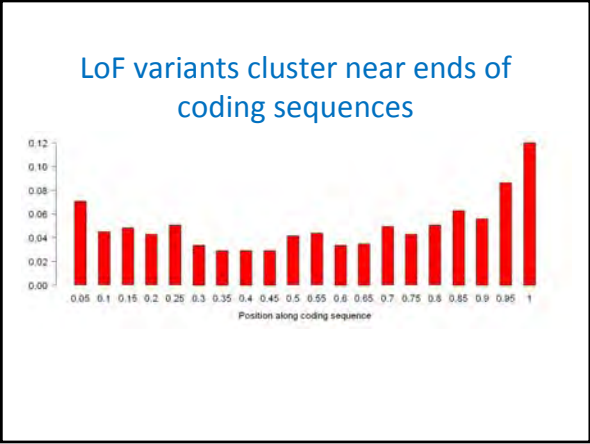
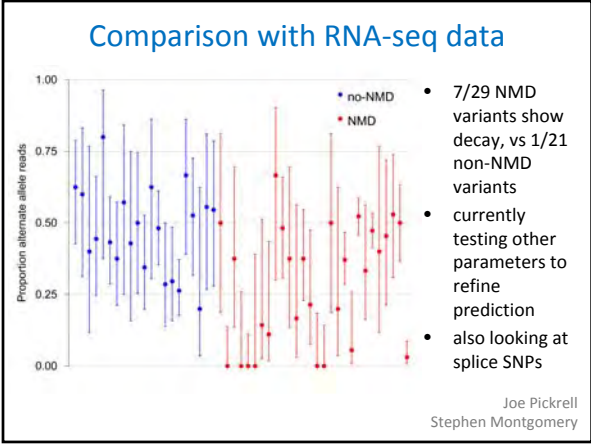
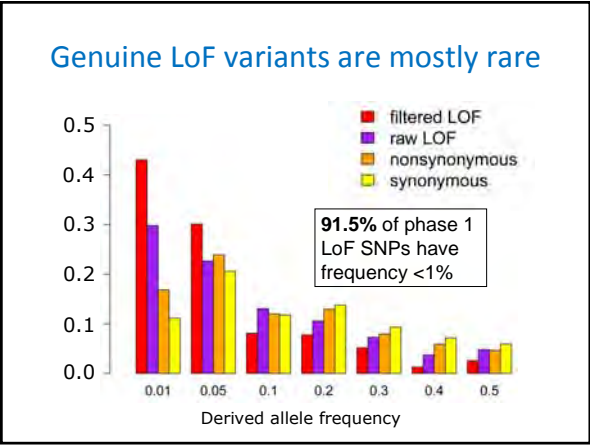
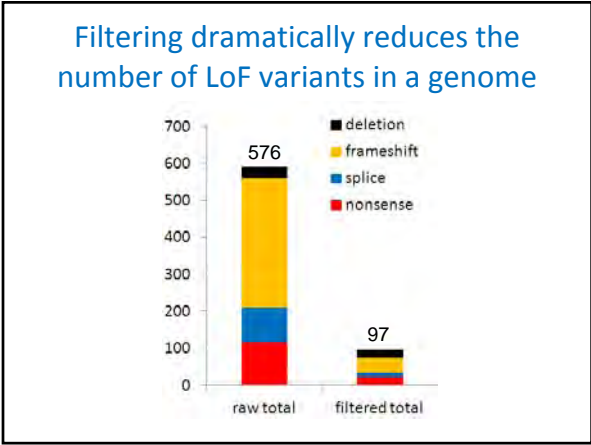


- Two apparent frameshift deletions in the *CASP8AP2* gene (one 17 bp, one 1 bp) on the same haplotype
- Overall effect is in-frame deletion of six amino acids

Variant "rescue" by nearby sequence



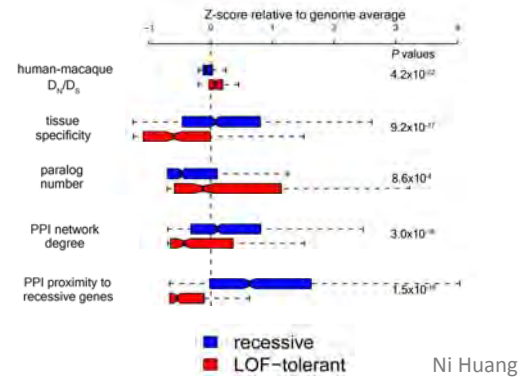
- Four-base deletion spanning a splice site in *CHIT1* is "rescued" by intronic sequence
- Deleted allele has fully intact splice site, and only synonymous substitution



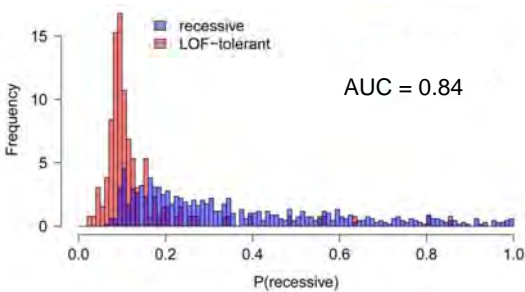
Properties of genuine LoF variants

- Mostly rare: 51.8% of pilot LoF and 91.5% of phase 1 LoF variants have frequency <1%
- Common LoF variants:
 - found in genes that are less conserved, have more paralogues and fewer interactions
 - have minor effect on complex disease risk
- Rare LoF variants:
 - found in more functionally constrained genes
 - include 25 known severe disease mutations, 21 likely (197 known, >500 likely in phase 1 data)
- ~16% of stop variants reduce gene expression

LoF-tolerant vs recessive disease genes



Distinguishing between LoF-tolerant and recessive disease genes



What next?

- Validate LoF annotation/filtering pipeline
- Building a catalogue of high-confidence LoF variants seen in large-scale sequencing studies
- LoF content on custom genotyping arrays
 - Illumina and Affymetrix exome chips
- Genotyping variants in large numbers of phenotyped individuals to characterise effects on human traits and disease risk

Thanks!

1000 Genomes Functional Interpretation Group

Bryndis Yngvadottir, Qasim Ayub, Chris Tyler-Smith, **Sanger**
 Suganthi Balasubramanian, Lukas Habegger, Mark Gerstein, **Yale**
 Adam Frankish and Jennifer Harrow, **Sanger**
 James Morris, Luke Jostins and Jeff Barrett, **Sanger**
 Ni Huang, Klaudia Walter, Don Conrad and Matt Hurles, **Sanger**
 Kees Albers and Richard Durbin, **Sanger**
 Joe Pickrell and Jonathan Pritchard, **U. Chicago**
 Stephen Montgomery and Manolis Dermitzakis, **U. Geneva**
 Mark DePristo and Eric Banks, **Broad Institute**

1000 Genomes Project

Issam Zineh, PharmD, MPH
Associate Director of Genomics
Office of Clinical Pharmacology
United States Food and Drug Administration
Co-Director, Biomarker Qualification Program
Office of Translational Sciences
Silver Spring, Maryland

Biography

Dr. Zineh is Associate Director for Genomics in the Office of Clinical Pharmacology and Co-Director of the Biomarker Qualification Program, CDER, U.S. FDA. He is an experienced clinical pharmacist who was formerly on the faculty of the University of Florida (UF) Department of Pharmacy Practice and Associate Director of the UF Center for Pharmacogenomics. Dr. Zineh received his PharmD from Northeastern University and completed his clinical residency at Duke University Medical Center. He did a fellowship in cardiovascular pharmacogenomics at UF where he also obtained his MPH in Health Policy and Management. He is a recognized expert in the field of clinical pharmacology, pharmacotherapy, and pharmacogenomics. Dr. Zineh is also the chair of the Molecular Pharmacology and Pharmacogenetics Scientific Section of the American Society for Clinical Pharmacology and Therapeutics, chair of the Coriell Personalized Medicine Collaborative Pharmacogenomics Advisory Group, and sits on the Centers for Disease Control and Prevention's EGAPP steering committee. Dr. Zineh currently leads 7 direct reports in CDER's efforts to enhance drug development and regulatory decision making through applied clinical pharmacology, pharmacogenetics, and biomarker science.

Notes



Biomarkers in drug development and regulation: a paradigm for clinical implementation of personalized medicine

The post-genomic era has been hallmarked by significant enthusiasm for therapeutic individualization through the use of pharmacogenomic and other biomarkers. This enthusiasm has been dampened by limited examples of widespread clinical adoption. The current clinical implementation paradigm may not be adequate to facilitate uptake of pharmacogenetics for a variety of reasons. This paper discusses certain limitations of the classical clinical implementation paradigm and describes the drug development paradigm as an additional, powerful mechanism to facilitate clinical implementation of individualized therapeutics.

KEYWORDS: biomarker ■ companion diagnostics ■ pharmacogenetics ■ pharmacogenomics ■ regulatory science ■ US FDA

The 'genomics revolution' has brought with it significant enthusiasm for the betterment of human health through use of biomarkers. Advances in genomic and related technologies have led to parallel advances in environmental health and bioprocessing, forensic science, and agriculture and livestock management. Since the completion of the human genome sequence, no area of genomics has received more attention than that of molecular medicine. The use of genomic and other biomarkers has been touted as the pathway for individualizing healthcare, with endorsement by agencies such as the US FDA and the NIH among many others [1]. The promise of the Human Genomic Project's completion was expected to be realized through better prediction of increased disease risk, improved disease diagnosis, enhanced design of therapeutic products (e.g., drugs and gene therapy), development of customized drugs and enhanced therapeutic decision-making capability.

Despite the potential for therapeutic individualization, sciences like pharmacogenetics have not enjoyed widespread adoption. Attitudinal factors cannot fully explain this lack of clinical translation. For example, a recent survey by Haga and colleagues suggests that the US public are quite supportive of using pharmacogenetic testing to select treatment, guide dosing and predict risk of adverse drug events, especially if the disclosure of information without permission could be avoided [2]. There is also significant agreement among clinicians that knowing a patient's genetic profile could help to inform treatment decisions, with the caveat that many clinicians feel they currently do not have the training and education to use and interpret test

results appropriately [3,4]. We submit that educational, attitudinal or cultural factors cannot be singly responsible for the lack of pharmacogenetic implementation into the clinic. Rather, the current clinical implementation paradigm (herein referred to as the 'classical' paradigm) may not be adequate to fully facilitate uptake of pharmacogenetics for a multitude of reasons.

This article highlights the limitations of the classical clinical implementation paradigm and further highlights the drug development paradigm as an additional, powerful mechanism to facilitate clinical implementation of individualized therapeutics. It also discusses regulatory issues relevant to the use of biomarkers in medicine, including the development of companion diagnostics for treatment decisions. While this article focuses on pharmacogenetic biomarkers for discussion's sake, the following principles and assumptions may be applied to other biomarkers.

'Classical' clinical implementation paradigm

The 'classical' paradigm for translation of pharmacogenetic findings into practice is contingent on biomarker research conducted largely by academic investigators, with subsequent dissemination of information in peer reviewed journals, evidence assessment by technology assessors (e.g., insurance companies, professional societies and various third parties), guideline development and subsequent deployment in the clinic (with or without reimbursement). There are substantial obstacles to the implementation of any new intervention [5,6]. These obstacles may limit the application of biomarkers in medical

Issam Zineh* & Shiew-Mei Huang

Office of Clinical Pharmacology, Office of Translational Sciences, Center for Drug Evaluation and Research, US FDA, 10903 New Hampshire Avenue, Building 51 – Room 3182, Silver Spring, MD 20993, USA

**Author for correspondence:*

Tel.: +1 301 796 4756

Fax: +1 301 847 8720

Issam.zineh@fda.hhs.gov

future
medicine part of fsg

decision-making to varying degrees depending on the clinical and healthcare contexts. We explore some of these barriers, largely with respect to the structure of current healthcare systems and practices in the USA.

A clinician's ability to fully consider and apply a biomarker test for a given patient may be limited by organizational and logistical barriers. For example, clinicians with high patient volumes and a limited amount of time per patient may not be able to meaningfully tailor treatment and counseling decisions based on full consideration of a clinically useful biomarker. This is compounded in health systems and practices that are not process engineered to make test ordering easy. Furthermore, clinicians often have limited training and confidence in their ability to order, interpret and apply the results of a biomarker test [3]. Clinicians' discomfort in their knowledge of genomic test availability and their ability to interpret genomic test results could clearly create an aversion to the use of these biomarkers in practice, thereby limiting implementation.

Another obstacle to implementation is the general paucity of consensus guidelines for the use of pharmacogenetic information in practice [7]. While some examples of treatment guidelines that incorporate pharmacogenetic test recommendations exist (e.g., in the field of HIV and cancer), the numbers are limited. The limited number of professional practice guidelines and consensus statements that advance biomarker use in medicine may be a consequence, in part, of differences in opinion as to the clinical utility of a particular biomarker. These differences are likely due to the variable frameworks for evidence assessment used by different evidence assessors. For illustration, we describe two paradigmatically different approaches to genomic biomarker utility assessment.

Perhaps the most well-known framework for evidence assessment in genomic medicine

is the one defined by the Centers for Disease Control and Prevention-supported Evaluation of Genomic Applications in Practice and Prevention (EGAPP) working group. EGAPP has published its hierarchy for evidence assessment in the domains of analytical validity, clinical validity and clinical utility [8]. Definitions for these three domains, which are widely used in health policy discussions, are given in TABLE 1.

To date, the EGAPP working group has disseminated eight evidence reports in the areas of breast and ovarian cancer, colorectal cancer and Lynch syndrome, cardiovascular and thromboembolic disease, and depression [101]. The majority of these evidence evaluations and recommendations have been in the area of genetic screening for disease and prognosis, as opposed to pharmacogenetics. Notwithstanding, in most cases, EGAPP's reports have concluded that there is insufficient evidence to recommend for or against use of the tests assessed. These reports have stimulated other groups to develop their own models for evidence assessment and guideline development in order to catalyze the transition from research to practice. For example, the NIH-supported Clinical Pharmacogenetics Implementation Consortium (CPIC) is a recently formed cadre of multidisciplinary scientists and clinicians with the goal of generating guidelines and recommendations for the use of pharmacogenetic information in the clinic [9]. CPIC focuses solely on pharmacogenetics rather than disease genetics. CPIC also uses a fundamentally different assumption in its utility assessments than EGAPP; while EGAPP essentially asks whether data support wholesale implementation of genetic testing for a given test, CPIC assumes that the test information will be available. That is, EGAPP asks "is testing likely to be useful?" while CPIC asks "given that you have the test results, is the information likely to be useful?" By taking this different perspective on the key

Table 1. Dimensions of biomarker utility assessment.

| Domain | Definition | Associated metrics |
|---------------------|--|--|
| Analytical validity | Ability of the test to accurately/reliably measure the genotype/analyte in the clinical laboratory and in specimens representative of the population of interest | Analytical sensitivity, specificity and reproducibility |
| Clinical validity | Ability of the test to accurately/reliably predict the phenotype of interest | Clinical sensitivity, specificity, negative predictive value and positive predictive value |
| Clinical utility | Evidence of the ability of the test to improve clinical outcomes and its usefulness/added value to patient management decisions compared with current management without testing | Test effectiveness, net benefit. May encompass dimensions of personal, family or community utility |

Definitions derived from the Evaluation of Genomic Applications in Practice and Prevention Initiative [8].

question, CPIC has been able to generate two published pharmacogenetic guidelines to date: one on the use of *TPMT* genotype information to guide thiopurine dosing and another on the use of *CYP2C19* genotype information to manage antiplatelet therapy [10,11], with several others forthcoming. CPIC uses a hybrid approach to assessment and recommendation by adopting elements from the National Academy of Clinical Biochemistry and NIH for quality of evidence and strength of recommendation, respectively [9]. These two methods for evidence assessment (i.e., EGAPP's and CPIC's approach) are two of many. The dialog is an order of magnitude more complicated when pharmacoeconomic metrics are entered into the assessment framework [12].

Even when a consensus exists for the utility of a biomarker's use in practice, clinicians may not be able to successfully integrate use of the biomarker without interpretive algorithms that provide simple recommendations. In the absence of these decision-support tools that direct clinicians when to order a test and how to interpret the results in a patient-specific way, clinicians will be unlikely to be able to deal with the rapid proliferation of pharmacogenetic information. Decision-support tools must be integrated into the clinician's flow of activities, provide ease of test ordering, and report interpretable, jargon-free results. Informatics capability is likely to be a key determinant of the successful implementation of therapeutic individualization [13,14].

The classical clinical implementation paradigm has had limited success to date for one simple reason: it largely fails to holistically address the above obstacles to implementation and instead tends to focus only on one or two aspects. Increasing clinician education on pharmacogenetics, for example, may be meaningless without the ability for clinicians to easily order tests. Making it easier to order pharmacogenetic tests may not be fully satisfactory without the decision-support tools to interpret and translate the results in a patient-specific manner. Evidence assessment may be an exercise in futility if the data are not generated from populations relevant to the population of the health system or practice interested in implementing the technology. We present idealized system requirements for the successful implementation of pharmacogenetic and other biomarkers in practice in FIGURE 1. Innovations in research design and methodology are needed to generate user-friendly information portable to a health system or practice interested in assessing evidence in its own environment. For example, a combination of randomized

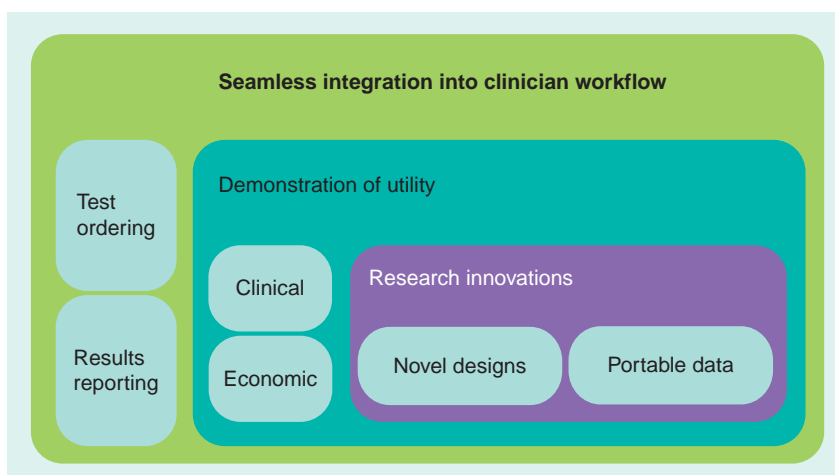


Figure 1. Genomics in healthcare. Innovations in research design and methodology may be needed to generate user-friendly information portable to a health system or practice interested in assessing evidence in the context of its own environment. The system or practice may then generate its own evidence as to the clinical and economic utility of biomarker information. Test ordering and results reporting for useful biomarkers can then be integrated into the clinician workflow to allow for individualization of therapy.

controlled trials, high-quality observational studies and targeted mechanistic (e.g., clinical pharmacology) studies can be used to generate a portfolio of evidence to be rated by a health system in terms of relevance to its own practices, patient populations, drug use data and other factors. The system or practice may then generate its own data on the clinical and economic utility of the biomarker, using what it determines to be important metrics. This creates a more meaningful, microcosmic assessment of utility; it also bypasses the need for national or international guideline development prior to implementation. Test ordering and results reporting for biomarkers vetted as useful by a given institution or system can then be integrated into its clinicians' workflow to allow for individualization of therapy.

Drug development as an additional paradigm

Despite its limitations, the classical implementation paradigm is important in the advancement of personalization. A complementary mechanism (not without its own limitations, discussed later), which allows integration of biomarkers into patient selection, dose optimization and risk minimization, is the direct incorporation of biomarkers into the drug development enterprise.

Zineh and Pacanowski have previously described the myriad ways in which biomarkers can be used in drug development [15]. Prior to drug approval, biomarkers are often used to demonstrate a drug candidate's preclinical activity

and toxicity, help establish proof of concept that the drug is active in people, provide early insights into a drug's safety in people and elucidate the drug's dose- or exposure-response relationship. In addition, biomarkers may be used to select patients that are at a higher risk of experiencing the clinical event of interest (prognostic enrichment) or have a higher likelihood of experiencing drug benefit (predictive enrichment) [16]. Prognostic enrichment may allow for increased power to demonstrate a drug benefit (if one exists) by increasing event rates or shortening the time to events. Predictive enrichment is based on some patient feature that makes them more likely to experience pharmacological benefit (e.g., they express the molecular target of the drug or have the drug metabolizing phenotype necessary to ensure adequate drug concentrations). Genomic and other biomarkers may be the enrichment factor of interest, and if the benefit of the drug is successfully demonstrated in the subpopulation of interest (e.g., biomarker-positive individuals), there would be significant implications for using the biomarker test and the drug together when prescribing.

Successful incorporation of patient-selection biomarkers into drug development addresses some of the limitations of the classical implementation paradigm. If the pharmacogenetic (or other) biomarker is intended to select patients for treatment, its utility will, by definition, be subject to conservative, rigorous, multidisciplinary assessment as part of the New Drug Application or Biologics License Application review. For a submission that includes a pharmacogenetic biomarker to select patients for treatment, the New Drug Application and Biologics License Application review teams are comprised of clinicians, pharmacogeneticists, nonclinical pharmacologists and toxicologists, clinical pharmacologists, biostatisticians, chemists, *in vitro* diagnostic assessors and other scientists [17]. These reviewers are, by statute, charged with determining whether the drug's benefits and risks, and the diagnostic's performance, are adequate to approve the drug and test for use in the public. In essence, the FDA's product oversight and review process is a built-in evidence assessment mechanism that ensures broad perspectives on the utility of a drug test pairing by a multidisciplinary team whose sole focus is the protection and promotion of the public's health.

In the classical implementation paradigm, consensus groups try to synthesize information from disparate data sources in order to generate practical recommendations. Data can range from

those derived through multiple randomized clinical trials, meta-analyses, a single randomized trial, observational studies or case studies. The heterogeneity of these data sources makes it difficult to provide clear recommendations for the use of biomarkers in medicine, especially if there is conflicting evidence from multiple data sources. In the drug development paradigm, this problem is circumvented to some degree. The evidence assessors (in this case FDA reviewers) are actively involved in consultation and advice with drug and diagnostic companies during the Investigational New Drug and Investigational Device Exemption review stages. In these stages, FDA reviewers provide drug or diagnostic companies with a scientific perspective on what evidence would be required to support the utility of the biomarker for regulatory decision-making purposes (e.g., approval). In this respect, evidence assessors (FDA reviewers) and evidence generators (drug and diagnostic companies) are trying to achieve a consensus on what data are needed to support biomarker utility before the totality of the data are generated. This *a priori* consideration of utility assessment parameters could be valuable if a similar mechanism can be established in the classical implementation paradigm.

Another advantage of prospective incorporation of selection biomarkers in drug development is that drug approvals will be accompanied by clear instructions for prescribing in the form of the drug product label (package insert). By law, prescribing information in the form of the label must meet three key general requirements intended to provide accurate and balanced information to prescribers. Specifically, labeling must: contain a summary of the essential scientific information needed for the safe and effective use of the drug; be informative and accurate and neither promotional in tone nor false or misleading in any particular; and be based whenever possible on data derived from human experience. No implied claims or suggestions of drug use may be made if there is inadequate evidence of safety or a lack of substantial evidence of effectiveness [102]. When appropriate, the label must define subpopulations with altered benefit/risk balance and describe whether a test is needed to best use the drug product. Provision of this information in the drug package insert may limit the ambiguity on how a test should be used to guide treatment decisions, a pervasive problem in the classical paradigm of translation.

There are other advantages to the simultaneous co-development of a drug and biomarker

test for patient selection. If the biomarker test is a true companion diagnostic (i.e., is required for patient selection), then testing would be a condition of use and there would not be a need to wait for community vetting in the form of guideline development by professional societies, which may be a protracted process. In addition, approval of a drug test pair may trigger a series of downstream events that could make it easier for clinicians to implement personalization. For example, tests of sufficient analytical quality would have to be available for use, which may not always be the case in the classical implementation paradigm. Additionally, FDA-cleared or -approved tests may be more easily reimbursed than tests that have not been subject to FDA review. Finally, decision-support tools might be more easily deployed when there is a drug whose use is clearly predicated on biomarker test results.

The drug development-based implementation paradigm is not without its own limitations. While investigator-initiated research in the classical paradigm is driven by scientific interest, the choice of which drugs to co-develop with a pharmacogenetic test will be driven largely by commercial interests. There is nothing inherently problematic with this impetus. It should be appreciated, however, that there may be scientifically viable candidates for co-development that may not be pursued because of the low likelihood for commercial success. Another limitation of the drug development/FDA review paradigm is that the FDA does not have unlimited capacity to provide pharmacogenetic guidance to drug developers throughout the various phases of drug development for all drug development programs. As a metric of growth in the field, the Genomics Group in the FDA's Office of Clinical Pharmacology has experienced a nearly 400% increase in review volume in recent years [15,17]. Unlike in the classical paradigm, where there is a whole scientific community involved in the peer review of a biomarker's utility, the FDA is limited by its review staff capacity, although this may change in the future.

Other limitations of the drug development paradigm should be mentioned. The volume of data in support of a drug's effectiveness and safety prior to approval is likely to be considerably lower than the data that emerge in the postapproval setting. This may also be true in terms of the data in support of a biomarker's utility. Therefore, the selection biomarker used to support the original approval of a drug in a targeted subset of the population may not ultimately be the most useful selection biomarker that emerges

after the drug is approved. As one example, while not described as a selection marker *per se* in the 'indications and usage' section of the original erlotinib label, EGFR protein expression status as a correlate of drug response was prominently featured in the 'Clinical studies' section of the label [103]. Inclusion in the label reflected the contemporary thinking supported by evidence at the time that knowing a tumor's EGFR expression level as measured by immunohistochemistry was important. Most of the description of basing erlotinib benefit on EGFR expression, however, was removed in subsequent versions of the drug label (2010), reflecting less confidence in EGFR expression status as an important determinant of efficacy with evolving data [103].

This moving target of biomarker utility certainly presents an implementation problem when the practicing community is looking to the drug package insert for guidance. Notwithstanding, the clinical community (especially in highly specialized therapeutic areas with significant morbidity or mortality rates such as cancer, HIV and hepatitis C virus) may be on the leading edge of individualization, even in the absence of guidance from the drug label. For example, the oncology community endorsed *KRAS* mutation testing of certain cancers prior to treatment with anti-EGFR monoclonal antibodies before the FDA's approval to update the cetuximab and panitumumab labels with this information [18]. The oncology guidelines also recommend *EGFR* mutation testing prior to the use of EGFR tyrosine kinase inhibitors in patients with non-small-cell lung cancer in certain settings, while there is no recommendation for testing in the current FDA-approved tyrosine kinase inhibitor labels [19,20]. The clinical practice guidelines for HIV management recommended HLA testing prior to abacavir treatment because of the known pharmacogenetic association between *HLA-B*5701* and abacavir-induced hypersensitivity in advance of the FDA label update [21]. While in some cases, the FDA has updated the label with pharmacogenetic information in advance of treatment guideline recommendations (e.g., clopidogrel and *CYP2C19* gene variants), these examples are rare.

One final limitation that should be noted has to do with drug labeling for already approved agents. In the context of prospective co-development, a diagnostic test with clear performance metrics and definitions for positivity and negativity is typically used. However, when postapproval label updates are undertaken, data from disparate sources are used. These data, often

from the biomedical literature, are heterogeneous with respect to what genes and alleles are tested, how phenotype (e.g., poor, intermediate and extensive metabolizer) is derived from genotype information, how race and ethnicity are treated in analyses, and other critical issues. It may, therefore, be a challenge to optimally capture this information in drug product labels in a way that makes the information digestible and actionable. In this regard, postapproval label updates may be limited in their ability to directly facilitate implementation.

For the reasons above, the drug development paradigm should be considered an important additional mechanism to advance implementation of biomarkers in medicine, but cannot be considered the sole mechanism. Co-development of drugs and diagnostics and the classical implementation paradigm should be considered complementary.

Companion diagnostics

While there are many ways in which biomarker information may be used to enhance therapeutic decisions, one of the most promising is the ability to determine which therapy a patient should receive based on a biomarker profile. Regardless of the implementation paradigm (i.e., classical or drug development), if a test is required before therapeutic selection can be made it is critical that the test be well characterized so as to limit the possibility that patients who should get the drug do not and *vice versa*. Tests that are required for the safe and effective use of treatments are considered by the FDA to be companion diagnostics.

The FDA recently released guidance that represents its current policy on companion diagnostics [104]. In the drug development context, the FDA has taken the general position that if a companion diagnostic is required for therapeutic selection, then an FDA-approved or -cleared test will be required at the same time that the drug is approved. To this end, the FDA has recently simultaneously approved a *BRAF* V600 mutation test with vemurafenib, a drug with demonstrated activity in patients with melanomas that harbor specific *BRAF* mutations. The FDA has also recently simultaneously approved a diagnostic that tests for *ALK* translocations along with crizotinib, a drug with demonstrated activity in non-small-cell lung cancer patients whose tumors harbor specific *ALK* genomic aberrations.

While the FDA expects that a companion diagnostic be approved or cleared at the same time as the drug, there are two notable exceptions. First, the FDA may approve a drug in

the absence of an FDA-approved or -cleared companion diagnostic if the drug is intended to treat an unmet medical need associated with significant morbidity or mortality, such that the need for the therapeutic outweighs the potential risk of not having the FDA-approved or -cleared test at the time of drug approval. Second, if a drug is already approved, the FDA may update the label with appropriate test-related information if such information represents an important public health communication. Notably, the FDA does not define tests that are “intended to provide information that is useful to the physician regarding the use of a therapeutic product, but that are not a determining factor in the safe and effective use of the product” as companion diagnostics [104]. Whether or not a biomarker test is FDA cleared or approved, tests should be analytically valid and clinically actionable in order to maximize the impact of personalized medicine.

Conclusion

Despite significant potential for transforming patient care, the use of biomarkers (particularly pharmacogenetic biomarkers) to make treatment decisions is not routine. This has led to earnest efforts to overcome barriers to translation. These barriers, which may be educational, logistical, organizational, financial, evidentiary or informatic in nature, will need to be addressed in a holistic fashion rather than in dissected parts. The classical clinical implementation paradigm will need to be supplemented with additional paradigms to advance the field of therapeutic individualization. Prospective incorporation of patient selection biomarkers in drug development, although not without its limitations and caveats, represents an important additional paradigm for consideration. In all likelihood, a systematized approach to innovative research, coupled with user-friendly applications that translate findings into actionable clinical recommendations, will be required.

Future perspective

There have been significant advances in our understanding of underlying disease mechanisms. Through the first quarter of 2011, there have been over 1300 published genetic associations with over 200 disease and drug response traits from genome-wide association studies [105]. We can only expect our understanding of disease biology to increase, and with that so will our options of ‘druggable’ targets. Targeting therapies to molecular subtypes of disease has seen its biggest successes in oncology, and is evidenced

by the most recent accelerated approvals of vemurafenib and crizotinib in specific patient populations based on their tumor profiles. We expect an increasing number of drug development programs in oncology to be targeted to specific populations based on molecular abnormalities that can be biomarker defined. The oncology successes will not be isolated. We predict that there will be increasing examples of co-development in other therapeutic areas. The next therapeutic area ripe for biomarker-based innovation in trial designs is likely to be hepatitis C virus drug development. While the viral genetic determinants of antiviral therapy have been well appreciated, it is only recently that we have become aware of the importance of host genetics in drug response [22,106]. Oncology and antiviral pharmacotherapy may serve as prototypes for personalized medicine in the future.

Drug development success rates have been abysmal for diseases that exhibit marked heterogeneity or for which the exact pathobiology has not been characterized. Understanding the genetic architecture of these diseases is critical to allow for biomarker-based enrichment strategies to be employed for prognostic or predictive purposes in the future. Where the optimal biomarker–drug pair is not exactly known, trials with adaptive designs may be increasingly used to increase efficiency of biomarker science.

Progress in drug development will not occur in a vacuum. The classical and drug development-based implementation paradigms will undoubtedly feed into one another. Precompetitive industry and academic research in genetic epidemiology and emerging sciences such as epigenetics and metabolomics will enhance our understanding of disease biology and drug action. Furthermore,

Executive summary

Background

- The use of genomic and other biomarkers has been touted by the US FDA and others as a way to individualize healthcare.
- The classical clinical pharmacogenetics implementation paradigm may not be adequate to fully facilitate uptake of genomic biomarkers in medicine for a multitude of reasons.

'Classical' implementation paradigm

- The classical paradigm is contingent on biomarker research conducted largely by academic investigators, with subsequent dissemination of information in peer-reviewed journals, evidence assessment by technology assessors, guideline development and subsequent deployment in the clinic.
- Obstacles to implementation under this paradigm include logistical barriers, difficulty of test ordering, limited clinician training, paucity of consensus guidelines and limited decision-support tools.
- System requirements for successful implementation are likely to require incorporation of test ordering and interpretation of results seamlessly into clinician's workflow.

Drug development as an additional paradigm

- Another mechanism (not without its own limitations) by which to integrate biomarkers into patient selection, dose optimization and risk minimization is through direct incorporation of biomarkers into drug development.
- Drug development and regulatory review of co-developed products may represent a paradigm that circumvents some of the limitations of the classical implementation paradigm.
- Limitations of the drug development paradigm also exist and it should, therefore, be considered an important additional mechanism to advance implementation of biomarkers in medicine, but not replace the classical paradigm.

Companion diagnostics

- If a test is required before therapeutic selection can be made it is critical that the test be well characterized.
- Tests that are required for the safe and effective use of treatments are considered by the FDA to be companion diagnostics.
- The FDA recently released a guidance that represents its current policy on companion diagnostics. With two exceptions, the FDA generally expects that a companion diagnostic be approved or cleared at the same time as the drug.

Conclusion

- Barriers to implementation (educational, logistical, organizational, financial, evidentiary and informatic) will need to be addressed in a holistic fashion. The classical clinical implementation paradigm will need to be supplemented with additional paradigms to advance the field of therapeutic individualization.

Future perspective

- We expect an increasing number of drug development programs to be targeted to specific populations.
- Understanding the genetic architecture of diseases is critical to allow for biomarker-based enrichment strategies to be employed for prognostic or predictive purposes in the future.
- As biomarker science advances, so too will the multidimensional nature of utility assessment. Community standards for evidence generation and assessment will be different in 5 years from what they are today.

advances in bioinformatic analyses of publicly available data will lead to rational repurposing of already approved drugs for new indications, or previously abandoned compounds [23,24].

Finally, as biomarker science advances, so too will the multidimensional nature of utility assessment. Complex frameworks of risk–benefit assessment for targeted therapies that include economic indices as a standard matter may be forthcoming, especially in economically constrained health systems. Community standards for evidence generation and assessment will be different in 5 years from what they are today, as today they are different from 5 years ago. Ultimately, the next decade is likely to determine

whether the promise of the postgenomics era can be counted among the hyperbole of discoveries past, or watershed advancement in translational science.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as:

▪ of interest

▪▪ of considerable interest

- Hamburg MA, Collins FS. The path to personalized medicine. *N. Engl. J. Med.* 363, 301–304 (2010).
- Haga SB, O'Daniel JM, Tindall GM *et al.* Survey of US public attitudes toward pharmacogenetic testing. *Pharmacogenomics J.* doi:10.1038/tpj.2011.1 (2011) (Epub ahead of print).
- Shields AE, Lerman C. Anticipating clinical integration of pharmacogenetic treatment strategies for addiction: are primary care physicians ready? *Clin. Pharmacol. Ther.* 83, 635–639 (2008).
- Feero WG, Green ED. Genomics education for healthcare professionals in the 21st century. *JAMA* 306, 989–990 (2011).
- Pirmohamed M. Acceptance of biomarker-based tests for application in clinical practice: criteria and obstacles. *Clin. Pharmacol. Ther.* 88, 862–866 (2010).
- Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 362, 1225–1230 (2003).
- Zineh I, Lesko JL. Pharmacogenetics in medicine: barriers, critical factors and a framework for dialog. *Personalized Med.* 6, 359–361 (2011).
- **Highlights perceived barriers to clinical implementation of pharmacogenetics and proposes a question-based framework for clinical utility assessment.**
- Teutsch SM, Bradley LA, Palomaki GE *et al.* The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP working group. *Genet. Med.* 11, 3–14 (2009).
- **Elucidates the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) evidence assessment methodology.**
- Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin. Pharmacol. Ther.* 89, 464–467 (2011).
- **Introduces the fairly newly formed Clinical Pharmacogenetics Implementation Consortium (CPIC), which uses pre-emptive genotyping as its underlying assumption in its clinical utility assessments.**
- Relling MV, Gardner EE, Sandborn WJ *et al.* Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing. *Clin. Pharmacol. Ther.* 89, 387–391 (2011).
- Scott SA, Sangkuhl K, Gardner EE *et al.* Clinical Pharmacogenetics Implementation Consortium Guidelines for cytochrome P450–2C19 (CYP2C19) genotype and clopidogrel therapy. *Clin. Pharmacol. Ther.* 90, 328–332 (2011).
- Beaulieu M, de Denus S, Lachaine J. Systematic review of pharmacoeconomic studies of pharmacogenomic tests. *Pharmacogenomics* 11, 1573–1590 (2010).
- Overby CL, Tarczy-Hornoch P, Hoath JI *et al.* Feasibility of incorporating genomic knowledge into electronic medical records for pharmacogenomic clinical decision support. *BMC Bioinformatics* 11(Suppl. 9), S10 (2010).
- Swen JJ, Wilting I, de Goede AL *et al.* Pharmacogenetics: from bench to byte. *Clin. Pharmacol. Ther.* 83, 781–787 (2008).
- Zineh I, Pacanowski MA. Pharmacogenomics in the assessment of therapeutic risks versus benefits: inside the United States Food and Drug Administration *Pharmacotherapy* 31, 729–735 (2011).
- **Transparently highlights the application of pharmacogenomic and biomarker principles as seen in the US FDA's recent experience.**
- Temple R. Enrichment of clinical study populations. *Clin. Pharmacol. Ther.* 88, 774–778 (2010).
- Zineh I, Woodcock J. The clinical pharmacogeneticist: an emerging regulatory scientist at the US Food and Drug Administration. *Hum. Genomics* 4, 221–225 (2010).
- Allegra CJ, Jessup JM, Somerfield MR *et al.* American Society of Clinical Oncology provisional clinical opinion: testing for *KRAS* gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *J. Clin. Oncol.* 27, 2091–2096 (2009).
- Keedy VL, Temin S, Somerfield MR *et al.* American Society of Clinical Oncology provisional clinical opinion: epidermal growth factor receptor (*EGFR*) mutation testing for patients with advanced non-small-cell lung cancer considering first-line *EGFR* tyrosine kinase inhibitor therapy. *J. Clin. Oncol.* 29, 2121–2127 (2011).
- Pirker R, Herth FJ, Kerr KM *et al.* Consensus for *EGFR* mutation testing in non-small cell lung cancer: results from a European workshop. *J. Thorac. Oncol.* 5, 1706–1713 (2010).
- Lai-Goldman M, Faruki H. Abacavir hypersensitivity: a model system for pharmacogenetic test adoption. *Genet. Med.* 10, 874–878 (2008).
- Clark PJ, Thompson AJ, McHutchison JG. Genetic variation in *IL28B*: impact on drug

- development for chronic hepatitis C infection. *Clin. Pharmacol. Ther.* 88, 708–711 (2010).
- 23 Dudley JT, Sirota M, Shenoy M *et al.* Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci. Transl. Med.* 3(96), 96ra76 (2011).
- 24 Sirota M, Dudley JT, Kim J *et al.* Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med.* 3(96), 96ra77 (2011).
- **Example of how innovations in informatics are being used to uncover potential new uses for already approved therapies.**
- 101 Evaluation of Genomic Applications in Practice and Prevention (EGAPP). www.egappreviews.org/workingrp/reports.htm
- 102 Code of Federal Regulation (CFR). Title 21. www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=201.56
- 103 Erlotinib drug label. www.accessdata.fda.gov/scripts/cder/drugsatfda/ (Accessed 31 October 2011)
- 104 Draft guidance for industry and Food and Drug Administration staff: *in vitro* companion devices. www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM262327.pdf
- 105 Hindorff LA, Junkins HA, Hall PN, Mehta JP, Manolio TA. A catalog of published genome-wide association studies. www.genome.gov/gwastudies/ (Accessed 4 October 2011)
- 106 Guidance for industry chronic Hepatitis C virus infection: developing direct-acting antiviral agents for treatment. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM225333.pdf

Hakan Sakul, PhD
Executive Director and Head of Diagnostics
Worldwide R&D, Pfizer
San Diego, California

Biography

Hakan is currently an Executive Director in the Clinical Research and Precision Medicine Group in Pfizer's Development Operations where he oversees Pfizer's diagnostics needs across the R&D Portfolio. Hakan received his BS and MS degrees from Ankara University in Turkey, and his PhD in Quantitative Genetics from the University of Minnesota as a Rotary Foundation Scholar. After conducting his postdoctoral studies at the University of California-Davis, he worked in the biotech industry in human genetics, pharmacogenomics and statistical genetics fields. Hakan's tenure at Pfizer started at Parke-Davis Pharmaceuticals where he directed human genetics, statistical genetics and pharmacogenetics programs. Subsequently, he served as Director and Site Head for Clinical Pharmacogenomics in Groton/New London Laboratories before taking the role of Senior Director and Global Head of Companion Diagnostics for about four years to oversee Pfizer's companion diagnostics needs across the pharmaceutical portfolio. In 2010, Hakan moved to Pfizer's Oncology Business Unit to serve as Program Manager for the Xalkori (Crizotinib) Companion Diagnostics Development program, leading to simultaneous FDA approvals of both the drug and the companion diagnostic test. Hakan is keenly interested in technologies for identification of patient sub-groups for targeted treatment and the development of companion diagnostics to advance Precision Medicine for the improvement of individualized healthcare.

Notes

Thomas J. Hudson, MD
President and Scientific Director
Ontario Institute for Cancer Research
MaRS Centre, South Tower
Toronto, Ontario, Canada

Biography

Dr. Thomas J. Hudson is President and Scientific Director of the Ontario Institute for Cancer Research (OICR), an Institute created to support multidisciplinary teams needed to effectively translate research discoveries into interventions for better prevention, detection, diagnosis and treatment of cancer. Since its inception, OICR has launched several large-scale programs including the Ontario Health Study, the One Millimetre Cancer Challenge, the Cancer Stem Cell Program, the Pancreatic Cancer Genome Project (which is part of the International Cancer Genome Consortium), the Terry Fox Research Institute/OICR Selective Therapies Program and High Impact Clinical Trials.

Dr. Hudson is internationally renowned for his work in genomics and human genome variation. Past positions include leadership roles as Director of the McGill University and Genome Quebec Innovation Centre and Assistant-Director of the Whitehead/MIT Center for Genome Research, where he led a team that generated physical and gene maps of the human and mouse genomes. Dr. Hudson has been a founding member of the International Haplotype Map Consortium, the Public Population Project in Genomics (P3G) and the International Cancer Genome Consortium. Dr. Hudson's laboratory at OICR is involved in the study of genome variation that affects cancer predisposition, progression, and response to therapy. His main project focuses on the genetic architecture of loci associated with risk to colorectal cancer.

In 2007, Dr. Hudson was appointed to the rank of Professor in the Department of Molecular Genetics at the University of Toronto. Dr. Hudson is a fellow of the Royal Society of Canada. He is editor-in-chief of the journal *Human Genetics*. Dr. Hudson has co-authored over 200 peer-reviewed scientific publications.

Notes

Jeffrey M. Trent, PhD
Professor and Research Director
Translational Genomics Research Institute (TGen)
Phoenix, Arizona
Van Andel Research Institute (VARI)
Grand Rapids, Michigan

Biography

Jeffrey M. Trent, Ph.D., F.A.C.M.G., is founding President and Research Director of the Translational Genomics Research Institute (TGen), and recently added the role of President and Research Director of the Van Andel Research Institute in Grand Rapids, Michigan. He specializes in developing and integrating novel “omic” technologies, supporting studies of molecular changes related to the predisposition to, and progression of, human cancers and other complex diseases. Prior TGen, he served for 10 years as the Scientific Director of the National Human Genome Research Institute (NHGRI) at the National Institutes of Health in Bethesda, Maryland.

jtrent@tgen.org

Notes

Daniel D. Von Hoff, MD
Physician-in-Chief, Distinguished Professor
Translational Genomics Research Institute (TGen)
Professor of Medicine, Mayo Clinic and Chief Scientific Officer
Scottsdale Healthcare and US Oncology

Biography

Daniel D. Von Hoff, M.D., F.A.C.P. is currently Physician in Chief, Distinguished Professor and Director of Clinical Translational Research Division at TGen (Translational Genomics Research Institute) in Phoenix, Arizona. He is also Chief Scientific Officer for US Oncology and for Scottsdale Healthcare's Clinical Research Institute. He holds an appointment as Professor of Medicine, Mayo Clinic, Scottsdale, AZ.

Dr. Von Hoff's major interest is in the development of new anticancer agents, both in the clinic and in the laboratory. He and his colleagues were involved in the beginning of the development of many of the agents we now use routinely, including: mitoxantrone, fludarabine, paclitaxel, docetaxel, gemcitabine, irinotecan, nelarabine, capecitabine, lapatinib and others. At present, he and his colleagues are concentrating on the development of molecularly targeted therapies particularly for patients with advanced pancreatic cancer.

Dr. Von Hoff has published more than 600 papers, 135 book chapters and over 1000 abstracts. Dr. Von Hoff received the 2010 David A. Karnofsky Memorial Award from the American Society of Clinical Oncology for his outstanding contributions to cancer research leading to significant improvement in patient care.

Dr. Von Hoff was appointed to President Bush's National Cancer Advisory Board in 2004-2010. Dr. Von Hoff is the past President of the American Association for Cancer Research (the world's largest cancer research organization), a Fellow of the American College of Physicians, and a member and past board member of the American Society of Clinical Oncology. He is a founder of ILEX™ Oncology, Inc. (acquired by Genzyme after Ilex had 2 agents, alemtuzumab and clofarabine approved by the FDA for patients with leukemia). Dr. Von Hoff is founder and the Editor Emeritus of Investigational New Drugs – The Journal of New Anticancer Agents; and, Editor-in-Chief of Molecular Cancer Therapeutics. He is a co-founder of the AACR/ASCO Methods in Clinical Cancer Research Workshop. He is also proud to have been a mentor and teacher for multiple medical students, medical oncology fellows, graduate students, and post-doctoral fellows.

Notes

Elaine R. Mardis, PhD

Professor, Genetics and Molecular Microbiology

Co-Director of The Genome Institute

Washington University

Washington University School of Medicine at Washington University Medical Center

St. Louis, Missouri

Biography

Dr. Elaine Mardis graduated Phi Beta Kappa from the University of Oklahoma with a B.S. degree in zoology. She then completed her Ph.D. in Chemistry and Biochemistry in 1989, also at Oklahoma. Following graduation, Dr. Mardis was a senior research scientist for four years at BioRad Laboratories in Hercules, CA.

In 1993, Dr. Mardis joined The Genome Institute at Washington University School of Medicine. As Director of Technology Development, she helped create methods and automation pipelines for sequencing the Human Genome. She currently orchestrates the Center's efforts to explore next generation and third generation sequencing technologies and to transition them into production sequencing capabilities.

Dr. Mardis has research interests in the application of DNA sequencing to characterize cancer genomes. She also is interested in facilitating the translation of basic science discoveries about human disease into the clinical setting.

Dr. Mardis serves on several NIH study sections, is an editorial board member of *Genome Research*, and acts as a reviewer for *Nature* and *Genome Research*. She serves as chair of the Basic and Translational Sciences Committee for the American College of Surgeons Oncology Group, an NCI funded cooperative group. She serves on the scientific advisory boards of Pacific Biosciences, Inc. and Edge Biosciences, Inc. Dr. Mardis received the Scripps Translational Research award for her work on cancer genomics in 2010, and was named a Distinguished Alumni of the University of Oklahoma College of Arts and Sciences for 2011.

Notes

Fred H. Gage, PhD

Professor and Vi and John Adler Chair for Research on Age-Related
Neurodegenerative Diseases
Laboratory of Genetics
Salk Institute for Biological Studies
La Jolla, California

Biography

Fred H. Gage, PhD is the head of the Laboratory of Genetics at the Salk Institute for Biological Studies, which he joined in 1995, and Vi and John Adler Chair for Research on Age-Related Neurodegenerative Diseases and an Adjunct Professor in the Department of Neurosciences, UCSD, La Jolla. He received his Ph.D. in 1976 from The Johns Hopkins University. Dr. Gage's work concentrates on the adult central nervous system and unexpected plasticity and adaptability to environmental stimulation that remains throughout the life of all mammals. In addition, his studies focus on the cellular, molecular, as well as environmental influences that regulate neurogenesis in the adult. He has won numerous prizes and awards for his work including the IPSEN Prize for Neuroplasticity, the Charles A. Dana Award, Metropolitan Life Research Award and the Keio Medical Science Prize. He serves on many health related boards, and was President of the Society for Neuroscience and is President-Elect of the International Society for Stem Cell Research. He is a Fellow of the American Association for the Advancement of Science, a Member of the National Academy of Sciences and the Institute of Medicine, a Member of the American Academy of Arts and Sciences, an Associate Member of the European Molecular Biology Organization, and a member of the American Philosophical Society.

Notes

Modeling Human Brain Disease with patient specific iPS cells

Fred H. Gage, PhD

Psychiatric disorders, including schizophrenia, bipolar disorder and autism spectrum disorders are extremely heritable complex genetic disorders. It is now possible to directly reprogram fibroblasts from psychiatric patients into human induced pluripotent stem cells (hiPSCs) and subsequently differentiate these disorder-specific hiPSCs into neurons. This means that researchers can now generate nearly limitless quantities of live human neurons with genetic backgrounds known to result in psychiatric disorders, even without knowing which genes are interacting to produce the disease state in each patient. With these new cell-based human models, scientists can investigate the precise cell types affected in these disorders and elucidate the cellular and molecular defects that contribute to disease initiation and progression. I present a short review of experiments from my lab using hiPSCs to study the pathways underlying psychiatric disorders.

Notes

L1 retrotransposition in neurons is modulated by MeCP2

Alysson R. Muotri^{1*}, Maria C. N. Marchetto^{2*}, Nicole G. Coufal², Ruth Oefner², Gene Yeo³, Kinichi Nakashima⁴ & Fred H. Gage²

Long interspersed nuclear elements-1 (LINE-1 or L1s) are abundant retrotransposons that comprise approximately 20% of mammalian genomes^{1–3}. Active L1 retrotransposons can impact the genome in a variety of ways, creating insertions, deletions, new splice sites or gene expression fine-tuning^{4–6}. We have shown previously that L1 retrotransposons are capable of mobilization in neuronal progenitor cells from rodents and humans and evidence of massive L1 insertions was observed in adult brain tissues but not in other somatic tissues^{7,8}. In addition, L1 mobility in the adult hippocampus can be influenced by the environment⁹. The neuronal specificity of somatic L1 retrotransposition in neural progenitors is partially due to the transition of a Sox2/HDAC1 repressor complex to a Wnt-mediated T-cell factor/lymphoid enhancer factor (TCF/LEF) transcriptional activator^{7,10}. The transcriptional switch accompanies chromatin remodelling during neuronal differentiation, allowing a transient stimulation of L1 transcription⁷. The activity of L1 retrotransposons during brain development can have an impact on gene expression and neuronal function, thereby increasing brain-specific genetic mosaicism^{11,12}. Further understanding of the molecular mechanisms that regulate L1 expression should provide new insights into the role of L1 retrotransposition during brain development. Here we show that L1 neuronal transcription and retrotransposition in rodents are increased in the absence of methyl-CpG-binding protein 2 (MeCP2), a protein involved in global DNA methylation and human neurodevelopmental diseases. Using neuronal progenitor cells derived from human induced pluripotent stem cells and human tissues, we revealed that patients with Rett syndrome (RTT), carrying MeCP2 mutations, have increased susceptibility for L1 retrotransposition. Our data demonstrate that L1 retrotransposition can be controlled in a tissue-specific manner and that disease-related genetic mutations can influence the frequency of neuronal L1 retrotransposition. Our findings add a new level of complexity to the molecular events that can lead to neurological disorders.

In neural stem cells, the repressor complex on the L1 promoter region (L1 5'UTR) includes the transcriptional factor Sox2 and the histone deacetylase 1 protein (HDAC1)⁷, a MeCP2 partner^{13,14}. MeCP2 has been shown to interfere with the L1 5'UTR promoter activity in transformed cell lines¹⁵. To investigate the role of MeCP2 in the activity of L1 promoter in neural stem cells, we cloned the L1 promoter region upstream to the luciferase gene, generating the L1 5'UTR–Luc plasmid⁷. Methylation of the L1 5'UTR–Luc reduced the promoter activity in neural stem cells (Fig. 1a and Supplementary Fig. 1a). Reduction of MeCP2 levels using siRNAs led to an increase in luciferase activity (Fig. 1b and Supplementary Fig. 1b). Transfection of the L1 5'UTR–Luc methylated plasmid in mouse neuroepithelial cells revealed that the L1 promoter activity was approximately four times more active in the MeCP2 knockout (KO) background than in wild-type (Fig. 1c and Supplementary Fig. 1c). Ectopic MeCP2 expression reduced the luciferase activity in MeCP2 KO cells close to wild-type levels (Fig. 1c).

We repeated the luciferase assay using neuroepithelial cells from a sibling MBD1 KO animal¹⁶. MBD1 (methyl-CpG binding domain protein 1) is part of the methyl-binding protein family and has differential

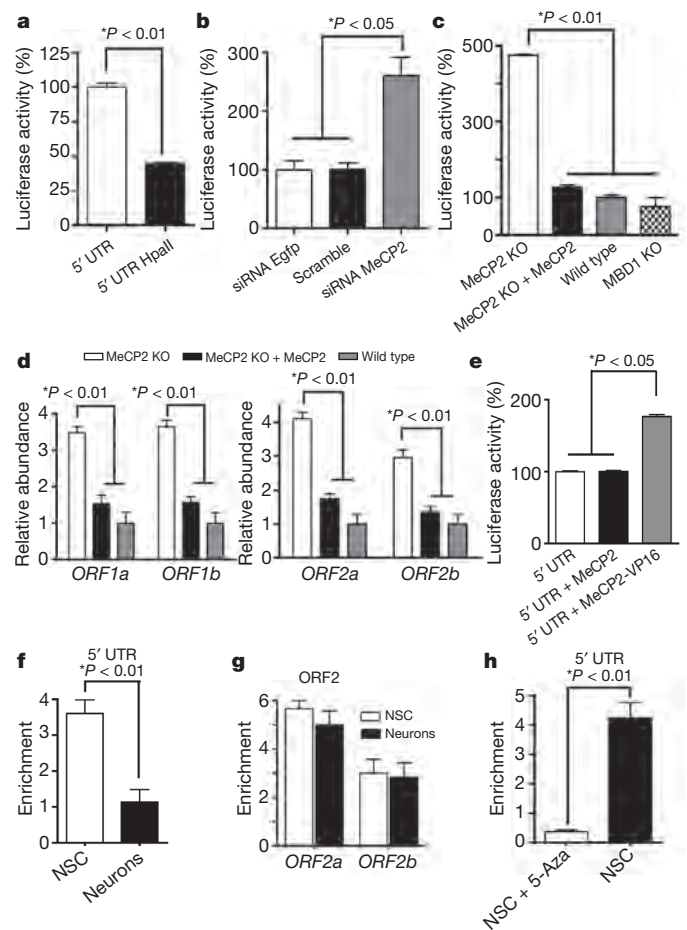


Figure 1 | MeCP2 silences L1 expression. **a**, Methylation of the L1 5'UTR–Luc reduced its transcriptional activity. **b**, Reduction of MeCP2 transcripts correlates with increased L1 promoter activity. **c**, Increased L1 promoter activity in the absence of MeCP2 but not MBD1. **d**, L1 RNA levels correlate with MeCP2 expression. **e**, Expression of the MeCP2–VP16 increased the activity of the L1 5'UTR promoter. **f**, **g**, Recruitment of MeCP2 on L1 sequences by ChIP in neural stem cells (NSC) or neurons, using 5'UTR primers (**f**) and two ORF2 regions (**g**). **h**, Occupancy of MeCP2 on the L1 promoter requires DNA methylation. Removal of DNA methylation with 5-azacytidine (5-Aza) reduced MeCP2 association to L1 promoter. ChIP–qPCR shows enrichment over IgG control precipitation. All experiments show experimental triplicates. Error bars in all panels show s.e.m.

¹University of California San Diego, School of Medicine, Department of Pediatrics/Rady Children's Hospital San Diego, Department of Cellular & Molecular Medicine, Stem Cell Program, 9500 Gilman Drive, La Jolla, California 92093-0695, USA. ²Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. ³University of California San Diego, School of Medicine, Department of Cellular & Molecular Medicine, Stem Cell Program, 9500 Gilman Dr, La Jolla, California 92093-0695, USA. ⁴Laboratory of Molecular Neuroscience, Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma 630-0101, Japan.

*These authors contributed equally to this work.

DNA specificity when compared to MeCP2¹⁷. The L1 promoter was not activated in MBD1 KO background, a finding that is consistent with the idea that L1 transcriptional repression is specific to MeCP2 (Fig. 1c). Moreover, the promoter activity correlated well with the level of L1 RNA, as measured by qPCR (Fig. 1d). Ectopic MeCP2 expression reduced L1 RNA levels in the MeCP2 KO background (Fig. 1d). We co-transfected neural stem cells with the methylated L1 5'UTR–Luc and a plasmid containing either the MeCP2 cDNA or the MeCP2 fused with the transactivator domain VP16. The overexpression of MeCP2 alone did not change the luciferase levels, but the MeCP2–VP16 fusion increased luciferase levels twofold (Fig. 1e).

Using chromatin immunoprecipitation (ChIP) followed by quantitative PCR (qPCR), we detected high levels of MeCP2 in association with endogenous L1 promoter regions in neural stem cells compared to neurons (Fig. 1f). MeCP2 was also associated with other L1 regions (ORF2), but this association did not change during differentiation (Fig. 1g; see controls for ChIP experiments in Supplementary Fig. 1d, e). After treatment with 5-azacytidine, the MeCP2 ChIP signal was reduced and L1 expression increased (Fig. 1h and Supplementary Fig. 1f). A set of the CpG sites within the L1 promoter had a tendency to demethylate during neuronal differentiation, indicating that DNA methylation may silence L1 expression in neural stem cells by attracting MeCP2 (Supplementary Fig. 1g, h).

To study L1 regulation *in vivo*, we compared the brains of the L1–EGFP (enhanced green fluorescent protein) transgenic mice in wild-type and MeCP2 KO backgrounds. L1–EGFP transgenic mice have a L1 indicator cassette that will only activate the EGFP reporter after retrotransposition⁷ (Supplementary Fig. 2a). The numbers of EGFP-positive cells in the brains of MeCP2 KO mice were significantly higher than in wild type (Fig. 2a, b). EGFP-positive cells were also observed in the germ line of MeCP2 KO at similar frequency as in wild-type animals, but not in other somatic tissues (Supplementary Fig. 2b). To visualize the distribution of EGFP-positive cells, we generated high-resolution, three-dimensional maps of both MeCP2 KO and wild-type brains. Although MeCP2 KO brain sections had an average of 3.5-fold more EGFP-positive cells than wild type, certain brain structures were more prone to L1 retrotransposition (Fig. 2b, c). Specifically, the cerebellum, striatum, cortex, hippocampus and olfactory bulb contained 4.2-, 5.3-, 2.8-, 6.3- and 3.8-fold more EGFP-positive neurons, respectively, in the MeCP2 KO genetic background than in wild type (Supplementary Fig. 3 and Supplementary Movie). More EGFP-positive cells may suggest an increased rate of L1 retrotransposition and/or higher rate of MeCP2 KO cell proliferation with the newly retrotransposed EGFP reporter. We found no evidence that neuroepithelial cells from the MeCP2 KO genetic background had a higher rate of division than wild type (Supplementary Fig. 4a).

We next asked whether endogenous L1 retrotransposition was also increased in the MeCP2 KO brain. New insertions from retroelements can be quantified using a qPCR approach^{8,18}. To determine the activity of endogenous L1 elements, we developed a technique based on single-cell genomic qPCR that measures the frequency of mouse L1 sequences within the genome (Fig. 3a). We proposed that MeCP2 KO-derived neuroepithelial cells would have increased genomic content of L1 sequences compared to wild-type cells. Neuroepithelial cells from wild-type and MeCP2 KO sibling mouse embryos were synchronized in G1 phase and karyotyped, to avoid interference during genomic L1 detection (Supplementary Fig. 4b, c). Finally, single-cell amplification using primers for ORF2 from active L1 families confirmed the presence of the expected amplicons (Supplementary Fig. 4d). MeCP2 KO-derived neuroepithelial cells displayed significantly more ORF2 genomic copies than wild-type cells (Fig. 3b). Specific primers for the L1 5'UTR were also tested in neuroepithelial cells and did not reveal an increase in copy number in MeCP2 KO background (Fig. 3c). This lack of difference can be explained by the fact that, upon retrotransposition, the 5' region of the L1 sequence is frequently truncated^{19,20}. Also, no difference between genetic backgrounds was observed when using primers for non-mobile

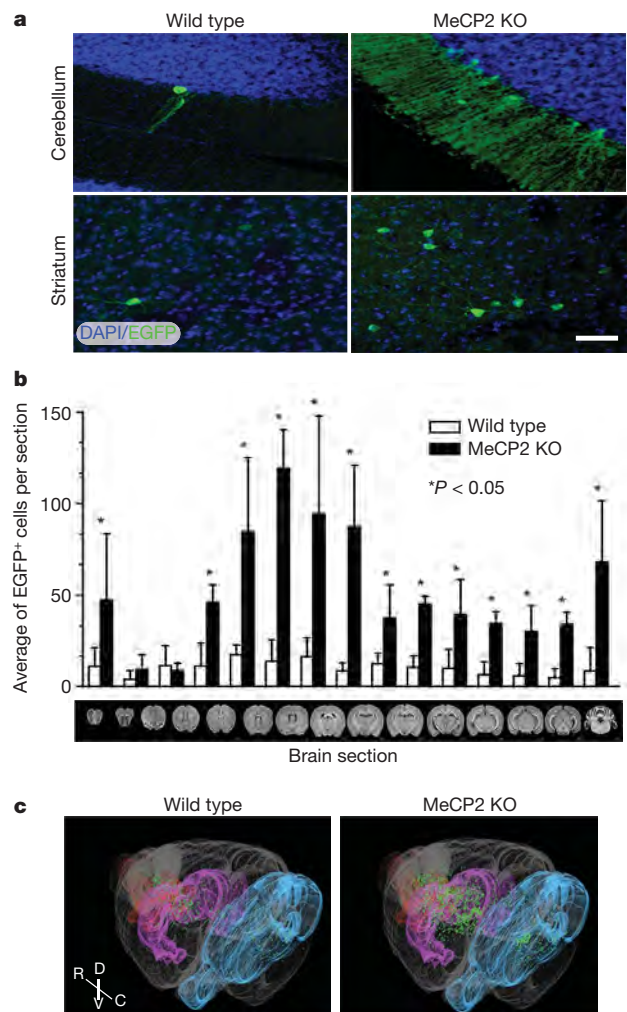


Figure 2 | MeCP2 modulates neuronal L1 retrotransposition *in vivo*.

a, EGFP-positive cells, indicating *de novo* L1 retrotransposition, were found in several regions of the brain. The images were taken from sections that were highly affected by L1 retrotransposition. Bar, 30 μ m. **b**, Quantification of brain sections in MeCP2 KO background revealed more EGFP-positive cells compared to wild type ($n = 6$ animals for each group). Error bars show s.d. **c**, Representative images from a three-dimensional reconstruction of wild-type and MeCP2 KO brains carrying the L1–EGFP transgene. Single dots (green) represent neurons that supported L1–EGFP retrotransposition. Olfactory bulb is shown in red, striatum in magenta and cerebellum in cyan. R, rostral; C, caudal; D, dorsal and V, ventral.

5S ribosomal RNA repetitive sequences (Fig. 3d). Another control experiment was performed using fibroblasts isolated from the two backgrounds (Fig. 3e). We did not observe a highly significant increase in L1 copy number in MeCP2 KO compared to wild type fibroblasts.

Mutations on the MeCP2 gene cause RTT, characterized by arrested development in early childhood and autistic behaviour at different levels of intensity²¹. To determine if L1 retrotransposition could occur in neuronal progenitor cells (NPC) derived from RTT patients, we generated induced pluripotent stem cells (iPSC) from a RTT patient's fibroblasts carrying a frameshift MeCP2 mutation and from a control, non-affected individual. All clones were pluripotent and able to produce NPC and neurons (Supplementary Fig. 5). Thus, we tested if the iPSC-derived NPC supported L1 retrotransposition.

NPC from both wild-type and RTT iPSC expressed the neural markers Sox1, Musashi1, Nestin and Sox2 at similar rates at the time of the experiment (Supplementary Fig. 6a, b). RTT and wild-type cells were electroporated with the L1_{RE3}–EGFP reporter construct^{22,23} (Fig. 4a). EGFP expression was detected in both wild-type and RTT cells (Fig. 4b). The frequency of EGFP-positive cells was approximately

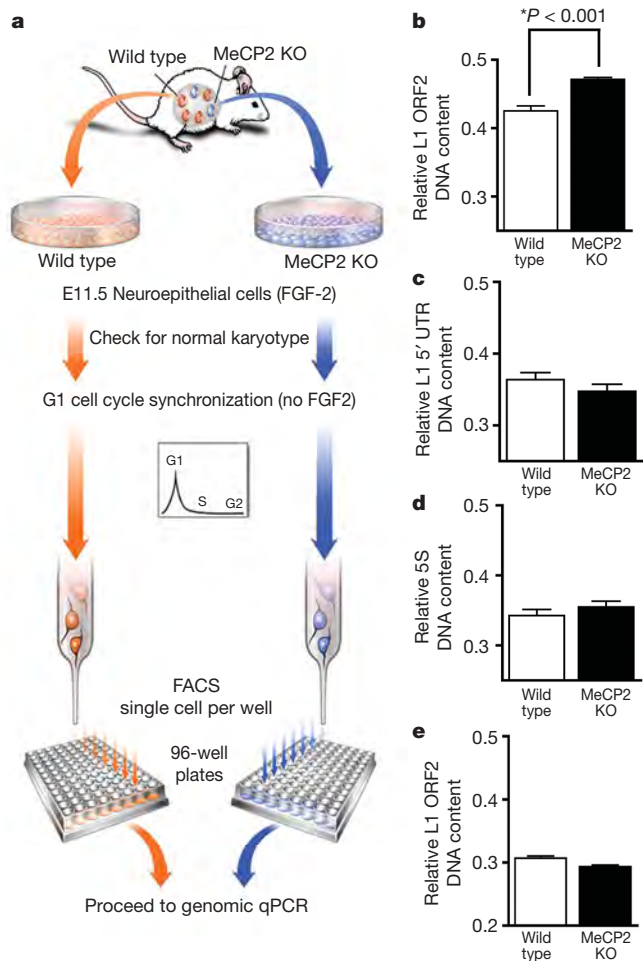


Figure 3 | Endogenous L1 retrotransposition in mouse neuroepithelial cells. **a**, Neuroepithelial cells harvested from embryonic day 11.5 (E11.5) sibling embryos were synchronized and sorted in individual wells followed by qPCR. **b**, Neuroepithelial cells in the MeCP2 KO background had higher L1 ORF2 DNA content than wild-type cells ($P < 0.001$). **c**, L1 5' UTR primers did not reveal a significant increase in copy number in MeCP2 KO background. **d**, Non-mobile 5S ribosomal genes were used as controls. **e**, The difference in the amount of L1 ORF2 DNA in fibroblasts from the different genetic backgrounds was smaller than in the neural lineage. All experiments show experimental triplicates ($n = 192$ cells for each primer pair). Error bars in all panels show s.e.m.

twofold higher in RTT than in control cells. Moreover, MeCP2 complementation reduced the levels of EGFP-positive cells in RTT NPC (Fig. 4b, c and Supplementary Fig. 6c). PCR confirmed the presence of the retrotransposed EGFP and sequencing confirmed the precise splicing of the intron (Supplementary Fig. 6d). We concluded that L1 activity could be facilitated by loss of MeCP2 function in human cells. We extended the iPSC findings *in vivo* using post mortem human tissues. To analyse the amounts of L1 retrotransposition in RTT patients and controls, brain and heart tissue was obtained from the same individuals. After genomic DNA extraction, a qPCR was used to compare the number of L1 ORF2 sequences normalized by four distinct non-mobile repetitive sequences. The number of L1 ORF2 sequences in the brains of RTT patients was significantly higher than in age/gender-matched controls (Fig. 4d). Moreover, the number of ORF2 sequences was higher in brain tissues in both controls and RTT patients when compared to heart tissue from the same individuals.

Our findings support previous data demonstrating that L1 5' UTR sequences are MeCP2 targets that may be subjected to methylation-dependent repression^{15,17}. However, we cannot exclude an indirect effect of MeCP2 in regulating genes involved in L1 expression and/

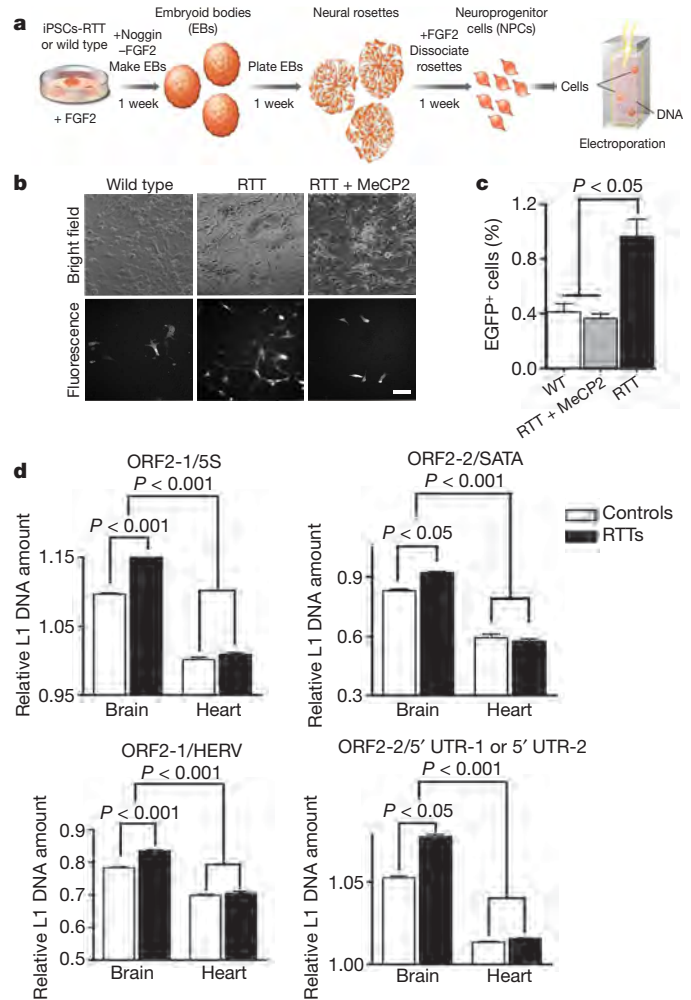


Figure 4 | L1 retrotransposition in RTT patients. **a**, Schematic view of the NPC differentiation from iPSC followed by L1_{RE3}-EGFP electroporation. **b**, Representative images of iPSC-derived NPC expressing EGFP after L1 retrotransposition. Bar, 30 μ m. **c**, Quantification of the EGFP-positive cells after transfection. **d**, Primers for ORF2 were used to multiplex with primers for control sequences, such as the 5S ribosomal gene (5S), the satellite alpha (SATA) region, the human endogenous retrovirus H (HERV) sequence and the 5' UTR. The inverse ratio of ORF2/5S represents the amount of L1 ORF2 sequence in each sample ($n = 5$ individuals per group). Similar results were obtained when different primers/probe for ORF2 (ORF2-2) were multiplex/normalized to other control sequences, using two pair of primers (5' UTR-1 or 5' UTR-2). Error bars show s.e.m, and the experiments were performed in triplicate.

or in changing the chromatin epigenetic landscape to facilitate *de novo* L1 insertions. An additive effect of multiple mechanisms is likely. Using different strategies, we have shown that L1 retrotransposition can be modulated by MeCP2. First, we demonstrated that MeCP2 can downregulate L1 promoter activity. Second, L1 retrotransposition from the L1-EGFP transgenic mice was significantly higher in the brains of a MeCP2 KO background than in a wild-type sibling animal. The L1-EGFP indicator system underestimates the actual capacity of retrotransposition and does not take into account insertions that truncate or silence the reporter cassette, *in trans* retrotransposition of *Alu* sequences or other RNAs²⁴⁻²⁶. Third, we developed a new technique based on single-cell genomic qPCR to measure the relative abundance of L1 sequences, revealing that MeCP2 KO neuroepithelial cells have more L1 sequences in the genome than wild-type cells. Lastly, RTT-NPC showed a higher L1 retrotransposition frequency than control cells. A qPCR experiment extended these observations to human brain samples from RTT patients compared to controls.

Our data provide evidence of a role for DNA methylation-dependent MeCP2 activity in controlling L1 mobility in the nervous system. Re-activation of MeCP2 expression was shown to reverse some of the neurological symptoms in MeCP2 KO mice²⁷. The high rates of neuronal retrotransposition in the MeCP2 KO mice and RTT patients may be a consequence, rather than a cause, of the disease process. Nonetheless, new somatic insertions, especially at early developmental stages, may contribute to the genetic and epigenetic status of mature neurons at later stages of life. Early developmental structural and functional modulations could have potential consequences for RTT, where the detrimental effects of MeCP2 mutation occur at later postnatal stages. It is plausible to conclude that the RTT process leads to an increased rate of somatic mutations in the brain. Increased L1 neuronal retrotransposition is a novel and unexpected characteristic of RTT pathology. Our findings add a new layer of complexity to the understanding of genomic plasticity and may have direct implications for individual variation and for neurological diseases.

METHODS SUMMARY

For the *luciferase* activity experiments, rat neural stem cells were isolated, characterized and cultured as described²⁸. Neuroepithelial cells from time-pregnant midgestation (embryonic day 11.5) telencephalons from male wild-type, MBD1 KO, and MeCP2 KO sibling mouse embryos, from the same genetic background (C57BL/6J) were isolated. Cells were cultured for two to three passages in Dulbecco's modified Eagle's medium (DMEM) F12 media with N2 supplement and fibroblast growth factor 2 (FGF2) as described elsewhere²⁹. Plasmid and siRNA transfections were performed by electroporation (Lonza/Amaxa Biosystem). Luciferase activity was measured with the Dual-Luciferase reporter assay system (Promega) according to the manufacturer's protocol. Chromatin immunoprecipitation (ChIP) assays were performed following the manufacturer's protocol using a kit from Millipore/Upstate. Antibodies used were anti-MeCP2 and IgG (Upstate). After immunoprecipitation, recovered chromatin fragments were subjected to PCR using primers for the rat L1 sequence. qPCR values were normalized to the IgG precipitation and shown as fold enrichment. For human iPSC derivation, RTT and control fibroblasts were infected with retroviral vectors containing the *Oct4*, *c-Myc*, *Klf4* and *Sox2* human cDNAs as described previously by Yamanaka's group³⁰. iPSC-derived neural progenitors were electroporated (Lonza/Amaxa Biosystem) with L1-EGFP plasmid and FACS sorted for EGFP to quantify L1 *de novo* insertions. Single-cell genomic quantitative PCR (qPCR) was performed in cell-cycle-arrested neuroepithelial cells and fibroblasts from wild-type and MeCP2 KO mice. The plates containing one cell per well were then snap frozen at -80°C until the day of the qPCR. The qPCR was performed using the protocol available on the manufacturer's website (Applied Biosystems). Briefly, a solution containing forward/reverse primers and SYBR Green PCR Master Mix was added to the previously sorted cells and the detection of DNA products was carried out in an ABI PRISM 7900HT Sequence Detection System. For multiplex genomic qPCR in human tissues the qPCR strategy and L1 copy estimation were done as previously described⁸.

Received 27 April; accepted 30 September 2010.

- Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Gibbs, R. A. *et al.* Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
- Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Kazazian, H. H. Jr. Mobile elements and disease. *Curr. Opin. Genet. Dev.* **8**, 343–350 (1998).
- Han, J. S., Szak, S. T. & Boeke, J. D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429**, 268–274 (2004).
- Perepelitsa-Belancio, V. & Deininger, P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nature Genet.* **35**, 363–366 (2003).

- Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**, 903–910 (2005).
- Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127–1131 (2009).
- Muotri, A. R., Zhao, C., Marchetto, M. C. & Gage, F. H. Environmental influence on L1 retrotransposons in the adult hippocampus. *Hippocampus* **19**, 1002–1007 (2009).
- Kuwabara, T. *et al.* Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nature Neurosci.* **12**, 1097–1105 (2009).
- Muotri, A. R. & Gage, F. H. Generation of neuronal variability and complexity. *Nature* **441**, 1087–1093 (2006).
- Singer, T., McConnell, M. J., Marchetto, M. C., Coufal, N. G. & Gage, F. H. LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends Neurosci.* **33**, 345–354 (2010).
- Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389 (1998).
- Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genet.* **19**, 187–191 (1998).
- Yu, F., Zingler, N., Schumann, G. & Stralting, W. H. Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Res.* **29**, 4493–4501 (2001).
- Zhao, X. *et al.* Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. *Proc. Natl Acad. Sci. USA* **100**, 6777–6782 (2003).
- Klose, R. J. *et al.* DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol. Cell* **19**, 667–678 (2005).
- Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).
- Grimaldi, G., Skowronski, J. & Singer, M. F. Defining the beginning and end of *KpnI* family segments. *EMBO J.* **3**, 1753–1759 (1984).
- Moran, J. V. & Gilbert, N. in *Mobile DNA II*, Vol. 2 (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) Ch. 35, 836–869 (ASM Press, 2002).
- Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2. *Nature Genet.* **23**, 185–188 (1999).
- Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 (1996).
- Ostertag, E. M., Prak, E. T., DeBerardinis, R. J., Moran, J. V. & Kazazian, H. H. Jr. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* **28**, 1418–1423 (2000).
- Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate pseudogenes. *Nature Genet.* **24**, 363–367 (2000).
- Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. *Nature Genet.* **35**, 41–48 (2003).
- Wei, W. *et al.* Human L1 retrotransposition: *cis* preference versus *trans* complementation. *Mol. Cell Biol.* **21**, 1429–1439 (2001).
- Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of neurological defects in a mouse model of Rett syndrome. *Science* **315**, 1143–1147 (2007).
- Palmer, T. D., Takahashi, J. & Gage, F. H. The adult rat hippocampus contains primordial neural stem cells. *Mol. Cell Neurosci.* **8**, 389–404 (1997).
- Nakashima, K. *et al.* Synergistic signaling in fetal brain by STAT3-Smad1 complex bridged by p300. *Science* **284**, 479–482 (1999).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A.R.M. is supported by the National Institutes of Health through the NIH Director's New Innovator Award Program, 1-DP2-OD006495-01 and by the Emerald Foundation. F.H.G. is supported by the Mathers Foundation, Lookout Fund, and NIH/NINDS R01MH088485. The authors would like to thank A. Huynh, B. Aimone, K. Stecker, B. Berg and D. Sepp for help during the 3D brain model assembly, J. Moran and J. Garcia-Perez for discussion and critical review of the manuscript, M. Gage for editorial comments, and B. Moddy and G. Peng for experimental assistance.

Author Contributions A.R.M. and M.C.N.M. are the leading authors. They contributed to the concept, designed and performed the experiments, analysed the data, and wrote the manuscript. N.G.C. performed and analysed qPCR experiments. R.O. performed tissue culture experiments and quantification. G.Y. helped with statistical analysis and data interpretation. K.N. contributed reagents, and performed data analyses and manuscript revision. F.H.G. contributed to the concept, analysed the data and revised the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.R.M. (muotri@ucsd.edu) or F.H.G. (gage@salk.edu).

Modelling schizophrenia using human induced pluripotent stem cells

Kristen J. Brennand¹, Anthony Simone^{1*}, Jessica Jou^{1*}, Chelsea Gelboin–Burkhart^{1*}, Ngoc Tran^{1*}, Sarah Sangar¹, Yan Li¹, Yangling Mu¹, Gong Chen², Diana Yu¹, Shane McCarthy³, Jonathan Sebat⁴ & Fred H. Gage¹

Schizophrenia (SCZD) is a debilitating neurological disorder with a world-wide prevalence of 1%; there is a strong genetic component, with an estimated heritability of 80–85%¹. Although post-mortem studies have revealed reduced brain volume, cell size, spine density and abnormal neural distribution in the prefrontal cortex and hippocampus of SCZD brain tissue² and neuropharmacological studies have implicated dopaminergic, glutamatergic and GABAergic activity in SCZD³, the cell types affected in SCZD and the molecular mechanisms underlying the disease state remain unclear. To elucidate the cellular and molecular defects of SCZD, we directly reprogrammed fibroblasts from SCZD patients into human induced pluripotent stem cells (hiPSCs) and subsequently differentiated these disorder-specific hiPSCs into neurons (Supplementary Fig. 1). SCZD hiPSC neurons showed diminished neuronal connectivity in conjunction with decreased neurite number, PSD95-protein levels and glutamate receptor expression. Gene expression profiles of SCZD hiPSC neurons identified altered expression of many components of the cyclic AMP and WNT signalling pathways. Key cellular and molecular elements of the SCZD phenotype were ameliorated following treatment of SCZD hiPSC neurons with the antipsychotic loxapine. To date, hiPSC neuronal pathology has only been demonstrated in diseases characterized by both the loss of function of a single gene product and rapid disease progression in early childhood^{4–6}. We now report hiPSC neuronal phenotypes and gene expression changes associated with SCZD, a complex genetic psychiatric disorder.

Four SCZD patients were selected: patient 1, diagnosed at 6 years of age, had childhood-onset SCZD; patients 2, 3 and 4 were from families in which all offspring and one parent were affected with psychiatric disease (Supplementary Fig. 3a). Primary human fibroblasts were reprogrammed using inducible lentiviruses⁷. Control and SCZD hiPSCs expressed endogenous pluripotency genes, repressed viral genes and were indistinguishable in assays for self-renewal and pluripotency (Fig. 1 and Supplementary Fig. 2). SCZD hiPSCs had no apparent defects in generating neural progenitor cells (NPCs) or neurons (Fig. 1 and Supplementary Fig. 3). Most hiPSC neurons were presumably glutamatergic and expressed VGLUT1 (also known as SLC17A7; Supplementary Fig. 8a). Approximately 30% of neurons were GAD65/67-positive (also known as GAD1/2; GABAergic) (Supplementary Fig. 8c, d) whereas less than 10% of neurons were tyrosine hydroxylase-positive (dopaminergic) (Supplementary Fig. 7).

Neuronal connectivity was assayed using trans-neuronal spread of rabies; *in vivo*, rabies transmission occurs via synaptic contacts and is strongly correlated with synaptic input strength⁸. Primary infection was restricted by replacing the rabies coat protein with envelope A (ENVA), which infects only via the avian tumour virus A (TVA) receptor; viral spread was limited to monosynaptically connected neurons by deleting the rabies glycoprotein gene (ΔG)⁹. Neurons were first transduced with a lentivirus expressing histone 2B (H2B)–green fluorescent protein

(GFP) fusion protein, TVA and G from the synapsin (SYN) promoter (LV-SYNP-HTG). One week later, neurons were transduced with modified rabies virus (Rabies-ENVA ΔG -RFP). Primary infected cells were positive for both H2B–GFP and RFP (red fluorescent protein); neurons monosynaptically connected to primary cells were GFP-negative but RFP-positive (Supplementary Fig. 4a). Transduction with Rabies-ENVA ΔG -RFP alone resulted in no RFP-positive cells, whereas transduction with Rabies-ENVA ΔG -RFP following lentiviral transduction without rabies glycoprotein (SYNP-HT) led to only single GFP⁺RFP⁺ cells, indicating that *in vitro* rabies infection and spread are dependent on TVA expression and G trans-complementation, respectively (Supplementary Fig. 4c, d).

There was decreased neuronal connectivity in SCZD hiPSC neurons (Fig. 2; Supplementary Figs 4b, c, 5 and 6). Fluorescence-activated cell sorting (FACS) analysis confirmed differences in neuronal connectivity and demonstrated that comparable numbers of β III-tubulin-positive neurons were labelled with LV-SYNP-HTG. Though the mechanism of rabies trans-neuronal tracing is not fully understood, the presynaptic protein NCAM has been implicated¹⁰; NCAM expression is decreased in SCZD hiPSC neurons (Supplementary Table 3). Rabies trans-neuronal tracing occurs in functionally immature hiPSC neurons (Supplementary Fig. 4e) and in the presence of the voltage-gated sodium channel blocker tetrodotoxin (TTX) (1 μ M), depolarizing KCl (50 mM) or the calcium channel blocker ryanodine (10 μ M) (Supplementary Fig. 4f). Decreased

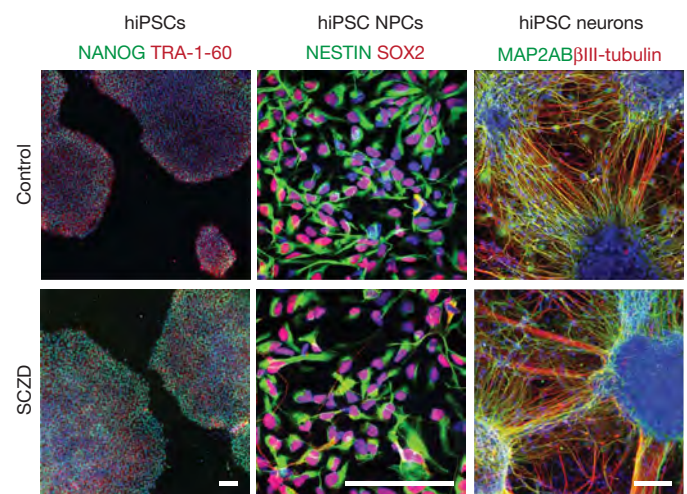


Figure 1 | Patient-specific hiPSCs, NPCs and neurons. Left, hiPSCs express NANOG (green) and TRA-1-60 (red). DAPI (blue). $\times 100$, scale bar 2.5 μ m. Centre, hiPSC neural progenitor cells (NPCs) express NESTIN (green) and SOX2 (red). DAPI (blue). $\times 630$, scale bar 2.5 μ m. Right, hiPSC neurons express β III-tubulin (red) and the dendritic marker MAP2AB (green). DAPI (blue). $\times 200$, scale bar 2.5 μ m.

¹Salk Institute for Biological Studies, Laboratory of Genetics, 10010 North Torrey Pines Road, La Jolla California 92037, USA. ²Department of Biology, Pennsylvania State University, 201 Life Science Building, University Park, Pennsylvania 16802, USA. ³Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. ⁴University of California San Diego, Department of Psychiatry and Department of Cellular and Molecular Medicine, University Of California, San Diego, La Jolla, California 92093, USA.

*These authors contributed equally to this work.

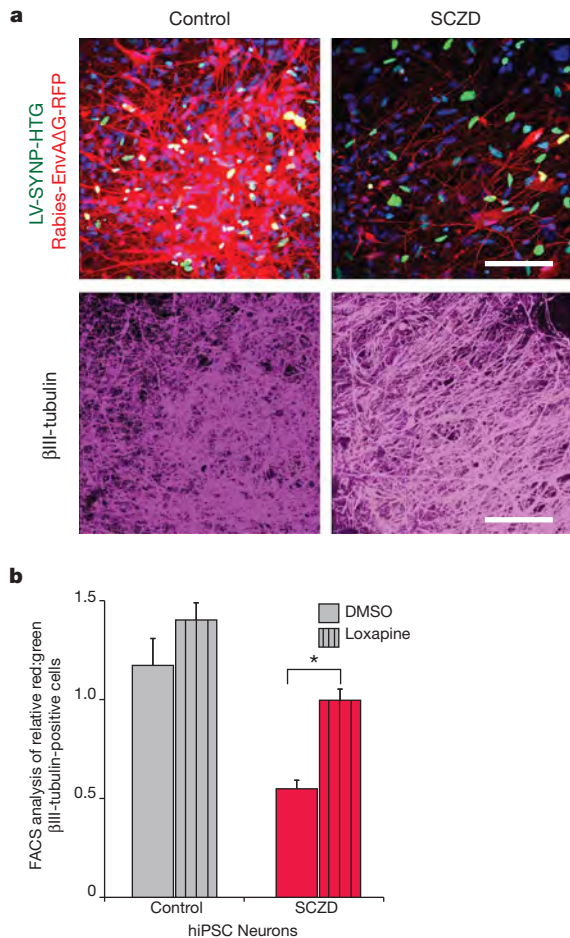


Figure 2 | Decreased neuronal connectivity in SCZD hiPSC neurons.

a, Representative images of control and SCZD hiPSC neurons cotransduced with LV-SYNP-HTG and Rabies-ENVAAG-RFP, 10 days post rabies transduction. All images were captured using identical laser power and gain settings. β III-tubulin staining (purple) of the field is shown below each panel. $\times 400$, scale bar 2 μ m. **b**, Graph showing treatment of SCZD hiPSC neurons with Loxapine resulted in a statistically significant improvement in neuronal connectivity. Error bars are s.e., $*P < 0.05$

trans-neuronal tracing is evidence of decreased neuronal connectivity, but not necessarily decreased synaptic function, in SCZD hiPSC neurons.

We tested the ability of five antipsychotic drugs to improve neuronal connectivity *in vitro*. Clozapine, loxapine, olanzapine, risperidone and thioridazine were administered for the final 3 weeks of neuronal differentiation. Only loxapine significantly increased neuronal connectivity in hiPSC neurons from all patients (Fig. 2b and Supplementary Fig. 5). Optimization of the concentration and timing of drug administration may improve the effects of the other antipsychotic medications.

Reduced dendritic arborization has been observed in post-mortem SCZD brains¹¹ and in animal models¹². SCZD hiPSC neurons show a decrease in the number of neurites (Fig. 3a and Supplementary Fig. 9a, b). Synaptic genes are associated with SCZD¹³ (Supplementary Fig. 9d) and impaired synaptic maturation occurs in a number of mouse models¹². hiPSC neurons express dense puncta of synaptic markers that co-stain for both pre- and post-synaptic markers (Supplementary Fig. 8a, b). Whereas we observed decreased PSD95 protein (also known as DLG4, levels relative to microtubule associated protein 2ab (MAP2AB) in SCZD hiPSC neuronal dendrites (Fig. 3b; Supplementary Fig. 9h), the levels of SYN, VGLUT1, GLUR1 (also known as GRIA1), VGAT (also known as SLC32A1) and GEPH (also known as GPHN) were unaffected (Supplementary Fig. 9e–j). Decreased PSD95 synaptic density

in SCZD hiPSC neurons failed to reach statistical significance (Fig. 3c and Supplementary Fig. 9c).

We used electrophysiology and calcium transient imaging to measure spontaneous neuronal activity (Fig. 3d–k and Supplementary Fig. 10). SCZD hiPSC neurons showed normal transient inward sodium currents and sustained outward potassium currents in response to membrane depolarizations (Fig. 3d), action potentials to somatic current injections (Fig. 3e), spontaneous excitatory postsynaptic currents (EPSCs) (Fig. 3f) and spontaneous inhibitory postsynaptic currents (IPSCs) (Fig. 3g). The amplitude and rate of spontaneous calcium transients were unaffected (Fig. 3h–j and Supplementary Fig. 10a–d) and there was no difference in synchronicity of spontaneous calcium transients (Fig. 3k and Supplementary Fig. 10e–g).

Increased *NRG1* expression has been observed in post-mortem SCZD brain tissue¹³. *NRG1* expression was increased in SCZD hiPSC neurons (Fig. 4d–f) but not SCZD fibroblasts, hiPSCs or NPCs (Fig. 4e), demonstrating the importance of studying gene expression changes in the cell type relevant to disease. In all, 596 unique genes (271 upregulated and 325 downregulated) showed greater than 1.30-fold-expression changes between SCZD and control hiPSC neurons ($P < 0.05$) (Supplementary Fig. 11a, b and Supplementary Table 3). Of these genes, 13% (74) have published associations with SCZD and 16% (96) have been linked to SCZD by post-mortem gene expression profiles available through the Stanley Medical Research Institute¹⁴ (Supplementary Table 3); in total 25% (149) of our differentially expressed genes have been previously implicated in SCZD. Gene ontology (GO) analysis identified significant perturbations of glutamate, cAMP and WNT signalling (Fig. 4a–c, Supplementary Table 4 and Supplementary Fig. 11c), pathways required for activity-dependent refinement of synaptic connections and long-term potentiation^{15–17}. Sixteen of 17 candidate genes from these families were validated by quantitative PCR (Supplementary Table 2; Fig. 4f and Supplementary Fig. 11e).

Copy number variants (CNVs) are rare, highly penetrant structural disruptions. SCZD patients have a 1.15-fold increase in CNV burden, but how this translates into illness is unknown. Patient 4 had four CNVs involving genes previously associated with SCZD or bipolar disorder (BD)^{13,18,19}; of these, neuronal expression of *NRG3* and *GALNT11*, but not of *CYP2C19* or *GABRB2/GABRA6* (also known as *GABRB2* and *GABRA6*, respectively) was affected (Supplementary Fig. 12 and Supplementary Table 5). A second analysis of CNVs unbiased by previous genome wide association studies (GWAS) identified 42 genes affected by CNVs in our four SCZD patients (Supplementary Table 5). Although twelve of these genes showed altered neuronal expression consistent with genotype ($P < 0.05$), most changes were extremely small and only three (*CSMD1*, *MYH1*, *MYH4*) showed >1.3 -fold effects (Supplementary Table 5). Well-established SCZD CNVs occur at 1q21.1, 15q11.2, 15q13.3, 16p11.2 and 22q11.2 (refs 13, 18, 19), but the relevant genes remain unidentified. Our patients had no evidence of CNVs at these regions, and gene expression of the best candidate genes in each region, such as *GJA8* (1q21.1), *CYFIP1* (15q11.1), *CHRFAM7A* (15q13.3), *PRODH* (22q11.2), *COMT* (22q11.2) and *ZDHC8* (22q11.2)^{18,20}, was not affected in our SCZD hiPSC neurons (Supplementary Table 6).

Consistent with published reports, loxapine increased *NRG1* expression in neurons²¹. Loxapine also increased expression of several glutamate receptors, including *GRIK1*, *GRM7* and *GRIN2A*, and ameliorated expression of *ADCY8*, *PRKCA*, *WNT7A* and *TCF4* (Fig. 4f and Supplementary Fig. 11e).

SCZD hiPSC neurons from heterogeneous patients had similar deficits, replicating some but not all aspects of the cellular and molecular phenotypes observed in post-mortem human studies and animal models (Supplementary Table 1). We observed decreased neuronal connectivity in SCZD hiPSC neurons, but not defects in synaptic function; this may reflect technical limitations of our synaptic activity assays. Due to the heterogeneity of our patient cohort and small sample size, our findings might not generalize to all subtypes of SCZD and our microarray

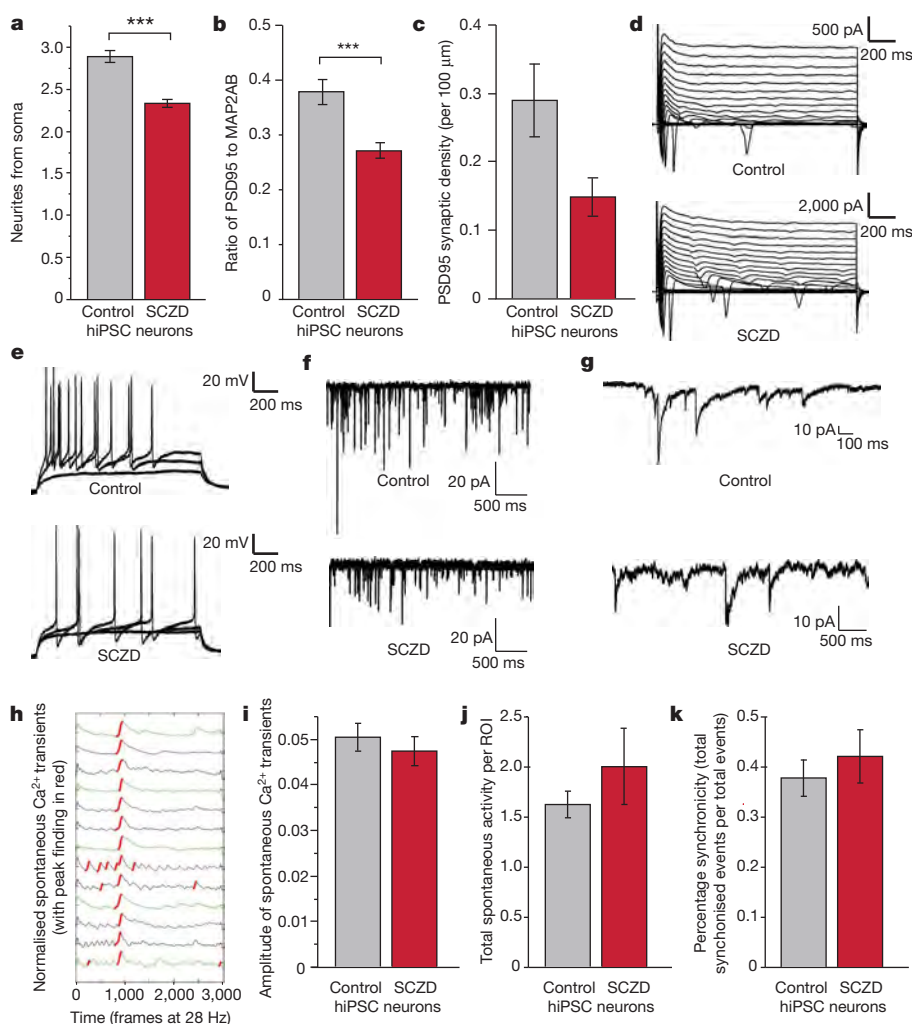


Figure 3 | Decreased neurites and synaptic protein levels but normal electrophysiological and spontaneous calcium transient activity in SCZD hiPSC neurons. **a**, Decreased neurites in SCZD hiPSC neurons. **b**, Decreased PSD95 protein relative to MAP2AB for SCZD hiPSC neurons. **c**, Trend of decreasing PSD95 synaptic density in SCZD hiPSC neurons. **d–g**, Electrophysiological characterization. hiPSC neurons grown on astrocytes show normal sodium and potassium currents when voltage-clamped (**d**), normal induced action potentials when current-clamped (**e**), and spontaneous excitatory (**f**) and inhibitory (**g**) synaptic activity.

h–k, Spontaneous calcium transient imaging. Representative spontaneous Fluo-4AM calcium traces of fluorescent intensity versus time generated from 3-month-old hiPSC neurons (**h**). There is no difference between the spike amplitude of spontaneous calcium transients of control and SCZD hiPSC neurons (**i**), no difference between the total numbers of spontaneous calcium transients per total number of regions of interest in cultures of control and SCZD hiPSC neurons (**j**), and no change in percentage synchronicity per calcium transient in control and SCZD hiPSC neurons (**k**). Error bars are s.e., $***P < 0.001$.

comparisons of SCZD and control hiPSC neurons are necessarily preliminary. Gene expression studies of hiPSC neurons permit straightforward comparisons of antipsychotic treatments on live, genetically identical neurons from patients with known clinical treatment outcomes, eliminating many confounding variables of post-mortem analysis such as treatment history, drug or alcohol abuse, and cause of death. For example, although loxapine is characterized as a high affinity antagonist of serotonin 5-HT₂ receptors and dopamine D1, D2 and D4 receptors²², treatment of SCZD hiPSC neurons resulted in altered gene expression and increased neuronal connectivity.

Of the 596 unique genes differentially expressed in our SCZD hiPSC neurons (>1.30-fold, $P < 0.05$), 25% have been previously implicated in SCZD (Supplementary Table 3). Although our gene expression profiles of SCZD hiPSC neurons confirm and extend the major hypotheses generated by pharmacological and GWAS studies of SCZD, they also identify some pathways not before linked to SCZD, such as NOTCH signalling, SLIT/ROBO axon guidance, EFNA mediated

axon growth, cell adhesion and transcriptional silencing (Supplementary Table 4). Many of the genes most affected in SCZD hiPSC neurons belong to pathways previously associated with SCZD, although they have not yet been singled out as SCZD genes. For example, whereas *PDE4B* is a well-characterized SCZD gene, we observed significant misexpression of *PDE1C*, *PDE3A*, *PDE4D*, *PDE4DIP*, *PDE7B*, *ADCY7* and *ADCY8*. Additionally, although some key SCZD/BD genes, including *NRG1* and *ANK3*, were misexpressed in all of our SCZD hiPSC neurons, many others, including *ZNF804A*, *GABRB1*, *ERBB4*, *DISC1* and *PDE4B*, were aberrantly expressed in some but not all patients (Fig. 4d). Our data support the “watershed model”²³ of SCZD whereby many different combinations of gene misfunction may disrupt the key pathways affected in SCZD. We predict that, as the number of SCZD cases studied using hiPSC neurons increases, a diminishing number of genes will be consistently affected across the growing patient cohort; instead, evidence will accumulate that a handful of essential pathways can be disrupted in diverse ways to result in SCZD.

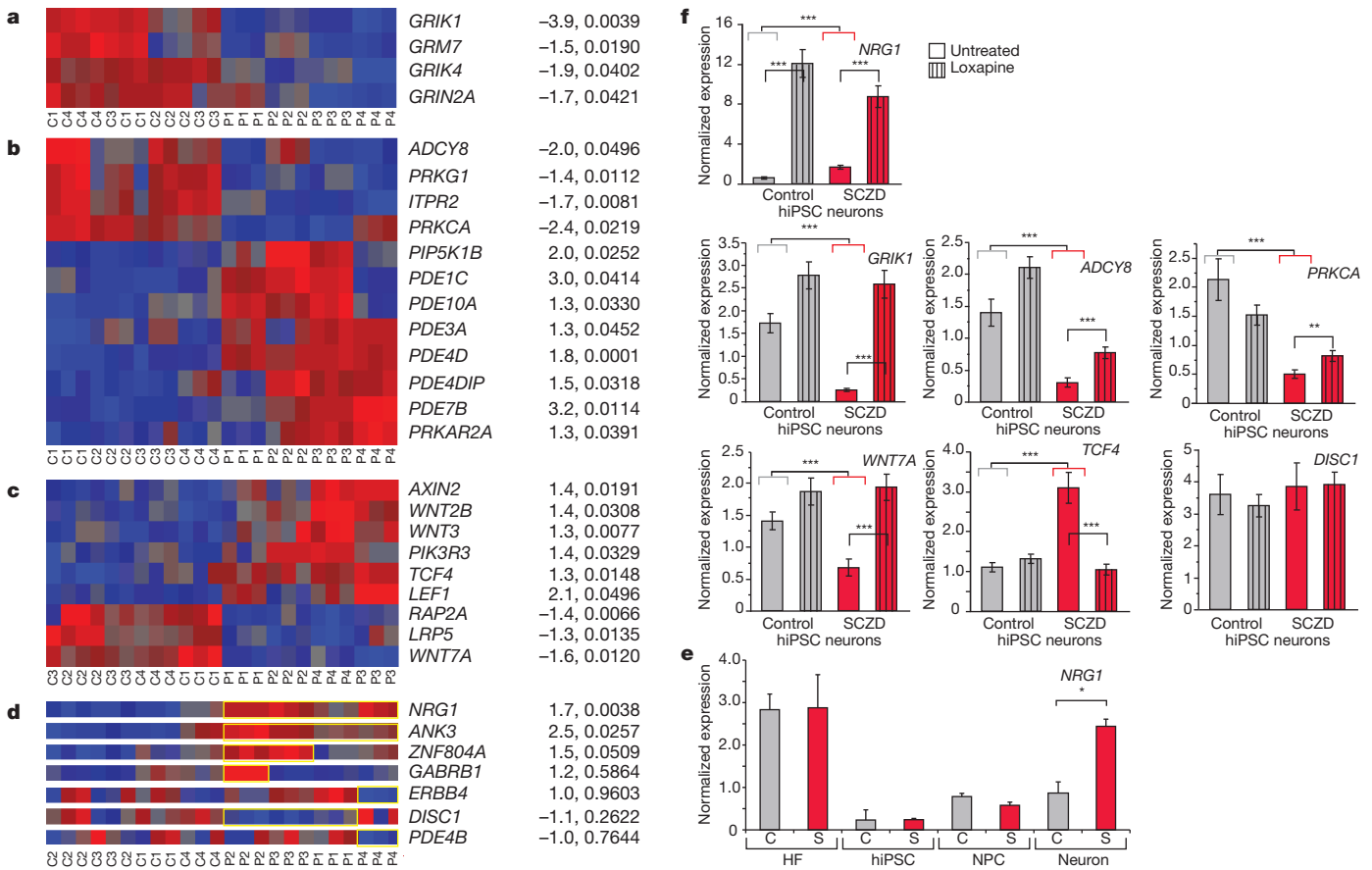


Figure 4 | RNA expression analysis of control and SCZD hiPSC neurons. **a–c**, Heat maps showing microarray expression profiles of altered expression of glutamate receptors (**a**), cAMP signalling (**b**) and WNT signalling (**c**) genes in SCZD hiPSC neurons. Fold-change and *P*-values (diagnosis) are shown on the right. **d**, Heat maps showing perturbed expression (highlighted in yellow) of *NRG1* and *ANK3* in all four SCZD patients, as well as altered expression of *ZNF804A*, *GABRB1*, *ERBB4*, *DISC1* and *PDE4B* in some but not all patients.

METHODS SUMMARY

Reprogramming hiPSCs. Control and SCZD human fibroblasts were obtained from cell repositories and were reprogrammed with tetracycline-inducible lentiviruses expressing the transcription factor genes *OCT4* (also known as *POU5F1*), *SOX2*, *KLF4*, *cMYC* and *LIN28* (ref. 7). Lentiviruses were packaged in 293T HEK cells transfected with polyethylenimine (PEI) (Polysciences). Human fibroblasts were transduced and then split onto mouse embryonic fibroblasts (mEFs). Cells were switched to HUES medium (KO-DMEM (Invitrogen), 10% KO-Serum Replacement (Invitrogen), 10% Plasminate (Talecris), 1× GlutaMAX (Invitrogen), 1× NEAA (Invitrogen), 1× 2-β-mercaptoethanol (Sigma) and 20 ng ml⁻¹ FGF2 (Invitrogen)) and 1 μg ml⁻¹ doxycycline (Sigma) was added to HUES medium for the first 21–28 days of reprogramming. hiPSCs were generally grown in HUES medium: early passage hiPSCs were split through manual passaging, whereas at higher passages hiPSCs could be enzymatically passaged with 1 mg ml⁻¹ collagenase (Sigma).

hiPSC differentiation to NPCs and neurons. Embryoid bodies were generated from hiPSCs and then transferred to non-adherent plates (Corning). Colonies were maintained in suspension in N2 medium (DMEM/F12 (Invitrogen), 1× N2 (Invitrogen)) for 7 days and then plated onto polyornithine (PORN)/laminin-coated plates. Visible rosettes formed within 1 week and were manually dissected and cultured in NPC medium (DMEM/F12, 1× N2, 1× B27-RA (Invitrogen), 1 μg ml⁻¹ laminin (Invitrogen) and 20 ng ml⁻¹ FGF2 (Invitrogen)). NPCs are maintained at high density, grown on PORN/laminin-coated plates in NPC medium and split approximately 1:4 every week with Accutase (Millipore). For neural differentiations, NPCs were dissociated with Accutase and plated at low density in neural differentiation medium (DMEM/F12-Glutamax, 1× N2, 1× B27-RA, 20 ng ml⁻¹ BDNF (Peprotech), 20 ng ml⁻¹ GDNF (Peprotech),

1 mM dibutyl-cyclic AMP (Sigma), 200 nM ascorbic acid (Sigma) onto PORN/laminin-coated plates.

Assays for neuronal connectivity, neurite outgrowth, synaptic protein expression, synaptic density, electrophysiology, spontaneous calcium transient imaging and gene expression were used to compare control and SCZD hiPSC neurons.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1 mM dibutyl-cyclic AMP (Sigma), 200 nM ascorbic acid (Sigma) onto PORN/laminin-coated plates.

Assays for neuronal connectivity, neurite outgrowth, synaptic protein expression, synaptic density, electrophysiology, spontaneous calcium transient imaging and gene expression were used to compare control and SCZD hiPSC neurons.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 July 2010; accepted 10 February 2011.

Published online 13 April 2011.

- Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60**, 1187–1192 (2003).
- Wong, A. H. & Van Tol, H. H. Schizophrenia: from phenomenology to neurobiology. *Neurosci. Biobehav. Rev.* **27**, 269–306 (2003).
- Javitt, D. C., Spencer, K. M., Thaker, G. K., Winterer, G. & Hajos, M. Neurophysiological biomarkers for drug development in schizophrenia. *Nature Rev. Drug Discov.* **7**, 68–83 (2008).
- Ebert, A. D. et al. Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* **457**, 277–280 (2009).
- Lee, G. et al. Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* **461**, 402–406 (2009).
- Marchetto, M. C. et al. A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* **143**, 527–539 (2010).
- Maherali, N. et al. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55–70 (2007).
- Ugolini, G. Use of rabies virus as a transneuronal tracer of neuronal connections: implications for the understanding of rabies pathogenesis. *Dev. Biol. (Basel)* **131**, 493–506 (2008).

9. Wickersham, I. R. *et al.* Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53**, 639–647 (2007).
10. Lafon, M. Rabies virus receptors. *J. Neurovirol.* **11**, 82–87 (2005).
11. Selemon, L. D. & Goldman-Rakic, P. S. The reduced neuropil hypothesis: a circuit based model of schizophrenia. *Biol. Psychiatry* **45**, 17–25 (1999).
12. Jaaro-Peled, H., Ayhan, Y., Pletnikov, M. V. & Sawa, A. Review of pathological hallmarks of schizophrenia: comparison of genetic models with patients and nongenetic models. *Schizophr. Bull.* **36**, 301–313 (2010).
13. Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539–543 (2008).
14. Higgs, B. W., Elashoff, M., Richman, S. & Barci, B. An online database for brain disease research. *BMC Genomics* **7**, 70 (2006).
15. Patil, S. T. *et al.* Activation of mGlu2/3 receptors as a new approach to treat schizophrenia: a randomized Phase 2 clinical trial. *Nature Med.* **13**, 1102–1107 (2007).
16. Patterson, S. L. *et al.* Some forms of cAMP-mediated long-lasting potentiation are associated with release of BDNF and nuclear translocation of phospho-MAP kinase. *Neuron* **32**, 123–140 (2001).
17. Freyberg, Z., Ferrando, S. J. & Javitch, J. A. Roles of the Akt/GSK-3 and Wnt signaling pathways in schizophrenia and antipsychotic drug action. *Am. J. Psychiatry* **167**, 388–396 (2010).
18. Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).
19. The International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).
20. Karayiorgou, M. & Gogos, J. A. The molecular genetics of the 22q11-associated schizophrenia. *Brain Res. Mol. Brain Res.* **132**, 95–104 (2004).
21. Wang, X. D., Su, Y. A., Guo, C. M., Yang, Y. & Si, T. M. Chronic antipsychotic drug administration alters the expression of neuregulin 1 β , ErbB2, ErbB3, and ErbB4 in the rat prefrontal cortex and hippocampus. *Int. J. Neuropsychopharmacol.* **11**, 553–561 (2008).
22. Kapur, S. *et al.* PET evidence that loxapine is an equipotent blocker of 5-HT₂ and D₂ receptors: implications for the therapeutics of schizophrenia. *Am. J. Psychiatry* **154**, 1525–1529 (1997).
23. Cannon, T. D. & Keller, M. C. Endophenotypes in the genetic analyses of mental disorders. *Annu. Rev. Clin. Psychol.* **2**, 267–290 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements L. Moore, B. Miller, K. Stecker, J. Jepsen, D. Sepp, S. Larkin and L. Johnson provided technical assistance. T. Berggren directs, and M. Lutz manages, the Salk Stem Cell facility. D. Gibbs directs the Salk Viral Vector Core. J. Nguyen and L. Ouyang provided gene expression support. D. Chambers and J. Barrie provided FACS support. E. Callaway and I. Wickersham provided rabies trans-neuronal tracing viruses and invaluable advice and scientific feedback. M. Lawson provided assistance with statistical analysis. Thanks to G. Yeo, M. McConnell, S. Aigner, C. Marchetto and L. Boyer for advice and conversation. K.J.B. is supported by a training grant from the California Institute for Regenerative Medicine. The Gage Laboratory, and this project, is partially funded by CIRM Grant RL1-00649-1, The Lookout and Mathers Foundation, the Helmsley Foundation as well as Sanofi-Aventis.

Author Contributions K.J.B. designed the experiments with F.H.G. K.J.B. completed the experiments and wrote the manuscript. A.S. contributed to the microarray analysis and qPCR experiments. J.J. established the synaptic density assay and completed the calcium transient experiments. C.G.-B. and S.S. performed most of the synaptic protein experiments. N.T. analysed the rabies data. N.T. and S.S. together counted neurites. Y.L., Y.M. and G.C. performed electrophysiology. D.Y. established the calcium transient assay. S.M.C. and J.S. completed the CNV analysis.

Author Information The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE25673. As per our agreement with Coriell Cell Repository, all hiPSC lines generated from control and schizophrenic fibroblasts will only be available from Coriell. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.H.G. (gage@salk.edu).

METHODS

Description of SCZD patients. All patient samples were obtained from the Coriell collection. Patients were selected based on the high likelihood of a genetic component to disease. Patient 1 (GM02038, male, 22 years of age, Caucasian) was diagnosed with SCZD at 6 years of age and committed suicide at 22 years of age. Patient 2 (GM01792, male, 26 years of age, Jewish Caucasian) displayed episodes of agitation, delusions of persecution, and fear of assassination. His sister, patient 3 (GM01835, female, 27 years of age, Jewish Caucasian) had a history of schizoaffective disorder and drug abuse. Patient 4 (GM02497, male, 23 years of age, Jewish Caucasian) was diagnosed with SCZD at age 15 and showed symptoms including paralogical thinking, affective shielding, splitting of affect from content, and suspiciousness. His sister, patient 5 (GM02503, female, 27 years of age, Jewish Caucasian), was diagnosed with anorexia nervosa in adolescence and with schizoid personality disorder (SPD) as an adult. SPD has an increased prevalence in families with SCZD and is a milder diagnosis characterized not by psychosis, but rather by a lack of interest in social relationships and emotional coldness²⁴. Although we show data from SPD patient 5 as an interesting point of comparison, we do not consider patient 5 to belong to either the 'control' or 'SCZD' groups.

Preliminary experiments were controlled using BJ fibroblasts from ATCC (CRL-2522). These fibroblasts were expanded from foreskin tissue of a newborn male. They are readily reprogrammed, low passage, karyotypically normal and extremely well-characterized primary fibroblast line cells. Age- and ancestry-matched controls were obtained from three Coriell collections: apparently healthy individuals with normal psychiatric evaluations, apparently healthy non-fetal tissue and gerontology research centre cell cultures. hiPSCs were generated from GM02937 (male, 22 years of age), GM03440 (male, 20 years of age), GM03651 (female, 25 years of age), GM04506 (female, 22 years of age), AG09319 (female, 24 years of age) and AG09429 (female, 25 years of age).

Generation of lentivirus. Lentivirus was packaged in 293T HEK cells grown in 293T medium (IMEM (Invitrogen), 10% FBS (Gemini), 1× GlutaMAX (Invitrogen)). 293T cells were transfected with polyethylenimine (PEI) (Polysciences). Per 15-cm plate, the following solution was prepared, incubated for 5 min at room temperature and added drop-wise to plates: 12.2 µg lentiviral DNA, 8.1 µg MDL-gagpol, 3.1 µg Rev-RSV, 4.1 µg CMV-VSVG, 500 µl of IMDM and 110 µl PEI (1 µg µl⁻¹) and vortexed lightly. Medium was changed after 3 h and the virus was harvested at 48 and 72 h post transfection.

hiPSC derivation. Human fibroblasts were cultured on plates treated with 0.1% gelatin (in milli-Q water) for a minimum of 30 min and grown in HF medium (DMEM (Invitrogen), 10% FBS (Gemini), 1× GlutaMAX (Invitrogen), 5 ng ml⁻¹ FGF2 (Invitrogen)).

Human fibroblasts were infected daily for 5 days with tetracycline-inducible lentiviruses expressing *OCT4*, *SOX2*, *KLF4*, *cMYC* and *LIN28*, driven by a sixth lentivirus expressing the reverse tetracycline transactivator (rtTA)⁷. Cells from a single well of a 6-well dish were split onto a 10-cm plate containing 10⁶ mouse embryonic fibroblasts (mEFs). Cells were switched to HUES medium (KO-DMEM (Invitrogen), 10% KO-Serum Replacement (Invitrogen), 10% Plasmanate (Talecris), 1× GlutaMAX (Invitrogen), 1× NEAA (Invitrogen), 1× 2-β-mercaptoethanol (Sigma) and 20 ng ml⁻¹ FGF2 (Invitrogen)). Doxycycline (1 µg ml⁻¹; Sigma) was added to HUES medium for the first 21–28 days of reprogramming.

hiPSC colonies were picked manually and clonally plated onto 24-well mEF plates. hiPSC lines were either maintained on mEFs in HUES medium or on Matrigel (BD) in TeSR medium (StemCell Technologies). At early passages, hiPSCs were split through manual passaging. At higher passages, hiPSC could be enzymatically passaged with collagenase (1 mg ml⁻¹ in DMEM; Sigma). Cells were frozen in freezing medium (DMEM, 10% FBS, 10% DMSO).

Karyotyping analysis was performed by Cell Line Genetics or by M. Dell'Aquila. Teratoma analysis was performed by injecting hiPSCs into the kidney capsules of isoflurane-anesthetized NOD-SCID mice. Teratomas were harvested eight weeks post-injection, paraffin-embedded and stained with haematoxylin and eosin.

hiPSC differentiation to NPCs and neurons. hiPSCs grown in HUES medium on mEFs were incubated with collagenase (1 mg ml⁻¹ in DMEM) at 37 °C for 1–2 h until colonies lifted from the plate and were transferred to a non-adherent plate (Corning). Embryoid bodies were grown in suspension in N2 medium (DMEM/F12-GlutaMAX (Invitrogen), 1× N2 (Invitrogen)). After 7 days, embryoid bodies were plated in N2 medium with 1 µg ml⁻¹ laminin (Invitrogen) onto polyornithine (PORN)/laminin-coated plates. Visible rosettes formed within one week and were manually dissected onto PORN/laminin-coated plates. Rosettes were cultured in NPC medium (DMEM/F12, 1x N2, 1x B27-RA (Invitrogen), 1 µg ml⁻¹ Laminin and 20 ng ml⁻¹ FGF2) and dissociated in TrypLE (Invitrogen) for 3 min at 37 °C. NPCs are maintained at high density, grown on PORN/laminin-coated plates in NPC medium and split approximately 1:4 every week with Accutase (Millipore).

For neural differentiations, NPCs were dissociated with Accutase and plated in neural differentiation medium (DMEM/F12, 1× N2, 1× B27-RA, 20 ng ml⁻¹ BDNF (Peprotech), 20 ng ml⁻¹ GDNF (Peprotech), 1 mM dibutyryl-cyclic AMP (Sigma), 200 nM ascorbic acid (Sigma) onto PORN/laminin-coated plates. Density is critical and the following guidelines were used: 2-well Permanox slide, 80,000–100,000 cells per well; 24-well, 40,000–60,000 cells per well; 6-well, 200,000 cells per well. hiPSC-derived neurons were differentiated for 1–3 months. Notably, synapse maturation occurs most robustly *in vitro* when hiPSC neurons are cultured together with wild-type human cerebellar astrocytes (ScienCell). FBS (0.5%) was supplemented into neural differentiation media for all astrocyte coculture experiments.

It is difficult to maintain healthy neurons for 3 months of differentiation and some cultures invariably fail or become contaminated. When even one SCZD patient neural culture failed, the experiments were abandoned as all assays were conducted on neurons cultured in parallel. If, however, only a control neural culture failed, and at least three control samples remained, analysis was completed. For this reason, although patients are consistently numbered throughout the manuscript, controls are not, and are instead listed in numerical order (BJ, GM02937, GM03651, GM04506, AG09319, AG09429).

Antipsychotic drugs were added for the final 3 weeks of a 3-month differentiation on astrocytes and for the final two weeks of a 6-week differentiation on PORN/laminin alone. Drugs were resuspended in DMSO at the following concentrations: clozapine (5 µM), loxapine (10 µM), olanzapine (1 µM), risperidone (10 µM) and thioridazine (5 µM).

Immunohistochemistry. Cells were fixed in 4% paraformaldehyde in PBS at 4 °C for 10 min. hiPSCs and NPCs were permeabilized at room temperature for 15 min in 1.0% Triton in PBS. All cells were blocked in 5% donkey serum with 0.1% Triton at room temperature for 30 min. The following primary antibodies and dilutions were used: mouse anti-Oct4 (Santa Cruz), 1:200; goat anti-Sox2 (Santa Cruz), 1:200; goat anti-Nanog (R&D), 1:200; mouse anti-Tral-60 (Chemicon), 1:100; mouse anti-human Nestin (Chemicon), 1:200; rabbit anti-βIII-tubulin (Covance), 1:200; mouse anti-βIII-tubulin (Covance), 1:200; rabbit anti-cow-GFAP (Dako) 1:200; mouse anti-MAP2AB (Sigma), 1:200; rabbit anti-synapsin (Synaptic Systems), 1:500; mouse anti-PSD95 (UCDavis/NIH Neuromab), 1:500; rabbit anti-PSD95 (Invitrogen), 1:200; rabbit anti-VGLUT1 (Synaptic; Systems), 1:500; rabbit anti-gephyrin, (Synaptic Systems), 1:500; mouse anti-vGAT (Synaptic Systems), 1:500; rabbit anti-vGat (Synaptic Systems), 1:500; rabbit anti-GLUR1 (Oncogene), 1:100; rabbit anti-GABA (Sigma), 1:200; rabbit anti-GAD65/67 (Sigma), 1:200.

Secondary antibodies were Alexa donkey 488, 555 and 647 anti-rabbit (Invitrogen), Alexa donkey 488 and 555 anti-mouse (Invitrogen), and Alexa donkey 488, 555, 568 and 594 anti-goat (Invitrogen); all were used at 1:300. To visualize nuclei, slides were stained with 0.5 µg ml⁻¹ DAPI (4',6-diamidino-2-phenylindole) and then mounted with Vectashield. Images were acquired using a Bio-Rad confocal microscope.

FACS. For sorting of dissociated two-month-old hiPSC neurons, cultures were dissociated with Accutase for 5 min, washed in DMEM, centrifuged at 500g and resuspended in PBS. Cells were fixed in 4% paraformaldehyde in PBS at 4 °C for 10 min. Cells were washed in PBS and aliquoted into 96-well conical plates. Cells were blocked in 5% donkey serum with 0.1% saponin at room temperature for 30 min. The following primary antibodies and dilutions were used for one hour at room temperature: rabbit anti-βIII-tubulin (Sigma), 1:200; mouse anti-βIII-tubulin (Covance), 1:200; rabbit anti-GAD56/67 (Sigma), 1:200. Cells were washed and then incubated with secondary antibodies at 1:200 for 30 min at room temperature: Alexa donkey 647 anti-rabbit (Invitrogen), and Alexa donkey 488 anti-mouse (Invitrogen). Cells were washed three times in PBS and stained with 0.5 µg ml⁻¹ DAPI. Cells were resuspended in PBS with 5% donkey serum and 0.1% detergent saponin. The homogeneous solution was filtered through a 250-µm nylon sieve and run in a BD FACSCalibur. Data were analysed using FloJo.

Rabies virus trans-neuronal tracing. Rabies virus trans-neuronal tracing was performed on 3-month-old hiPSC neurons grown together with wild-type human astrocytes (ScienCell) on acid-etched glass coverslips and then transduced with LV-SYNP-HTG or LV-SYNP-HT. Cultures were transduced with Rabies-ENVAΔG-RFP after at least a week to allow expression of ENVA and rabies G. Either 5, 7 or 10 days later, hiPSC neurons were fixed with 4% paraformaldehyde in PBS for fluorescent microscopy or FACS analysis; cells for FACS analysis were first dissociated with Accutase before fixation.

Neurite analysis. Neurite analysis was performed on 3-month-old hiPSC neurons grown together with wild-type human astrocytes (ScienCell) on acid-etched glass coverslips. Low titre transduction of a lentivirus driving expression of GFP from the SYN promoter (LV-SYNP-GFP) occurred at least 7 days before assay. LV-SYNP-GFP was used to image and count branching neurites from single neurons

(Fig. 3a). The number of neurites extending from the soma of 691 single LV-SYNP-GFP-labelled neurons was determined by a blinded count.

Synaptic protein staining analysis. Synaptic protein staining was performed on 3-month-old hiPSC neurons grown together with wild-type human astrocytes (ScienCell) on acid-etched glass coverslips. To calculate ratios of MAP2AB-positive dendrites and synaptic proteins, confocal images were taken at $\times 630$ magnification and $\times 4$ zoom. Using NIH ImageJ, images were thresholded and the integrated pixel density was determined for each image. Integrated pixel density measurement is the product of area (measured in square pixels) and mean grey value (the sum of the grey values of all the pixels in the selection divided by the number of pixels).

Synapse density. Synapse density analysis was performed on 3-month-old hiPSC neurons grown together with wild-type human astrocytes (ScienCell) on acid-etched glass coverslips. Manual counts of synaptic density were done in three steps using NIH ImageJ. First, the colocalization plugin was used to identify colocalization of VGLUT1 and PSD95. Second, the particle analysis function was used to restrict size 50-infinity. Third, dendrites were traced using the NeuronJ plugin. The mask generated by particle analysis was overlaid on the trace generated by NeuronJ and synapses were manually counted.

Electrophysiology. Whole-cell perforated patch recordings were performed on SCZD ($n = 30$) and control ($n = 20$) 3-month-old hiPSC neurons grown together with wild-type human astrocytes (ScienCell) on acid-etched coverslips and typically transduced with LV-SYNP-GFP. The recording micropipettes (tip resistance 3–6 M Ω) were tip-filled with internal solution composed of 115 mM K-gluconate, 4 mM NaCl, 1.5 mM MgCl₂, 20 mM HEPES, and 0.5 mM EGTA (pH 7.4) and then back-filled with the same internal solution containing 200 $\mu\text{g ml}^{-1}$ amphotericin B (Calbiochem). Recordings were made using an Axopatch 200B amplifier (Axon Instruments). Signals were sampled and filtered at 10 kHz and 2 kHz, respectively. The whole-cell capacitance was fully compensated, whereas the series resistance was uncompensated but monitored during the experiment by the amplitude of the capacitive current in response to a 5-mV pulse. The bath was constantly perfused with fresh HEPES-buffered saline composed of 115 mM NaCl, 2 mM KCl, 10 mM HEPES, 3 mM CaCl₂, 10 mM glucose and 1.5 mM MgCl₂ (pH 7.4). For voltage-clamp recordings, cells were clamped at -60 to -80 mV; Na⁺ currents and K⁺ currents were stimulated by voltage step depolarizations. Command voltage varied from -50 to $+20$ mV in 10 mV increments. For current-clamp recordings, induced action potentials were stimulated with current steps from -0.2 to $+0.5$ nA. All recordings were performed at room temperature.

Spontaneous calcium transients. Calcium imaging analysis was performed on 2.5- to 3-month-old hiPSC neurons grown together with wild-type human astrocytes (ScienCell) on acid-etched glass coverslips. Culture medium was removed and hiPSC cultures were incubated with 0.4 μM Fluo-4AM (Molecular Probes) and 0.02% Pluronic F 127 detergent in Krebs HEPES buffer (KHB) (10 mM HEPES, 4.2 mM NaHCO₃, 10 mM dextrose, 1.18 mM MgSO₄·2H₂O, 1.18 mM KH₂PO₄, 4.69 mM KCl, 118 mM NaCl, 1.29 mM CaCl₂; pH 7.3) for 1 h at room temperature. Cells were washed with KHB buffer, incubated for 2 min with Hoechst dye diluted 1:1,000 in KHB, and allowed to incubate for an additional 15 min in KHB to equilibrate intracellular dye concentration. Time-lapse image sequences ($\times 100$ magnification) were acquired at 28 Hz using a Hamamatsu ORCA-ER digital camera with a 488-nm (FITC) filter on an Olympus IZ81 inverted fluorescence confocal microscope. Images were acquired with MetaMorph.

In total, eight independent neural differentiations were tested per patient, 210 movies of spontaneous calcium transients (110 control and 100 SCZD) were generated and 2,676 regions of interest (1,158 control and 1,518 SCZD) were analysed. Up to four 90-s videos of Fluo-4AM fluorescence were recorded per neural differentiation per patient with a spinning disc confocal microscope at 28 frames per second (Supplementary Fig. 5A). Using ImageJ software, regions of interest can be manually selected and the mean pixel intensity of each region of interest can be followed over time, generating time trace data for each region of interest. The data were analysed in Matlab where background subtraction was performed by normalizing traces among traces of the sample, and spike events were identified based on the slope and amplitude of the time trace.

The amplitude of spontaneous calcium transients was calculated by measuring the change in total pixel intensity for each normalized calcium transient trace. The

rate was determined by dividing the total number of spontaneous calcium transients for any regions of interest by the total length of the movie (90 s). The synchronicity of spontaneous calcium transients was determined by two independent calculations. First, to determine the percentage synchronicity per calcium transient, the total number of synchronized calcium transients, defined as three or more simultaneous peaks, was divided by the total number of spontaneous calcium transients identified. Second, to calculate the maximum percentage synchronicity, the maximum number of regions of interest involved in a single synchronized event was divided by the total number of regions of interest identified.

CNV analysis. Cells were lysed in DNA lysis solution (100 mM Tris, pH 8.5, 5 mM EDTA, 200 mM NaCl, 0.2% (w/v) sarcosyl and 100 $\mu\text{g ml}^{-1}$ fresh proteinase K) overnight at 50 °C. DNA was precipitated by the addition of an equal volume of NaCl-ethanol mixture (150 μl of 5 M NaCl in 10 ml cold 95% ethanol) and then washed three times in 70% ethanol before resuspension in water with RNase A overnight at 4 °C.

Genome scans were performed using NimbleGen HD2 arrays (NimbleGen Systems) according to the manufacturer's instructions using a standard reference genome SKN1. NimbleGen HD2 dual-colour intensity data were normalized in a two-step process: first, a 'spatial' normalization of probes was performed to adjust for regional differences in intensities across the surface of the array; second, the Cy5 and Cy3 intensities were adjusted to a fitting curve by invariant set normalization, preserving the variability in the data. The log₂ ratio for each probe was then estimated using the geometric mean of normalized and raw intensity data²⁵.

CNV analysis was completed to identify deletions and duplications present within our patients. By using a virtual 'genotyping' step whereby individual CNV segment probe ratios were converted into z-scores, a distribution of median z-scores was generated, outliers of which were considered to be true CNVs. In doing so, we better filtered out common artefacts and false-positive CNVs and generated a list of CNVs unbiased by previous genetic studies of SCZD.

Patient fibroblasts were used for CNV analysis. Lymphocytes were available for patients 4 and 5 and their parents, allowing us to validate the CNVs identified for patient 4 and also to determine the parent of origin for each mutation; many were inherited from the unaffected mother (Supplementary Table 7).

Gene expression analysis. Unless otherwise specified, gene expression analysis was performed on 6-week-old hiPSC neurons without astrocyte coculture. Cells were lysed in RNA BEE (Tel-test). RNA was chloroform-extracted, pelleted with isopropanol, washed with 70% ethanol and resuspended in water. RNA was treated with RQ1 RNase-free DNase (Promega) for 30 min at 37 °C and then the reaction was inactivated by incubation with EGTA stop buffer at 65 °C for 10 min.

For gene expression microarrays, three independent neural differentiations for each of the four SCZD patients as well as four control subjects were compared using Affymetrix Human 1.0ST arrays as specified by the manufacturer.

Gene expression microarray analysis was completed using Partek Genomics Suite software. Intensity values were generated as follows: RMA background correction, quantile normalization, log₂ transformation and mean polished probeset summarization. Pathway analysis was performed using Metacore GeneGo.

For qPCR, cDNA was synthesized using Superscript III at 50 °C for 1–2 h, inactivated for 15 min at 70 °C and then treated with RNase H for 15 min at 37 °C, inactivated with EDTA and heated to 70 °C for 15 min. qPCR was performed using SYBR Green. Primers used are listed in Supplementary Table 8.

Statistical analysis. Statistical analysis was performed using JMP. Box-Cox transformation of raw data was performed to correct non-normal distribution of the data and residuals. Improvements were assessed by Shapiro–Wilk *W* test of the transformed data and residuals. Means were compared within diagnosis by one-way analysis using both Student's *t*-test and Tukey–Kramer HSD. Finally, a nested analysis of values for individual patients was performed using standard least squares analysis comparing means for all pairs using Student's *t*-test for specific pairs and Tukey–Kramer HSD for multiple comparisons.

24. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, fourth edition (DSM-IV)*. (American Psychiatric Press, 1994).
25. McCarthy, S. E. *et al.* Microduplications of 16p11.2 are associated with schizophrenia. *Nature Genet.* **41**, 1223–1227 (2009).

A Model for Neural Development and Treatment of Rett Syndrome Using Human Induced Pluripotent Stem Cells

Maria C.N. Marchetto,^{1,5} Cassiano Carromeu,^{2,5} Allan Acab,² Diana Yu,¹ Gene W. Yeo,³ Yangling Mu,¹ Gong Chen,⁴ Fred H. Gage,¹ and Alysson R. Muotri^{2,*}

¹The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA

²University of California San Diego, School of Medicine, Department of Pediatrics, Rady Children's Hospital San Diego, Department of Cellular and Molecular Medicine, Stem Cell Program, 9500 Gilman Drive, La Jolla, CA 92093, USA

³University of California San Diego, School of Medicine, Department of Cellular & Molecular Medicine, Stem Cell Program, 9500 Gilman Drive, La Jolla, CA 92093, USA

⁴Pennsylvania State University, Department of Biology, 201 Life Science Building, University Park, PA 6802, USA

⁵These authors contributed equally to the work

*Correspondence: muotri@ucsd.edu

DOI 10.1016/j.cell.2010.10.016

SUMMARY

Autism spectrum disorders (ASD) are complex neurodevelopmental diseases in which different combinations of genetic mutations may contribute to the phenotype. Using Rett syndrome (RTT) as an ASD genetic model, we developed a culture system using induced pluripotent stem cells (iPSCs) from RTT patients' fibroblasts. RTT patients' iPSCs are able to undergo X-inactivation and generate functional neurons. Neurons derived from RTT-iPSCs had fewer synapses, reduced spine density, smaller soma size, altered calcium signaling and electrophysiological defects when compared to controls. Our data uncovered early alterations in developing human RTT neurons. Finally, we used RTT neurons to test the effects of drugs in rescuing synaptic defects. Our data provide evidence of an unexplored developmental window, before disease onset, in RTT syndrome where potential therapies could be successfully employed. Our model recapitulates early stages of a human neurodevelopmental disease and represents a promising cellular tool for drug screening, diagnosis and personalized treatment.

INTRODUCTION

Autism spectrum disorders (ASD) are complex neurodevelopmental diseases affecting 1 in 150 children in the United States (Autism and Developmental Disabilities Monitoring Network Surveillance Year 2000 Principal Investigators; Centers for Disease Control and Prevention, 2007). Such diseases are mainly characterized by impaired social interaction and repetitive behavior. Family history and twin studies suggest that, in

some cases, these disorders share genetic roots, but the degree to which environmental and genetic patterns account for individual differences within ASD is currently unknown (Piven et al., 1997; Ronald et al., 2006). A different combination of genetic mutations is likely to play a role in each individual. Nevertheless, the study of mutations in specific genes can help to identify molecular mechanisms responsible for subtle alterations in the nervous system, perhaps pointing to common mechanisms for ASD.

Rett syndrome (RTT) is a progressive neurological disorder caused by mutations in the X-linked gene encoding MeCP2 protein (Amir et al., 1999). RTT patients have a large spectrum of autistic characteristics and are considered part of the ASD population (Hammer et al., 2002; Samaco et al., 2005, 2004; Zappella et al., 2003). These individuals undergo apparently normal development until 6–18 months of age, followed by impaired motor function, stagnation and then regression of developmental skills, hypotonia, seizures and autistic behavior (Amir et al., 1999). MeCP2 may be involved in the epigenetic regulation of target genes, by binding to methylated CpG dinucleotides within promoters, and may function as a transcriptional repressor, although this view has been challenged recently (Chahrouh et al., 2008; Yasui et al., 2007).

Pluripotent human embryonic stem cells (hESCs) have been successfully generated from early stage human embryos and can differentiate into various cell types (Thomson et al., 1998). However, to develop cellular models of human disease, it is necessary to generate cell lines with genomes predisposed to diseases. Recently, reprogramming of somatic cells to a pluripotent state by overexpression of specific genes (induced pluripotent stem cells, iPSCs) has been accomplished (Takahashi and Yamanaka, 2006; Yu et al., 2007). Isogenic pluripotent cells are attractive not only for their potential therapeutic use with lower risk of immune rejection but also for understanding complex diseases (Marchetto et al., 2010; Muotri, 2008). Although iPSCs have been generated for several neurological diseases (Dimos et al., 2008; Ebert et al., 2009; Hotta et al., 2009; Lee et al., 2009; Park et al., 2008; Soldner et al., 2009), the demonstration

of disease-specific pathogenesis and phenotypic rescue in relevant cell types is a current challenge in the field (Marchetto et al., 2010).

We have developed a human model of RTT by generating iPSCs from fibroblasts of RTT patients carrying different MeCP2 mutations and unaffected individuals. We show that RTT-iPSCs retained the capacity to generate proliferating neural progenitor cells (NPCs) and functional neurons that underwent X-inactivation. We observed a reduced number of dendritic spines and synapses in iPSC-derived neurons. Moreover, we detected an altered frequency of intracellular calcium spikes and electrophysiological defects in RTT-derived neuronal networks, revealing potential new biomarkers for RTT pathology. Gain and loss of function experiments in iPSC-derived neurons confirmed that some of the alterations observed were related to MeCP2 expression levels. Finally, we used the iPSC system to test candidate drugs to rescue synaptic deficiency in RTT neurons. Together, our results suggest that RTT and other complex CNS diseases can be modeled using the iPSC technology to investigate the cellular and molecular mechanisms underlying their abnormalities.

RESULTS

Generation of iPSCs from RTT Patients and Normal Individuals

Nonaffected control fibroblasts and cells carrying four distinct MeCP2 mutations (Figure 1A and Table S1 available online) isolated from clinically affected female patients with RTT symptoms were infected with retroviral reprogramming vectors (Sox2, Oct4, c-Myc and Klf4), as described elsewhere (Takahashi et al., 2007). After 2 to 3 weeks, compact iPSC colonies emerged from a background of fibroblasts (Figures 1B and 1C). Colonies were manually picked and transferred to matrigel (Figures 1D and 1E). We obtained at least 10 clones from each control (WT)-iPSC and RTT-iPSC that continuously expressed pluripotent markers such as Nanog, Lin28, Tra-1-81 and Sox2 (Figures 1F and 1G and Figures S1A–S1C). All iPSC clones used in this study maintained a normal karyotype (Figure 1H). Teratomas containing derivatives from all 3 embryonic germ layers confirmed that the iPSCs were able to differentiate in vivo (Figure 1I). PCR fingerprinting confirmed their derivation from respective fibroblasts (data not shown). Next, we asked if the global molecular signatures of RTT-iPSC clones carrying the two distinct MeCP2 mutations (1155del32 and Q244X) and WT-iPSC clones (from AG09319) resembled those of available hESC lines (HUES6). Gene expression profiles measured using human genome Affymetrix Gene Chip arrays were grouped by hierarchical clustering, and correlation coefficients were computed for all pair-wise comparisons (GEO accession number GSE21037). We observed that the WT-iPSC and RTT-iPSC clones were almost indistinguishable. The results clearly revealed that the iPSC and hESC lines were more similar to each other than to the respective original fibroblasts (Figure S1D). These findings, combined with manual inspection of the gene expression of known pluripotent- and fibroblast-related genes (Figures S1E and S1F), indicated that the reprogramming was

successful. In Table S2 we present a summary of all iPSC subjects and clones utilized for each experiment.

Neural Induction of iPSCs

Our protocol for neuronal differentiation is outlined in Figure 2A. We initiated neural differentiation by plating embryoid bodies (EBs). After a week, EB-derived rosettes became apparent (Figure 2B). Rosettes were then manually collected, dissociated and re-plated. The NPCs derived from rosettes formed a homogeneous population after a couple of passages. NPCs were positive for early neural precursor markers, such as Nestin, Sox1-2 and Musashi1 (Figure 2C). To obtain mature neurons, EBs in the presence of retinoic acid (RA) were dissociated and re-plated (Figure 2B). At this stage, cells were positive for Tuj1 (β -III-Tubulin) and Map2 (Microtubule-associated protein 2) (Figure 2D). Moreover, we detected expression of GABA (γ -amino butyric acid) and VGLUT1 (vesicular glutamate transporter-1). We also observed synapsin puncta outlining Map2-positive neurites (Figure 2D). We did not detect a significant alteration in RTT neuronal survival when compared to controls, as measured by Map2 staining (Figure 2E and Figure S2A). In addition, infection with a lentivirus expressing the *DsRed* gene under the control of *Synapsin* promoter (Syn::DsRed) did not reveal any difference in neuronal survival between RTT and controls (Figure 2E and Figure S2B). Interestingly, the number of GABA-positive neurons was also not affected between RTT and controls (Figure 2F and Figure S2C).

X-Inactivation during Neuronal Differentiation of RTT-iPSCs

In female hESCs, both chromosomes should be active, but one X chromosome becomes silenced upon differentiation (Dhara and Benvenisty, 2004). Similar to ESCs, female mouse iPSCs have shown reactivation of a somatically silenced X chromosome and have undergone random X-inactivation upon differentiation (Maherali et al., 2007). Because MeCP2 is an X-linked gene, we examined the ability of our RTT-iPSCs clones to reset the X chromosome (i.e., to erase X-inactivation) and whether X-inactivation would take place again after neuronal differentiation (Figure 3A). We stained RTT-iPSCs clones and their respective fibroblasts with an antibody against trimethylated histone 3 Lysine 27 (me3H3K27), an epigenetic silencing marker present on the inactive X chromosome in interphase nuclei (Silva et al., 2003). Some, but not all, undifferentiated RTT-iPSCs clones displayed diffuse immunoreactivity throughout the nucleus, similar to some hESCs, showing that the memory of the previous inactivation state had been erased (Figure 3B). For further analysis, we only selected clones that displayed a diffuse me3H3K27 pattern to differentiate into neurons. Upon neuronal differentiation, intense nuclear foci staining, a prominent diagnostic of the inactive X, was found in 80% of neurons labeled by the infection of a lentivirus carrying the neuron-specific *Synapsin* promoter driving the EGFP reporter (Syn::EGFP). Nuclear foci were also present in RTT fibroblasts before reprogramming (Figure 3B). We quantified the percentage of cells displaying either a diffuse or intense X-inactivation (nuclear foci) (Figure 3C). Our data suggest that the majority of cells in selected clones from both hESCs (99%) and iPSCs (95%) have a diffuse pattern. In contrast,

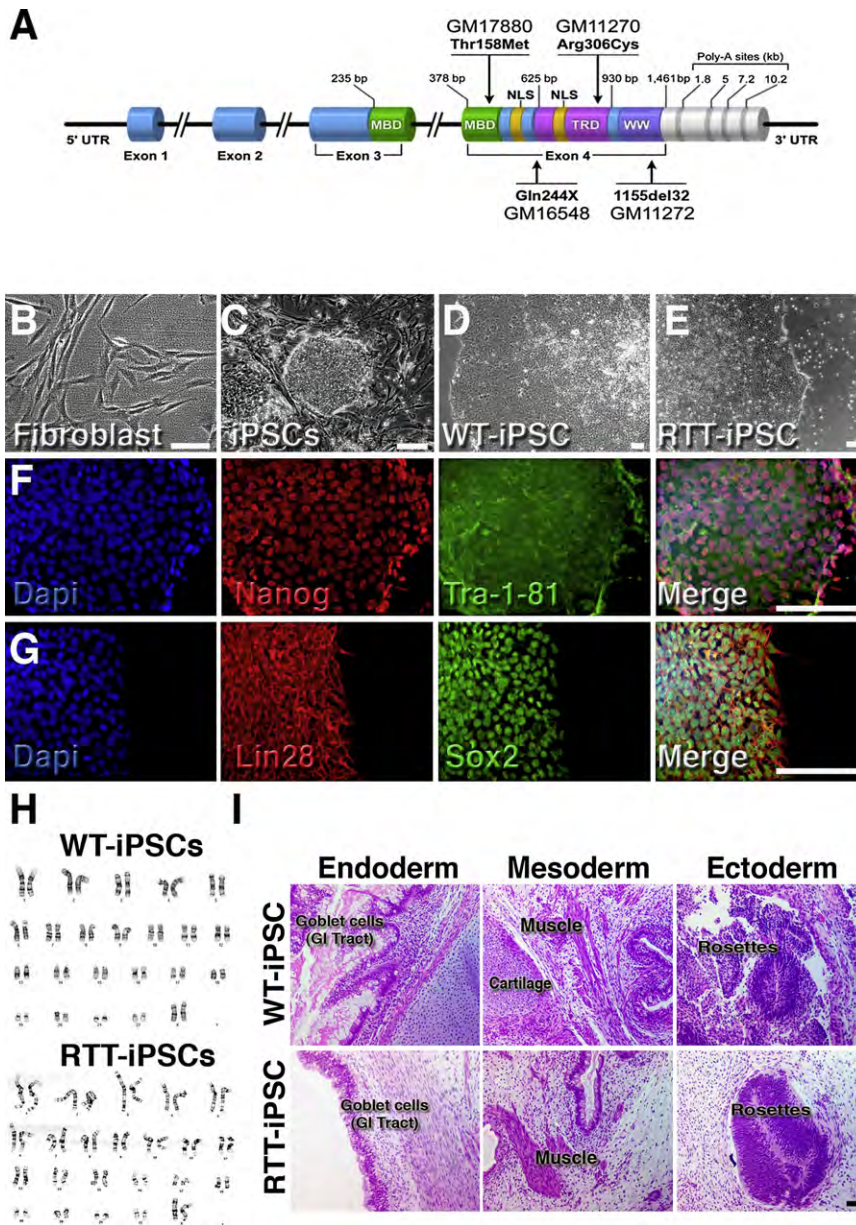


Figure 1. Generation of iPSCs

(A) Schematic representation of the MeCP2 gene structure and mutations used in this study. UTR, untranslated region; MBD, methyl-CpG binding domain; NLS, nuclear localization signal; Poly-A, polyadenylation signal; TRD, transcriptional repression domain; WW, domain-containing WW; X, stop codon. Respective cell-line codes are shown close to their mutations.

(B) Morphology of human fibroblasts before retroviral infection.

(C) Aspect of iPSCs colonies 14 days after infection.

(D and E) Representative images of established iPSC colonies.

(F and G) Representative images of RTT-iPSCs showing expression of pluripotent markers.

(H) No karyotypic abnormalities were observed.

(I) Representative images of teratoma sections.

The scale bar represents 100 μ m.

See also Figure S1.

differentiated populations of fibroblasts and iPSC-derived neurons have me3H3K27 nuclear foci staining, indicating X-inactivation.

We also used fluorescent in situ hybridization (FISH) to visualize Xist RNA, a noncoding transcript involved in X chromosome silencing that physically wraps the inactive X (Lucchesi et al., 2005). Before reprogramming, the majority of fibroblasts exhibit a clear Xist cloud. The signal is lost after reprogramming, indicating that selected iPSC clones have two active X chromosomes in our culture conditions. A Xist cloud is also observed in iPSC-derived neurons (Figure 3D). Fluorescent in situ hybridization (FISH) analysis using a centromeric X chromosome probe in iPSC-derived NPCs and neurons showed the presence of two X

chromosomes (Figure 3E). As a consequence of both X-chromosomes' activation after reprogramming, the MeCP2 protein can be detected in undifferentiated iPSCs from RTT patients (Figure 3F). However, after differentiation, RTT-iPSC-derived neurons recapitulated X-inactivation and the population became mosaic regarding MeCP2 expression. Immunostaining was performed on several RTT-iPSC clones, and a representative example of MeCP2 expression after differentiation is shown in Figure 3F. Clones obtained from RTT fibroblasts carrying the 1155del32 MeCP2 mutation do not produce a full-length MeCP2 protein (Traynor et al., 2002). Next, we selected one WT-iPSC clone (WT-33 C1) and one RTT-iPSC clone (1155del32 C15) to determine whether the RTT-iPSC-derived neuronal population showed reduced MeCP2 protein levels. As expected, we observed a reduction in the full-length

MeCP2 protein amounts in both fibroblasts and neurons derived from the RTT-iPSC clone (Figure 3G). We tested the original fibroblasts and iPSC-derived neurons from this patient for X-inactivation using standard methodology for the androgen receptor locus (Allen et al., 1992). RTT fibroblasts carrying the 1155del32 MeCP2 mutation had a 55:45 distribution, but RTT-derived neurons showed highly skewed X-inactivation, with a 96:4 distribution (Figure S3). The outcome of the X-inactivation process, measured by the androgen receptor locus, seems to be consistent within the same clone. An independent differentiation of the same clone (RTT-1155del32 C15) yielded a 98:2 distribution. Unfortunately, androgen receptor locus analysis was not conclusive for the MeCP2 mutation Q244X

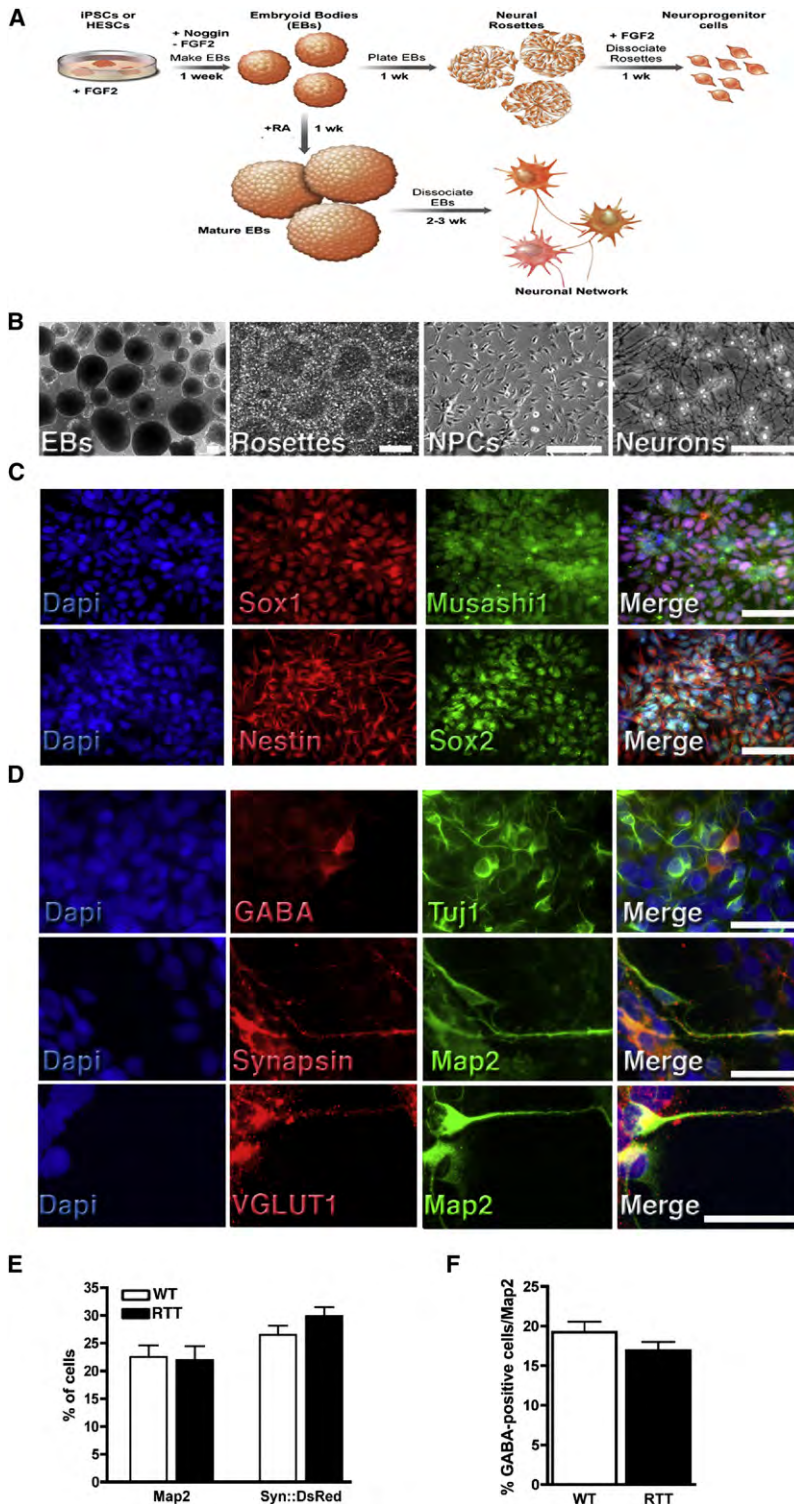


Figure 2. Neural Differentiation of iPSCs

(A) Schematic view of the neural differentiation protocol. (B) Representative images depicting morphological changes during neuronal differentiation. The scale bar represents 100 μm. (C) NPCs are positive for neural precursor markers: Sox1, Sox2, Musashi1, and Nestin. The scale bar represents 50 μm. (D–F) (D) Representative images of cells after neuronal differentiation. iPSC-derived neurons express mature neuronal markers: GABA, Map2 and Synapsin. The scale bar represents 20 μm. Similar numbers of Map2-positive and Syn::DsRed-positive (E) as well as GABA-positive (F) neurons from WT and RTT cultures. Data shown as mean ± SEM. See also Figure S2.

Our data show that X-inactivation was erased in selected reprogrammed RTT-iPSC clones and subsequently restored during neuronal differentiation. Importantly, the recapitulation of X-inactivation produces mosaic neuronal cultures with different ratios of cells expressing normal MeCP2 levels, mimicking what is observed in RTT patients’ brains. Our data do not preclude that partial reprogramming from a single fibroblast or retention of the X-inactivation would lead to clones with highly skewed X-inactivation, where neurons would express only the normal or mutant form of MeCP2. In fact, we do observe WT and RTT-iPSC clones retaining X-inactivation after reprogramming. The RTT-T158M C3-derived neurons showed 100:0 distribution. The expression of the mutant MeCP2 allele was confirmed by sequencing.

Normal Cellular Proliferation from RTT-iPSC-Derived NPCs

An increased incidence of large head size has been reported in autism (Piven et al., 1995). Other studies have suggested that the autistic brain is smaller at birth, followed by rapid head growth during early development and then a period of reduced brain growth (Courchesne et al., 2003). Head growth deceleration has also been reported for RTT patients (Hagberg et al., 2001). Since the cellular mechanism behind this phenomenon is unknown, we investigated whether a perturbed NPC replication cycle was affected in RTT. NPCs derived from RTT-iPSCs, WT-iPSCs and hESCs (Cyth25 and HUES6) were generated and kept under proliferating conditions in the presence of FGF2. NPCs were derived using the same protocol described above, had identical

cells. However, a reduction of 50% in the amount of MeCP2 protein level (Figure S4E) is consistent with a random X-inactivation. We have not analyzed the distribution for RTT-R306C clones.

passage numbers and were analyzed for cell cycle by flow cytometry. Our results showed no significant differences in any cycle phase between HESC-, WT-iPSC- and RTT-iPSC-derived NPCs (Figure 4A), though we cannot exclude the possibility that

altered head growth in RTT patients is caused by eventual abnormal NPC proliferation in another developmental stage. We then investigated potential phenotypic changes in RTT neurons compared to controls.

Reduced Glutamatergic Synapse Number and Morphological Alterations in RTT Neurons

Strong evidence implicates synapse alteration in ASD, including RTT (Zoghbi, 2003). Loss of MeCP2 and doubling of MeCP2 dosage in mice have opposite effects on excitatory synapse numbers in individual neurons (Chao et al., 2007). These results suggest that MeCP2 may be a rate-limiting factor in regulating glutamatergic synapse formation and indicate that changes in excitatory synaptic strength may underlie global network alterations in RTT. Therefore, we determined whether excitatory synapse numbers were reduced in human RTT neurons. After 8 weeks of differentiation, glutamatergic neurons were identified using antibodies against VGLUT1 (Takamori et al., 2000), and dendrites were labeled with Map2 (Figure 4B). To confirm the specificity of glutamatergic neurons in our cultures, we showed that VGLUT1 puncta were mostly adjacent to the postsynaptic density-95 (Psd95) protein (Niethammer et al., 1996) (Figure S4A). We found a reduction in the density of VGLUT1 puncta from RTT-iPSCs clones carrying 3 different MeCP2 mutations compared to HUES6 and distinct WT-iPSCs-derived Map2-positive neurons, suggesting a specific defect in glutamate transport in RTT cultures (Figure 4B and Figure S4B). Since neurons carrying different MeCP2 mutations showed reduced VGLUT1 puncta in our cultures, we tested whether loss of function of MeCP2 was directly related to the number of glutamatergic synapses in our neuronal cultures. We cloned an shRNA against MeCP2 in a lentiviral vector that is able to knock-down both isoforms of MeCP2 (Figure S4C). Neurons derived from WT-iPSCs expressing the shMeCP2 showed a similar reduction in VGLUT1 puncta when compared to control neurons expressing a scramble shRNA (shControl) (Figure 4C and Figure S4B). Overexpression of MeCP2 using a lentiviral vector (Figure S4C) increased the number of VGLUT1 puncta in WT and RTT neurons (Figure 4D and Figure S4B). Our data strongly suggest that MeCP2 is a rate-limiting factor in regulating glutamatergic synapse number in human neurons.

We also investigated whether RTT neurons displayed any morphological alteration when compared to controls. To visualize neuronal anatomy, we infected the cultures with the Syn::EGFP lentivirus. Morphological analysis of RTT neurons revealed that the number of spines in RTT neurites was reduced when compared to WT neurons and after ectopic expression of shMeCP2 (Figure 4E). Consistent with this observation, the number of spines in dendrites of neurons from postmortem RTT patients' brains was previously reported to be lower than that in normal individuals (Chapleau et al., 2009). Finally, we documented that the cell soma sizes from neurons derived from the RTT-iPSCs carrying different MeCP2 mutations were smaller when compared to controls (reduction of $14.31 \pm 4.83\%$). Similarly, loss of function using the shMeCP2 knock-down strategy in WT neurons reduced soma size at levels comparable to RTT levels (reduction of $14.52 \pm 4.31\%$) (Figure 4F and Figure S4D).

Rescuing a RTT Neuronal Phenotype

Recent studies have shown that re-activation of MeCP2 expression knockout mice led to a prolonged life span and delayed onset or reversal of certain neurological symptoms (Giacometti et al., 2007; Guy et al., 2007). These reports suggest that some RTT phenotypes can be rescued in vivo. We used our model to analyze the effect of selected compounds that may revert the neuronal phenotype in culture as a validation for future high-throughput drug screening platforms. Administration of IGF1 was recently described to promote a partial reversal of the RTT-like symptoms in a mouse model (Tropea et al., 2009). We treated RTT-derived neurons carrying different MeCP2 mutations in culture with IGF1 and observed an increase in glutamatergic synapse number, suggesting that the drug treatment could correct the RTT neuronal phenotype (Figure 4B and Figure S4B).

Around 60% of MeCP2 mutations in RTT are nonsense mutations (Laccone et al., 2001). Thus, we tested whether we could increase MeCP2 expression levels in affected neurons by suppressing the nonsense mutation (Q244X) with read-through of the premature stop codon using pharmacological treatments. High concentrations of aminoglycosides antibiotics, such as gentamicin, can bind to the 16S rRNA, impairing ribosomal proofreading (Kellermayer, 2006). As a consequence, a full-length protein is produced by incorporating a random amino acid at the stop codon position. We treated RTT-Q244X clones 3- and 4-derived neurons with two different doses of gentamicin and found that MeCP2 protein levels and glutamatergic synapse numbers were increased after 1 week (Figure 4G and Figure S4E). Treatment with a higher gentamicin dose (400ug/ml) for the same period did not rescue RTT neurons and lowered the number of VGLUT1 puncta in control neurons (Figure 4G).

The finding that RTT patient-derived neurons displayed changes in neuronal morphology and in number of synapses prompted us to explore putative circuit alterations in vitro.

Altered Activity-Dependent Calcium Transients in RTT-iPSC-Derived Cells

Early in neural development, spontaneous electrical activity leads to increases in intracellular calcium levels and activation of signaling pathways that are important in regulating several neuronal processes (Spitzer et al., 2004). Recently, a disturbance in calcium homeostasis during early postnatal development was reported in a MeCP2 knockout model (Mironov et al., 2009). Moreover, several studies showed that functional mutations in genes encoding voltage-gated calcium channels and in genes whose activity is modulated by calcium, such as MeCP2, could lead to ASD (Splawski et al., 2006; Zhou et al., 2006). Neuronal activity-induced calcium influx can trigger the calcium/calmodulin-dependent protein kinase (CamK). CamK activation has been reported to induce phosphorylation of MeCP2, which was further postulated to regulate neuronal spine maturation (Tao et al., 2009; Zhou et al., 2006). Although these studies raised an interesting link between neuronal activity and spine maturation, the extent of cellular alteration in human ASD neurons was never characterized. To test if RTT-iPSCs-derived neuronal networks are affected in our system, we preloaded the cells with the calcium indicator fluo-4AM and highlighted neurons using

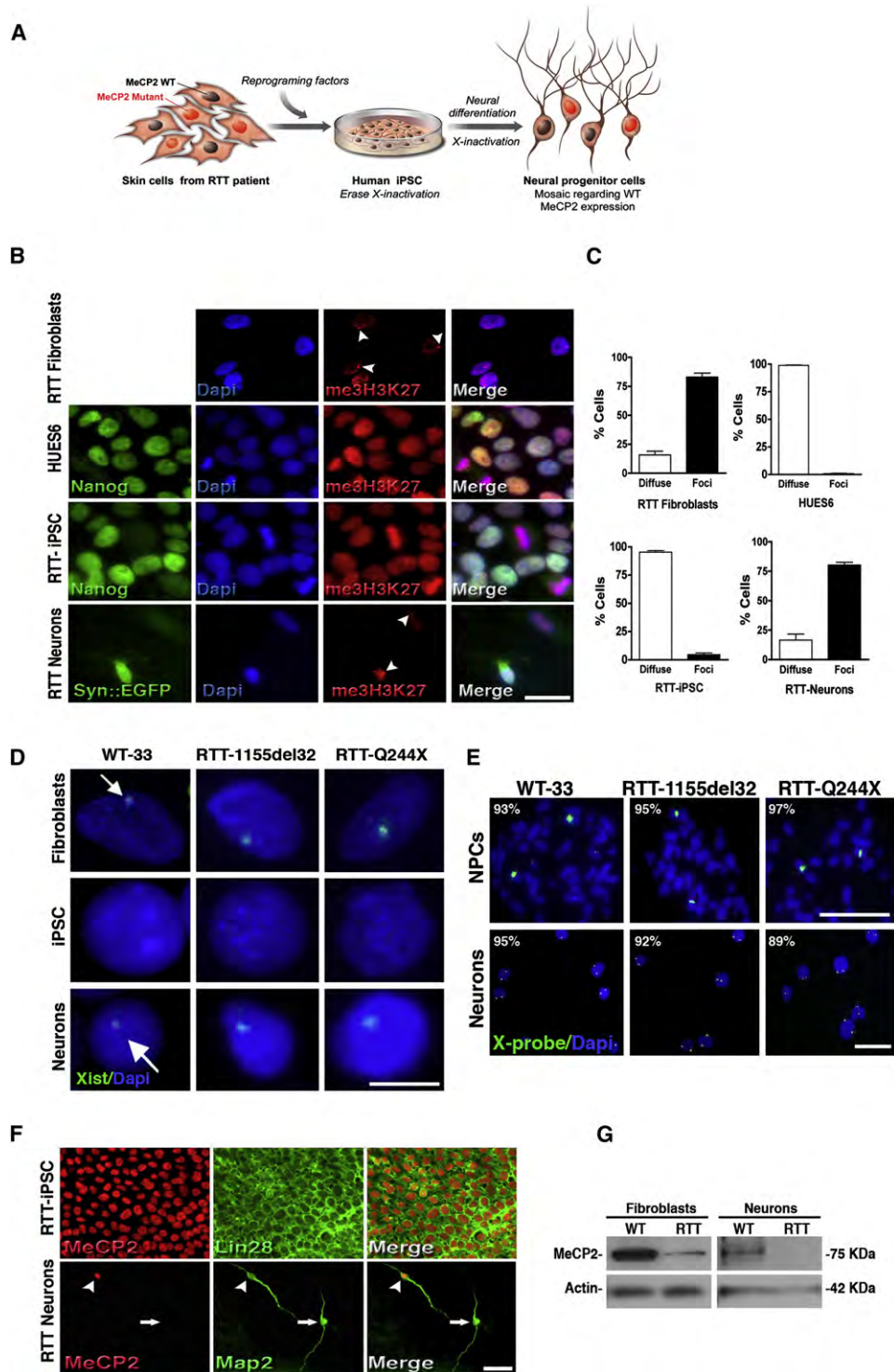


Figure 3. RTT-iPSC Clones Undergo X-Inactivation during Differentiation

(A) Schematic representation of X-inactivation dynamics during reprogramming and further neural differentiation. RTT fibroblasts are mosaic for the MeCP2 WT gene expression. During reprogramming, X-inactivation is erased and iPSCs express both MeCP2 alleles. Upon neuronal differentiation, X-inactivation is re-established and the resultant cells are mosaic for MeCP2 WT gene expression.

(B) Immunofluorescence for me3H3K27 in fibroblasts, pluripotent cells (Nanog-positive) and after neuronal differentiation (Syn::EGFP-positive). Pluripotent cells (hESCs and iPSCs) show diffuse staining whereas differentiated cells (fibroblasts and neurons) exhibit prominent me3H3K27 nuclear foci (arrowheads). Cells were counterstained with Dapi. The scale bar represents 15 μ m.

the Syn::DsRed vector. Cultures with similar cell density and numbers of DsRed-positive neurons were used (Figure S2B). Spontaneous calcium transients were analyzed from WT and RTT neuronal networks in several independent experiments over time (Figure 5).

In our analyses, we only considered calcium transients generated by synaptic activity. Neurons were selected after confirmation that calcium transients were blocked with TTX or with the glutamate receptor antagonists CNQX (AMPA) and APV (NMDA) treatments, indicating neuronal signaling dependence on local synaptic connections (Figures S5A, S5B, and S5D). Gabazine, an antagonist of GABA_A receptors, increased the number of calcium transients in the networks, indicating the presence of glutamatergic and gabaergic synapses in our system (Figure S5C, D). A representative example of calcium tracing in control and RTT neurons is depicted in Figure 5A and shows a sharp increase in amplitude followed by a decrease over time. The frequency of calcium oscillations in RTT neurons and in WT neurons expressing shMeCP2 was abnormally decreased when compared to controls, suggesting a deficiency in the neuronal network connectivity and activity dynamics (Figures 5B and 5C and Figures S5E and S5F). The deficiency in connectivity was further corroborated by a decrease in the percentage of Syn::DsRed-positive neurons exhibiting calcium transients in the RTT cultures when compared to controls (Figure 5D and Figure S5F).

Decreased Frequency of Spontaneous Postsynaptic Currents in RTT Neurons

Next we determined the functional maturation of the iPSC-derived neurons using electrophysiological methods. Whole-cell recordings were performed from cells that had differentiated for at least 6 weeks in culture. Neurons were visualized by infection with the Syn::EGFP viral vector (Figure 6A). Both WT and RTT neurons showed similar transient sodium inward currents, sustained potassium outward currents in response to voltage step depolarizations, and action potentials evoked by somatic current injections (Figure 6B). Therefore, our data indicated that WT and RTT reprogramming did not affect the ability of WT-iPSC- and RTT-iPSC-derived neurons to mature and become electrophysiologically active. We also recorded spontaneous excitatory and inhibitory postsynaptic currents (sEPSCs and sIPSCs) as a way of measuring intercellular connectivity and network formation (Figures 6B and 6C). Cumulative probability plots of amplitudes and inter-event intervals of spontaneous postsynaptic currents revealed that RTT neurons have a significant decrease in frequency and amplitude when compared to WT neurons (Figures 6D and 6E). Together, our data suggest that the neuronal network is altered in RTT iPSC-derived cultures.

DISCUSSION

The lack of detectable symptoms in female RTT patients until 6–18 months of age and the apparent phenotypic reversibility of some RTT phenotypes in MeCP2 knockout animals indicate that MeCP2 is not essential for early wiring of the nervous system but instead may only be required at late stages. It is possible that RTT patients have aberrant excitatory synaptic strength at very early stages, when the disease phenotype is not yet clearly observed. In fact, increasing evidence from clinical studies and mouse models indicates the presence of alterations during the so-called presymptomatic developmental phase (Charman et al., 2002; De Filippis et al., 2009; Kerr et al., 1987; Picker et al., 2006; Santos et al., 2007).

To study human RTT neurons in culture, we derived iPSCs from RTT fibroblasts. RTT iPSCs are pluripotent and able to recapitulate X-inactivation upon neuronal differentiation. Even though the ratio of neurons expressing mutant MeCP2 due to X-inactivation was variable, the phenotypes described here for all RTT-derived neurons are similar. One interpretation could be that astrocytes, or other nonneuronal cells, carrying MeCP2 mutations present in our cultures could also affect neurons expressing the normal MeCP2 protein. In fact, the non-cell-autonomous influence was recently described for RTT, indicating that glial cells carrying MeCP2 mutations can distress healthy neurons (Ballas et al., 2009; Kishi and Macklis, 2010; Maezawa et al., 2009).

Using human neurons carrying MeCP2 mutations, we showed that RTT glutamatergic neurons have a reduced number of synapses and dendritic spines when compared to nonaffected controls. Moreover, electrophysiological recordings from RTT neurons showed a significant decrease in the frequency and amplitude of spontaneous synaptic currents compared to WT neurons. The reduced frequency in RTT neurons could reflect the presence of fewer release sites or a decreased release probability. The results of electrophysiology recordings are consistent with the decreased VGLUT1 puncta observed in Map2-positive dendrites from RTT neurons. Also consistent with these findings, the frequency of intracellular calcium transients was decreased in RTT neurons when compared to controls. Our data indicate a potential imbalance in the neuronal networks associated with RTT pathology. The observations described here provide valuable information for RTT and, potentially, ASD patients, since they suggest that presymptomatic defects may represent novel biomarkers to be exploited as diagnostic tools and that early intervention may be beneficial.

Therapies aiming at earlier stages of development may attenuate the downstream consequences of MeCP2 mutations. Restoring protein levels may be challenging, since MeCP2 levels are tightly regulated and chronically overdosing neurons with the

(C) Quantification of cells with diffused or foci me3H3K27 nuclear staining. Data shown as mean \pm SEM.

(D) RNA FISH shows that Xist RNA domains are present in the original fibroblasts before reprogramming. iPSCs show no Xist expression. Neurons derived from normal and RTT iPSCs show clear Xist clouds, indicating transcriptional silencing of the X chromosome (arrows). The scale bar represents 5 μ m.

(E) Two DNA FISH signals are evident in the nuclei of iPSC-derived NPCs and neurons, revealing the presence of two X chromosomes. The scale bar represents 10 μ m.

(F) RTT-iPSCs (1155del32) expressed WT MeCP2 but derived neurons displayed mosaicism regarding WT (arrowhead) and mutant (arrow) MeCP2 forms. The scale bar represents 50 μ m.

(G) RTT-derived fibroblasts and neurons have reduced levels of WT MeCP2 protein by Western blot. See also Figure S3.

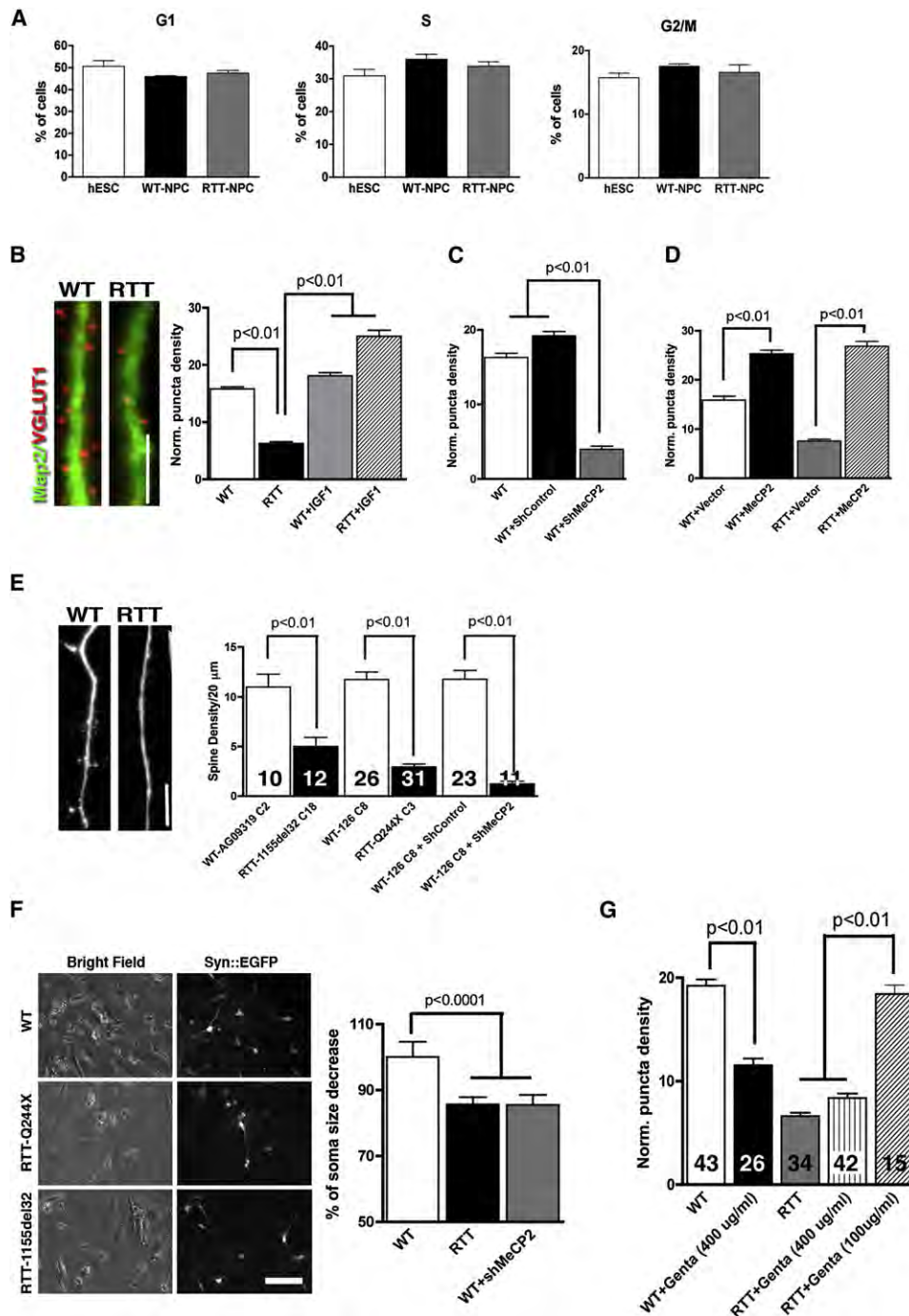


Figure 4. Alterations in RTT Neurons

(A) Proliferating RTT NPCs displayed no signal of aberrant cell cycle when compared to controls.
 (B) Representative images of neurons showing VGLUT1 puncta on Map2 neurites. Bar graphs show synaptic density in RTT and WT neurons. IGF1 treatment increased VGLUT1 puncta number in RTT-derived neurons. The scale bar represents 5 μ m.
 (C) Reduction of MeCP2 expression decreased the number of glutamatergic synapses in WT neurons.
 (D) Overexpression of MeCP2 increased the number of glutamatergic synapses.
 (E) Representative images of neurites of different genetic backgrounds. Bar graph shows the spine density from independent experiments using different RTT backgrounds and controls and after expression of shMeCP2. The scale bar represents 5 μ m.
 (F) Representative images of neuronal cell body size. Bar graph shows the percentage of soma size decrease in RTT compared to WT neurons. Neuronal morphology was visualized using the Syn::EGFP lentiviral vector. The scale bar represents 50 μ m.
 (G) A lower dose of gentamicin was able to rescue glutamatergic synapses in RTT neurons. Numbers of neurons analyzed (n) are shown within the bars in graphs (E) and (G). For all clones and mutations used refer to Figure S4 and Table S2. Data shown as mean \pm SEM.

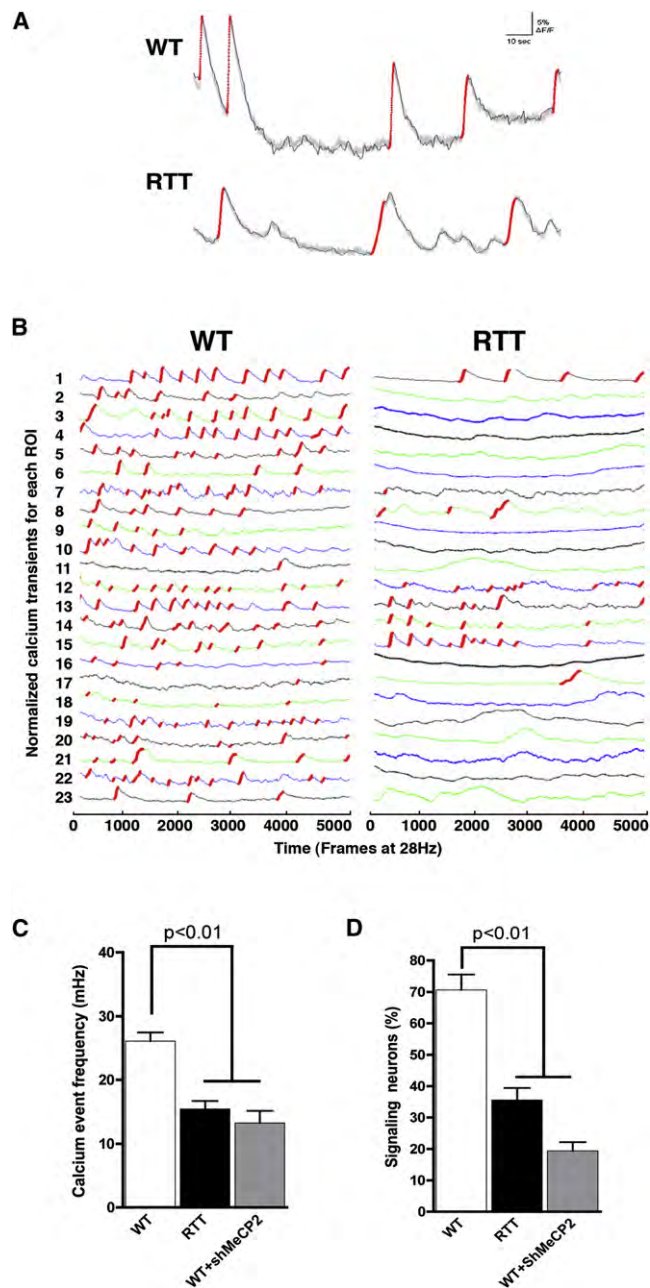


Figure 5. Altered Activity-Dependent Calcium Transients in RTT-Derived Neurons

(A) Representative examples of WT and RTT calcium signal traces. Red traces correspond to the calcium rise phase detected by the algorithm used (see [Extended Experimental Procedures](#)).

(B) Fluorescence intensity changes reflecting intracellular calcium fluctuations in RTT and WT neurons in different Regions of Interest (ROI).

(C) RTT neurons show a lower average of calcium spikes when compared to WT control neurons.

(D) The percentage of Syn::DsRed-positive neurons signaling in the RTT neuronal network is significantly reduced when compared to controls. Data shown as mean \pm SEM. See also [Figure S5](#).

WT allele can be as harmful as a loss of expression (Collins et al., 2004; Ramocki et al., 2009; Van Esch et al., 2005). Thus, we tested pharmacological treatment as a way to recover the RTT neuronal phenotype. We investigated the use of IGF1 in human neuronal cultures. Although it likely acts in a nonspecific manner, IGF1 is considered to be a candidate for pharmacological treatment of RTT and potentially other CNS disorders in a future clinical trial (Tropea et al., 2009). While IGF1 treatment increased synapse number in some clones, it stimulated glutamatergic RTT neurons above normal levels. Our data indicate that the IGF1 dose and timing parameters need to be precisely tuned in future clinical trials to avoid side effects. In a different approach, we tested a read-through drug (gentamicin) to rescue neurons derived from iPSCs carrying a nonsense MeCP2 mutation. A lower dosage of gentamicin was enough to increase full-length MeCP2 levels in RTT neurons, rescuing glutamatergic synapses. New drugs with reduced toxicity and enhanced suppression of premature stop codon mutations might be good therapeutic candidates (Nudelman et al., 2009; Welch et al., 2007).

Control of glutamatergic synapse number and the other neuronal phenotypes analyzed here may be caused by loss of MeCP2 function in the cell. Alternatively, significant experimental and genomic variability in our system could be directly responsible for the RTT differences displayed in our data. Our gain and loss of function data strongly suggest that MeCP2 is indeed the causative agent of the cellular phenotypes reported here that might be relevant to the clinical features of RTT.

Our data indicate that iPSCs not only can recapitulate some aspects of a genetic disease but also can be used to better design and anticipate results from translational medicine. This cellular model has the potential to lead to the discovery of new compounds to treat RTT and other forms of ASD. Finally, other CNS diseases may be modeled in vitro using a similar approach.

EXPERIMENTAL PROCEDURES

Cell Culture and Retrovirus Infection

Female RTT and control fibroblasts were generated from explants of dermal biopsies following informed consent under protocols approved by the University of California San Diego. The Syn::EGFP or DsRed reporter vector was obtained by cloning the Synapsin-1 promoter (a gift from Dr. G. Thiel, Hamburg, Germany) in a lentivirus backbone. The shRNA against a target sequence on the human MeCP2 gene was cloned in the LentiLox3.7 lentivirus vector. Retrovirus vectors containing the Oct4, c-Myc, Klf4 and Sox2 human cDNAs from Yamanaka's group (Takahashi et al., 2007) were obtained from Addgene. Two days after infection, fibroblasts were plated on mitotically inactivated mouse embryonic fibroblasts (Chemicon) with hESC medium. After 2 weeks, iPSC colonies were directly transferred to feeder-free conditions on matrigel-coated dishes (BD) using mTeSR1 (StemCell Technologies), and passed manually. The detailed protocols to obtain NPCs and mature neurons are described in the supplemental material. For the rescue experiments, 10 ng/mL of IGF1 (Peprotech) or Gentamicin (Invitrogen; at 100 or 400 μ g/mL) was added to neuronal cultures for 1 week. Protocols were previously approved by the University of California San Diego and Salk Institute Institutional Review Board and the Embryonic Stem Cell Research Oversight Committee.

Immunocytochemistry and Neuronal Morphology Quantification

Cells were briefly fixed in 4% paraformaldehyde and then permeabilized with 0.5% Triton X-100 in PBS. Cells were then blocked in PBS containing 0.5% Triton X-100 and 5% donkey serum for 1 hr before incubation with primary antibody overnight at 4°C. After three washes with PBS, cells were incubated

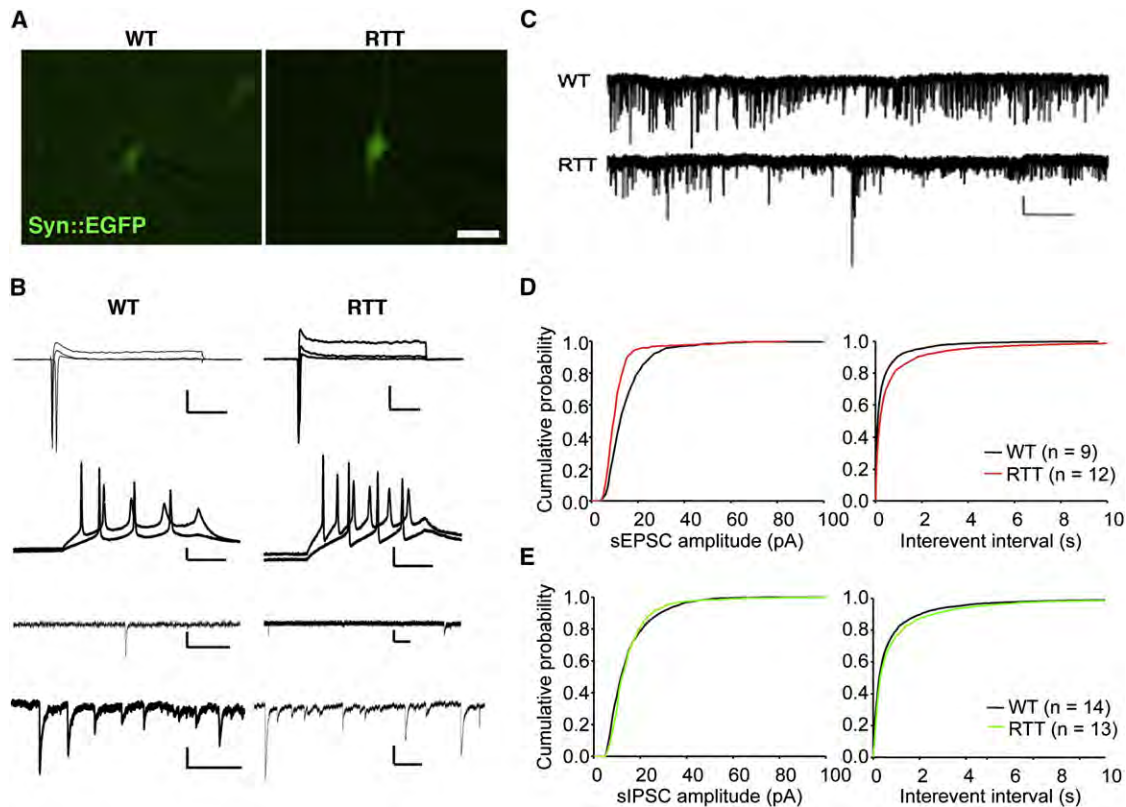


Figure 6. Decreased Frequency of Spontaneous Postsynaptic Currents in RTT Neurons

(A) Fluorescence micrographs of representative WT and RTT neurons. The scale bar represents 10 μm .

(B) Electrophysiological properties of WT and RTT neurons. From top to bottom: Transient Na^+ currents and sustained K^+ currents in response to voltage step depolarizations (command voltage varied from -20 to $+30$ mV in 5 mV increments when cells were voltage-clamped at -70 mV, Bars = 400 pA and 50 ms). Action potentials evoked by somatic current injections (cells current-clamped at around -60 mV, injected currents from 10 to 40 pA, Bars = 20 mV and 100 ms), sEPSCs (Bars = right, 20 pA, 100 ms; left: 10 pA, 500 ms), and sIPSCs (Bars = right, 20 pA, 500 ms; left: 20 pA, 400 ms).

(C) Sample 4 min recordings of spontaneous currents when the cells were voltage-clamped at -70 mV (Bars = 20 pA and 25 s).

(D) Cumulative probability plot of amplitudes (left panel, 1 pA bins; $p < 0.001$) and inter-event intervals (right panel, 20 ms bins; $p < 0.05$) of sEPSCs from groups of WT (black) and RTT (red) cells, respectively.

(E) Cumulative probability plot of amplitudes (left panel, 1 pA bins; $p < 0.05$) and inter-event intervals (right panel, 20 ms bins; $p < 0.05$) of sIPSCs from each group (WT, black; RTT, green).

with secondary antibodies (Jackson ImmunoResearch) for 1 hr at room temperature. Fluorescent signals were detected using a Zeiss inverted microscope and images were processed with Photoshop CS3 (Adobe Systems). Primary antibodies used in this study are described in the supplemental information. Cell soma size was measured in bright field using ImageJ software after identification of neurons using the Syn::EGFP. The morphologies of neuronal dendrites and spines were studied from an individual projection of z-stacks optical sections and scanned at 0.5- μm increments that correlated with the resolution valued at z-plane. Each optical section was the result of 3 scans at 500 fps followed by Kalman filtering. For synapse quantification, images were taken by a z-step of 1 μm using Biorad radiance 2100 confocal microscope. Synapse quantification was done blinded to genotype. Only VGLUT1 puncta along Map2-positive processes were counted. Statistical significances were tested using Two-way ANOVA test and Bonferroni post-test.

Cell Cycle Analysis

One million NPCs were fixed in 70% EtOH for at least 2 hr at 4°C . After PBS washing, cells were stained with 1 ml of propidium iodide (PI) solution (50 $\mu\text{g}/\text{mL}$ PI in 3.8 M sodium citrate) and treated with 20 $\mu\text{L}/\text{mL}$ of RNaseA. Cells were analyzed by fluorescence-activated cell sorting (FACS) on a Becton

Dickinson LSRI and cell cycle gating was examined using FLOWJO - Flow Cytometry Analysis Software.

RNA Extraction and RT-PCR

Total cellular RNA was extracted from $\sim 5 \times 10^6$ cells using the RNeasy Protect Mini kit (QIAGEN, Valencia, CA), according to the manufacturer's instructions, and reverse transcribed using the SuperScript III First-Strand Synthesis System RT-PCR from Invitrogen. The cDNA was amplified by PCR using Accuprime Taq DNA polymerase system (Invitrogen). Primer sequences used are described in Supplemental information.

Teratoma Formation in Nude Mice

Around $1-3 \times 10^6$ fibroblasts or iPSCs were injected subcutaneously into the dorsal flanks of nude mice (CByJ.Cg-Foxn1nu/J) anesthetized with isoflurane. Five to six weeks after injection, teratomas were dissected, fixed overnight in 10% buffered formalin phosphate and embedded in paraffin. Sections were stained with hematoxylin and eosin for further analysis. Control mice injected with RTT fibroblasts failed to form teratomas. Protocols were previously approved by the University of California San Diego Institutional Animal Care and Use Committee.

Karyotyping and DNA Fingerprinting

Standard G-banding chromosome and DNA fingerprinting analysis was performed by Cell Line Genetics (Madison, WI).

DNA and RNA FISH

Xist RNA exon 6 probes (GenBank U80460: 75081-78658 – a gift from Dr. Jeannie T. Lee, Massachusetts General Hospital, Harvard Medical School) were transcribed by using T7 RNA polymerase (Roche) with AlexaFluor 488-5-UTP. X chromosome probe and Xist slide hybridization were performed by Molecular Diagnostic Services, Inc. (San Diego, CA).

Protein Isolation and Western Blot Analysis

Cells were isolated, suspended in 1× RIPA lyses buffer (Upstate) supplemented with 1% protease inhibitor cocktail (Sigma), triturated and centrifuged at 10,000 × *g* for 10 min at 4°C. Twenty micrograms of total protein was separated on 12% SDS-polyacrylamide gel, transferred to a nitrocellulose membrane and probed with a primary antibody against MeCP2 (1:5,000; Diagenode), followed by horseradish-peroxidase-conjugated secondary antibody (1:5,000; Promega), and then visualized using ECL chemiluminescence (Amersham). As a control, membranes were stripped and re-probed for β-actin (1:10,000; Ambion) or α-tubulin (1:5,000, Ambion). For semiquantitative analysis, MeCP2 signal intensity was analyzed and corrected with respect to β-actin.

Microarray Analysis

The Affymetrix Power Tools (APT) suite of programs and Affymetrix Human Gene 1.0 ST Arrays library files and annotation were obtained from <http://www.affymetrix.com/support> and details of the analysis are available in Supplemental information.

Calcium Imaging

Neuronal networks derived from human iPSCs were previously infected with the lentiviral vector carrying the Syn:DsRed reporter construct. Cell cultures were washed twice with sterile Krebs HEPES Buffer (KHB) and incubated with 2–5 μM Fluo-4AM (Molecular Probes/Invitrogen, Carlsbad, CA) in KHB for 40 min at room temperature. Excess dye was removed by washing twice with KHB and an additional 20 min incubation was done to equilibrate intracellular dye concentration and allow de-esterification. Time-lapse image sequences (100× magnification) of 5000 frames were acquired at 28 Hz with a region of 336 × 256 pixels, using a Hamamatsu ORCA-ER digital camera (Hamamatsu Photonics K.K., Japan) with a 488 nm (FITC) filter on an Olympus IX81 inverted fluorescence confocal microscope (Olympus Optical, Japan). Images were acquired with MetaMorph 7.7 (MDS Analytical Technologies, Sunnyvale, CA). Images were subsequently processed using ImageJ (<http://rsbweb.nih.gov/ij/>) and custom written routines in Matlab 7.2 (Mathworks, Natick, MA). Detailed quantitative analysis of calcium transients is available in the Supplemental material.

Electrophysiology

Whole-cell patch clamp recordings were performed from cells co-cultured with astrocytes after 6 weeks of differentiation. The bath was constantly perfused with fresh HEPES-buffered saline (see supplemental methods for recipe). The recording micropipettes (tip resistance 3–6 MΩ) were filled with internal solution described in the Supplemental materials. Recordings were made using Axopatch 200B amplifier (Axon Instruments). Signals were filtered at 2 kHz and sampled at 5 kHz. The whole-cell capacitance was fully compensated. The series resistance was uncompensated but monitored during the experiment by the amplitude of the capacitive current in response to a 10 mV pulse. All recordings were performed at room temperature and chemicals were purchased from Sigma. Frequency and amplitude of spontaneous postsynaptic currents were measured with the Mini Analysis Program software (Synaptosoft, Leonia, NJ). Statistical comparisons of WT and RTT groups were made using the nonparametric Kolmogorov-Smirnov two-tailed test, with a significance criterion of *p* = 0.05. EPSCs were blocked by CNQX or DNQX (10–20 μM) and IPSPs were inhibited by bicuculline (20 μM).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, and two tables and can be found with this article online at doi:10.1016/j.cell.2010.10.016.

ACKNOWLEDGMENTS

The work was supported by the Emerald Foundation and by the National Institutes of Health through the NIH Director's New Innovator Award Program, 1-DP2-OD006495-01. F.H.G. is supported by California Institute for Regenerative Medicine RL1-00649-1 and RC1-00115-1, The Lookout Fund and the Mathers Foundation. C.C. is a fellow from the International Rett Syndrome Foundation. M.C.N.M. is a Christopher and Danna Reeve Foundation fellow. G.C. was supported by the Glenn Foundation. We would like to thank Monica Coenraads for critical discussion; the Greenwood Genetic Center clinical diagnostic laboratory for X-inactivation analysis; Dr. Jeannie T. Lee for the Xist probe; and M.L. Gage for editorial comments.

Received: February 9, 2010

Revised: August 4, 2010

Accepted: October 8, 2010

Published: November 11, 2010

REFERENCES

- Allen, R.C., Zoghbi, H.Y., Moseley, A.B., Rosenblatt, H.M., and Belmont, J.W. (1992). Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am. J. Hum. Genet.* 51, 1229–1239.
- Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* 23, 185–188.
- Autism and Developmental Disabilities Monitoring Network Surveillance Year 2000 Principal Investigators; Centers for Disease Control and Prevention. (2007). Prevalence of autism spectrum disorders—autism and developmental disabilities monitoring network, six sites, United States, 2000. *MMWR Surveill Summ* 56, 1–11.
- Ballas, N., Lioy, D.T., Grunseich, C., and Mandel, G. (2009). Non-cell autonomous influence of MeCP2-deficient glia on neuronal dendritic morphology. *Nat. Neurosci.* 12, 311–317.
- Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T., Qin, J., and Zoghbi, H.Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* 320, 1224–1229.
- Chao, H.T., Zoghbi, H.Y., and Rosenmund, C. (2007). MeCP2 controls excitatory synaptic strength by regulating glutamatergic synapse number. *Neuron* 56, 58–65.
- Chapleau, C.A., Calfa, G.D., Lane, M.C., Albertson, A.J., Larimore, J.L., Kudo, S., Armstrong, D.L., Percy, A.K., and Pozzo-Miller, L. (2009). Dendritic spine pathologies in hippocampal pyramidal neurons from Rett syndrome brain and after expression of Rett-associated MECP2 mutations. *Neurobiol. Dis.* 35, 219–233.
- Charman, T., Cass, H., Owen, L., Wigram, T., Slonims, V., Weeks, L., Wisbeach, A., and Reilly, S. (2002). Regression in individuals with Rett syndrome. *Brain Dev.* 24, 281–283.
- Collins, A.L., Levenson, J.M., Vilaythong, A.P., Richman, R., Armstrong, D.L., Noebels, J.L., David Sweatt, J., and Zoghbi, H.Y. (2004). Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum. Mol. Genet.* 13, 2679–2689.
- Courchesne, E., Carper, R., and Akshoomoff, N. (2003). Evidence of brain overgrowth in the first year of life in autism. *JAMA* 290, 337–344.
- De Filippis, B., Ricceri, L., and Laviola, G. (2009). Early postnatal behavioral changes in the Mecp2-308 truncation mouse model of Rett syndrome. *Genes Brain Behav.* 9, 213–223.

- Dhara, S.K., and Benvenisty, N. (2004). Gene trap as a tool for genome annotation and analysis of X chromosome inactivation in human embryonic stem cells. *Nucleic Acids Res.* *32*, 3995–4002.
- Dimos, J.T., Rodolfa, K.T., Niakan, K.K., Weisenthal, L.M., Mitsumoto, H., Chung, W., Croft, G.F., Saphier, G., Leibel, R., Golland, R., et al. (2008). Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* *321*, 1218–1221.
- Ebert, A.D., Yu, J., Rose, F.F., Jr., Mattis, V.B., Lorson, C.L., Thomson, J.A., and Svendsen, C.N. (2009). Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* *457*, 277–280.
- Giacometti, E., Luikenhuis, S., Beard, C., and Jaenisch, R. (2007). Partial rescue of MeCP2 deficiency by postnatal activation of MeCP2. *Proc. Natl. Acad. Sci. USA* *104*, 1931–1936.
- Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. *Science* *315*, 1143–1147.
- Hagberg, G., Stenbom, Y., and Engerström, I.W. (2001). Head growth in Rett syndrome. *Brain Dev.* *23 (Suppl 1)*, S227–S229.
- Hammer, S., Dorrani, N., Dragich, J., Kudo, S., and Schanen, C. (2002). The phenotypic consequences of MECP2 mutations extend beyond Rett syndrome. *Ment. Retard. Dev. Disabil. Res. Rev.* *8*, 94–98.
- Hotta, A., Cheung, A.Y., Farra, N., Vijayaragavan, K., Séguin, C.A., Draper, J.S., Pasceri, P., Maksakova, I.A., Mager, D.L., Rossant, J., et al. (2009). Isolation of human iPS cells using EOS lentiviral vectors to select for pluripotency. *Nat. Methods* *6*, 370–376.
- Kellermayer, R. (2006). Translational readthrough induction of pathogenic nonsense mutations. *Eur. J. Med. Genet.* *49*, 445–450.
- Kerr, A.M., Montague, J., and Stephenson, J.B. (1987). The hands, and the mind, pre- and post-regression, in Rett syndrome. *Brain Dev.* *9*, 487–490.
- Kishi, N., and Macklis, J.D. (2010). MeCP2 functions largely cell-autonomously, but also non-cell-autonomously, in neuronal maturation and dendritic arborization of cortical pyramidal neurons. *Exp. Neurol.* *222*, 51–58.
- Laccone, F., Huppke, P., Hanefeld, F., and Meins, M. (2001). Mutation spectrum in patients with Rett syndrome in the German population: Evidence of hot spot regions. *Hum. Mutat.* *17*, 183–190.
- Lee, G., Papapetrou, E.P., Kim, H., Chambers, S.M., Tomishima, M.J., Fasano, C.A., Ganat, Y.M., Menon, J., Shimizu, F., Viale, A., et al. (2009). Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPS cells. *Nature* *461*, 402–406.
- Lucchesi, J.C., Kelly, W.G., and Panning, B. (2005). Chromatin remodeling in dosage compensation. *Annu. Rev. Genet.* *39*, 615–651.
- Maezawa, I., Swanberg, S., Harvey, D., LaSalle, J.M., and Jin, L.W. (2009). Rett syndrome astrocytes are abnormal and spread MeCP2 deficiency through gap junctions. *J. Neurosci.* *29*, 5051–5061.
- Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* *1*, 55–70.
- Marchetto, M.C., Winner, B., and Gage, F.H. (2010). Pluripotent stem cells in neurodegenerative and neurodevelopmental diseases. *Hum. Mol. Genet.* *19 (R1)*, R71–R76.
- Mironov, S.L., Skorova, E., Hartelt, N., Mironova, L.A., Hasan, M.T., and Kügler, S. (2009). Remodelling of the respiratory network in a mouse model of Rett syndrome depends on brain-derived neurotrophic factor regulated slow calcium buffering. *J. Physiol.* *587*, 2473–2485.
- Muotri, A.R. (2008). Modeling epilepsy with pluripotent human cells. *Epilepsy Behav.* *14 (Suppl.)*, 81–85.
- Niethammer, M., Kim, E., and Sheng, M. (1996). Interaction between the C terminus of NMDA receptor subunits and multiple members of the PSD-95 family of membrane-associated guanylate kinases. *J. Neurosci.* *16*, 2157–2163.
- Nudelman, I., Rebibo-Sabbah, A., Cherniavsky, M., Belakhov, V., Hainrichson, M., Chen, F., Schacht, J., Pilch, D.S., Ben-Yosef, T., and Baasov, T. (2009). Development of novel aminoglycoside (NB54) with reduced toxicity and enhanced suppression of disease-causing premature stop mutations. *J. Med. Chem.* *52*, 2836–2845.
- Park, I.H., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., Lensch, M.W., Cowan, C., Hochedlinger, K., and Daley, G.Q. (2008). Disease-specific induced pluripotent stem cells. *Cell* *134*, 877–886.
- Picker, J.D., Yang, R., Ricceri, L., and Berger-Sweeney, J. (2006). An altered neonatal behavioral phenotype in Mecp2 mutant mice. *Neuroreport* *17*, 541–544.
- Piven, J., Arndt, S., Bailey, J., Haverkamp, S., Andreasen, N.C., and Palmer, P. (1995). An MRI study of brain size in autism. *Am. J. Psychiatry* *152*, 1145–1149.
- Piven, J., Palmer, P., Jacobi, D., Childress, D., and Arndt, S. (1997). Broader autism phenotype: evidence from a family history study of multiple-incidence autism families. *Am. J. Psychiatry* *154*, 185–190.
- Ramocki, M.B., Peters, S.U., Tavyev, Y.J., Zhang, F., Carvalho, C.M., Schaaf, C.P., Richman, R., Fang, P., Glaze, D.G., Lupski, J.R., and Zoghbi, H.Y. (2009). Autism and other neuropsychiatric symptoms are prevalent in individuals with MeCP2 duplication syndrome. *Ann. Neurol.* *66*, 771–782.
- Ronald, A., Happé, F., Bolton, P., Butcher, L.M., Price, T.S., Wheelwright, S., Baron-Cohen, S., and Plomin, R. (2006). Genetic heterogeneity between the three components of the autism spectrum: a twin study. *J. Am. Acad. Child Adolesc. Psychiatry* *45*, 691–699.
- Samaco, R.C., Hogart, A., and LaSalle, J.M. (2005). Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3. *Hum. Mol. Genet.* *14*, 483–492.
- Samaco, R.C., Nagarajan, R.P., Braunschweig, D., and LaSalle, J.M. (2004). Multiple pathways regulate MeCP2 expression in normal brain development and exhibit defects in autism-spectrum disorders. *Hum. Mol. Genet.* *13*, 629–639.
- Santos, M., Silva-Fernandes, A., Oliveira, P., Sousa, N., and Maciel, P. (2007). Evidence for abnormal early development in a mouse model of Rett syndrome. *Genes Brain Behav.* *6*, 277–286.
- Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev. Cell* *4*, 481–495.
- Soldner, F., Hockemeyer, D., Beard, C., Gao, Q., Bell, G.W., Cook, E.G., Hargus, G., Blak, A., Cooper, O., Mitalipova, M., et al. (2009). Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* *136*, 964–977.
- Spitzer, N.C., Root, C.M., and Borodinsky, L.N. (2004). Orchestrating neuronal differentiation: patterns of Ca²⁺ spikes specify transmitter choice. *Trends Neurosci.* *27*, 415–421.
- Splawski, I., Yoo, D.S., Stotz, S.C., Cherry, A., Clapham, D.E., and Keating, M.T. (2006). CACNA1H mutations in autism spectrum disorders. *J. Biol. Chem.* *281*, 22085–22091.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* *131*, 861–872.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* *126*, 663–676.
- Takamori, S., Rhee, J.S., Rosenmund, C., and Jahn, R. (2000). Identification of a vesicular glutamate transporter that defines a glutamatergic phenotype in neurons. *Nature* *407*, 189–194.
- Tao, J., Hu, K., Chang, Q., Wu, H., Sherman, N.E., Martinowich, K., Klose, R.J., Schanen, C., Jaenisch, R., Wang, W., and Sun, Y.E. (2009). Phosphorylation of MeCP2 at Serine 80 regulates its chromatin association and neurological function. *Proc. Natl. Acad. Sci. USA* *106*, 4882–4887.
- Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* *282*, 1145–1147.

- Traynor, J., Agarwal, P., Lazzeroni, L., and Francke, U. (2002). Gene expression patterns vary in clonal cell cultures from Rett syndrome females with eight different MECP2 mutations. *BMC Med. Genet.* 3, 12.
- Tropea, D., Giacometti, E., Wilson, N.R., Beard, C., McCurry, C., Fu, D.D., Flannery, R., Jaenisch, R., and Sur, M. (2009). Partial reversal of Rett Syndrome-like symptoms in MeCP2 mutant mice. *Proc. Natl. Acad. Sci. USA* 106, 2029–2034.
- Van Esch, H., Bauters, M., Ignatius, J., Jansen, M., Raynaud, M., Hollanders, K., Lugtenberg, D., Bienvenu, T., Jensen, L.R., Gecz, J., et al. (2005). Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am. J. Hum. Genet.* 77, 442–453.
- Welch, E.M., Barton, E.R., Zhuo, J., Tomizawa, Y., Friesen, W.J., Trifillis, P., Paushkin, S., Patel, M., Trotta, C.R., Hwang, S., et al. (2007). PTC124 targets genetic disorders caused by nonsense mutations. *Nature* 447, 87–91.
- Yasui, D.H., Peddada, S., Bieda, M.C., Vallero, R.O., Hogart, A., Nagarajan, R.P., Thatcher, K.N., Farnham, P.J., and Lasalle, J.M. (2007). Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. *Proc. Natl. Acad. Sci. USA* 104, 19416–19421.
- Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318, 1917–1920.
- Zappella, M., Meloni, I., Longo, I., Canitano, R., Hayek, G., Rosaia, L., Mari, F., and Renieri, A. (2003). Study of MECP2 gene in Rett syndrome variants and autistic girls. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 119, 102–107.
- Zhou, Z., Hong, E.J., Cohen, S., Zhao, W.N., Ho, H.Y., Schmidt, L., Chen, W.G., Lin, Y., Savner, E., Griffith, E.C., et al. (2006). Brain-specific phosphorylation of MeCP2 regulates activity-dependent Bdnf transcription, dendritic growth, and spine maturation. *Neuron* 52, 255–269.
- Zoghbi, H.Y. (2003). Postnatal neurodevelopmental disorders: meeting at the synapse? *Science* 302, 826–830.

EXTENDED EXPERIMENTAL PROCEDURES

Cell Culture and Retrovirus Infection

Female RTT (1155del32, GM11272; Q244X, GM16548; T158M, GM17880 and R306C, GM11270; from Coriell Institute) and WT fibroblasts (AG09319; from Coriell Institute, CRL2529; from ATCC). WT-126, ADRC40 and WT-33 were cultured in Minimum Essential Medium (Invitrogen) supplemented with 10% fetal bovine serum (HyClone Laboratories).

The hESC Cyth25 (CyThera Inc., San Diego) and HUES6 (Harvard) cell lines were cultured as previously described (Muotri et al., 2005). Recombinant viruses were produced by transient transfection in 293T cells, as previously described (Muotri et al., 2005).

To obtain NPCs, EBs were formed by mechanical dissociation of cell clusters and plating onto low-adherence dishes in hESC medium without FGF2 for 5–7 days. After that, EBs were plated onto poly-ornithine/laminin (Sigma)-coated dishes in DMEM/F12 (Invitrogen) plus N2. Rosettes were visible to collect after 7 days. Rosettes were then dissociated with accutase (Chemicon) and plated again onto coated dishes with NPC media (DMEM/F12; 0.5X N2; 0.5X B27 and FGF2). Homogeneous populations of NPCs were achieved after 1–2 passages with accutase in the same condition. To obtain mature neurons, floating EBs were treated with 1 μ M of retinoic acid for 3 more weeks (total time of differentiation 4 weeks). Mature EBs were then dissociated with Papain and DNase (Worthington) for 1 hr at 37°C and plated in poly-ornithine/laminin-coated dishes in NPC media without FGF2.

Primary Antibodies Used for Immunofluorescence in This Study

Primary antibodies used in this study were TRA-1-60, TRA-1-81 (1:100, Chemicon); Nanog and Lin28 (1:500, R&D Systems); human Nestin (1:100, Chemicon); Tuj-1 (1:500, Covance); Map2 (1:100; Sigma); MeCP2 (1:1000, Sigma); VGLUT1 (1:200, Synaptic Systems); Psd95 (1:500, Synaptic Systems), GFP (1:200, Molecular Probes-Invitrogen); Sox1 (1:250, BD Biosciences), Musashi1 (1:200, Abcam) and me3H3K27 (1:500, Millipore).

Oligonucleotide Sequences Used in This Study

ShRNA against the human MeCP2 gene (5'-GGAGTCTTCTATCCGATCTGT-3') was cloned in the LentiLox3.7 lentivirus vector, which forms a part of hairpin loop 5'-GGAGTCTTCTATCCGATCTGTTC AAGAGACAGATCGGATAGAAGACCTCC-3'.

The primer sequences were: hOct4-f: gggaggggaggagctagg and hOct4-R: tccaaccagttgcccacaac; hSox2-F: tgggaggggtgcaaaa-gagg and hSox2-R: gagtgtggatgggattgtgtg; hNanog-F: cctatgcctgtgattgtgtg and hNanog-R: ctgggacctgtctctcttt; hMSX1-F: 5' aggaccccgtggatgcagag and hMSX1-R: 5' ggccatcttcagctctccag; hGAPDH-Fw: 5' accacagtcctatgcatcac 3', hGAPDH-Rv: 5' tcca ccaccctgtgctgta 3'. PCR products were separated by electrophoresis on a 2% agarose gel, stained with ethidium bromide and visualized by UV illumination.

Microarray Analysis

Gene-level signal estimates were derived from the CEL files by RMA-sketch normalization as a method in the apt-probeset-summarize program. Hierarchical clustering of the full dataset by probeset values was performed by complete linkage using Euclidean distance as a similarity metric in Matlab.

Quantification of Calcium Transients

The recipe for Krebs HEPES Buffer (KHB) used for calcium imaging was: 10 mM HEPES, 4.2 mM NaHCO₃, 10 mM dextrose, 1.18 mM MgSO₄·2H₂O, 1.18 mM KH₂PO₄, 4.69 mM KCl, 118 mM NaCl, 1.29 mM CaCl₂; pH 7.3). Neurons were selected after the confirmation that calcium transients were blocked with 1 μ M of tetrodotoxin (TTX) or the glutamate receptor antagonists CNQX/APV (6-cyano-7-nitroquinoxaline-2,3-dione at 10 μ M / (2R)-amino-5-phosphonovaleric acid; (2R)-amino-5-phosphonopentanoate at 20 μ M, respectively) treatments. Calcium transients increased after 30 μ M Gabazine treatment. For quantification of calcium transients, ImageJ, an NIH-funded open source, JAVA-based morphometric application, was used to allow manual selection of individual neurons on the Syn::DsRed image that correspond to each calcium movie using circular regions of interest (ROI) of 4 pixels (~5 μ m) in diameter. Each cell was considered as an individual ROI and the average fluorescence intensity was calculated for each ROI through the entire acquired image sequence. Quantitative signal analysis and processing were done with custom-written Matlab routines. Individual temporal fluorescence intensity signals indicative of intracellular calcium fluctuations were filtered using power spectrum calculated from Fourier transforms to reduce noise. Amplitude of signals was presented as relative fluorescence changes ($\Delta F/F$) after background subtraction. A first-derivative filter was used to identify regions of increase in calcium signal and a calcium event was identified by a positive derivative value of 2 SD or more above background with a rise phase that persisted a minimum of 5 consecutive frames (~70ms).

Electrophysiology

Recipe for HEPES-buffered saline: 115 mM NaCl, 2 mM KCl, 10 mM HEPES, 3 mM CaCl₂, 10 mM glucose and 1.5 mM MgCl₂ (pH 7.4).

Recipe for solution inside the recording micropipettes (tip resistance 3–6 M Ω): 140 mM K-gluconate, 5 mM KCl, 2 mM MgCl₂, 10 mM HEPES and 0.2 mM EGTA, 2.5 mM Na-ATP, 0.5 mM Na-GTP, 10 mM Na₂-phosphocreatine (pH 7.4).

SUPPLEMENTAL REFERENCES

Mnatzakanian, G.N., Lohi, H., Munteanu, I., Alfred, S.E., Yamada, T., MacLeod, P.J., Jones, J.R., Scherer, S.W., Schanen, N.C., Friez, M.J., et al. (2004). A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. *Nat. Genet.* 36, 339–341.

Muotri, A.R., Nakashima, K., Toni, N., Sandler, V.M., and Gage, F.H. (2005). Development of functional human embryonic stem cell-derived neurons in mouse brain. *Proc. Natl. Acad. Sci. USA* 102, 18644–18648.

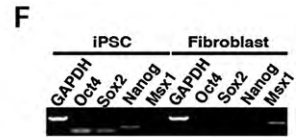
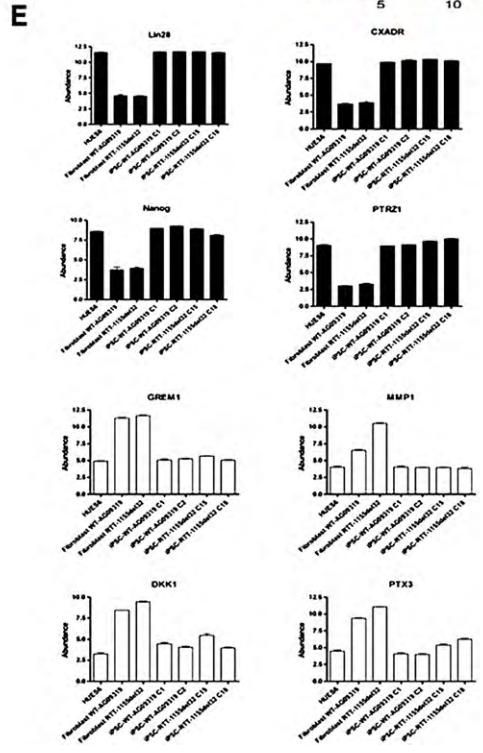
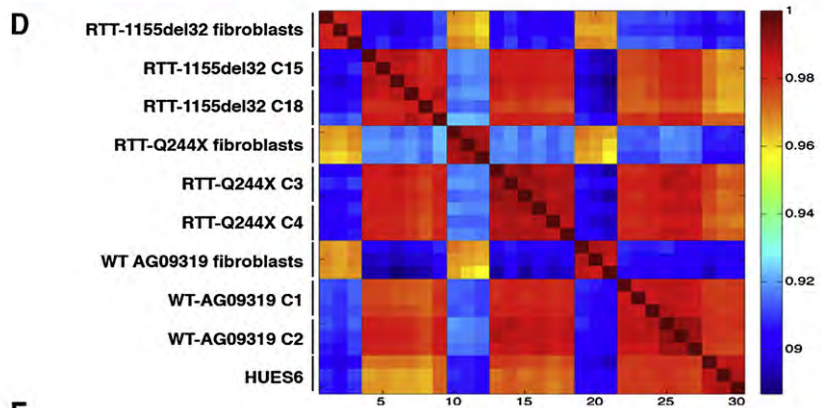
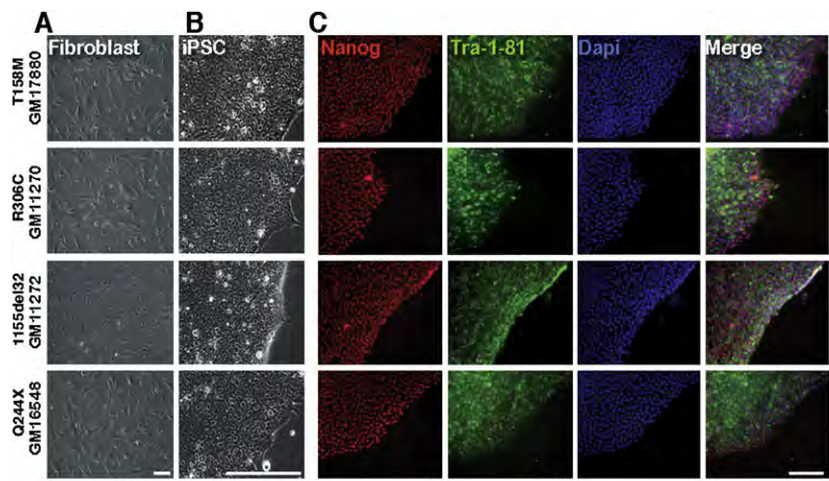


Figure S1. Generation of iPSCs Derived from RTT Patients' Fibroblasts Carrying Distinct Mutations in the MeCP2 Gene, Related to Figure 1

(A) Morphology of fibroblasts before retroviral infection. (B) Aspect of iPSCs colonies growing in the absence of feeder layer. Colonies are compact and have well-defined borders. Cells display high nucleus-to-cytoplasm ratio and are morphologically similar to hESCs. (C) Representative immunofluorescence analysis of RTT-iPSC clones. Expression of pluripotent markers such as Nanog and Tra-1-81 is observed. The scale bar represents 100 μm . (D) Hierarchical clustering and correlation coefficients of microarray profiles of triplicate WT Fibroblasts, RTT Fibroblasts, WT-iPSC clone 1, WT-iPSC clone 2, RTT-iPSC clones 15 and 18 (1155del32), RTT-iPSC clones 1 and 2 (Q244X) and the hESC line HUES6. Color bar indicates the level of correlation (from 0 to 1), with color bar reporting log₂ normalized expression values (green/red indicates high/low relative expression). (E), Reprogrammed iPSCs showed expressions similar to hESC-enriched genes (Lin28, CXADR, Nanog and PTRZ1; black bars) and showed distinct differences from fibroblast-enriched genes (GREM1, MMP1, DKK1 and PTX3; white bars). (F) RT-PCR from reprogrammed iPSCs showed endogenous expressions of hESC-enriched genes (Oct4, Sox2 and Nanog) but not from a fibroblast-enriched gene (Msx1).

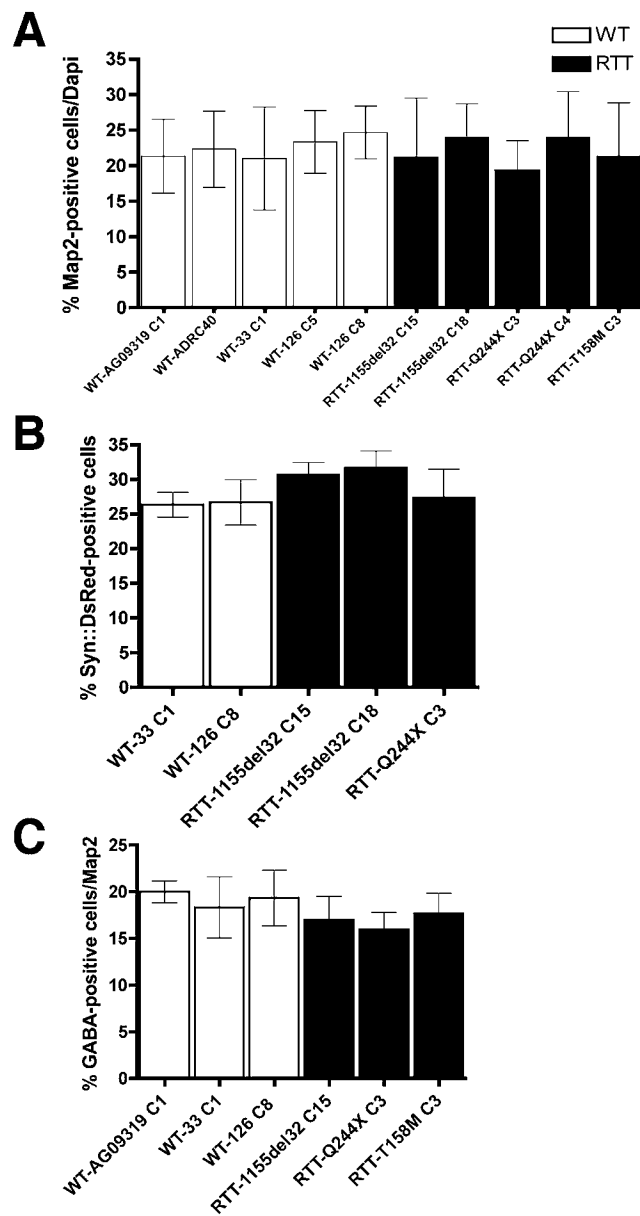
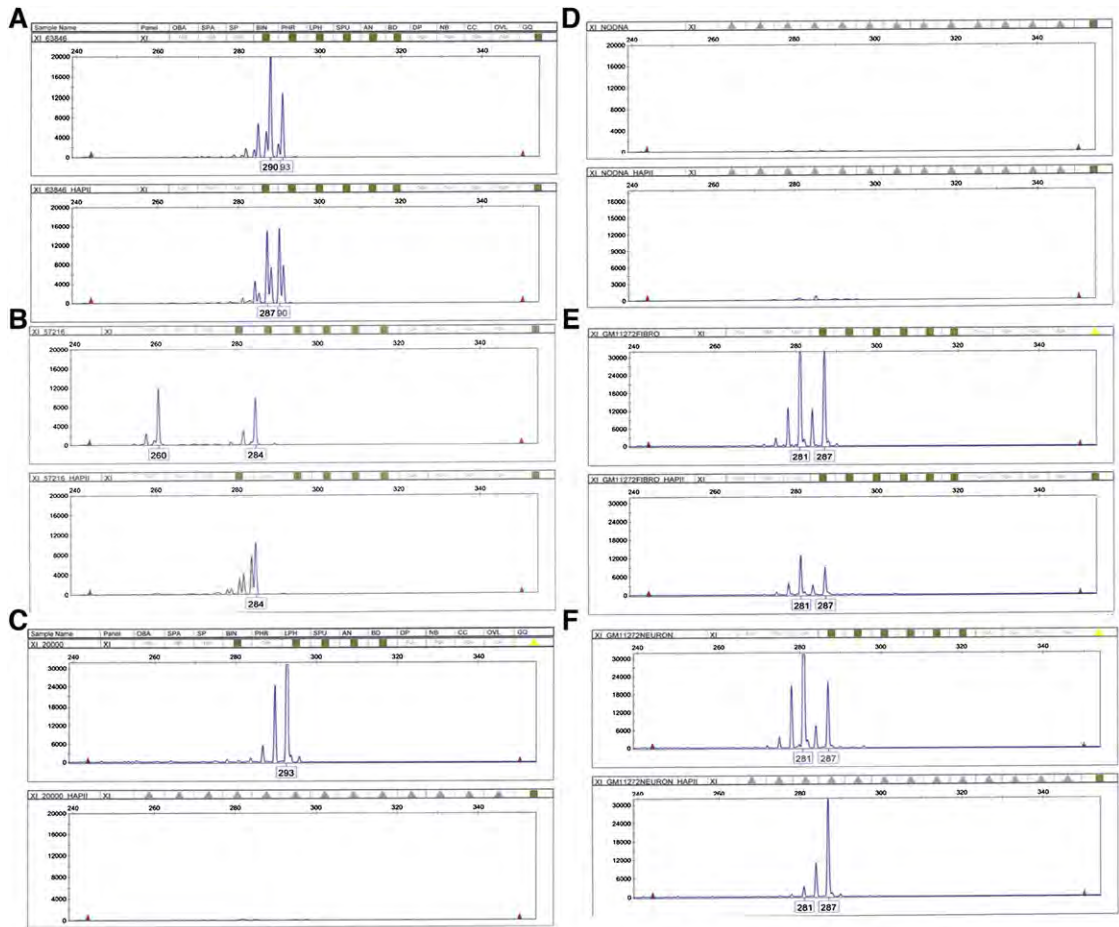


Figure S2. Neuronal Differentiation from Individual WT and RTT-iPSC Clones, Related to Figure 2

Clones from WT and RTT-iPSCs were differentiated into neurons for approximately 1 month. (A) Neurons were stained with the Map2 neuronal marker. (B) Neurons were infected with a lentiviral vector expressing the DsRed reporter under the control of the Synapsin promoter region. (C) Inhibitory neurons were revealed in the cultures after staining with anti-GABA antibody. Each bar represents 3 independent experiments for each individual clone. Data shown as mean \pm s.d.m.



| Sample Name | Peak Area1 | Peak Area2 | Corr. Factor | Corr. Peak2 (Hpa) | Corr. Total | Peak1 % | Peak2 % | XI:Xa |
|------------------------|------------|------------|--------------|-------------------|-------------|---------|---------|------------------------|
| XI_63846 | 156508 | 87390 | 1.7909143 | | | | | |
| XI_63846_HapII | 100309 | 103847 | | 185981.0765 | 286290.0765 | 35 | 65 | 65:35 (Random Control) |
| XI_57216 | 87250 | 73588 | 1.1856553 | | | | | |
| XI_57216_HapII | 79106 | | | 0 | 79106 | 100 | 0 | 100 (Highly Control) |
| XI_GM11272Fibro | 273822 | 245617 | 1.1148333 | | | | | |
| XI_GM11272Fibro_HapII | 85193 | 61995 | | 69114.08775 | 154307.0877 | 55.2 | 44.8 | 55:45 (Random) |
| XI_GM11272Neuron | 330180 | 150314 | 2.1966018 | | | | | |
| XI_GM11272Neuron_HapII | 19851 | 229017 | | 503059.1499 | 522910.1499 | 3.8 | 96.2 | 96:4 (Highly) |
| XI_20000 | 325981 | | | | | | | |
| XI_20000_HapII | | | | | | | | Male Control |
| XI_NoDNA | | | | | | | | |
| XI_NoDNA_HpaII | | | | | | | | No DNA |

Figure S3. Androgen Receptor Analysis, Related to Figure 3

Example of X-inactivation analysis using the X-linked androgen receptor locus for the RTT-1155del32 C15 genomic DNA. After the PCR, 2 different-sized amplicons were detected (different peaks) and digested with a methylation-sensitive restriction enzyme (HpaII). The PCR using undigested DNA shows if two distinct alleles are present and also allows a correction factor due to the advantage on the amplification of the smaller allele. When the template DNA is digested, amplification occurs if the restriction sites are methylated. If the site is unmethylated, digestion will occur between the flanking oligonucleotides and amplification will not be possible. The peak areas after HpaII restriction digestion of genomic DNA are used to distinguish each parental X chromosome. (A) When random inactivation is present, the maternal and paternal alleles are represented at similar proportions. (B) In contrast, in a condition where nonrandom inactivation is present, the more commonly inactive allele will be preferentially amplified and this will be detected by a stronger peak. (C) A male control is displayed showing a single peak before HpaII digestion. (D) A PCR was run without DNA template as a control. (E) Fibroblasts carrying the 1155del32 MeCP2 mutant (GM11272) displayed random X-inactivation. (F) RTT-1155del32-derived neurons showed highly skewed X-inactivation.

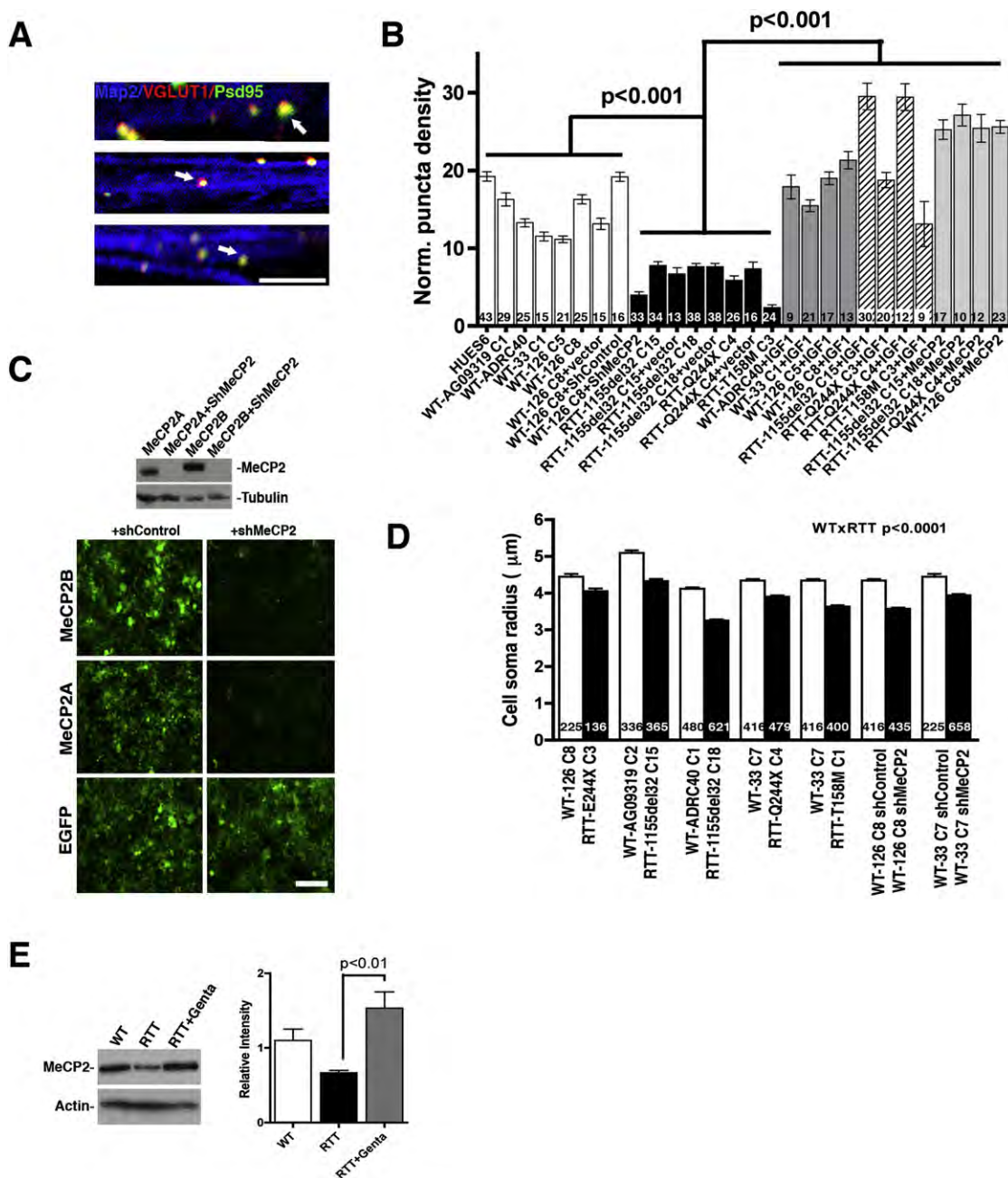


Figure S4. Phenotypic Analysis of iPSC-Derived Neurons from Several Clones, Related to Figure 4

(A) Representative images showing co-localization between VGLUT1 and Psd95 (arrows). The scale bar represents 5 μm . (B) Experimental and clonal variation of VGLUT1 puncta quantification in different individuals. (C) Efficient expression and knockdown of both MeCP2 isoforms by a specific shRNA against MeCP2. The scale bar represents 50 μm . Two alternatively spliced MeCP2 transcripts have been characterized, isoforms A and B, which differ only in their most 5' regions. The MeCP2 isoform B is more prevalent in the brain and during neuronal differentiation (Mnatzakanian et al., 2004). (D) Graph shows cell soma radius for several RTT and WT clones. (E) WT MeCP2 protein levels detected in control and RTT neurons (Q244X). Gentamicin treatment in RTT neurons increased protein levels after 2 weeks. Numbers of neurons analyzed (n) are shown within the bars in graphs (B) and (D). Data shown as mean \pm SEM.

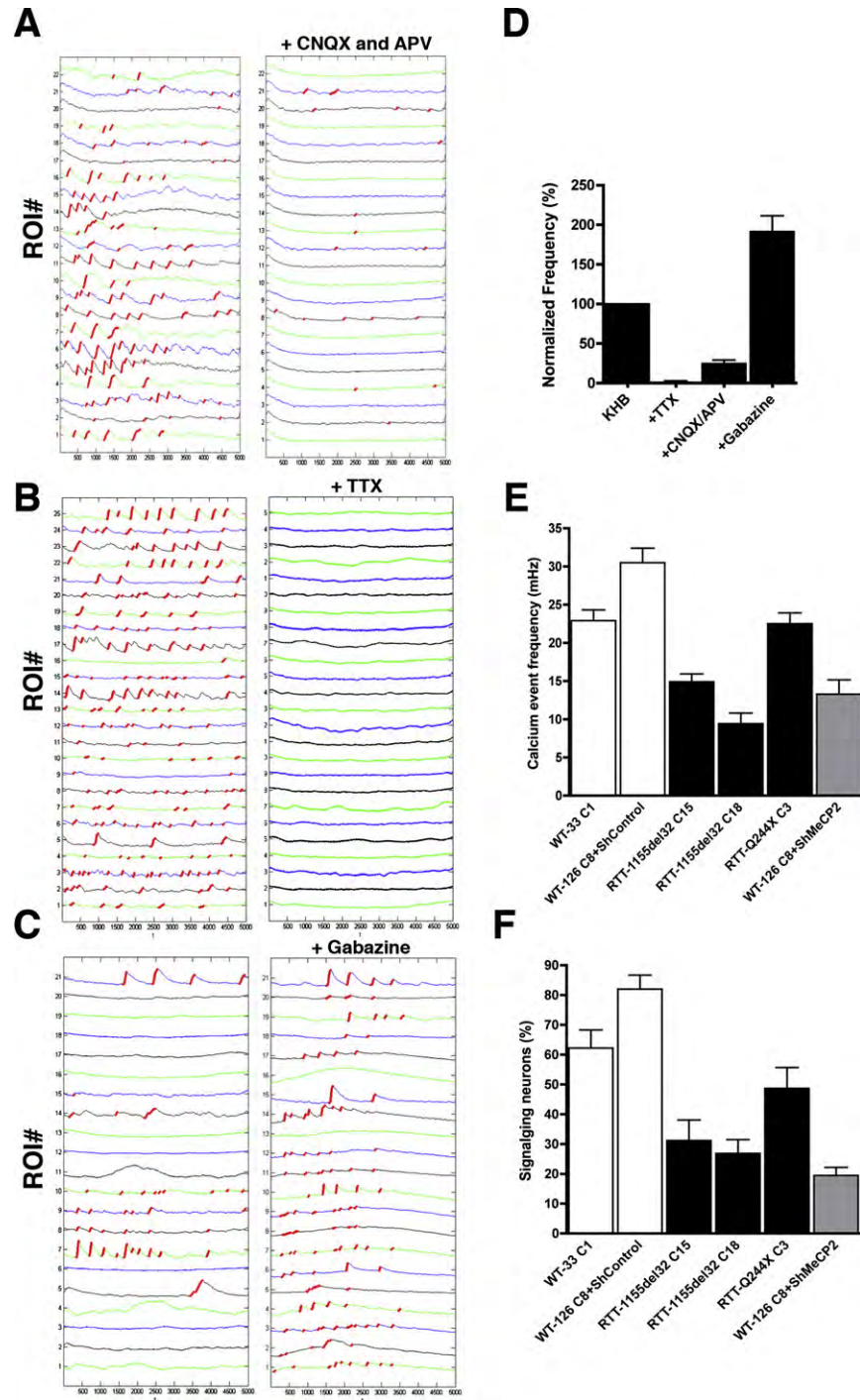


Figure S5. Calcium Transient Analysis in iPSC-Derived Neurons, Related to Figure 5

Neurons were selected after the confirmation that calcium transients were blocked with 1 μ M of TTX or the glutamate receptor antagonists CNQX/APV treatments. (A) Blocking glutamatergic signaling in the neuronal network using CNQX and APV resulted in significant reduction in intracellular calcium transients. (B) Blocking voltage-gated sodium channels using TTX prohibited the generation of action potentials and resulted in complete elimination of neuronal intracellular calcium transients. (C) Gabazine increased the number of calcium transients in the iPSC-derived neuronal networks. Red traces correspond to the calcium rise phase detected by the algorithm used. (D) Bar graph shows the normalized frequency of neurons with calcium transients after drug treatments. (E) Bar graph shows the event frequency decrease in RTT and shMeCP2-treated WT neurons compared to WT control neurons. (F) Bar graph shows the percentage of signaling neurons in RTT and shMeCP2-treated WT neurons compared to WT control neurons.

Bruce D. Gelb, MD

The Gogel Family Professor and Director, Child Health and Development Institute
Professor of Pediatrics and Genetics & Genomic Sciences
Mount Sinai School of Medicine
New York, New York

Biography

Bruce D. Gelb, MD is the Director of the Child Health and Development Institute at the Mount Sinai School of Medicine. He is the Gogel Family Professor of Child Health and Development as well as Professor of Pediatrics and Genetics & Genomic Sciences. Dr. Gelb completed a pediatric residency and pediatric cardiology fellowship at Babies Hospital of Columbia-Presbyterian Medical Center and Texas Children's Hospital at the Baylor College of Medicine, respectively. He joined the faculty at Mount Sinai in 1991 after fellowship and has remained there since. He developed an extensive program in genomics/gene discovery for congenital heart disease. Dr. Gelb has received the E. Mead Johnson Award from the Society for Pediatric Research and the Norman J. Siegel New Member Outstanding Science Award from the American Pediatric Society. He has been elected to the American Society of Clinical Investigation and the Institute of Medicine. In addition to his research, he directs the Cardiovascular Genetics Program at Mount Sinai.

Notes

The RASopathies: Studying inherited human disease via iPSC

Bruce D. Gelb, MD

Child Health and Development Institute, Departments of Pediatrics and Genetics & Genomic Sciences, Mount Sinai School of Medicine, New York, NY, USA

Noonan syndrome (NS) and related disorders are relatively common autosomal dominant traits (aggregate prevalence of 1:1000 live births) with pleiomorphic features. The seminal features are short stature, particular facial dysmorphism, cardiac disease, and neurocognitive deficits. Cardiovascular involvement, observed in up to 90% of affected individuals, includes a range of congenital heart defects (CHD), with pulmonary valve stenosis and atrial septal defects being most common, as well as early-onset hypertrophic cardiomyopathy (HCM). There is also hematologic involvement including leukemias, principally juvenile myelomonocytic leukemia (JMML), platelet numerical and functional deficits and coagulopathies.

Starting with the discovery that gain-of-function mutations in *PTPN11*, which encodes the non-membranous protein tyrosine phosphatase SHP-2, mutations in numerous genes encoding proteins critical for RAS/mitogen-activated protein kinase (MAPK) signaling have been discovered for NS as well as LEOPARD, cardio-facio-cutaneous, and Costello syndromes. These disorders now constitute a disease family termed the RASopathies. Careful characterization of the RASopathy phenotypes has permitted the establishment of correlations with genotypes, including those relevant for CHD and HCM. Modeling of several mutations in animals (*Drosophila*, *C. elegans*, and mice) causing RASopathy phenotypes has provided insights into the pathogenesis of disease. In particular, relevant signaling perturbations related to HCM associated with *PTPN11* and *RAF1* mutations have been established and the feasibility of therapy that prevents or reverses HCM has been documented.

Our research group, in collaboration with that of Dr. Ihor Lemischka, initiated a research program to use induced pluripotent stem cell (iPSC) lines developed from individuals with RASopathies due to mutations in several of the relevant genes in order to model seminal aspects of these disorders. For this presentation, Dr. Gelb will review the progress in this project to date with particular emphasis on the modeling of hypertrophic cardiomyopathy using iPSC-derived cardiomyocytes, which resulted in the first published human iPSC-based model of a cardiovascular disease. In addition, preliminary data from ongoing studies of perturbed myeloid development obtained using iPSCs from individuals with Noonan syndrome who developed JMML will be presented.

Bibliography

Tartaglia M, Niemeyer CM, Fragale A, Song X, Buechner J, Hahlen K, Hasle H, Jung A, Licht JD, **Gelb BD**. Somatic *PTPN11* mutations in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat Genet* 2003, **34**:148-150.

Carvajal-Vergara X, Sevilla A, D'Souza SL, Ang Y-S, Schaniel C, Lee D-F, Yang L, Kaplan AD, Adler ED, Ge Y, Cohen N, Edelmann LJ, Chang B, Waghray A, Su J, Pardo S, Lichtenbelt KD, Tartaglia M, **Gelb BD**, Lemischka I. Patient-specific induced pluripotent stem cell derived models of LEOPARD syndrome. *Nature* 2010, **465**:808-812.

Josowitz R, Carvajal-Vergara X, Lemischka IR, **Gelb BD**. Induced pluripotent stem cell-derived cardiomyocytes as models for genetic cardiovascular disorders. *Curr Contents Cardiol* 2011, **26**: 223-229.

Tartaglia M, **Gelb BD**. Noonan syndrome and clinically related disorders. *Best Pract Res Clin Endocrinol Metab* 2011, **25**:161-179.

Gelb BD, Tartaglia M. Ras signaling pathway mutations and hypertrophic cardiomyopathy: Getting into and out of the thick of it. *J Clin Invest* 2011, **121**:844-847.

Notes

LETTERS

Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome

Xonia Carvajal-Vergara^{1,2}, Ana Sevilla^{1*}, Sunita L. D'Souza^{1*}, Yen-Sin Ang¹, Christoph Schaniel¹, Dung-Fang Lee¹, Lei Yang¹, Aaron D. Kaplan³, Eric D. Adler³, Roye Rozov¹, YongChao Ge⁴, Ninette Cohen⁵, Lisa J. Edelmann⁵, Betty Chang¹, Avinash Waghray¹, Jie Su¹, Sherly Pardo^{5,6}, Klaske D. Lichtenbelt⁷, Marco Tartaglia⁸, Bruce D. Gelb^{5,6,9*} & Ihor R. Lemischka^{1*}

The generation of reprogrammed induced pluripotent stem cells (iPSCs) from patients with defined genetic disorders holds the promise of increased understanding of the aetiologies of complex diseases and may also facilitate the development of novel therapeutic interventions. We have generated iPSCs from patients with LEOPARD syndrome (an acronym formed from its main features; that is, lentigines, electrocardiographic abnormalities, ocular hypertelorism, pulmonary valve stenosis, abnormal genitalia, retardation of growth and deafness), an autosomal-dominant developmental disorder belonging to a relatively prevalent class of inherited RAS-mitogen-activated protein kinase signalling diseases, which also includes Noonan syndrome, with pleomorphic effects on several tissues and organ systems^{1,2}. The patient-derived cells have a mutation in the *PTPN11* gene, which encodes the SHP2 phosphatase. The iPSCs have been extensively characterized and produce multiple differentiated cell lineages. A major disease phenotype in patients with LEOPARD syndrome is hypertrophic cardiomyopathy. We show that *in vitro*-derived cardiomyocytes from LEOPARD syndrome iPSCs are larger, have a higher degree of sarcomeric organization and preferential localization of NFATC4 in the nucleus when compared with cardiomyocytes derived from human embryonic stem cells or wild-type iPSCs derived from a healthy brother of one of the LEOPARD syndrome patients. These features correlate with a potential hypertrophic state. We also provide molecular insights into signalling pathways that may promote the disease phenotype.

Approximately 90% of LEOPARD syndrome cases, and 45% of Noonan syndrome cases, are caused by missense mutations in the *PTPN11* gene which encodes the protein tyrosine phosphatase SHP2. *PTPN11* is ubiquitously expressed, essential for normal development, and somatic mutations in this gene contribute to leukaemogenesis in children^{3,4}. For LEOPARD syndrome, two mutations, T468M and Y279C, are most recurrent⁵. Hypertrophic cardiomyopathy is the most common life-threatening cardiac anomaly in LEOPARD syndrome². Animal models of LEOPARD syndrome have been generated in *Drosophila* and zebrafish^{6,7}, but the molecular pathogenesis of LEOPARD syndrome remains obscure.

Ectopic expression of four transcription factors (*OCT4* (also called *POU5F1*), *SOX2*, *KLF4* and *MYC*) in adult human dermal fibroblasts can generate pluripotent iPSCs⁸⁻¹⁰. Together with defined *in vitro* differentiation protocols, this suggests the possibility of developing

reliable disease models¹¹⁻¹⁴. We have established iPSC lines from two LEOPARD syndrome patients, a 25-year-old female (L1) and a 34-year-old male (L2). A heterozygous T468M substitution mutation in *PTPN11* is present in both.

Fibroblasts were transduced with *OCT4*-, *SOX2*-, *KLF4*- and *MYC*-encoding VSV-pseudotyped Moloney-based retroviral vectors. Compact embryonic stem (ES)-cell-like colonies emerged after 2 weeks and TRA1-81-positive colonies were clonally expanded to create stable LEOPARD syndrome iPSC lines (Supplementary Fig. 1)⁸. Three iPSC lines per patient were used for preliminary characterization: L1-iPS1, L1-iPS6, L1-iPS13, L2-iPS6, L2-iPS16 and L2-iPS18.

To verify that the iPSCs originated from patient-derived fibroblasts, we performed DNA fingerprinting analysis (Supplementary Fig. 2a). All iPSCs had normal karyotypes of 46,XX (L1) and 46,XY (L2) (Supplementary Fig. 2b and data not shown). In addition, they carried the expected T468M mutation (Supplementary Fig. 3a). Restriction fragment length polymorphism analysis of a polymerase chain reaction with reverse transcription (RT-PCR) amplicon containing the mutation with BsmFI showed biallelic expression of *PTPN11* (Supplementary Fig. 3b). PCR and Southern blot analysis indicated the presence of all four transgene proviruses in the LEOPARD syndrome iPSCs (Supplementary Fig. 4), and quantitative RT-PCR (qRT-PCR) results confirmed efficient transgene silencing (Supplementary Fig. 5).

To characterize further the LEOPARD syndrome iPSC clones, expression of several human ES-cell markers in two LEOPARD syndrome iPSC lines from each patient (L1-iPS1, L1-iPS13, L2-iPS6 and L2-iPS16) was analysed and compared to the HES2 human ES cell line and a wild-type iPSC line, BJ-iPSB5, derived in our laboratory from a normal human fibroblast line (BJ). The BJ-iPSB5 cell line was also karyotypically normal (46,XY), and contained all four transgene proviruses, which were silenced (Supplementary Figs 2b, 4b and 5b). All LEOPARD syndrome and control iPSC lines showed high alkaline phosphatase activity and expressed pluripotency markers, including surface antigens TRA1-81, TRA1-60 and SSEA4, as well as the nuclear transcription factors *OCT4* and *NANOG* (Supplementary Fig. 6). Activation of a series of endogenous stemness genes (*OCT4*, *NANOG*, *SOX2*, *GDF3*, *DPPA4*, *REX1* and *TERT*) in iPSCs was confirmed by qRT-PCR (Fig. 1a and Supplementary Fig. 7a). Extensive demethylation of CpG dinucleotides in the *OCT4* and *NANOG* promoters compared to their parental fibroblasts was confirmed by bisulphite sequencing (Fig. 1b).

¹Department of Gene and Cell Medicine, Department of Developmental and Regenerative Biology, Black Family Stem Cell Institute, Mount Sinai School of Medicine, New York, New York 10029, USA. ²Department of Regenerative Cardiology, Centro Nacional de Investigaciones Cardiovasculares, 28029 Madrid, Spain. ³Department of Medicine, Cardiovascular Institute, Mount Sinai School of Medicine, New York, New York 10029, USA. ⁴Department of Neurology, Mount Sinai School of Medicine, New York, New York 10029, USA.

⁵Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York 10029, USA. ⁶Child Health and Development Institute, Mount Sinai School of Medicine, New York, New York 10029, USA. ⁷Department of Medical Genetics, University Medical Center Utrecht, 3584 EA Utrecht, the Netherlands. ⁸Dipartimento di Ematologia, Oncologia e Medicina Molecolare, Istituto Superiore di Sanità, 00161 Rome, Italy. ⁹Department of Pediatrics, Mount Sinai School of Medicine, New York, New York 10029, USA.

*These authors contributed equally to this work.

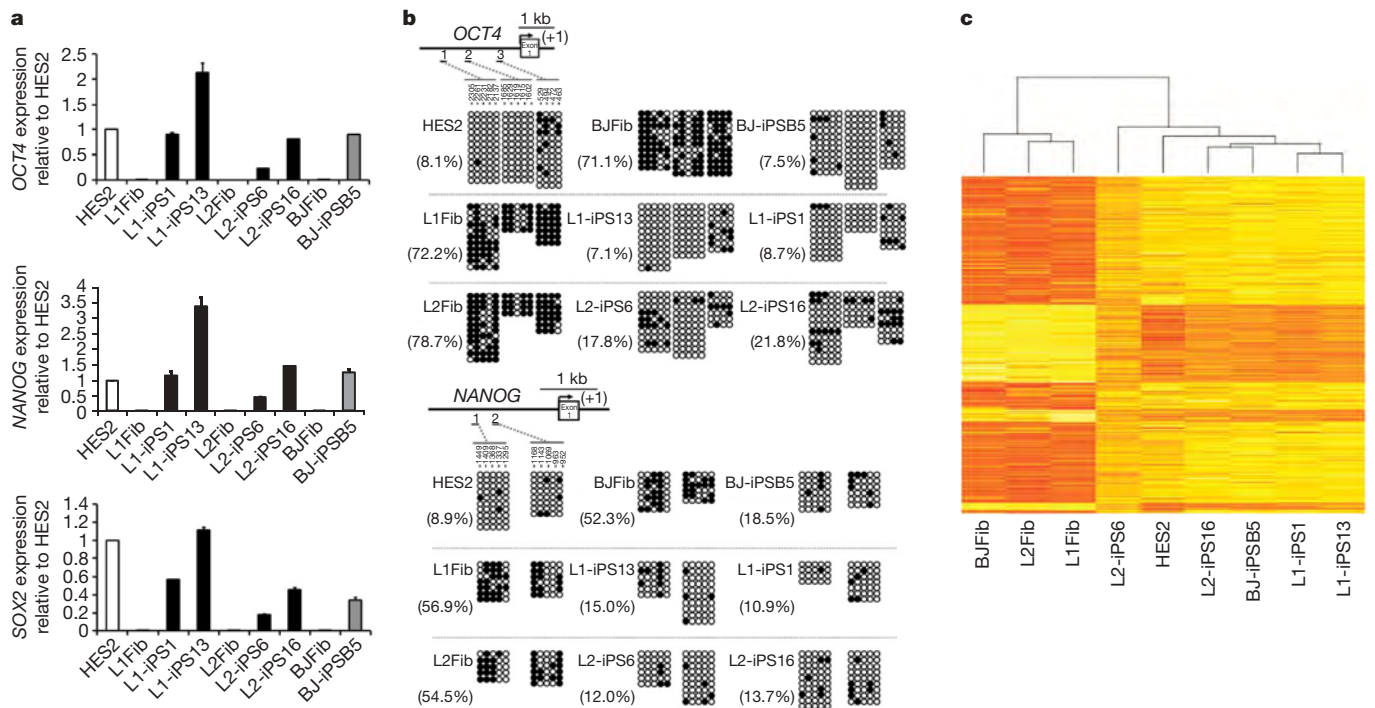


Figure 1 | Gene expression profile in LEOPARD syndrome iPSCs is similar to human ES cells. a, Quantitative real-time PCR assay for the expression of endogenous human *OCT4*, *NANOG* and *SOX2* in iPSCs and parental fibroblasts (Fib). PCR reactions were normalized against β -actin and plotted relative to expression levels in HES2. Error bars indicate \pm s.d. of triplicates. **b**, Bisulphite sequencing analyses of the *OCT4* and *NANOG* promoters. The

cell line and the percentage of methylation are indicated to the left of each cluster. **c**, Heat map showing hierarchical clustering of 3,657 genes with at least twofold expression change between the average of the three fibroblast cell lines versus all the iPSC lines/human ES cell samples. Expression levels are represented by colour; red indicates lower and yellow higher expression.

We next examined genome-wide messenger RNA expression profiles of two LEOPARD syndrome iPSC lines from each patient, the BJ-iPSB5 cell line, parental fibroblasts and HES2 cells. The resulting heat map and scatter-plot analyses indicated that iPSC lines shared a higher degree of similarity with HES2 cells than with their parental fibroblast cell lines (Fig. 1c and Supplementary Fig. 7b).

Pluripotent human ES cells can differentiate into cell types representative of all three germ layers. We tested the differentiation

abilities of our iPSCs using an *in vitro* floating embryoid body (EB) system, followed by replating on gelatin-coated dishes^{10,15}. Immunocytochemistry analyses detected expression of α -smooth muscle actin (α -SMA, mesoderm), desmin (mesoderm), α -fetoprotein (AFP, endoderm), vimentin (mesoderm), glia fibrillary acidic protein (GFAP, ectoderm) and β III-tubulin (ectoderm) markers (Fig. 2a and Supplementary Fig. 8). To determine pluripotency *in vivo*, we injected LEOPARD syndrome iPSCs, BJ-iPSB5 and HES2 cells into

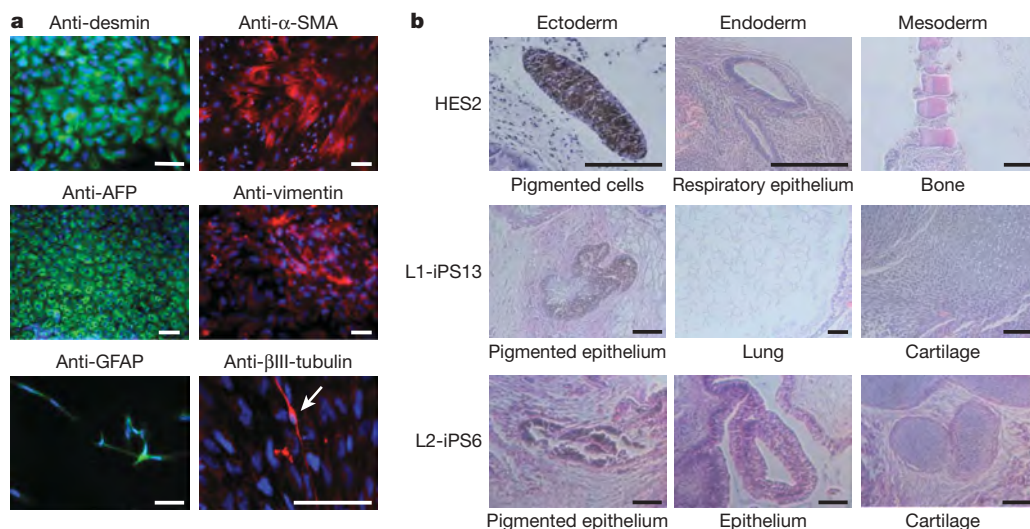


Figure 2 | LEOPARD syndrome iPSCs differentiate *in vitro* and *in vivo* into all three germ layers. a, L2-iPS6 cells were differentiated as floating EBs for 8 days and then plated onto gelatin-coated dishes and allowed to differentiate for another 8 days. Immunocytochemistry showed cell types positively stained for differentiation markers including desmin/ α -SMA (mesoderm), AFP (endoderm), vimentin (mesoderm) and GFAP/ β III-

tubulin (ectoderm). The arrow indicates a β III-tubulin-positive cell. Scale bar, 100 μ m. **b**, HES2, L1-iPSC and L2-iPSC were injected subcutaneously into the right hind leg of immuno-compromised NOD-SCID mice. The resulting teratomas were stained with haematoxylin and eosin and tissues representative of all three germ layers were observed. Scale bar, 100 μ m.

immuno-compromised NOD-SCID mice. Histological analyses of the resulting teratomas showed cell types representative of the three germ layers, including pigmented cells (ectoderm), lung, respiratory and gut-like epithelia (endoderm), and mesenchyme, adipose tissue and cartilage (mesoderm) (Fig. 2b and data not shown).

As mentioned previously, hypertrophic cardiomyopathy is one of the major features of LEOPARD syndrome, affecting 80% of the patients. In addition, affected individuals occasionally manifest haematological complications such as myelodysplasia and leukaemia^{16,17}. Therefore, we asked if LEOPARD syndrome iPSCs were able to differentiate into haematopoietic and cardiac lineages. iPSCs from both patients differentiated into a variety of haematopoietic cell types including early haematopoietic progenitors (CD41⁺)¹⁸, early erythroblasts (CD71⁺/CD235a⁺)¹⁹ and macrophages (CD11b⁺)²⁰ (Supplementary Fig. 9 and data not shown). The cardiac hypertrophic response includes induction of immediate-early genes (such as *JUN*, *FOS* and *MYC*), an increase in cell size, and organization of contractile proteins into sarcomeric units^{21,22}. To have an appropriate control cell line to analyse some of these parameters, besides human ES cells, we generated a wild-type iPSC line (S3-iPS4) from fibroblasts obtained from an unaffected brother of L1 without the T468M mutation (Supplementary Figs 10 and 11). Using a well-established cardiac differentiation protocol²³, we observed contracting EBs emerging around day 11 of differentiation (Supplementary Movies 1–7). To monitor cardiac development, we analysed cardiac troponin T (cTNT) expression on day 18 of differentiation by flow cytometry (data not shown). Replated cells from beating EBs were processed as described in Methods. Briefly, cells were fixed, immunostained for cTNT (Fig. 3b) and 50 cardiomyocytes were randomly chosen from each sample for surface area measurement using a computerized morphometric system (ImageJ software, NIH). Cardiomyocytes derived from LEOPARD syndrome iPSC lines L1-iPS13, L1-iPS6 and L2-iPS10 (Supplementary Fig. 10 and Supplementary Fig. 11) had a significantly increased median surface area compared to wild-type iPSC cardiomyocytes (1.8 times, 2.5 times and 4.8 times larger, respectively), whereas the area median of the cardiomyocytes obtained from human ES cells was similar to wild-type iPSC cardiomyocytes (Fig. 3a). We also observed increased sarcomere assembly in L1-iPS6 and L2-iPS10 cells when compared to wild-type S3-iPS4 cells (Fig. 3b). Recently, the calcineurin–NFAT pathway has been shown to be an important regulator of cardiac hypertrophy. Active calcineurin dephosphorylates NFAT transcription factors, resulting in their nuclear translocation^{22,24}. We analysed the localization of NFATC4 using immunocytochemistry in 50 cTNT-positive cardiomyocytes derived from the L2-iPS10 cell line, which produced the largest cardiomyocytes, and wild-type S3-iPS4. We observed a significantly higher proportion of LEOPARD syndrome cardiomyocytes with nuclear NFATC4 (~80% versus ~30%, respectively; Fig. 3c, d).

To identify potential molecular targets that could be affected by the T468M *PTPN11* mutation, protein extracts from LEOPARD syndrome iPSCs, wild-type BJ-iPSB5 and HES2 cells were analysed using a phosphoproteomic microarray chip containing approximately 600 pan and phospho-specific antibodies (Kinexus Bioinformatics Corporation). We established eight groups for comparison, each of the LEOPARD syndrome iPSC lines versus one control cell line, either HES2 or wild-type iPSCs. Proteins with a 1.5-fold change were filtered, and those that were conserved in most of the groups were represented in a heat map (Fig. 4a). Some of the proteins were more abundantly present in LEOPARD syndrome iPSCs when compared to either HES2 cells (DDR2, TYK2 and haspin) or wild-type iPSCs (p-MARCKs, p-synapsin 1, p-GRIN2B, p-RPS6KA5, p-RSK1/3 and p-p53). The phosphorylation of other proteins was increased (p-caveolin 2, p-MEK1, p-EGFR and p-FAK) or decreased (p-vinculin, p-S6 and p-LCK) in LEOPARD syndrome iPSCs when compared to control cell lines. To eliminate false positives, we verified the phosphoproteomic results by western blot for three of the most altered proteins (p-S6, p-EGFR and p-MEK1) in four LEOPARD syndrome

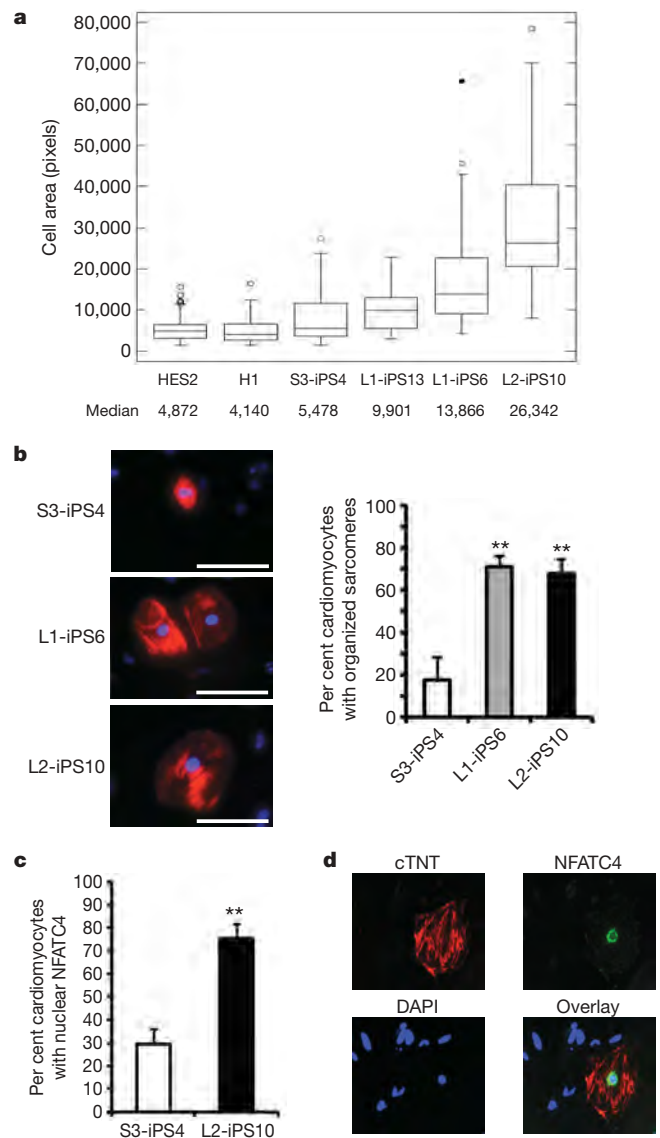


Figure 3 | Cardiomyocytes derived from LEOPARD syndrome iPSCs show hypertrophic features. **a**, HES2, H1, wild-type S3-iPS4 and three LEOPARD syndrome iPSC clones were differentiated into cardiac lineage. Cell areas of 50 random cTNT-positive cardiomyocytes of each cell line were measured using ImageJ. Boxes show the span from the median (50th percentile) to the first and third quartiles. The lines represent the largest/smallest sizes that are no more than 1.5 times the median to quartile distance. Additional points drawn represent extreme values, ≥ 1.5 times (open circles) and ≥ 3 times (filled circle) the median to quartile distance. **b**, Sarcomeric organization was assessed in 50 cTNT-positive (red) cardiomyocytes. Data are presented as mean \pm s.d. $n = 3$; $**P < 0.01$ (Student's *t*-test). **c**, S3-iPS4 and L2-iPS10 cell-derived cardiomyocytes were re-stained with NFATC4 antibody, and the nuclear versus cytosolic expression was analysed. $n = 3$; $**P < 0.01$ (Student's *t*-test). **d**, Nuclear localization of NFATC4 protein in a cTNT-positive cell from L2-iPS10 is shown.

iPSC lines, in comparison to wild-type iPSCs. Although we did not confirm a major change in the phosphorylation status of S6 protein (data not shown), western blot confirmed that the phosphorylation of EGFR and MEK1 proteins was considerably increased in the LEOPARD syndrome iPSC samples (Fig. 4b).

RAS–mitogen-activated protein kinase (MAPK) represents the major signalling pathway deregulated by SHP2 mutants. Noonan syndrome mutants increase basal and stimulated phosphatase activity, whereas LEOPARD syndrome mutants are catalytically impaired and have dominant-negative effects, inhibiting growth-factor-evoked ERK1/2 activation²⁵. We analysed the ability of LEOPARD syndrome

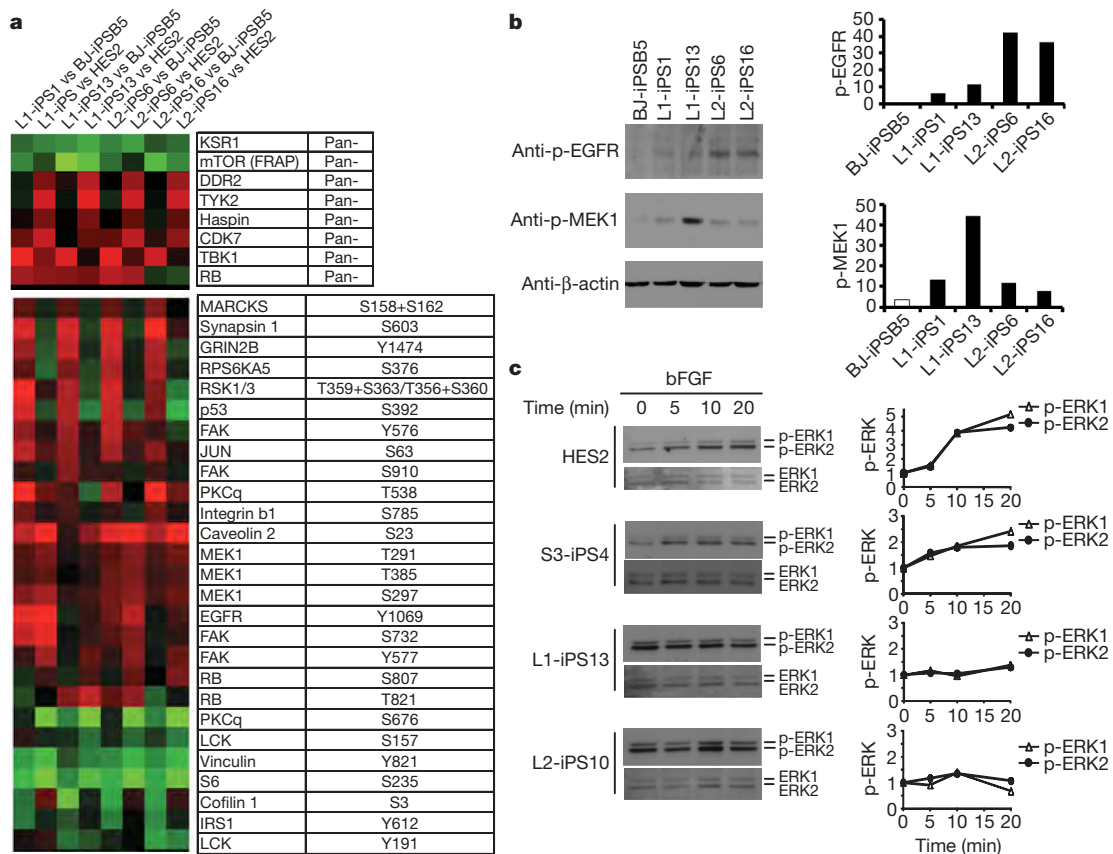


Figure 4 | Phosphoproteomic and MAPK activation analyses. **a**, Protein extracts of two iPSCs from each LEOPARD syndrome patient (L1 and L2), wild-type iPSCs (BJ-iPSB5) and HES2 cells were hybridized to an antibody microarray. The heat map represents the most significant protein changes preserved in all the comparison groups. **b**, p-MEK1 and p-EGFR expression was confirmed by western blot using phospho-specific antibodies. Band density was measured (ImageJ software) and normalized to β-actin. **c**, HES2,

wild-type S3-iPS4 and LEOPARD syndrome iPSCs were serum- and bFGF-starved for 6 h and then treated with bFGF (20 ng ml^{-1}) for the indicated time. Phosphorylated ERK1/2 (p-ERK1/2) and total ERK were assessed by immunoblotting and quantified. p-ERK1/2 levels were compared to the untreated p-ERK1/2 level in each sample, normalized to the total ERK1/2 and represented graphically at the right of each panel.

iPSCs to respond to external growth factors. We used basic fibroblast growth factor (bFGF), the main growth factor in the maintenance of human ES cells, to induce the stimulation of the MAPK signalling pathway. bFGF treatment increased the phosphorylation of ERK1/2 (p-ERK) levels over time in HES2 and wild-type S3-iPS4 cells (Fig. 4c). Although the LEOPARD syndrome iPSCs expressed the four FGF receptor (FGFR) family members (Supplementary Fig. 12a), bFGF stimulation did not cause any substantial change in p-ERK levels (Fig. 4c and Supplementary Fig. 12b, c). However, the LEOPARD syndrome iPSC lines had higher basal p-ERK levels compared to HES2 and S3-iPS4 cells (Fig. 4c and Supplementary Fig. 12b, c), in accordance with the increased p-MEK1—the upstream kinase of ERK—levels found in LEOPARD syndrome iPSC samples in phosphoproteomic array results.

We have generated and characterized LEOPARD syndrome patient-specific iPSCs, providing a new system for the study of disease pathogenesis. Some of the standard procedures to analyse cardiomyocyte hypertrophy (for example, protein synthesis rate, re-activation of the fetal gene program) could not be reliably assessed due to the variably mixed population of cells obtained using this cardiac differentiation procedure (for example, endothelial, cardiomyocytes), and the lack of a reliable cell surface marker for cardiomyocyte purification. However, we observed increases in cell size, sarcomeric organization and nuclear NFATC4 localization in LEOPARD syndrome iPSC-derived cardiomyocytes, when compared to human ES cell and wild-type iPSC-derived cardiomyocytes. These results would be consistent with cardiac hypertrophy, a condition commonly found in LEOPARD syndrome patients², and indicate that this abnormality occurs through a

cell-autonomous mechanism due to the *PTPN11* mutation. Because many human cell types, such as cardiomyocytes, cannot be propagated readily in cell culture, iPSC-derived cells exhibiting disease-relevant phenotypes provide the requisite resource for precisely elucidating pathogenesis and pursuing novel therapeutic strategies.

In our studies, we also attempted to provide insights into the molecular events that could be affected by the *PTPN11* mutation in the pluripotent iPSCs using antibody microarrays. We found that the phosphorylation of certain proteins was increased in LEOPARD syndrome iPSCs when compared to wild-type human ES cells and iPSCs. Further analysis will be required to elucidate if these proteins/signalling pathways are involved in the development of the disease phenotype. Notably, one of the more upregulated phosphoproteins was MEK1, the upstream kinase of ERK1/2, the gene of which is sometimes mutated in the related disorder cardiofaciocutaneous syndrome. *PTPN11* mutations underlie 45% and 90% of Noonan syndrome and LEOPARD syndrome, respectively. It is not well understood how mutations that provoke opposite effects on SHP2 phosphatase activity cause syndromes with similar features²⁶. In concordance with observations in the *Drosophila* LEOPARD syndrome model⁷, basal p-ERK levels were increased in LEOPARD syndrome iPSCs. Of note, receptor tyrosine kinase stimulation with bFGF in LEOPARD syndrome iPSCs failed to elicit further activation of ERK, as previously observed in a different cellular model²⁵. This result demonstrates that RAS–MAPK signal transduction is perturbed in LEOPARD syndrome as early as the pluripotent stem-cell stage. Taken together, this is, to our knowledge, the first described human model of an inherited RAS pathway disorder.

METHODS SUMMARY

Cell culture. Dermal fibroblast lines were obtained from skin biopsies, collected under an Institutional Review Board-approved protocol and with informed consent. Fibroblasts and GP2 cells were maintained in Dulbecco's modified Eagle medium (DMEM) containing 10% fetal bovine serum and penicillin/streptomycin (Invitrogen). HES2 cells and iPSCs were maintained on irradiated Swiss-Webster mouse embryonic feeder cells (MEFs), in a serum-free human ES cell medium containing 20 ng ml⁻¹ bFGF (R&D Systems).

LEOPARD syndrome iPSC generation. *OCT4*, *SOX2*, *KLF4* and *MYC* transcription factors were introduced in dermal fibroblasts derived from two patients with LEOPARD syndrome via the pMXs retroviral vector. In parallel, the pMXs-EGFP vector was used to estimate the infection efficiency (data not shown). Six days after infection, fibroblasts were seeded onto MEFs. The following day the medium was replaced with the human ES cell medium and bFGF.

Microarray analysis. Gene-level mRNA abundance measures were extracted using the Affymetrix GeneChip Exon 1.0 ST array, and Robust Multi-Array (RMA)-normalized using the Affymetrix Expression Console software. Subsequently, these genes were clustered and a heat map was generated against a background subset of genes showing at least twofold change between sample averages of iPSC/HES2 cells and fibroblast samples.

In vitro differentiation. For non-lineage-specific and haematopoietic differentiation we used previously described protocols^{10,15,27}, with certain modifications²⁸. For cardiomyocyte induction, we used a well-established assay²³.

Phosphoproteomics. HES2 cells and iPSCs were incubated overnight in human ES cell culture medium deprived of bFGF and knockout serum replacement (KSR). Protein lysates were quantified by Bradford assay and sent to Kinexus Bioinformatics Corporation for antibody microarray screening. The proteins with at least 1.5-fold change between the LEOPARD syndrome iPSC samples and control sample (either HES2 or wt-iPSC), and conserved in the majority of the comparison groups, were represented in a heat map.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 December 2009; accepted 8 March 2010.

- Gorlin, R. J., Anderson, R. C. & Moller, J. H. The Leopard (multiple lentiginos) syndrome revisited. *Birth Defects Orig. Artic. Ser.* **7**, 110–115 (1971).
- Sarkozy, A., Digilio, M. C. & Dallapiccola, B. Leopard syndrome. *Orphanet J. Rare Dis.* **3**, 13 (2008).
- Loh, M. L. *et al.* Mutations in PTPN11 implicate the SHP-2 phosphatase in leukemogenesis. *Blood* **103**, 2325–2331 (2004).
- Tartaglia, M. *et al.* Genetic evidence for lineage-related and differentiation stage-related contribution of somatic PTPN11 mutations to leukemogenesis in childhood acute leukemia. *Blood* **104**, 307–313 (2004).
- Tartaglia, M. *et al.* Diversity and functional consequences of germline and somatic PTPN11 mutations in human disease. *Am. J. Hum. Genet.* **78**, 279–290 (2006).
- Jopling, C., van Geemen, D. & den Hertog, J. Shp2 knockdown and Noonan/LEOPARD mutant Shp2-induced gastrulation defects. *PLoS Genet.* **3**, e225 (2007).
- Oishi, K. *et al.* Phosphatase-defective LEOPARD syndrome mutations in PTPN11 have gain-of-function effects during Drosophila development. *Hum. Mol. Genet.* **18**, 193–201 (2008).
- Lowry, W. E. *et al.* Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc. Natl Acad. Sci. USA* **105**, 2883–2888 (2008).
- Park, I. H. *et al.* Disease-specific induced pluripotent stem cells. *Cell* **134**, 877–886 (2008).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Ebert, A. D. *et al.* Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* **457**, 277–280 (2009).
- Lee, G. *et al.* Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* **461**, 402–406 (2009).
- Raya, A. *et al.* Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature* **460**, 53–59 (2009).

- Ye, Z. *et al.* Human induced pluripotent stem cells from blood cells of healthy donors and patients with acquired blood disorders. *Blood* **114**, 5473–5480 (2009).
- Dimos, J. T. *et al.* Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* **321**, 1218–1221 (2008).
- Laux, D., Kratz, C. & Sauerbrey, A. Common acute lymphoblastic leukemia in a girl with genetically confirmed LEOPARD syndrome. *J. Pediatr. Hematol. Oncol.* **30**, 602–604 (2008).
- Ucar, C., Calyskan, U., Martini, S. & Heinritz, W. Acute myelomonocytic leukemia in a boy with LEOPARD syndrome (*PTPN11* gene mutation positive). *J. Pediatr. Hematol. Oncol.* **28**, 123–125 (2006).
- Mikkola, H. K., Fujiwara, Y., Schlaeger, T. M., Traver, D. & Orkin, S. H. Expression of CD41 marks the initiation of definitive hematopoiesis in the mouse embryo. *Blood* **101**, 508–516 (2003).
- Wu, C. J. *et al.* Evidence for ineffective erythropoiesis in severe sickle cell disease. *Blood* **106**, 3639–3645 (2005).
- Fan, S. T. & Edgington, T. S. Coupling of the adhesive receptor CD11b/CD18 to functional enhancement of effector macrophage tissue factor response. *J. Clin. Invest.* **87**, 50–57 (1991).
- Aoki, H., Sadoshima, J. & Izumo, S. Myosin light chain kinase mediates sarcomere organization during cardiac hypertrophy *in vitro*. *Nature Med.* **6**, 183–188 (2000).
- Buitrago, M. *et al.* The transcriptional repressor Nab1 is a specific regulator of pathological cardiac hypertrophy. *Nature Med.* **11**, 837–844 (2005).
- Yang, L. *et al.* Human cardiovascular progenitor cells develop from a KDR⁺ embryonic-stem-cell-derived population. *Nature* **453**, 524–528 (2008).
- Molkentin, J. D. Calcineurin-NFAT signaling regulates the cardiac hypertrophic response in coordination with the MAPKs. *Cardiovasc. Res.* **63**, 467–475 (2004).
- Kontaridis, M. I., Swanson, K. D., David, F. S., Barford, D. & Neel, B. G. PTPN11 (Shp2) mutations in LEOPARD syndrome have dominant negative, not activating, effects. *J. Biol. Chem.* **281**, 6785–6792 (2006).
- Edouard, T. *et al.* How do Shp2 mutations that oppositely influence its biochemical activity result in syndromes with overlapping symptoms? *Cell. Mol. Life Sci.* **64**, 1585–1590 (2007).
- Kennedy, M., D'Souza, S. L., Lynch-Kattman, M., Schwantz, S. & Keller, G. Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures. *Blood* **109**, 2679–2687 (2007).
- Grigoriadis, A. E. *et al.* Directed differentiation of hematopoietic precursors and functional osteoclasts from human ES and iPS cells. *Blood* **115**, 2769–2776 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. James, X. Niu and D. York for their technical support and laboratory management, and B. MacArthur for support in microarray analysis. We also would like to thank K. Moore and her laboratory, and S. Mulero-Navarro from B.D.G.'s laboratory for their help, and V. Fuster and A. Bernad for their support. This research was funded by grants from the National Institutes of Health (NIH) to I.R.L. (5R01GM078465), the Empire State Stem Cell Fund through New York State Department of Health (NYSTEM) C024410 to I.R.L. and C.S., C024176 (HESC-SRF) to I.R.L. and S.L.D., C024407 to B.D.G., American College of Cardiology/Pfizer Research Fellowship to E.D.A., and ERA-Net for research programmes on rare diseases 2009 to M.T. X.C.-V. is a recipient of a Postdoctoral Fellowship from the Ministerio de Ciencia e Innovacion/Instituto de Salud Carlos III, D.-F.L. is a New York Stem Cell Foundation Stanley and Fiona Druckenmiller Fellow and S.P. is a recipient of a Ruth L. Kirschstein National Research Service Award (NRSA) Institutional Research Training Grant (T32).

Author Contributions X.C.-V. (iPSC establishment, project planning, experimental work and preparation of manuscript); A.S., S.L.D., Y.-S.A., L.Y., A.D.K., E.D.A., D.-F.L., A.W., B.C., J.S. and S.P. (experimental work); R.R. and Y.G. (microarray analysis); N.C. and L.J.E. (karyotype analysis); K.D.L. and M.T. (obtaining of fibroblast samples from patients); C.S. (project planning, experimental work); B.D.G. and I.R.L. (project planning, preparation of manuscript).

Author Information Microarray data have been deposited in NCBI-GEO under the accession number GSE20473. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.R.L. (ihor.lemischka@mssm.edu) or X.C.-V. (xcarvajal@gmail.com).

METHODS

Cell culture. Fibroblasts derived from LEOPARD syndrome patients, BJ fibroblasts (American Type Culture Collection) and GP-2 cells were maintained in DMEM 10% FBS medium. iPSCs and human ES cells were maintained on mitotically inactivated MEFs in human ES cell medium composed of DMEM/F12 (Cellgro, Mediatech) containing 20% (vol/vol) KSR (Invitrogen), 5% (vol/vol) MEF-conditioned medium, penicillin/streptomycin, L-glutamine (L-Gln), non-essential amino acids (Invitrogen), β -mercaptoethanol (β -ME, Sigma-Aldrich) and bFGF (R&D Systems).

Plasmid construction. Full-length sequences of human *OCT4*, *SOX2*, *KLF4* and *MYC* transcription factors were obtained from Open Biosystems. The coding sequences were PCR amplified using Pfu Turbo (Stratagene) and cloned into pMXs vector and verified by sequencing. pMXs-EGFP vector was constructed by introducing the BamHI/NotI EGFP fragment from FUGW (provided by C. Lois) into pMXs vector. The latter vector was used to monitor the transfection and infection efficiency. Detailed primer and cloning information will be provided upon request.

Retroviral infection and human iPSC generation. GP-2 cells were plated at 8×10^6 cells per 10-cm dish and transfected with pMXs, VSV-G and Gag-Pol vectors using SuperFect transfection reagent on the following day. The same day, human fibroblasts were seeded at 8×10^5 cells per 10-cm dish. Twenty-four hours after transfection, the four retroviruses (*OCT4*, *SOX2*, *KLF4* and *MYC*) containing supernatants were collected, and equal amounts of each were mixed and filtered through a 0.45- μ m pore-size filter, and supplemented with 4 μ g ml⁻¹ polybrene. Retrovirus-containing medium was added to the fibroblasts plates. The following day, 48-h after transfection, the fibroblasts were re-infected following the same procedure as the day before. Six days after transduction, fibroblasts were transferred into four dishes coated with MEFs, at 50,000 fibroblasts per plate.

DNA fingerprinting analysis. To verify the genetic relatedness of the iPSCs to their parental fibroblasts, we PCR amplified across three discrete genomic loci containing highly variable numbers of tandem repeats. Genomic DNA (gDNA) was isolated with the Easy-DNA kit (Invitrogen). Fifty nanograms of genomic DNA was used per reaction. Primers are summarized in Supplementary Table 1.

Karyotype analysis. HES2 cells and iPSCs were grown on Matrigel-coated glass coverslip dishes (MatTek, Ma). The day of culture harvest, 20 μ l of colcemid (5 μ g ml⁻¹) was added to the *in situ* ES cell culture which was 30–50% confluent. The culture was re-incubated for 15 min at 37 °C. A robotic harvester (Tecan) was used, which included automatic addition of 2 ml of hypotonic solution (sodium citrate solution 0.8%) with incubation for 20 min at 25 °C, pre-fixation with addition of 2 ml of fixative (methanol: glacial acetic acid; 3:1), followed by addition of 4 cc of fixative, twice. The coverslip was dried completely at 37 °C with 45–50% humidity and mounted on a microscope slide and GTG-banded according to standard protocols. Metaphases were captured and karyotypes were prepared using the CytoVision software program (Version 3.92 Build 7, Applied Imaging).

qPCR and transgene integration. For quantitative real-time PCR (qPCR) analyses, total RNA was extracted from cells using Trizol Reagent (Invitrogen) and subsequently column-purified with the RNeasy kit (Qiagen) and treated with RNase-free DNase (Qiagen). One microgram of total RNA was reverse transcribed into cDNA using random primers and Superscript II Reverse Transcriptase (Invitrogen). PCR for transgene silencing was performed with Expand High Fidelity Enzyme Taq Polymerase (Roche). Real-time qPCR was performed on a StepOne Plus Real-Time PCR System (Applied Biosystems) with Fast SYBR Green Master Mix (Applied Biosystems). The results were analysed with the StepOne Software v2.0, normalized to β -actin gene expression, and compared to HES2 cell expression levels. To examine the presence of transgenes in the iPSC lines, gDNA was isolated with the Easy-DNA Kit (Invitrogen). PCR reactions were carried out with the Expand High Fidelity Enzyme Taq Polymerase (Roche). Primer sequences are described in Supplementary Table 1. Primers for FGF receptors expression analysis have been previously described²⁹.

Southern blot analyses. gDNA (2 μ g) was completely digested with BglII, separated on a 0.8% agarose gel, transferred to a positively charged nylon membrane, and hybridized with DIG-labelled human *OCT4*, *SOX2*, *KLF4* and *MYC* cDNA probes. After hybridization, membranes were washed, blocked with DIG blocking solution, and incubated with anti-DIG-AP Fab fragments (Roche). Probes were then incubated with chemiluminescent CDP-Star substrates (Roche) and detected via exposure to X-ray film.

Bisulphite sequencing. We treated 500 ng of purified gDNA with sodium bisulphite using the Zymo EZ-DNA Methylation Kit, following the manufacturer's instructions. The sequences of primers used for amplification of genomic fragments were previously published³⁰. PCR products were then size fractionated in 1% TAE-agarose, extracted using the Qiaquick gel extraction kit (Qiagen) and cloned into the pGEM-T Easy Vector system (Promega). Blue-white selection was applied to eliminate false positives, and 12 random clones were picked and

sequenced. Bisulphite conversion efficiency of non-CpG cytosines was >90% for all individual clones for each sample.

Immunocytochemistry, AP staining and FACS analysis. For *in vivo* immunostaining, HES2 cells and iPSCs were washed once with DMEM medium supplemented with 10% FBS and antibiotics (DMEM 10%), and incubated with biotin-TRA1-81 antibody (1:100, eBiosciences) for 2 h. Cells were washed three times with DMEM 10% and they were incubated with the secondary antibody streptavidin-FITC (1:100, eBiosciences) and the phycoerythrin TRA1-60 (PE-TRA1-60) antibody (1:100, eBiosciences) where indicated. All the incubations were performed in a humidified incubator at 37 °C with 5% CO₂. For intracellular staining, cells were fixed in 2% paraformaldehyde for 30 min, and blocked and permeabilized in PBS containing 10% donkey serum, 1% BSA and 0.1% Triton X-100 for 45 min. Cells were incubated with primary antibody in blocking solution overnight at 4 °C, washed and incubated with the corresponding Alexa donkey secondary antibody for 1 h at 25 °C. Then cells were washed and stained with DAPI (1 μ g ml⁻¹) for 20 min. The primary antibodies used for intracellular immunostaining were OCT4 (1:100, BioVision), NANOG (1:100, R&D Systems), desmin (1:100, Lab Vision), α -SMA (pre-diluted, DAKO), vimentin (1:100, Chemicon), AFP (1:500, DAKO), GFAP (1:1000, DAKO), β III-tubulin (1:100, Chemicon), or NFATC4 (1:100, Santa Cruz Biotechnology). All the secondary antibodies, Alexa 488 donkey anti-rabbit (1:100), Alexa 546 donkey anti-goat (1:100) and Alexa 546 donkey anti-mouse (1:100) were obtained from Invitrogen. Alkaline phosphatase staining was detected following the manufacturer's recommendations (Millipore). SSEA-4, tropinin T and haematopoietic marker expression were evaluated on a BD Biosciences LSRII FACS machine analyser. Primary antibodies SSEA-4-PE (R&D Systems), cardiac troponin T (Lab Vision), CD11b-APC (Caltag), and CD45-APC, CD45-PE, CD71-PE, CD41-PE and CD235a-APC were purchased from BD Biosciences.

Teratoma formation. All animal procedures were performed in accordance with the Mount Sinai Medical Center's Institutional Animal Care and Use Committee. Approximately $1-2 \times 10^6$ cells were injected subcutaneously into the right hind leg of immuno-compromised NOD-SCID mice (The Jackson Laboratory). Teratomas were excised 6–10 weeks after injection, fixed overnight in formalin, embedded in paraffin, sectioned and stained with haematoxylin and eosin by the Morphology and Assessment Core of the Department of Gene and Cell Medicine. Histological evaluation was performed using a Nikon TE2000-U microscope and ACT-1 software.

In vitro differentiation. For non-lineage-specific differentiation we used previously described protocols^{10,15} with certain modifications. For EB formation, HES2 cells and iPSCs were treated with collagenase B (Roche) for 10 min, and collected by scraping. After centrifuging, cell pellets were re-suspended in basic differentiation media, StemPro 34 (Invitrogen), containing 2 mM L-Gln, 4×10^{-4} monothioglycerol (MTG), 50 μ g ml⁻¹ ascorbic acid (Sigma) and 150 μ g ml⁻¹ transferrin (Sigma). EBs were grown in ultra-low-binding plates (Costar) and medium was changed every 3 days. After 8 days of differentiation, EBs were collected, re-suspended in DMEM 10% and transferred to gelatin-coated dishes to allow them to attach and differentiate for eight additional days before processing for immunocytochemistry analyses. For haematopoietic differentiation we used a described protocol²⁷ with certain modifications²⁸. For cardiomyocyte induction, we used a well-established protocol²³.

Microarray analysis. RNA probes were hybridized to Affymetrix GeneChip Exon 1.0ST array according to the manufacturer's protocols by the Genomics Core Lab in The Institute for Personalized Medicine at Mount Sinai Medical Center. Microarrays were scanned and data were analysed using the Affymetrix Expression Console software.

Cytology/cardiomyocyte size determination. On day 18 of differentiation, beating EBs were plated on gelatin-coated dishes. Three days after plating, EB outgrowths were trypsinized, filtered through a 40- μ m size pore-size filter, and single cells were replated at low density on gelatin-coated dishes. The following day, cells were fixed with 4% paraformaldehyde, permeabilized, blocked in PBS/1% BSA/0.1% Triton/10% donkey serum, and stained for cardiac troponin T (1:200, Lab Vision) overnight at 4 °C. Stained cells were washed three times with PBS, and then incubated with the Alexa Fluor 547 donkey-anti-mouse antibody (Invitrogen) for 1 h. The areas of human ES cell- and LEOPARD syndrome iPSC-derived cardiomyocytes were analysed using ImageJ software (NIH).

Phosphoproteomics and western blotting. We prepared a lysis buffer (pH 7.2) containing 20 mM MOPS pH 7.0, 2 mM EGTA, 5 mM EDTA, 30 mM sodium fluoride, 60 mM β -glycerophosphate, 20 mM sodium pyrophosphate and 1% Triton X-100. Protease and phosphatase inhibitors (1 mM phenylmethylsulphonyl fluoride, 3 mM benzamide, 10 μ g ml⁻¹ aprotinin, 10 μ M leupeptin, 5 μ M pepstatin, 1 mM dithiothreitol and 1 mM sodium orthovanadate) were added to the lysis buffer immediately before use. Protein extracts were sent to Kinexus Bioinformatics Corporation. The antibody microarray results were processed following the company recommendations. Western blot was carried out as previously described³¹. The primary antibodies used were: pS6 S235/236 (1:1,000, Cell Signaling), pEGFR Y1086 (1:1,000, Cell Signaling), pMEK1 S298 (1:1,000, Cell

- Signaling), β -actin (1:5,000, Abcam), p-ERK1/2 T202/Y204 (1:2,000 Cell Signaling) and ERK1 (1:2,500, Santa Cruz Biotechnology).
29. Dvorak, P. *et al.* Expression and potential role of fibroblast growth factor 2 and its receptors in human embryonic stem cells. *Stem Cells* **23**, 1200–1211 (2005).
 30. Freberg, C. T., Dahl, J. A., Timoskainen, S. & Collas, P. Epigenetic reprogramming of OCT4 and NANOG regulatory regions by embryonal carcinoma cell extract. *Mol. Biol. Cell* **18**, 1543–1553 (2007).
 31. Carvajal-Vergara, X. *et al.* Multifunctional role of Erk5 in multiple myeloma. *Blood* **105**, 4492–4499 (2005).

Joseph C. Wu, MD, PhD
Associate Professor of Medicine & Radiology
Institute of Stem Cell Biology
Stanford University School of Medicine
Stanford, California

Biography

Joseph C. Wu, MD, PhD is an Associate Professor in the Department of Medicine (Cardiology) and Department of Radiology at the Stanford School of Medicine. Dr. Wu received his M.D. from the Yale School of Medicine. He completed his cardiology fellowship training followed by a PhD in Molecular Pharmacology at UCLA. He has received numerous awards, including the Burroughs Wellcome Foundation Career Award for Medical Scientists (2007), the NIH Director's New Innovator Award (2008), the NIH Roadmap Transformative R01 Award (2009), and the Presidential Early Career Award for Scientists and Engineers (2010).

Dr. Wu's lab uses a combination of gene profiling, tissue engineering, physiological testing, and molecular imaging technologies to better understand molecular and pathophysiological processes. The lab works on biological mechanisms of adult stem cells, embryonic stem cells (ESCs), and induced pluripotent stem cells (iPSCs). For adult stem cells, we are interested in monitoring stem cell survival, proliferation, and differentiation. For ESCs, we are currently studying their tumorigenicity, immunogenicity, and differentiation. For iPSCs, we are working on novel derivation techniques for potential downstream clinical translation and using patient-specific cardiac iPSC lines for drug screening.

Notes

Derivation of Human Induced Pluripotent Stem Cells for Cardiovascular Disease Modeling

Kamileh Narsinh, Kazim H. Narsinh and Joseph C. Wu

Circulation Research 2011, 108:1146-1156

doi: 10.1161/CIRCRESAHA.111.240374

Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75214

Copyright © 2011 American Heart Association. All rights reserved. Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://circres.ahajournals.org/content/108/9/1146>

Subscriptions: Information about subscribing to *Circulation Research* is online at
<http://circres.ahajournals.org/subscriptions/>

Permissions: Permissions & Rights Desk, Lippincott Williams & Wilkins, a division of Wolters Kluwer Health, 351 West Camden Street, Baltimore, MD 21202-2436. Phone: 410-528-4050. Fax: 410-528-8550. E-mail:
journalpermissions@lww.com

Reprints: Information about reprints can be found online at
<http://www.lww.com/reprints>

Derivation of Human Induced Pluripotent Stem Cells for Cardiovascular Disease Modeling

Kamileh Narsinh,* Kazim H. Narsinh,* Joseph C. Wu

Abstract: The successful derivation of human induced pluripotent stem cells (hiPSCs) by dedifferentiation of somatic cells offers significant potential to overcome obstacles in the field of cardiovascular disease. hiPSC derivatives offer incredible potential for new disease models and regenerative medicine therapies. However, many questions remain regarding the optimal starting materials and methods to enable safe, efficient derivation of hiPSCs suitable for clinical applications. Initial reprogramming experiments were carried out using lentiviral or retroviral gene delivery methods. More recently, various nonviral methods that avoid permanent and random transgene insertion have emerged as alternatives. These include transient DNA transfection using plasmids or minicircles, protein transduction, or RNA transfection. In addition, several small molecules have been found to significantly augment hiPSC derivation efficiency, allowing the use of a fewer number of genes during pluripotency induction. We review these various methods for the derivation of hiPSCs, focusing on their ultimate clinical applicability, with an emphasis on their potential for use as cardiovascular therapies and disease-modeling platforms. (*Circ Res.* 2011;108:1146-1156.)

Key Words: induced pluripotent stem cells ■ somatic donor cells ■ derivation technique ■ cardiovascular disease ■ pluripotent stem cell-derived cardiomyocytes

Human embryonic stem cells (hESCs) initially generated much enthusiasm because of their self-renewing and pluripotent properties.¹ hESCs theoretically can generate an unlimited number of any somatic cell, given the proper culture conditions. Directed differentiation can be promoted by varying the concentrations of various growth factors.² Recently, the Food and Drug Administration approved 2 hESC-based clinical trials for treatment of acute spinal cord injury (Geron; <http://www.geron.com>) and Stargardt macular dystrophy (Advanced Cell Technology; <http://www.advancedcell.com>).

Although hESCs have been established as a renewable source of definitive cardiac tissue cells, no clinical cardiovascular application of hESCs has yet been realized. hESC-derived cardiomyocytes (hESC-CMs) can be exploited as a human in vitro modeling system, in which the study of cardiogenesis, myocardial-related pathology, drug targets, drug screening, and tissue engineering can be easily conducted,³ paving the way for their future application in vivo as bona fide transplantation cells. Of note, the hESC-CMs derived to date are more akin to fetal/neonatal rather than adult cardiomyocytes,^{4,5} and enhancing their in vitro maturation using electric and mechanical cues is an ongoing area of investigation.⁶ Despite this, the advantages of using the human-based modeling system provided by hESCs remain

appealing, given that nonhuman transgenic models may not accurately reflect all aspects of the human disease phenotype. However, in the United States, hESC research funding can be subject to the vagaries of the courts and Congress, as evidenced by the recently lifted ban on federal funding for hESC research.⁷ In addition, the issues of potential immunologic rejection^{8,9} and tumorigenicity¹⁰ continue to be challenges to clinical cell transplantation approaches. The recent discovery of human induced pluripotent stem cells (hiPSCs) has mitigated some of these concerns because hiPSCs can be generated autologously and do not require the destruction of ex utero embryos. hiPSCs are comparable to hESCs in morphology, feeder dependency, surface markers, gene expression, promoter methylation status, telomerase activities, in vitro differentiation potential, and in vivo teratoma-forming capacity.¹¹ Hence, hiPSCs have great potential to replace hESCs as disease models or prospective treatment options.

Although the results of these initial comparisons are promising, the functional equivalence and safety of hESCs and hiPSCs remain contested.¹² For instance, hESCs display chromosomal instability with extended in vitro culture,¹³ and hiPSCs undergo dynamic changes in copy number during reprogramming and culture.¹⁴ Also, reactivation of repro-

Original received January 6, 2011; revision received March 1, 2011; accepted March 17, 2011.

From the Department of Medicine (K.N., K.H.N., J.C.W.) and the Department of Radiology (K.N., K.H.N., J.C.W.), Stanford University School of Medicine Stanford, CA; the Department of Chemistry and Biochemistry, University of California, Los Angeles (K.N.); and the Institute for Stem Cell Biology and Regenerative Medicine (J.C.W.), Stanford University School of Medicine, Stanford, CA.

*Contributed equally to this work.

Correspondence to Joseph C. Wu, MD, PhD, 300 Pasteur Dr, Grant S140B, Stanford, CA 94305-5111. E-mail joewu@stanford.edu

© 2011 American Heart Association, Inc.

Circulation Research is available at <http://circres.ahajournals.org>

DOI: 10.1161/CIRCRESAHA.111.240374

gramming transgenes in iPSCs may enable malignant transformation of the cells after transplantation.¹⁵ Most recently, variations in gene expression profiles between different hESC and hiPSC lines have been investigated, with mounting evidence indicating that cell line-specific differences in epigenetic landscape underlie differences in differentiation propensity.^{12,16–21} In light of these variations in performance, it is becoming clear that choice of reprogramming technique plays a crucial role in the quality and ultimate utility of the resulting hiPSC lines.

Since the ability to directly reprogram somatic cells to pluripotency has been well established in adult mouse fibroblasts,^{22–27} attempts to generate hiPSCs have become more pronounced, and the logistics involved in appropriately generating and using hiPSCs is of significant current interest. Several techniques have evolved for derivation of hiPSCs suitable for cellular regenerative medicine and disease modeling. This review discusses the methods, benefits, and drawbacks of several techniques, with an emphasis on their potential for future cardiovascular clinical applications.

Integrating Methods

Efforts to reprogram human somatic differentiated cell types to a state that resembles hESCs began with the pioneering work of Takahashi and Yamanaka.¹¹ Their methods included retroviral integration of 4 vital reprogramming factors—*OCT3/4*, *SOX2*, *KLF4*, and *c-MYC*—into adult human dermal fibroblasts. These 4 transcription factors would later become known as the “Yamanaka factors,” and their roles in reprogramming are now known to be significant²⁸ but not collectively necessary.^{29–36} Often the omission of 1 or more of these reprogramming genes was contingent on the endogenous network of the donor cell type. For example, one study found that hiPSC derivation from keratinocytes required only 10 days, whereas neonatal skin fibroblasts required ≈30 days.³⁷ It was postulated that perhaps the keratinocytes’ higher endogenous expression levels of *c-MYC* and *KLF4* predispose them to quicker reprogramming.³⁸ Starting cell type is thus an important consideration in the derivation process and a topic that is more thoroughly discussed elsewhere.³⁹ Two other transcription factors, namely *NANOG* and *LIN28*, were initially shown to be able to substitute for *c-MYC* and *KLF4*, although a number of other different factor combinations have been subsequently demonstrated.^{26,29,36,40,41} In any event, several cocktails comprising any number of these 6 reprogramming factors, and in some cases, additional supplements such as small molecules and enzymes, have been shown to be capable of reprogramming cells to pluripotency. Some of the advances made in the generation of hiPSCs using such integrating methods are detailed in Table 1.

A chief aim of clinical hiPSC researchers is to achieve a high efficiency of derivation of hiPSCs, because current yields of bona fide hiPSCs can be as low as 0.001% to 0.1% of the starting cell population.⁴² Even in secondary reprogramming systems, using transgenic fibroblasts expressing all four transgenes simultaneously, the efficiency of pluripotency induction remains low, at 1% to 5%.^{37,43} Two mutually nonexclusive models have been proposed to explain the apparent resistance to pluripotency induction, termed the

Abbreviations and Acronyms

| | |
|-------------------------------|---|
| hiPSC | human induced pluripotent stem cell |
| hESC | human embryonic stem cell |
| hESC-CM | hESC-derived cardiomyocyte |
| hiPSC-CM | hiPSC-derived cardiomyocyte |
| RT-PCR | reverse-transcription polymerase chain reaction |
| TGF-β | transforming growth factor beta |
| VPA | valproic acid |
| bFGF | basic fibroblast growth factor |
| CTNT | cardiac troponin T |
| KDR | kinase insert domain receptor |
| GFP | green fluorescent protein |
| EB | embryoid body |

“elite” and “stochastic” models.⁴⁴ The elite model proposes that only a small percentage of somatic cells, presumably resident tissue progenitor cells, are amenable to reprogramming. In support of this notion is evidence that hematopoietic stem cells undergo more efficient reprogramming than their differentiated progeny.⁴⁵ However, reports of successful reprogramming of terminally differentiated cells such B-lymphocytes⁴⁶ and pancreatic β -islets⁴⁷ favor a stochastic model of reprogramming, in which successive cell divisions allow rare cells to acquire the stochastic changes necessary for conversion to full pluripotency.⁴⁸ Perhaps these seemingly contradictory hypotheses can be reconciled by a model in which adult stem/progenitor cells require fewer stochastic changes to undergo reprogramming than more differentiated cells. Further investigation of the reprogramming process using single-cell resolution imaging and other techniques will undoubtedly help yield further insight into these reprogramming roadblocks.

Clearly, the choice of gene delivery vector can change reprogramming efficiency by directly affecting the degree of expression of the reprogramming genes. Retroviral/lentiviral infection provides the benefit of high transgene expression levels in primary cells as compared with nonviral methods of reprogramming. However, retroviral/lentiviral methods for hiPSC generation have come under scrutiny because of concerns regarding their ultimate clinical safety. In particular, the random integration of transgenes into the human genome can potentially cause insertional mutagenesis, leading to malignant transformation of a clonal cell population and disastrous consequences.⁴⁹ Second, leaky expression caused by ineffective silencing of the transgenes may interfere with the physiological expression of the factors endogenously present within the cell, thereby potentially restricting differentiation propensity.⁵⁰ This residual expression may hamper the validity of in vitro hiPSC uses, such as in disease modeling, drug screening, and toxicology tests. Third, reactivation of *OCT4* and *c-MYC* has been shown to promote tumor formation in chimeric mice,^{26,51,52} prompting legitimate concern over posttransplantation tumorigenic risk if such methods were used in human patients.

Table 1. Summary of Different Techniques for Generating hiPSCs by Integration of Reprogramming Factors Into the Genome

| Vector | Factors (+ Conditions) | Human Starting Cell Type | Approximate Reprogramming Efficiency | Approximate Speed | Reference |
|------------|---|-------------------------------------|--------------------------------------|-------------------|-----------|
| Retroviral | OSKM | Fibroblasts | 0.02% | 25 d | 11 |
| | | | <0.01% | 31 d | 91 |
| Lentiviral | OSK | Keratinocytes | 0.8% | 10 d | |
| | | | 0.05% | 20 d | |
| | OSKMN | Fibroblasts | 0.001% | 31 d | 37 |
| | | Keratinocytes | 0.002% | 10 d | |
| | OSKM | Adipose-derived | 0.2% | 18 d | 92 |
| OSK | | 0.002%–0.02% | 29–42 d | 93 | |
| Retroviral | OSNL | Hematopoietic ESC-derived | 0.02% | 20 d | 41 |
| | OSN | Amnion-derived | 0.1% | >14 d | 94 |
| | OSK+VPA | Neonatal | 1% | 30 d | 32 |
| | | fibroblasts | 0.002% | | |
| | OSK+VPA+vitamin C | Adipose-derived | 7.06%* | 30–34 d | 76 |
| | | Fibroblasts | 3.7%* | | |
| | | Umbilical cord matrix-derived | 0.4% | 13–20 d | 77 |
| Lentiviral | OSKM+butyrate | Placental amniotic membrane-derived | 0.1% | | |
| | | Fetal fibroblasts | 16.3% | 18 d | 56 |
| | OSNL+hLIF+PD0325901+A-83-01+CHIR99021 | | 0.0025% | 28 d | 95, 96 |
| | OK+Parnate+CHIR99021 | Neonatal | 0.002% | 49 d | 30 |
| | OK+Parnate+CHIR99021+PD0325901+SB431542 | keratinocytes | 0.0075% | | |
| | O+PD0325901+butyrate+A-83-01+PS48 | | 0.00017% | 56 d | 79 |
| | O+Parnate+CHIR99021+PD0325901+butyrate+A-83-01+PS48 | Keratinocytes | 0.00025% | | |
| | Umbilical vein endothelium-derived | 0.01% | 42 d | | |
| | Amniotic fluid-derived | 0.004% | | | |

Retroviral and lentiviral methods are shown with associated enhancements due to addition of small molecules to reprogramming cell culture media.

*The parameter used to define pluripotency in reference 76 only includes positive testing of alkaline phosphatase. Other parameters used to define bona fide pluripotent stem cells were not used.

O indicates *OCT4*; S, *SOX2*; K, *KLF4*; M, *c-MYC*; N, *NANOG*; L, *LIN28*; VPA, valproic acid; and hLIF, human leukemia inhibiting factor.

In one of the initial forays into generating safer hiPSCs, Maherali et al³⁷ created a doxycycline-inducible lentiviral system, attempting to maintain the silencing of transcription factors and thus reduce the tumorigenicity of the cells after differentiation. Although this system is a step toward safer hiPSCs, the leakiness of the doxycycline-inducible promoter and the permanent incorporation of oncogenes into the host genome still warrant concern.

Viral Integration Followed by Excision: Cre-loxP

Soldner et al⁵³ generated viable hiPSCs free of exogenous reprogramming factors using doxycycline-inducible lentiviral vectors that integrated into the host genome but were subsequently excised by Cre recombinase. Fibroblasts were obtained via skin biopsies from 5 patients exhibiting sporadic Parkinson disease and transduced using 3 or all of the Yamanaka factors. The reported reprogramming efficiency

was 0.005% after transduction with the 3-factor combination and 0.01% with the 4-factor combination. Furthermore, the 3-factor transduced cells required 12 days of doxycycline exposure, as opposed to the 4-factor cells, which required only 8 days. Despite its lower reprogramming efficiency and temporal requirements, the 3-factor combination may be preferred over the 4-factor combination because the transduced cells are not overgrown by granulate colonies.^{11,31} Southern blot analysis demonstrated successful excision of the transgenes, and the resulting hiPSC lines maintained pluripotency independent of residual exogenous transcription factor expression.

Because the isolated hiPSC lines are patient-specific, they provide a system for investigating the proposed molecular and cellular mechanisms of the disease. Soldner et al⁵³ demonstrated successful derivation of dopaminergic neurons from Parkinson disease patients' cells, indicating that the

Table 2. Summary of Different Nonviral, Nonintegrating Techniques for Generating hiPSCs

| | Vector | Factors (+Conditions) | Starting Cell Type | Approximate Reprogramming Efficiency | Approximate Speed | Reference(s) |
|------------|--|-----------------------------|------------------------|--------------------------------------|-------------------|--------------|
| Excisional | Lentiviral | OSKM | Neonatal keratinocytes | 0.00001% | 22–35 d | 97 |
| | | | Fibroblasts | 0.0024% | 21–28 d | 53 |
| | PiggyBac transposition | OSKM | Embryonic fibroblasts | 0.008% | 14–28 d | 55 |
| | | | | 0.00625% | >14 d | 54 |
| | | | Bone marrow–derived | 0.16% | 24 d | 56 |
| Episomal | Plasmid DNA | OSKMNL, SVLT, EBNA1 | Fibroblasts | 0.00045% | 17 d | 57 |
| | Minicircle DNA | OSNL | Adipose-derived | 0.005% | 26–28 d | 61 |
| | | | Fetal fibroblasts | 0.0005% | | |
| Protein | Polyarginine tract tagged polypeptides | OSKM | Neonatal fibroblasts | 0.001% | 56 d | 65 |
| RNA | Modified synthetic mRNAs | OSKM | Fibroblasts | 1.4% | 17 d | 66 |
| | | OKSML + ↓ [O ₂] | | 4.4% | | |

Episomal, excisional, protein- and RNA-mediated methods are shown with associated enhancements due to addition of small molecules to reprogramming cell culture media.

O indicates *OCT4*; S, *SOX2*; K, *KLF4*; M, *c-MYC*; N, *NANOG*; L, *LIN28*; SVLT, SV40 large T gene; and EBNA1, Epstein-Barr nuclear antigen-1.

underlying age or disease of the donor most likely does not affect the ability of their cells to produce hiPSC-derived replacements *ex vivo*. Interestingly, in this study, factor-free hiPSCs were found to be more closely related to embryo-derived hESCs than provirus-carrying parental hiPSCs, based on gene expression analysis. Hence, basal expression of proviruses carried in conventional hiPSCs can affect the molecular characteristics of the hiPSCs. Although transgenes are expected to be completely silenced in bona fide hiPSC lines, residual sequences and chromosomal disruptions during and after viral integration may still result in harmful alterations that could pose clinical risks. As previously mentioned, reactivation of reprogramming transgenes after transplant can result in malignant transformation of the cells and formation of a tumor.¹⁵ Such unpredictable effects of incomplete transgene silencing on downstream hiPSC phenotype further highlight the need for transgene-free hiPSC derivation methods. Some recent advances in a variety of such methods are detailed in Table 2.

Nonviral Integration Followed by Excision: PiggyBac Transposition

Although Cre recombinase–driven excision utilizes a highly efficient and widely used system, small residual vector backbone sequences remain at the site of integration and may engender unpredictable downstream effects. Woltjen et al⁵⁴ and Kaji et al⁵⁵ demonstrated successful reprogramming of human embryonic fibroblasts using doxycycline-inducible reprogramming factors that were delivered as plasmids, stably integrated into the host genome, and subsequently excised using piggyBac transposition. Woltjen et al noted that successful transposon-based nonviral reprogramming has several advantages over traditional lentiviral integration-based reprogramming: (1) improved accessibility of reprogramming techniques through the use of plasmid DNA preparations and commercial transfection products, thereby eliminating the need for specialized biohazard containment facilities; (2) increased variety of reprogrammable donor cell types because

susceptibility to viral infection is no longer an important consideration; (3) feasibility of xeno-free production of hiPSCs; and (4) most importantly, near complete elimination of the expression of reprogramming factors after establishment of hiPSC lines by piggyBac transposase–mediated excision. However, successful excision of the reprogramming cassette was only achieved in approximately 2% of the bona fide hiPSCs exposed to piggyBac transposase, limiting the amount of vector-free hiPSCs that could potentially be produced. Of note, Mali et al⁵⁶ demonstrated the use of butyrate to enhance reprogramming efficiency 15- to 51-fold when used in conjunction with piggyBac transposase–driven integration and excision of the Yamanaka reprogramming factors. Genome-wide analysis of the effects of butyrate exposure at days 6 to 12 demonstrated significant changes in H3 acetylation and promoter methylation status in a variety of pluripotency-related genes, including DPPA2.

Nonviral Nonintegrating Methods

Despite the aforementioned progress toward safe and efficient reprogramming, the ideal method for hiPSC derivation would avoid temporary or permanent genomic modification. However, nonviral, nonintegrating reprogramming techniques typically have significantly lower reprogramming efficiencies than the aforementioned retroviral and lentiviral methods. Nonviral, nonintegrating methods for hiPSC generation include episomal plasmid DNA, minicircle DNA, protein, and synthetic RNA delivery. DNA- and RNA-based methods accomplish hiPSC induction by the transient expression of reprogramming factors, often through repetitive transfection protocols.

Episomal Plasmid

Yu et al⁵⁷ derived hiPSCs from human foreskin fibroblasts using nonintegrating episomal plasmids vectors. Repeated transient transfection of 3 plasmids expressing 7 reprogramming factors resulted in hiPSCs that were completely

transgene-free (as confirmed by reverse-transcription polymerase chain reaction [RT-PCR]) but maintained the proliferative and developmental potential of hESCs. The 7 factors include *OCT4*, *SOX2*, *c-MYC*, *KLF4*, *NANOG*, *LIN28*, SV40 large T antigen (SVLT), and Epstein-Barr nuclear antigen-1 (EBNA1). Of note, stable extrachromosomal replication of these plasmid vectors is enabled by the presence of the *cis*-acting oriP sequence and the transacting EBNA1 protein.^{58–60} Interestingly, coexpression of the 6 reprogramming factors (*OCT4*, *SOX2*, *NANOG*, *LIN28*, *KLF4*, and *c-MYC*) in initial experiments resulted in substantial cell toxicity and produced no hiPSC colonies. Only the inclusion of the SVLT and modification of the reprogramming factor ratio allowed for isolation of bona fide hiPSCs.⁵⁷ These results emphasize the importance of balanced absolute expression levels during the reprogramming process. Subsequently, the progressive loss of the episomal vectors enables isolation of hiPSCs devoid of vector or transgene sequences. However, reprogramming efficiency remained low, at approximately 3 to 6 colonies per 10^6 input cells.

Minicircle Vectors

Jia et al⁶¹ constructed a minicircle vector for reprogramming consisting of a cassette of the reprogramming factors *OCT4*, *SOX2*, *LIN28*, and *NANOG* and a green fluorescent protein (*GFP*) reporter gene, each separated by sequences encoding the 2A ribosomal slippage site. Though minicircles are also supercoiled derivatives of plasmids, they are unique in that they primarily consist of eukaryotic expression cassettes, which lack both a bacterial origin of replication and antibiotic resistance gene. These vectors are preferred to standard plasmids because they have shown comparatively enhanced and more persistent transgene expression both in vivo and in vitro, perhaps on account of their smaller sizes and minimal expression cassettes.^{62,63} Indeed, Jia et al⁶¹ reported that a plasmid vector-based protocol failed to generate hiPSCs in their system. In their minicircle protocol, the vectors were introduced into human adipose stromal cells 3 times: first via nucleofection, a proprietary electroporation system of Lonza, then twice via Lipofectamine. On days 14 to 16, colonies that were morphologically similar to hESC colonies were produced from human adipose stem cells and neonatal fibroblasts, whose reprogramming efficiency was reported at $\approx 0.005\%$ and $\approx 0.0005\%$, respectively.⁶⁴ Southern blot analysis confirmed the absence of genomic integration of the minicircle transgene in the subclones.

Protein Delivery

Direct delivery of reprogramming factor proteins into the cell is one way to entirely eliminate risks associated with viruses, genome manipulation, and DNA transfection. Kim et al⁶⁵ reported successful generation of hiPSCs from human newborn fibroblasts by direct delivery of the Yamanaka factor proteins fused with a cell-penetrating polyarginine peptide. After 6 protein treatment cycles over 8 weeks, 5 alkaline phosphatase-positive hiPSC-like colonies were derived from a starting cell population of 5×10^5 cells. Expansion of these hiPSC lines allowed confirmation of pluripotency by in vitro RT-PCR and in vivo teratoma formation assays. Despite the

hampered reprogramming process created by the need for repeated treatments and relative subdued reprogramming efficiency ($\approx 0.001\%$), reprogramming via protein transduction creates an attractive alternative by allowing generation of hiPSCs free of vectors or limitations caused by viral delivery. In conjunction with other methods such as use of purified recombinant mammalian proteins or concomitant use of small molecules, one could potentially further improve the reprogramming efficiency.

RNA Delivery

Nonintegrating DNA- and protein-based reprogramming strategies still have certain disadvantages. Even episomal DNA vectors used for hiPSC derivation entail a small risk of unintended genetic integration. Protein-based reprogramming requires relatively challenging production and purification of recombinant proteins. Recently, Warren et al⁶⁶ demonstrated the successful production of hiPSCs via repeated transfection of modified synthetic mRNAs. These modified mRNAs potentially bypass any innate immune response to foreign DNA elements. Tra-1 to 60^+ hiPSCs were derived from BJ fibroblasts by repeated transfection of a combination of 4 modified synthetic mRNAs encoding the Yamanaka factors at an efficiency of 1.4%. Retrovirally mediated reprogramming experiments conducted in parallel achieved an efficiency of 0.04%. Therefore, RNA-mediated reprogramming represented a 36-fold improvement in reprogramming efficiency. The kinetics of hiPSC generation using modified mRNAs was also noted to be accelerated approximately 2-fold. Addition of *LIN28* to the 4-factor cocktail and subjection of the cells to low oxygen culture conditions (5% O_2) generated a striking reprogramming efficiency of 4.4%. The significantly enhanced reprogramming efficiency of Warren et al may be attributable to several unique experimental advantages. First and foremost, transfected mRNA is translated into functional protein within several hours, whereas several days may elapse after lentiviral transduction before expression of functional protein. The ability to achieve stable intracellular stoichiometric concentrations of mRNA using transfection is presumably preferable to the somewhat variable expression patterns of transfected or transduced DNA constructs. Also, the immunogenicity of the foreign RNA was reduced by substitution of 5-methylcytidine and pseudouridine bases for cytidine and uridine, respectively, as well as media supplementation with the Vaccinia virus decoy protein B18R to allay any interferon-mediated immune response. These changes allowed for repetitive daily transfections for a period of ≈ 17 days without significant cytotoxicity. Such results represent a significant advance in our ability to derive transgene-free hiPSCs. In agreement with previous reports, the authors observed that transgene-free hiPSCs derived by RNA transfection more faithfully recapitulated the global gene expression profile of hESCs than retrovirally derived hiPSCs.

Small Molecules Augment Reprogramming Efficiency

Small molecules may improve the reprogramming efficiency or completely replace 1 or more reprogramming factors

during the hiPSC derivation process. Proponents of small molecules also assert that they are more timely in reaching their target, easier to control via concentration variations, simpler to use, and cheaper than the aforementioned reprogramming agents.⁶⁷ Chemical approaches to screening combinatorial small-molecule libraries have successfully identified several molecules that can augment pluripotency induction. O'Malley et al⁶⁸ separated these small molecules into 2 groups: (1) effectors of chromatin modification and (2) effectors of cell signaling pathways. This general classification system helps to organize and differentiate the targets of small molecules, allowing prediction of their effects and potential synergism. Detailed and comprehensive knowledge of the involved pathways is especially important because the effects of cell exposure to drugs may entail nonspecific and wide-ranging changes that unpredictably affect the safety and phenotype of the resulting cells. In addition, an understanding of the effects of small molecules on intracellular signaling mechanisms, developmental pathways, and cell fate may provide important insight into the molecular mechanisms of reprogramming.

Valproic Acid

Valproic acid (VPA) is a small-molecule histone deacetylase inhibitor that is already Food and Drug Administration–approved to treat epilepsy and bipolar disorder.^{69,70} VPA has been used to successfully reprogram neonatal foreskin fibroblasts when used in conjunction with *OCT4* and *SOX2*. Thus, it can effectively replace the oncogenic *c-MYC* and *KLF4* factors, albeit at the cost of a significantly reduced reprogramming efficiency (<0.005%).³² When VPA was used with *OCT4*, *SOX2*, and *KLF4*, a reprogramming efficiency of 1% was reported. Notably, reprogramming was accomplished in dermal fibroblasts, a differentiated cell type that does not endogenously express any of the reprogramming factors. Using VPA in conjunction with a cell type that endogenously expresses high levels of *c-MYC* or *KLF4*, such as keratinocytes or adipose stromal cells, might therefore significantly improve reprogramming efficiency.^{37,39}

SB431542, PD0325901, and Thiazovivin

Lin et al⁷¹ tested a combination of 3 molecules in an effort to precipitate speedier production of hiPSCs (created by retroviral transduction of fibroblasts [CRL2097 or BJ] with cDNAs encoding the Yamanaka factors). They concluded that dual inhibition of the transforming growth factor (TGF)- β and mitogen-activated protein kinase kinase pathways in a dose-dependent, temporal manner may guide and perhaps accelerate the kinetics of partially reprogrammed colonies to a fully pluripotent state. This is based on a \approx 100-fold improvement in reprogramming efficiency that occurred when TGF- β inhibitor SB431542 and mitogen-activated protein kinase kinase inhibitor PD0325901 were added to the cultures, promoting improved induction at earlier stages than expected, and a decrease in the number of granulate noniPSC colonies produced. These results highlight the importance of signal transduction pathways in the reprogramming process.

For example, TGF- β signaling can be active in tumorigenic suppression and immune response and cell migration regulation because members of this family primarily promote cellular senescence, differentiation, and apoptosis.⁴² TGF- β antagonists probably benefit reprogramming in other ways as well, including inhibition of the aforementioned cellular processes and promotion of the mesenchymal-to-epithelial transition by upregulating genes such as E-cadherin,⁷² a regulator of *NANOG* expression.⁷³

An additional impediment to high reprogramming efficiency is the poor survival of hiPSCs after dissociation into single cells by trypsinization. Addition of thiazovivin to hiPSC reprogramming culture media has been found to improve cell survival during protocols involving splitting via trypsinization.⁷¹ Remarkably, by 30 days after transduction, \approx 900 hESC-like colonies were produced from 10 000 seeded cells after a single 1:4 split on day 14 by using thiazovivin in conjunction with SB431542 and PD0325901. Independent of its demonstrated ability to promote single-cell survival, thiazovivin appears able to enhance reprogramming efficiency by approximately 2-fold when used in conjunction with SB431542 and PD0325901.

CHIR99021 and Parnate

CHIR99021 and Parnate are 2 other small molecules that have been used to successfully optimize reprogramming efficiency. Li et al³⁰ reported successful reprogramming of neonatal epidermal keratinocytes with only 2 transcription factors (*OCT4* and *KLF4*) in conjunction with CHIR99021 and Parnate. On average, a reprogramming efficiency of 0.002% was achieved. CHIR99021 is a glycogen synthase kinase-3 inhibitor that activates the Wnt signaling pathway, and Wnt3a-conditioned media is known to promote reprogramming in mouse embryonic fibroblasts.⁷⁴ Parnate is a monoamine oxidase inhibitor used as a second-line treatment for depression. It is also classified as an epigenetic modifier because of its inhibition of lysine-specific demethylase 1.⁷⁵ It is thereby thought to enhance reprogramming efficiency by inhibiting H3K4 demethylation. Interestingly, the use of Parnate and CHIR99021 in addition to SB431542 and PD0325901 did not result in any significant improvement in reprogramming efficiency because only 5 to 10 hiPSC colonies could be identified from 10^5 transduced cells. The pathways targeted by these additional molecules and their lack of synergy with other known small-molecule enhancers of reprogramming remain under investigation.

Vitamin C

Of particular interest is the effect of vitamin C on the reprogramming efficiency of hiPSCs. Esteban et al⁷⁶ found that vitamin C and other antioxidants had no effect on reprogramming efficiency when added to cell culture media supplemented with knockout serum replacement. However, when added to cell culture media supplemented with Dulbecco modified fetal bovine serum, vitamin C used in conjunction with VPA produced roughly 3 times as many alkaline phosphatase positive colonies as VPA alone (7.06% versus 2.10% efficiency). Improved reprogramming efficiency (maximum efficiency of 0.40%) using vitamin C was

also demonstrated by Cai et al,⁷⁷ using umbilical cord-derived cells. These cells were transduced with retroviruses using the Yamanaka factors and a chemical concoction of vitamin C, Dulbecco modified fetal bovine serum-based medium, and VPA. Given the known role of cellular senescence in impeding reprogramming, it is perhaps expected that vitamin C, a natural antioxidant, would enhance cellular reprogramming by reducing intracellular levels of reactive oxygen species.⁷⁸ However, alternative mechanisms for vitamin C's effects could include promotion of epigenetic modifications because vitamin C is a cofactor for many significant enzymes.

Butyrate, A-83-01, and PS48

Because keratinocytes express *KLF4* and *c-MYC* endogenously, it was believed that full reprogramming was possible with the sole transduction of *OCT4* in the presence of small molecules.³⁰ Only recently was this hypothesis definitively confirmed in a study by Zhu et al,⁷⁹ who used a mixture of sodium butyrate (sodium butyrate, a histone deacetylase inhibitor), PS48 (an inhibitor of phosphoinositide-dependent kinase-1), A-83-01 (TGF- β kinase/activin receptor-like kinase) inhibitor), PD0325901, CHIR99021, and Parnate to reprogram not only adult keratinocytes but umbilical vein endothelial and amniotic fluid derived cells at an average of 0.004% reprogramming efficiency. Interestingly, butyrate is believed to assist the reprogramming process epigenetically in mouse embryonic fibroblasts by modulating the flexibility of chromatin structure and repressing cancerous cell growth but only in the presence of *c-MYC*.⁸⁰ PS48 was found to upregulate glycolytic gene expression,⁷⁹ promoting a switch from mitochondrial oxidation, a metabolism used by somatic cells when they are no longer proliferating.⁸¹ The respective pathways affected by these molecules are critical in the reprogramming process, underscoring the need for further investigation into the mechanisms of dedifferentiation. Ideally, a chemical concoction without transcription factors would be sufficient to reprogram somatic cells to pluripotency, and progress toward this goal is readily apparent in recent work. Additional research into reprogramming enhancement using these and other small molecules is imperative, given the possibility that off-target effects of these small molecules may be detrimental to cell phenotype.

Cardiovascular Applications of iPSC-Derived Cells

Once derived, hiPSCs can serve as an inexhaustible source of more differentiated cardiovascular cells, including the cardiomyocyte, endothelial cell, and smooth muscle lineages. The adult heart has limited regenerative capacity, and pluripotent stem cell-derived lineages are ideal candidates for replacement therapies. Transplantation treatment of myocardial infarction may demand replacement of up to 1 billion damaged cardiomyocytes.⁸² Aside from the potential benefits that hiPSC-derived cardiomyocytes (hiPSC-CMs) can provide after myocardial infarction, endothelial cells generated from hiPSCs may help repair heart valves, vessels, and ischemic tissue,⁸³ perhaps by preventing cell death and providing vascular support in grafts. Obstacles to overcome before

applying hiPSC derivatives clinically include (1) standardization of lineage specification protocols to produce large quantities of pure, quality-controlled cells of the desired cell types, (2) methods to ensure their safe delivery, and (3) consideration of all possible adverse effects after treatment. As mentioned above, however, the hESC-CMs generated to date phenotypically resemble fetal/neonatal cardiomyocytes more closely than adult cardiomyocytes,^{4,5} and due care should be exercised in extrapolating results using hESC-CMs or hiPSC-CMs to the adult heart.

Generation of Cardiovascular Cells

Pioneering work by Gordon Keller's group (Yang et al²) has definitively demonstrated the cardiovascular potential of a specific population of kinase insert domain receptor ($KDR^{low}/C-KIT^{neg}$) cells derived from hESCs. Specifically, by mirroring developmental cues, they successfully induced generation of cardiovascular colony-forming cells from hESCs after temporal exposure to a combination of cytokines and signaling molecules, including bone-associated morphogenic protein 4, Activin A, basic fibroblast growth factor (bFGF), vascular endothelial growth factor, and Dickkopf-related protein 1. The $KDR^{low}/C-KIT^{neg}$ cells were able to differentiate into all 3 cardiovascular lineages, cardiomyocytes, smooth muscle cells, and endothelial cells. Cardiovascular progenitors were identified by gene expression profiling showing upregulation of cardiac genes (including *NKX2.5*, *ISL1*, *TBX5*, and *TBX20*), expression of cardiac troponin T (CTNT), and contractile ability. The cells were also found to be clonal and therefore could be sources of cardiovascular regeneration. Chemical augmentation of the Wnt signaling pathway was explored because Dickkopf-related protein 1, a Wnt pathway inhibitor, increased production of CTNT⁺ cells, whereas Wnt3a suppressed development. The largest amount of CTNT⁺ cells ($\approx 45\%$ of the developing embryoid bodies [EBs]) was produced with a combination of bone-associated morphogenic protein 4, bFGF, and Activin A. $KDR^{low}/C-KIT^{neg}/GFP$ cells transplanted into the hearts of nonobese diabetic/severe combined immunodeficient mice improved ejection fraction by 31%, and no further complications were reported. Other small-animal studies using transplanted hESC-CMs have revealed more modest short-term benefits,^{4,5} although grafted tissue is susceptible to acute donor cell death, tumorigenesis, and arrhythmogenesis.⁸²

More recently, Zhang et al⁸⁴ investigated the cardiovascular differentiation potential of hiPSC lines, potentially bypassing the aforementioned obstacles associated with hESC use. The cardiomyocytes derived from the EBs of hiPSCs lentivirally transduced with *OCT4*, *SOX2*, *NANOG*, and *LIN28* were comparable to cardiomyocytes derived from the EBs of hESCs. hiPSC-CMs and hESC-CMs both displayed relative downregulation of *OCT4* and *NANOG*, upregulation of cardiac genes, myofilament protein expression, and sarcomeric organization. hiPSC-CMs also proliferated robustly, generated atrial, nodal, and ventricular action potentials, and responded to electric and chemical stimulation of the β -adrenergic signaling pathway. There were somewhat fewer hiPSC-CMs that exhibited contractile ability than hESC-CMs, and ineffective silencing of the *OCT4* and *NANOG*

transgenes was demonstrated in RT-PCR analyses of the hiPSC-EBs. However, Zhang et al note that the differences between hiPSC-CMs and hESC-CMs are comparable to the differences already observed between hESC lines. Moreover, the aforementioned nonviral nonintegrating techniques to generate hiPSCs can bypass the problem presented by the transgene insertions in this study, thus showing promise for hiPSC lines as an alternative to hESC lines for a variety of applications.

Although they do not involve pluripotent stem cells, 2 recent studies warrant mention due to their novelty. Ieda et al⁸⁵ demonstrated direct transdifferentiation of murine fibroblasts into cardiomyocytes, entirely bypassing the pluripotent stage, using a 3-factor combination of developmental transcription factors. In vitro, 30% of cells exhibited CTNT expression 1 week after viral transduction with the GATA4, MEF2C, and TBX5 factors. Immunocytochemistry confirmed the presence of sarcomeric α -actinin and atrial natriuretic factor in some of these cells. Additionally, transduction of these factors into murine hearts induced cardiomyocyte differentiation after a single day, demonstrating a potential proof-of-concept therapeutic application. Although the induced cardiomyocytes were found to be epigenetically and electrophysiologically similar to wild-type cardiomyocytes, the cardiomyocyte-specific genes ACTC1, MYH6, RYR2, and GJA1 were not detected.

Direct reprogramming of murine fibroblasts to cardiomyocytes was subsequently been demonstrated using a slightly different transdifferentiation strategy.⁸⁶ By briefly overexpressing the Yamanaka factors and carefully controlling cardiogenic media supplementation, Efe et al⁸⁶ successfully transdifferentiated up to 40% of mouse fibroblasts into functional CTNT⁺ cardiomyocytes within 18 days. Importantly, small-molecule signaling played a pivotal role in modulating cell fate, for example, by inhibiting JAK/STAT-driven pluripotency induction. These recently developed protocols provide an important alternate platform for the production of patient-specific cardiomyocytes while entirely avoiding the pluripotent state and its concomitant tumorigenic risk.

Disease Modeling Using hiPSC-CMs

Several exciting demonstrations of the disease modeling capability of hiPSC-CMs have recently been published (Table 3). In all studies to date, the resulting patient-specific hiPSC-CMs have been found to at least partially exhibit the phenotype of the diseases under investigation: long-QT syndrome, Timothy syndrome, and LEOPORD syndrome.

A seminal study evaluated the use of hiPSCs in modeling LEOPARD syndrome, an autosomal-dominant developmental disorder of multiple organ systems resulting from a missense mutation in the *PTPN11* gene.⁸⁹ Compared with control hESC-CMs, diseased hiPSC-CMs were noted to have a higher mean cell surface area as well as nuclear translocation of the NFATC4 transcription factor, perhaps representing in vitro molecular surrogates of the disease's cardiac hypertrophy phenotype. In addition, phosphoproteomic comparison of the diseased and wild-type hiPSCs revealed that diseased hiPSCs did not respond to attempted MAPK activation by

Table 3. Studies Using hiPSC-CMs for Cardiac Disease Modeling

| Disease | Reprogramming Method | Methods for Evaluation of Phenotype | Reference |
|-------------------------|------------------------|---|-----------|
| LEOPARD syndrome | Retroviral integration | Microscopic morphometry Immunocytochemistry Phosphoproteomic analysis using antibody arrays and Western blots | 89 |
| Type 1 long-QT syndrome | Retroviral integration | Single-cell patch clamping Immunocytochemistry Adrenergic response | 87 |
| Type 2 long-QT syndrome | Retroviral integration | Single-cell patch clamping Microelectrode arrays Pharmacological response | 88 |
| Timothy syndrome | Retroviral integration | Single-cell patch clamping Calcium transient imaging Response to roscovitine | 90 |

hiPSC-CM indicates human induced pluripotent stem cell–derived cardiomyocyte.

bFGF. Although defects in MAPK activation are an expected result of mutation in the *PTPN11* gene encoding SHP2 tyrosine phosphatase, it is worthwhile to note that the standard procedures for analysis of cardiomyocyte hypertrophy, such as protein synthesis rate and activation of the fetal gene program, could not be reliably assessed in the mixed population of cells resulting from attempted cardiomyocyte differentiation.

A subsequent study evaluated type 1 long-QT syndrome in hiPSCs by comparing wild-type cells with patient-specific cells containing a [569G→A] missense mutation in the *KCNQ1* gene.⁸⁷ Once the patient-specific hiPSCs had been directed to the cardiac lineage, the ventricular and atrial myocyte action potentials had significantly longer QT intervals and slower repolarization velocity as compared with wild-type cardiomyocytes. Type 1 long-QT syndrome is attributed to a reduction in the I_{KS} (slow outward potassium current) responsible for mediating action potential repolarization, due to the defective *KCNQ1* channel. In agreement with this, single-cell electrophysiological analysis on the ventricular patient-specific hiPSC-CMs revealed a reduction in this current, confirming that the *KCNQ1* mutant interferes with the function of the wild-type subunit. Immunocytochemical tests of both populations of cardiomyocytes suggested that the phenotype is the result of a trafficking defect, in which the mutated *KCNQ1* protein fails to achieve membrane targeting. Last, stimulation of the ventricular patient-specific hiPSC-CMs using isoproterenol had little to no effect on the patient-specific hiPSC-CM repolarization and I_{KS} currents, whereas a significant reduction of both was produced in the

wild-type cells. Additionally, the patient-specific hiPSC-CMs had reduced action potential duration:action potential interval ratios. These adrenergically stimulating tests cumulatively suggest that long-QT syndrome hiPSC-CMs are predisposed to arrhythmic events.

Another study has extended the above-mentioned findings by modeling type 2 long-QT syndrome using similar methods.⁸⁸ These hiPSCs were derived from type 2 long-QT syndrome patients and contained a missense mutation in the *KCNH2* gene. The resulting hiPSC-CMs exhibited the electrophysiological hallmarks of the disease, including prolonged action potential duration and early after depolarizations in patch-clamping studies, as well as prolonged field potential duration in microelectrode array studies. Diseased hiPSC-CMs displayed the expected defect in I_{Kr} (delayed-rectifier potassium current) as well as increased susceptibility to pharmacologically induced arrhythmogenesis. The authors also demonstrate an important proof-of-concept drug screening experiment to evaluate the effects of nifedipine, pinacidil, and ranolazine on the electrophysiological properties of the diseased hiPSC-CMs.

Yazawa et al⁹⁰ derived hiPSCs from patients with Timothy syndrome, a disorder in which patients have long-QT syndrome, autism, immune deficiency, and syndactyly caused by a mutation in the *CACNA1C* gene encoding the $Ca_v1.2$ L-type channel. Beating hiPSC-derived EBs displayed irregular contraction rates, whereas single hiPSC-CMs displayed increased action potential duration as well as reduced voltage-dependent inactivation of the L-type calcium channel current. Interestingly, ventricular cardiomyocytes but not atrial cardiomyocytes displayed the prolonged action potential phenotype indicative of long-QT syndrome, in contrast to the results of Moretti et al.⁸⁷

Conclusion

The ability to safely and efficiently derive hiPSCs may be of decisive importance to the future of regenerative medicine, and it depends on researchers' eventual ability to generate hiPSCs free from foreign chemical or genomic elements while maintaining a stable cell line. Many techniques for hiPSC derivation have been developed in recent years, utilizing different starting cell types, vector delivery systems, and culture conditions. A refined or perfected combination of these techniques might prove to be the key to generating clinically applicable hiPSCs. Although this review is by no means meant to be exhaustive, we have attempted to highlight the significant developments in clinical translation of hiPSCs and their relevance to cardiovascular disease. Among these developments, the most exciting techniques involve enhancement of reprogramming by small molecules (Table 1), non-viral nonintegrating methods for hiPSC derivation (Table 2), and their subsequent application to disease modeling platforms (Table 3). Although outside the scope of this review, it bears noting the critical need for standardization of the parameters that define a cell as an hiPSC.³⁹ Without consensus on the criteria for identification of hiPSCs, comparison between the results of various investigators is difficult because the guidelines by which various investigators select their fully reprogrammed hiPSCs vary. Future endeavors to

advance hiPSCs as clinically relevant cardiovascular therapies may use an exciting combination of the methods reviewed here, with additional improvements sure to follow.

Acknowledgments

Because of space limitations, we are unable to include all of the important studies relevant to induced pluripotent stem cell derivation and application; we apologize to those investigators whom we omitted here.

Sources of Funding

This study was supported by National Institutes of Health New Innovator Award DP2OD004437, HL091453, HL089027, and Burroughs Wellcome Foundation Career Award Medical Scientists (J.C.W.) and the Howard Hughes Medical Institute (K.H.N.).

Disclosures

None.

References

1. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic stem cell lines derived from human blastocysts. *Science*. 1998;282:1145–1147.
2. Yang L, Soonpaa MH, Adler ED, Roepke TK, Kattman SJ, Kennedy M, Henckaerts E, Bonham K, Abbott GW, Linden RM, Field LJ, Keller GM. Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature*. 2008;453:524–528.
3. Zimmermann WH. Embryonic and embryonic-like stem cells in heart muscle engineering. *J Mol Cell Cardiol*. 2010;50:320–326.
4. Laflamme MA, Chen KY, Naumova AV, Muskheli V, Fugate JA, Dupras SK, Reinecke H, Xu C, Hassanipour M, Police S, O'Sullivan C, Collins L, Chen Y, Minami E, Gill EA, Ueno S, Yuan C, Gold J, Murry CE. Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nat Biotech*. 2007;25:1015–1024.
5. Cao F, Wagner RA, Wilson KD, Xie X, Fu J-D, Drukker M, Lee A, Li RA, Gambhir SS, Weissman IL, Robbins RC, Wu JC. Transcriptional and functional profiling of human embryonic stem cell-derived cardiomyocytes. *PLoS ONE*. 2008;3:e3474.
6. Chen MQ, Jin Y, Whittington RH, Wu JC, Kovacs GTA, Giovannardi L. Modeling conduction in host-graft interactions between stem cell grafts and cardiomyocytes. *Conf Proc IEEE Eng Med Biol Soc*. 2009;6014–6017.
7. Levine AD. Policy uncertainty and the conduct of stem cell research. *Cell Stem Cell*. 2011;8:132–135.
8. Swijnenburg RJ, Schrepfer S, Govaert JA, Cao F, Ransohoff K, Sheikh AY, Haddad M, Connolly AJ, Davis MM, Robbins RC, Wu JC. Immunosuppressive therapy mitigates immunological rejection of human embryonic stem cell xenografts. *Proc Natl Acad Sci U S A*. 2008;105:12991–12996.
9. Pearl JI, Lee AS, Leveson-Gower DB, Sun N, Ghosh Z, Lan F, Ransohoff J, Negrin RS, Davis MM, Wu JC. Short-term blockade of leukocyte costimulatory molecules promotes engraftment of embryonic and induced pluripotent stem cells. *Cell Stem Cell*. 2011;8:309–317.
10. Lee AS, Tang C, Cao F, Xie X, van der Bogt K, Hwang A, Connolly AJ, Robbins RC, Wu JC. Effects of cell number on teratoma formation by human embryonic stem cells. *Cell Cycle*. 2009;8:2608–2612.
11. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*. 2007;131:861–872.
12. Narsinh KH, Plews J, Wu JC. Comparisons of human induced pluripotent and embryonic stem cells: fraternal or identical twins? *Mol Ther*. 2011;19:635–638.
13. Xie X, Hiona A, Lee AS, Cao F, Huang M, Li Z, Cherry A, Pei X, Wu JC. Effects of long-term culture on human embryonic stem cell aging. *Stem Cells Dev*. 2011;20:127–138.
14. Hussein SM, Batada NN, Vuoristo S, Ching RW, Autio R, Narva E, Ng E, Sourour M, Hamalainen R, Olsson C, Lundin K, Mikkola M, Trokovic R, Peitz M, Brustle O, Bazett-Jones DP, Alitalo K, Lahesmaa R, Nagy A, Otonkoski T. Copy number variation and selection during reprogramming to pluripotency. *Nature*. 2011;471:58–62.

15. Miura K, Okada Y, Aoi T, Okada A, Takahashi K, Okita K, Nakagawa M, Koyanagi M, Tanabe K, Ohnuki M, Ogawa D, Ikeda E, Okano H, Yamanaka S. Variation in the safety of induced pluripotent stem cell lines. *Nat Biotech.* 2009;27:743–745.
16. Ghosh Z, Wilson KD, Wu Y, Hu S, Quertermous T, Wu JC. Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. *PLoS ONE.* 2010;5:e8975.
17. Feng Q, Lu S-J, Klimanskaya I, Gomes I, Kim D, Chung Y, Honig GR, Kim K-S, Lanza R. Hemangioblastic derivatives from human induced pluripotent stem cells exhibit limited expansion and early senescence. *Stem Cells.* 2010;28:704–712.
18. Hu BY, Weick JP, Yu J, Ma LX, Zhang XQ, Thomson JA, Zhang SC. Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc Natl Acad Sci U S A.* 2010;107:4335–4340.
19. Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T, Natesan S, Wagers AJ, Melnick A, Evans T, Hochedlinger K. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotech.* 2010;28:848–855.
20. Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, Kim J, Aryee MJ, Ji H, Ehrlich LIR, Yabuuchi A, Takeuchi A, Cunniff KC, Hongguang H, McKinney-Freeman S, Naveiras O, Yoon TJ, Irizarry RA, Jung N, Seita J, Hanna J, Murakami P, Jaenisch R, Weissleder R, Orkin SH, Weissman IL, Feinberg AP, Daley GQ. Epigenetic memory in induced pluripotent stem cells. *Nature.* 2010;467:285–290.
21. Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH, Gnirke A, Eggen K, Meissner A. Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell.* 2011;144:439–452.
22. Aoi T, Yae K, Nakagawa M, Ichisaka T, Okita K, Takahashi K, Chiba T, Yamanaka S. Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science.* 2008;321:699–702.
23. Boland MJ, Hazen JL, Nazor KL, Rodriguez AR, Gifford W, Martin G, Kupriyanov S, Baldwin KK. Adult mice generated from induced pluripotent stem cells. *Nature.* 2009;461:91–94.
24. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell.* 2006;126:663–676.
25. Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, Bernstein BE, Jaenisch R. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature.* 2007;448:318–324.
26. Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. *Nature.* 2007;448:313–317.
27. Zhao XY, Li W, Lv Z, Liu L, Tong M, Hai T, Hao J, Wang X, Wang L, Zeng F, Zhou Q. Viable fertile mice generated from fully pluripotent iPS cells derived from adult somatic cells. *Stem Cell Rev Rep.* 2010;6:390–397.
28. Liu X, Huang J, Chen T, Wang Y, Xin S, Li J, Pei G, Kang J, Yamanaka factors critically regulate the developmental signaling network in mouse embryonic stem cells. *Cell Res.* 2008;18:1177–1189.
29. Utikal J, Maherali N, Kulalert W, Hochedlinger K. Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells. *J Cell Sci.* 2009;122:3502–3510.
30. Li W, Zhou H, Abujarour R, Zhu S, Young Joo J, Lin T, Hao E, Schöler HR, Hayek A, Ding S. Generation of human-induced pluripotent stem cells in the absence of exogenous Sox2. *Stem Cell.* 2009;27:2992–3000.
31. Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotech.* 2008;26:101–106.
32. Huangfu D, Osafune K, Maehr R, Guo W, Eijkelenboom A, Chen S, Muhlestein W, Melton DA. Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nat Biotech.* 2008;26:1269–1275.
33. Jincho Y, Araki R, Hoki Y, Tamura C, Nakamura M, Ando S, Kasama Y, Abe M. Generation of genome integration-free induced pluripotent stem cells from fibroblasts of C57BL/6 mice without c-Myc transduction. *J Biol Chem.* 2010;285:26384–26389.
34. Kim JB, Greber B, Arauzo-Bravo MJ, Meyer J, Park KI, Zaehres H, Scholer HR. Direct reprogramming of human neural stem cells by Oct4. *Nature.* 2009;461:649–643.
35. Tsai SY, Clavel C, Kim S, Ang YS, Grisanti L, Lee DF, Kelley K, Rendl M. Oct4 and Klf4 reprogram dermal papilla cells into induced pluripotent stem cells. *Stem Cells.* 2010;28:221–228.
36. Eminli S, Utikal J, Arnold K, Jaenisch R, Hochedlinger K. Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. *Stem Cell.* 2008;26:2467–2474.
37. Maherali N, Ahfeldt T, Rigamonti A, Utikal J, Cowan C, Hochedlinger K. A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell.* 2008;3:340–345.
38. Yang S, Ha C, Jung M, Jin H, Lee M, Song H, Choi S, Oh W, Yang Y. Mesenchymal stem/progenitor cells developed in cultures from UC blood. *Cytotherapy.* 2004;6:476–486.
39. Sun N, Longaker MT, Wu JC. Human iPS cell-based therapy: considerations before clinical applications. *Cell Cycle.* 2010;9:880–885.
40. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA. Induced pluripotent stem cell lines derived from human somatic cells. *Science.* 2007;318:1917–1920.
41. Haase A, Olmer R, Schwanke K, Wunderlich S, Merkert S, Hess C, Zweigerdt R, Gruh I, Meyer J, Wagner S, Maier LS, Han DW, Glage S, Miller K, Fischer P, Schuler HR, Martin U. Generation of induced pluripotent stem cells from human cord blood. *Cell Stem Cell.* 2009;5:434–441.
42. Firestone AJ, Chen JK. Controlling destiny through chemistry: Small-molecule regulators of cell fate. *ACS Chem Biol.* 2009;5:15–34.
43. Hockemeyer D, Soldner F, Cook EG, Gao Q, Mitalipova M, Jaenisch R. A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell.* 2008;26:916–924.
44. Yamanaka S. Elite and stochastic models for induced pluripotent stem cell generation. *Nature.* 2009;460:49–52.
45. Eminli S, Foudi A, Stadtfeld M, Maherali N, Ahfeldt T, Mostoslavsky G, Hock H, Hochedlinger K. Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nat Genet.* 2009;41:968–976.
46. Hanna J, Markoulaki S, Schorderet P, Carey BW, Beard C, Wernig M, Creighton MP, Steine EJ, Cassady JP, Foreman R, Lengner CJ, Dausman JA, Jaenisch R. Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell.* 2008;133:250–264.
47. Stadtfeld M, Brennand K, Hochedlinger K. Reprogramming of pancreatic β cells into induced pluripotent stem cells. *Curr Biol.* 2008;18:890–894.
48. Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creighton MP, van Oudenaarden A, Jaenisch R. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature.* 2009;462:595–601.
49. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E, Sorensen R, Forster A, Fraser P, Cohen JI, de Saint Basile G, Alexander I, Wintergerst U, Frebourg T, Aurias A, Stoppa-Lyonnet D, Romana S, Radford-Weiss I, Gross F, Valensi F, Delabesse E, Macintyre E, Sigaux F, Soulier J, Leiva LE, Wissler M, Prinz C, Rabbitts TH, Le Deist F, Fischer A, Cavazzana-Calvo M. LMO2-associated clonal t cell proliferation in two patients after gene therapy for SCID-X1. *Science.* 2003;302:415–419.
50. Okita K, Yamanaka S. Induction of pluripotency by defined factors. *Exp Cell Res.* 2010;316:2565–2570.
51. Hochedlinger K, Yamada Y, Beard C, Jaenisch R. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell.* 2005;121:465–477.
52. Markoulaki S, Hanna J, Beard C, Carey BW, Cheng AW, Lengner CJ, Dausman JA, Fu D, Gao Q, Wu S, Cassady JP, Jaenisch R. Transgenic mice with defined combinations of drug-inducible reprogramming factors. *Nat Biotech.* 2009;27:169–171.
53. Soldner F, Hockemeyer D, Beard C, Gao Q, Bell GW, Cook EG, Hargus G, Blak A, Cooper O, Mitalipova M, Isacson O, Jaenisch R. Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell.* 2009;136:964–977.
54. Kaji K, Norrby K, Paca A, Mileikovsky M, Mohseni P, Woltjen K. Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature.* 2009;458:771–775.
55. Woltjen K, Michael IP, Mohseni P, Desai R, Mileikovsky M, Hamalainen R, Cowling R, Wang W, Liu P, Gertsenstein M, Kaji K, Sung HK, Nagy A. PiggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature.* 2009;458:766–770.
56. Mali P, Chou BK, Yen J, Ye Z, Zou J, Dowey S, Brodsky RA, Ohm JE, Yu W, Baylin SB, Yusa K, Bradley A, Meyers DJ, Mukherjee C, Cole

- PA, Cheng L. Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. *Stem Cells*. 2010;28:713–720.
57. Yu J, Hu K, Smuga-Otto K, Tian S, Stewart R, Slukvin II, Thomson JA. Human induced pluripotent stem cells free of vector and transgene sequences. *Science*. 2009;324:797–801.
 58. Yates J, Warren N, Reisman D, Sugden B. A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells. *Proc Natl Acad Sci U S A*. 1984;81:3806–3810.
 59. Yates JL, Guan N. Epstein-Barr virus-derived plasmids replicate only once per cell cycle and are not amplified after entry into cells. *J Virol*. 1991;65:483–488.
 60. Yates JL, Warren N, Sugden B. Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells. *Nature*. 1985;313:812–815.
 61. Jia F, Wilson KD, Sun N, Gupta DM, Huang M, Li Z, Panetta NJ, Chen ZY, Robbins RC, Kay MA, Longaker MT, Wu JC. A nonviral minicircle vector for deriving human iPS cells. *Nat Meth*. 2010;7:197–199.
 62. Huang M, Chen Z, Hu S, Jia F, Li Z, Hoyt G, Robbins RC, Kay MA, Wu JC. Novel minicircle vector for gene therapy in murine myocardial infarction. *Circulation*. 2009;120:S230–S237.
 63. Kay MA, He CY, Chen ZY. A robust system for production of minicircle DNA vectors. *Nat Biotech*. 2010;28:1287–1289.
 64. Narsinh KH, Jia F, Robbins RC, Kay MA, Longaker MT, Wu JC. Generation of adult human induced pluripotent stem cells using nonviral minicircle DNA vectors. *Nat Protocols*. 2010;6:78–88.
 65. Kim D, Kim CH, Moon JI, Chung YG, Chang MY, Han BS, Ko S, Yang E, Cha KY, Lanza R, Kim KS. Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell*. 2009;4:472–476.
 66. Warren L, Manos PD, Ahfeldt T, Loh YH, Li H, Lau F, Ebina W, Mandal PK, Smith ZD, Meissner A, Daley GQ, Brack AS, Collins JJ, Cowan C, Schlaeger TM, Rossi DJ. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell*. 2010;7:618–630.
 67. Anastasia L, Pelissero G, Venerando B, Tettamanti G. Cell reprogramming: expectations and challenges for chemistry in stem cell biology and regenerative medicine. *Cell Death Diff*. 2010;17:1230–1237.
 68. O'Malley J, Woltjen K, Kaji K. New strategies to generate induced pluripotent stem cells. *Curr Opin Biotechnol*. 2009;20:516–521.
 69. Leeman BA, Cole AJ. Advancements in the treatment of epilepsy. *Annu Rev Med*. 2008;59:503–523.
 70. Gottlicher M, Minucci S, Zhu P, Kramer OH, Schimpf A, Giavara S, Sleeman JP, Lo Coco F, Nervi C, Pelicci PG, Heinzel T. Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. *EMBO J*. 2001;20:6969–6978.
 71. Lin T, Ambudhan R, Yuan X, Li W, Hilcove S, Abujarour R, Lin X, Hahm HS, Hao E, Hayek A, Ding S. A chemical platform for improved induction of human iPSCs. *Nat Meth*. 2009;6:805–808.
 72. Thiery JP, Sleeman JP. Complex networks orchestrate epithelial-mesenchymal transitions. *Nat Rev Mol Cell Biol*. 2006;7:131–142.
 73. Chou YF, Chen HH, Eijpe M, Yabuuchi A, Chenoweth JG, Tesar P, Lu J, McKay RDG, Geijsen N. The growth factor environment defines distinct pluripotent ground states in novel blastocyst-derived stem cells. *Cell*. 2008;135:449–461.
 74. Marson A, Foreman R, Chevalier B, Bilodeau S, Kahn M, Young RA, Jaenisch R. Wnt signaling promotes reprogramming of somatic cells to pluripotency. *Cell Stem Cell*. 2008;3:132–135.
 75. Mimasu S, Sengoku T, Fukuzawa S, Umehara T, Yokoyama S. Crystal structure of histone demethylase LSD1 and tranylcypromine at 2.25 Å. *Biochem Biophys Res Commun*. 2008;366:15–22.
 76. Esteban MA, Wang T, Qin B, Yang J, Qin D, Cai J, Li W, Weng Z, Chen J, Ni S, Chen K, Li Y, Liu X, Xu J, Zhang S, Li F, He W, Labuda K, Song Y, Peterbauer A, Wolbank S, Redl H, Zhong M, Cai D, Zeng L, Pei D. Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell*. 2010;6:71–79.
 77. Cai J, Li W, Su H, Qin D, Yang J, Zhu F, Xu J, He W, Guo X, Labuda K, Peterbauer A, Wolbank S, Zhong M, Li Z, Wu W, So K-F, Redl H, Zeng L, Esteban MA, Pei D. Generation of human induced pluripotent stem cells from umbilical cord matrix and amniotic membrane mesenchymal cells. *J Biol Chem*. 2010;285:11227–11234.
 78. Banito A, Rashid ST, Acosta JC, Li S, Pereira CF, Geti I, Pinho S, Silva JC, Azuara V, Walsh M, Vallier L, Gil JS. Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev*. 2009;23:2134–2139.
 79. Zhu S, Li W, Zhou H, Wei W, Ambudhan R, Lin T, Kim J, Zhang K, Ding S. Reprogramming of human primary somatic cells by oct4 and chemical compounds. *Cell Stem Cell*. 2010;7:651–655.
 80. Liang G, Taranova O, Xia K, Zhang Y. Butyrate promotes induced pluripotent stem cell generation. *J Biol Chem*. 2010;285:25516–25521.
 81. Prigione A, Fauler B, Lurz R, Lehrach H, Adjaye J. The senescence-related mitochondrial/oxidative stress pathway is repressed in human induced pluripotent stem cells. *Stem Cell*. 2010;28:721–733.
 82. Gepstein L. Derivation and potential applications of human embryonic stem cells. *Circ Res*. 2002;91:866–876.
 83. Levenberg S, Ferreira LS, Chen-Konak L, Kraehenbuehl TP, Langer R. Isolation, differentiation and characterization of vascular cells derived from human embryonic stem cells. *Nat Protocol*. 2010;5:1115–1126.
 84. Zhang J, Wilson GF, Soerens AG, Koonce CH, Yu J, Palecek SP, Thomson JA, Kamp TJ. Functional cardiomyocytes derived from human induced pluripotent stem cells. *Circ Res*. 2009;104:e30–e41.
 85. Ieda M, Fu J-D, Delgado-Olguin P, Vedantham V, Hayashi Y, Bruneau BG, Srivastava D. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell*. 2010;142:375–386.
 86. Efe JA, Hilcove S, Kim J, Zhou H, Ouyang K, Wang G, Chen J, Ding S. Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. *Nat Cell Biol*. 2011;13:215–222.
 87. Moretti A, Bellin M, Welling A, Jung CB, Lam JT, Bott-Flügel L, Dorn T, Goedel A, Höhnke C, Hofmann F, Seyfarth M, Sinnecker D, Schömig A, Laugwitz K-L. Patient-specific induced pluripotent stem-cell models for long-QT syndrome. *N Engl J Med*. 2010;363:1397–1409.
 88. Itzhaki I, Maizels L, Huber I, Zwi-Dantsis L, Caspi O, Winterstern A, Feldman O, Gepstein A, Arbel G, Hammerman H, Boulos M, Gepstein L. Modelling the long QT syndrome with induced pluripotent stem cells. *Nature*. 2011;471:225–229.
 89. Carvajal-Vergara X, Sevilla A, D'Souza SL, Ang Y-S, Schaniel C, Lee D-F, Yang L, Kaplan AD, Adler ED, Rozov R, Ge Y, Cohen N, Edelmann LJ, Chang B, Waghay A, Su J, Pardo S, Lichtenbelt KD, Tartaglia M, Gelb BD, Lemischka IR. Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome. *Nature*. 2010;465:808–812.
 90. Yazawa M, Hsueh B, Jia X, Pasca AM, Bernstein JA, Hallmayer J, Dolmetsch RE. Using induced pluripotent stem cells to investigate cardiac phenotypes in Timothy syndrome. *Nature*. 2011;471:230–234.
 91. Aasen T, Raya A, Barrero MJ, Garreta E, Consiglio A, Gonzalez F, Vassena R, Bilic J, Pekarik V, Tiscornia G, Edel M, Boue S, Belmonte JCI. Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. *Nat Biotech*. 2008;26:1276–1284.
 92. Sun N, Panetta NJ, Gupta DM, Wilson KD, Lee A, Jia F, Hu S, Cherry AM, Robbins RC, Longaker MT, Wu JC. Feeder-free derivation of induced pluripotent stem cells from adult human adipose stem cells. *Proc Natl Acad Sci U S A*. 2009;106:15720–15725.
 93. Aoki T, Ohnishi H, Oda Y, Tadokoro M, Sasao M, Kato H, Hattori K, Ohgushi H. Generation of induced pluripotent stem cells from human adipose-derived stem cells without c-myc. *Tissue Engineering Part A*. 2010;16:2197–2206.
 94. Zhao HX, Li Y, Jin HF, Xie L, Liu C, Jiang F, Luo YN, Yin GW, Li Y, Wang J, Li LS, Yao YQ, Wang XH. Rapid and efficient reprogramming of human amnion-derived cells into pluripotency by three factors Oct4/Sox2/Nanog. *Differentiation*. 2010;80:123–129.
 95. Li W, Ding S. Generation of novel rat and human pluripotent stem cells by reprogramming and chemical approaches. In: Ding S, ed. *Cellular Programming and Reprogramming*. New York, Humana Press; 2010:293–300.
 96. Li W, Wei W, Zhu S, Zhu J, Shi Y, Lin T, Hao E, Hayek A, Deng H, Ding S. Generation of rat and human induced pluripotent stem cells by combining genetic reprogramming and chemical inhibitors. *Cell Stem Cell*. 2009;4:16–19.
 97. Carey BW, Markoulaki S, Hanna J, Saha K, Gao Q, Mitalipova M, Jaenisch R. Reprogramming of murine and human somatic cells using a single polycistronic vector. *Proc Natl Acad Sci U S A*. 2009;106:157–162.

Production of De Novo Cardiomyocytes: Human Pluripotent Stem Cell Differentiation and Direct Reprogramming

Paul W. Burridge,^{1,2} Gordon Keller,³ Joseph D. Gold,⁴ and Joseph C. Wu^{1,2,*}

¹Department of Medicine, Institute for Stem Cell Biology and Regenerative Medicine

²Cardiovascular Institute

Stanford University School of Medicine, Stanford, CA 94305, USA

³McEwen Centre for Regenerative Medicine, University Health Network, MaRS Centre, Toronto, ON MG5 1L7, Canada

⁴Neurobiology and Cell Therapies Research, Geron Corporation, Menlo Park, CA 94025, USA

*Correspondence: joewu@stanford.edu

DOI 10.1016/j.stem.2011.12.013

Cardiovascular disease is a leading cause of death worldwide. The limited capability of heart tissue to regenerate has prompted methodological developments for creating de novo cardiomyocytes, both in vitro and in vivo. Beyond uses in cell replacement therapy, patient-specific cardiomyocytes may find applications in drug testing, drug discovery, and disease modeling. Recently, approaches for generating cardiomyocytes have expanded to encompass three major sources of starting cells: human pluripotent stem cells (hPSCs), adult heart-derived cardiac progenitor cells (CPCs), and reprogrammed fibroblasts. We discuss state-of-the-art methods for generating de novo cardiomyocytes from hPSCs and reprogrammed fibroblasts, highlighting potential applications and future challenges.

Introduction

Heart disease is the most significant cause of morbidity and mortality in the United States, accounting for more than 800,000 deaths per year on average (equivalent to 1 death every 39 s) (Roger et al., 2011). Each year, nearly 6 million patients suffer heart failure and 1.25 million patients suffer a new or recurrent myocardial infarction. Both of these conditions result in cardiomyocyte death by apoptosis and/or necrosis. Dead cardiomyocytes are replaced by fibroblasts that divide and migrate into the damaged area to form scar tissue, leading to the development of a thin ventricular wall that no longer contracts properly. Formation of a fibroblastic scar initiates a series of events that lead to remodeling, hypertrophy, and ultimately heart failure and further cell death.

The persistence of scar tissue following myocardial infarction suggests that the heart has little if any capacity to generate new cardiomyocytes. Shortly after birth, myocardial growth transitions from a hyperplastic to a hypertrophic phase, characterized by the formation of binucleated cardiomyocytes that withdraw from the cell cycle (Pasumarthi and Field, 2002). This transition gave rise to the notion that adult cardiomyocytes are incapable of proliferating; that is, they are terminally differentiated.

Whereas the majority of adult cardiomyocytes do not proliferate, evidence exists indicating that the adult heart has limited regenerative capacity, although insufficient to compensate for the cell death caused by heart disease. Radiocarbon dating of postmortem cardiac tissue has demonstrated that human adult cardiomyocytes have a turnover rate of less than 1% per year and that 40% of the mature adult heart consists of postnatally generated cardiomyocytes (Bergmann et al., 2009). In addition, cardiac regeneration without scar tissue has been demonstrated in model organisms, such as zebrafish, after surgical removal of up to 20% of the ventricle (Poss et al., 2002). This regeneration

occurs through the proliferation and dedifferentiation of cardiomyocytes (Jopling et al., 2010), and similar regeneration has been demonstrated in mice, but only within the first week postpartum (Porrello et al., 2011). Work in adult mice provided evidence that limited numbers of new cardiomyocytes may be formed after a myocardial infarction, and lineage tracing experiments suggest these new cells are not derived from existing cardiomyocytes but rather develop from a progenitor population (Loffredo et al., 2011), some of which are derived from the epicardium (Smart et al., 2011). Finally, mononuclear rat cardiomyocytes have been shown to dedifferentiate and proliferate in vitro (Zhang et al., 2010). Growth factors from prenatal cardiac development, such as neuregulin1 (NRG1) (Bersell et al., 2009) and periostin (POSTN) (Kühn et al., 2007), can induce subpopulations of these cardiomyocytes to reenter the cell cycle in vitro. However, it is not clear to what extent this phenomenon occurs in vivo.

In clinical trials, transplantation of noncardiac stem cells such as skeletal muscle progenitors and bone-marrow-derived cells has resulted in minor improvement in left ventricular ejection fraction (LVEF) (Hansson et al., 2009), but also the induction of arrhythmias (Menasché et al., 2008). Because these transplanted populations do not generate new cardiomyocytes, the improvement may be due to paracrine effects that lead to enhanced vascularization (Gnecchi et al., 2005). The modest improvement provided by these nonmyocyte populations coupled with the limited regenerative capacity of the adult heart has led researchers to investigate new methods for the de novo derivation of cardiomyocytes. One potential source of these cells is the adult heart itself, in the form of CPCs, expertly reviewed elsewhere (Perino et al., 2008). With the discovery of human embryonic stem cells (hESCs) (Thomson et al., 1998), and more recently, human induced pluripotent stem cells (hiPSCs)

(Takahashi et al., 2007; Yu et al., 2007), many investigators have focused their efforts on developing strategies to efficiently and reliably direct stem cell differentiation to the cardiovascular lineage. Since the initial demonstration that contracting cardiomyocytes can be generated from both types of human pluripotent stem cells (hPSCs) (Itskovitz-Eldor et al., 2000; Zwi et al., 2009), the possibility of producing unlimited numbers of human cardiomyocytes to rebuild the heart has tantalized researchers. Substantial effort has been made to improve the efficiency and reproducibility of differentiation, while advancing the aims of progressing to defined conditions and producing cells on a clinically relevant scale. Advances in embryology and hPSC differentiation have offered key insights into the mechanisms of cardiopoiesis, providing hope that in the future injured hearts may be repaired through clinical applications of these cells.

A potential alternative source of de novo cardiomyocytes could result from the “direct reprogramming” of murine cardiac fibroblasts and other adult cell types into cardiomyocytes using cardiac-specific transcription factors (*Gata4*, *Mef2c*, and *Tbx5*) (Ieda et al., 2010). This approach offers the possibility of reprogramming cardiac fibroblasts in vivo for heart regeneration, using direct delivery of these transcription factors. A variation of the theme of reprogramming fibroblasts into cardiomyocytes has been recently described, in which fibroblasts are first partially reprogrammed using exogenous expression of pluripotency genes (*Oct4*, *Sox2*, and *Klf4*) (OSK) and then differentiated to the cardiac lineage (Efe et al., 2011).

Applications of De Novo Cardiomyocytes

One of the main long-term goals of de novo cardiomyocyte production is to provide a source of donor cardiomyocytes for cell replacement in damaged hearts. Many forms of heart disease, including congenital defects and acquired injuries, are irreversible because they are associated with the loss of nonregenerative, terminally differentiated cardiomyocytes. Current therapeutic regimes are palliative, and in the case of end-stage heart failure, transplantation remains the last resort. However, transplantation is limited by a severe shortage of both donor cells and organs. In cases of myocardial infarction, 1 billion cells would potentially need to be replaced (Laflamme and Murry, 2005), highlighting the need for high-throughput and reproducible methodologies for de novo cardiomyocyte production. A major challenge in this field is to establish the most efficient format for the transplantation of these substantial numbers of cells. Transplantation of single cell suspensions is easiest, but engraftment of 3D engineered constructs may be the best approach for replacing scar tissue with new working myocardium. In addition, concerns over cell survival, immune rejection, electrical maturation, electrical coupling, arrhythmia, and whether autologous hiPSCs possess immune privileges (a question that has recently been raised with murine iPSCs; Zhao et al., 2011) still need to be addressed.

A second application lies in novel cardiac drug discovery, development, and safety testing, a process that is collectively long, arduous, and expensive, and one which is confounded by the lack of economical and reliable methods to accurately mimic the human cardiac physiological response, among other challenges. Many drug discovery programs have failed because targets validated in animal models proved unreliable and non-

predictive in humans (Denning and Anderson, 2008). The pharmaceutical industry currently invests approximately \$1.5 billion to successfully develop a candidate drug from primary screening to market. Among the drugs that ultimately make it to market, many are later withdrawn due to side effects associated with electrophysiological alterations of the heart (Braam et al., 2010). The use of de novo human cardiomyocytes offers the pharmaceutical industry an invaluable tool for pre-clinical screening of candidate drugs to treat cardiomyopathy, arrhythmia, and heart failure, as well as therapeutics to combat secondary cardiac toxicities. Studies have already demonstrated that hiPSC-derived cardiomyocytes will react to cardioactive drugs with the expected response, indicating that these cells can be used in the context of larger predictive toxicology screens (Davis et al., 2011). The development of new screens using human cardiomyocytes should reduce the time and cost of bringing new drugs to market.

A third application is in developmental biology, disease modeling, and postgenomic personalized medicine. Deriving hiPSCs from patients with specific cardiac diseases, differentiating them to cardiomyocytes, and then performing electrophysiological and molecular analyses may provide a powerful tool for deciphering the molecular mechanisms of disease (Josowitz et al., 2011). Studies to date have largely concentrated on recapitulating genetic disease phenotypes in vitro, such as long QT syndromes (Itzhaki et al., 2011a; Matsa et al., 2011; Moretti et al., 2010), Timothy syndrome (Yazawa et al., 2011), and LEOPARD syndrome (Carvajal-Vergara et al., 2010). The possibility of modeling cardiac diseases without a known genetic element is another exciting prospect. The combination of novel drug discovery and efficacy testing with cardiomyocytes derived from patient-specific hiPSCs is a potentially groundbreaking option for personalized medicine. Additionally, the utility of pluripotent stem cells as a tool for modeling cardiac development is another important application that has recently been demonstrated for a variety of aneuploid syndromes (Li et al., 2012).

Cardiomyogenesis

To effectively serve the above applications, efficient and reproducible generation of cardiovascular cells in vitro must be developed. Because hPSC differentiation recapitulates embryonic development, understanding how the cardiac lineage is established in the early embryo is essential for differentiation and development strategies. Understanding cardiogenesis also allows access to the feedforward gene regulatory networks that occur during development and ultimately deriving physiologically relevant cells. Cardiomyogenesis begins with the generation of mesoderm via the process of gastrulation, which has been best studied in the mouse (Arnold and Robertson, 2009; Buckingham et al., 2005; Tam and Loebel, 2007) (Figure 1). Mesoderm induction begins with NODAL signaling in the proximal epiblast on mouse embryonic day 5 (E5.0), which maintains BMP4 expression in the extraembryonic ectoderm adjacent to the epiblast. BMP4 acts by inducing WNT3 expression in the proximal epiblast. Around E5.5, expression of the WNT antagonist *Dkk1* and the NODAL antagonists *Lefty1* and *Cer1* moves to the anterior visceral endoderm, restricting NODAL and WNT signaling to the posterior epiblast. At E5.75, WNT induces the

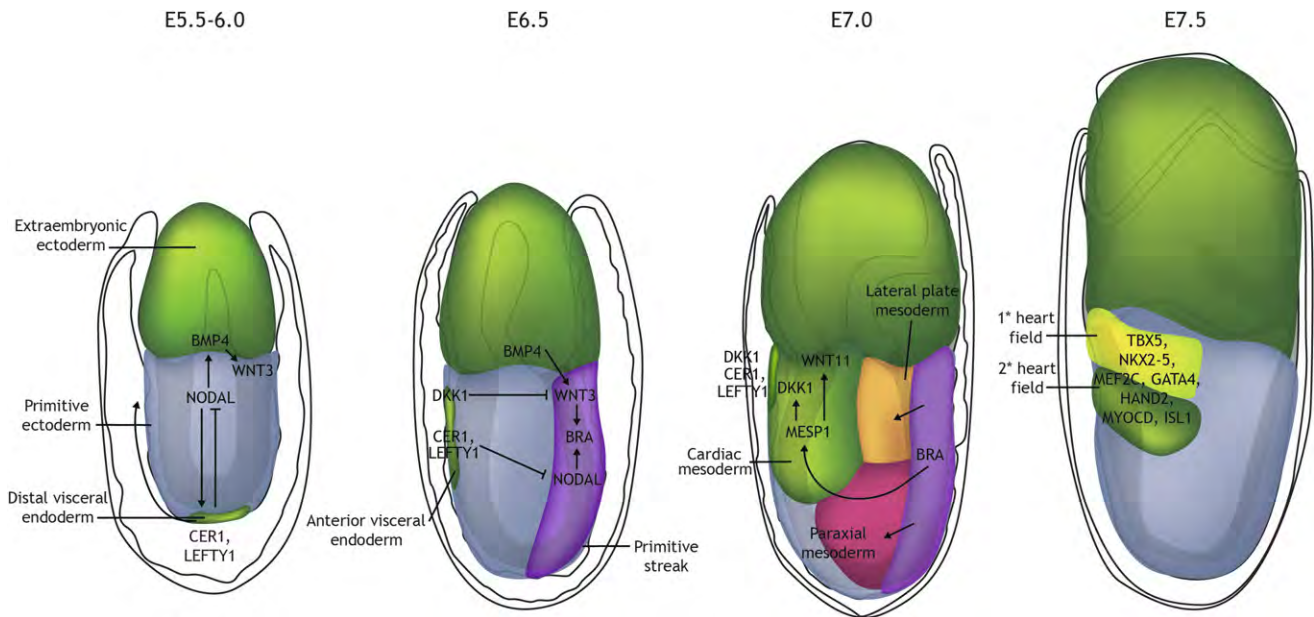


Figure 1. The Mouse as a Model for Human Cardiogenesis

(A) Gastrulation begins with the formation of the primitive streak. Anterior primitive ectoderm cells perform an epithelial to mesenchymal transition (EMT), pass through the primitive streak, and move laterally between the primitive ectoderm and visceral endoderm.
 (B) Cells in the most proximal portion of the primitive streak will form the extraembryonic mesoderm, midprimitive streak cells will form the embryonic tissues such as heart and blood, and cells of the distal portion of the primitive streak will form the endoderm.
 (C) In the case of cardiac progenitors, this midstreak mesoderm progresses laterally, then ventrally, around both sides of the embryo, becoming the lateral plate mesoderm from which the first heart field (FHF) is derived. The lateral plate mesoderm then delaminates to form the splanchnic mesoderm (on the ventral side), which forms the second heart field (SHF) and somatic mesoderm (on the dorsal side).
 (D) The FHF progenitors form the cardiac crescent, whereas SHF progenitors are found medial to the crescent. The cells of both heart fields then move to the midline, where the FHF progenitors form a linear heart tube that later contributes to the left ventricle. Cells of the SHF proliferate, migrate, and join with the cardiomyocytes of the FHF, resulting in the rightward looping of the cardiac tube, a process that culminates in a segmented structure and the formation of cardiac chambers.

expression of mesoendodermal markers such as *T (brachyury)* and *Eomes*. Subsequently, genes involved in mesoderm patterning and epithelial-mesenchymal transitions (EMTs), such as *Fgf4* and *Fgf8*, are expressed in the developing primitive streak. T and EOMES then induce the expression of MESP1 (Costello et al., 2011; David et al., 2011), which has been described as the “master regulator” of cardiac progenitor specification (Bondué et al., 2008).

MESP1 can drive cardiac differentiation via the DKK1-mediated inhibition of WNT signaling (David et al., 2008). It is now clear that canonical WNT/ β -catenin signaling has a biphasic effect on cardiogenesis (Naito et al., 2006; Ueno et al., 2007). Downstream of MESP1, cardiogenesis relies on a complex web of interacting factors and genes, including *Tbx5*, *Nkx2-5*, *Mef2c*, *Gata4*, *Hand2*, *Myocd*, *Isl1*, and *Foxh1* (Bondué and Blanpain, 2010). Once cardiogenic mesoderm is specified, canonical WNT and NOTCH signaling regulate CPC maintenance and differentiation, respectively (Qyang et al., 2007; Kwon et al., 2009).

The cardiac mesoderm gives rise to the endocardium, the first heart field (FHF, which forms the atria, left ventricle, and the nodal conduction system), the secondary heart field (SHF, which forms the right ventricle, outflow tract, and part of the atria), and the proepicardial mesenchyme (Buckingham et al., 2005). The FHF differentiates at the cardiac crescent stage, whereas the SHF (marked by the expression of *Isl1*) remains in an undiffer-

entiated progenitor state, due to inhibitory WNT signals from the midline, until incorporation into the heart (Kwon et al., 2007). Once the cardiac crescent is formed, it is exposed to BMP signaling from the underlying anterior ectoderm as well as to BMP, FGF, anticanonical WNT, and noncanonical WNT signaling from the overlying anterior endoderm (Solloway and Harvey, 2003). By E8.0, these cells form a primitive heart tube, consisting of an interior layer of endocardial cells and an exterior layer of myocardial cells. Once within the heart, FHF and SHF cells appear to proliferate in response to endocardial-derived signals, such as neuregulin1 (NRG1), which is driven by NOTCH signaling (Grego-Bessa et al., 2007), and epicardial signals, such as retinoic acid, that function via FGF signaling (Lavine et al., 2005). Further study of the possibility of maintaining and proliferating cardiac progenitors, along with the signaling that differentiates atrial from ventricular cells and decoupling this from signaling involved with morphological movement, is of great interest to the further development of reproducible de novo cardiomyocyte generation methodologies.

Cardiac Differentiation of hPSCs

The ability to differentiate hPSCs in a directed manner has progressed considerably in the past 10 years. The most reproducible and efficient strategies involve stage-specific activation and inhibition of different signaling pathways in defined culture conditions, recapitulating key steps in cardiac development

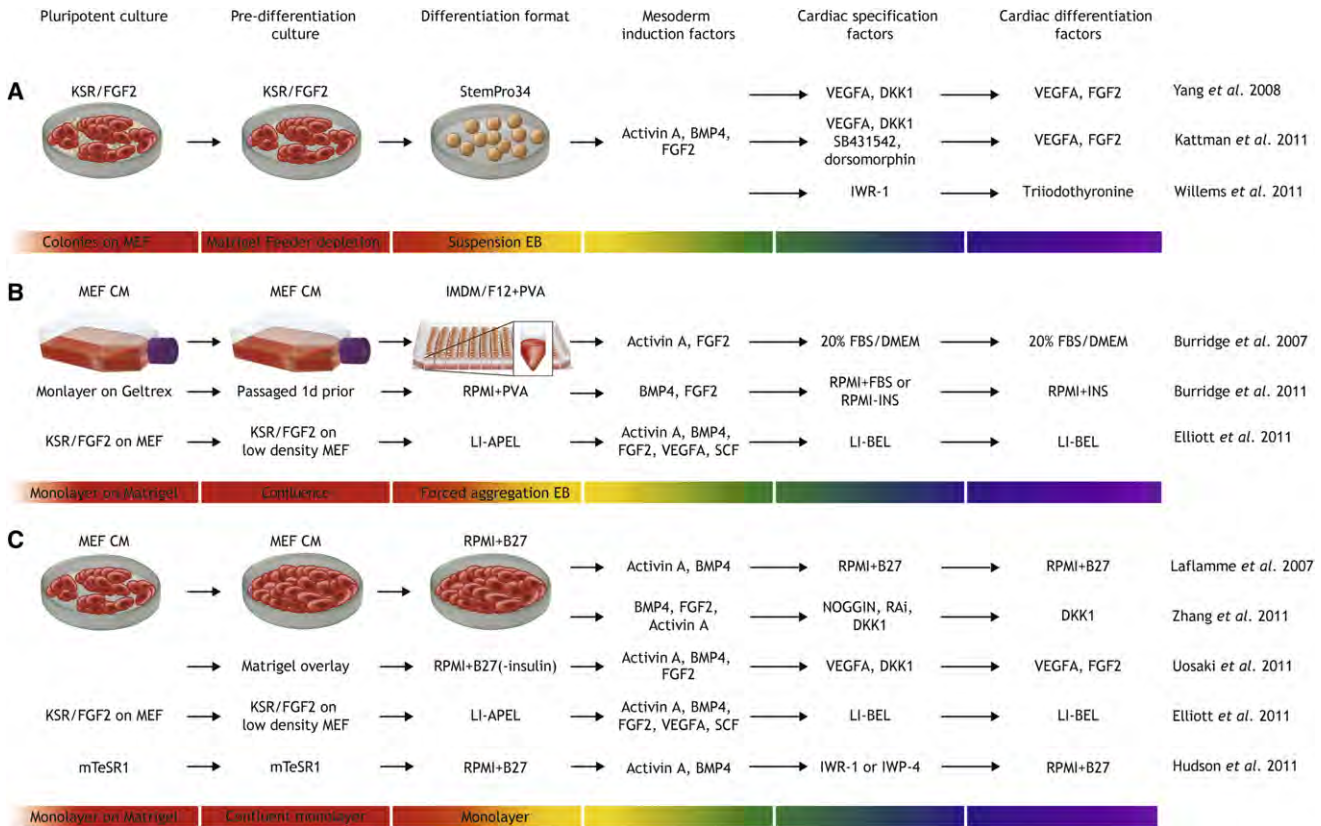


Figure 2. Methods for the Differentiation of Human Pluripotent Stem Cells

Three methods for differentiating hPSCs, highlighting commonalities at each of the six steps: pluripotent culture, predifferentiation culture, differentiation format, and treatment with mesoderm induction factors, cardiac specification factors, and cardiac differentiation factors. (A), Yang et al. (2008) suspension EBs in StemPro34; (B), Burridge et al. (2007) forced aggregation; (C), Laflamme et al. (2007) monolayer differentiation. Abbreviations: KSR, Knockout Serum Replacement; FGF2, fibroblast growth factor 2; StemPro34, proprietary medium from Invitrogen; BMP4, bone morphogenetic protein 4; VEGFA, vascular endothelial growth factor A; DKK1, dickkopf homolog 1; SB431542, TGF β /activin/NODAL signaling inhibitor (ALK4,5,7); dorsomorphin, BMP signaling inhibitor (ALK2,3,6); IWR-1, WNT signaling inhibitor; MEF CM, mouse embryonic fibroblast conditioned hESC medium; IMDM/F12+PVA, IMDM/F12-based media supplemented with polyvinyl alcohol; RPMI, Roswell Park Memorial Institute 1640 basal medium; FBS, fetal bovine serum; DMEM, basal media; RPMI+PVA, RPMI-based media supplemented with polyvinyl alcohol; RPMI-INS, RPMI-based media without insulin; B27, media supplement; NOGGIN, BMP signaling inhibitor; RAI, retinoic acid signaling inhibitor; LI-APEL, low insulin, AlbuCult, polyvinyl alcohol, essential lipids media; SCF, stem cell factor (KITLG); LI-BEL, low insulin, bovine serum albumin, essential lipids media; IWP-4, WNT signaling inhibitor.

in the early embryo. There is consensus that the cardiac differentiation process is very delicate, and the variability in each individual component of the cardiac differentiation strategy must be carefully optimized to reliably produce cardiomyocytes. One of the first directed differentiation methods involved hESCs cocultured with mouse visceral endoderm-like cells (END-2) (Mummery et al., 2003), which is relatively inefficient but has been shown to generate mostly (~85%) ventricular-like cardiomyocytes (Mummery et al., 2003). This protocol provided early insight into methods for improving differentiation efficiency, such as removing fetal bovine serum (FBS), adding L-ascorbic acid (Passier et al., 2005), and removing insulin between d0 and d4 (Freund et al., 2008). Two basic methods for the cardiac differentiation of hPSCs are now widely in use: the formation of embryoid bodies (EBs), and culturing hPSCs as a monolayer. Mature versions of each method rely on progressive sequential inductive environments using growth factors and/or small molecules. Figure 2 summarizes the principles of the reported strategies used for cardiomyocyte differentiation from hPSCs.

The EB methodology initially involved suspending hPSC colonies in media containing 20% FBS to form spherical aggregates (Kehat et al., 2001). This technique produces differentiated cell types of all three germ layers, including cardiomyocytes, when it is followed by adhesion at a later stage (~d4–d7). Although inefficient, nearly all cell lines (including hiPSCs) form some rhythmically contracting outgrowths in the 5%–15% range (Zwi et al., 2009). This protocol can be enhanced by adding BMP4 during d0–d4 (Takei et al., 2009), by the addition of WNT3A during d0–d2 (Tran et al., 2009), and via small molecule inhibition of WNT signaling by IWR-1 during d4–d6 (Ren et al., 2011). FBS- and insulin-free media can support EB cardiac differentiation, the efficiency of which can be enhanced by adding prostaglandin I2 and the MAP kinase inhibitor SB203580 (Xu et al., 2008), specifically between d4 and d6 (Gaur et al., 2010). Factors such as WNT3A (Bu et al., 2009), G-CSF (Shimoji et al., 2010), and L-ascorbic acid (Cao et al., 2011) can improve cardiac differentiation by enhancing CPC proliferation, whereas IGF1 and IGF2 can enhance hPSC-derived cardiomyocyte proliferation (McDevitt et al., 2005).

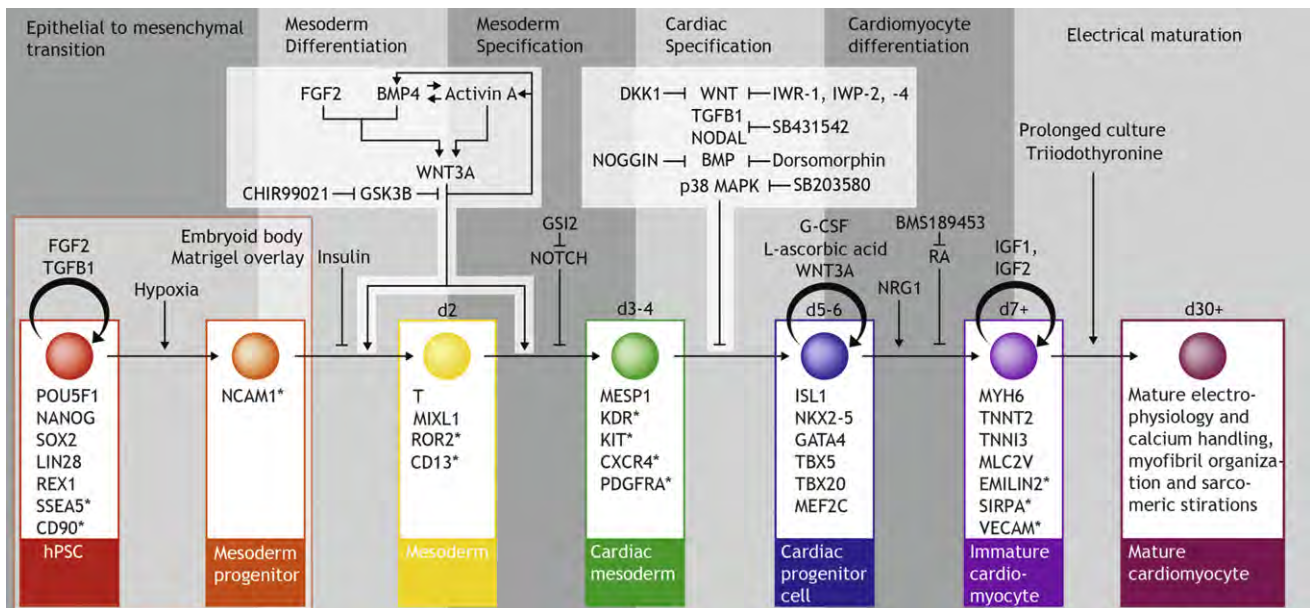


Figure 3. Schematic of Current Knowledge of Factors Involved in hPSC Cardiac Differentiation

Factors that influence the progression through each of the six major steps of hPSC cardiomyogenesis: epithelial to mesenchymal transition, mesoderm differentiation, mesoderm specification, cardiac specification, cardiomyocyte differentiation, and electrical maturation. Data shown are derived from developmental biology models that have been directly assessed and proved functional in hPSC cardiac differentiation, along with knowledge gained directly from hPSC differentiation. Below are the markers associated with each of the seven cell types during differentiation; surface markers are marked with an asterisk.

A serum-free, EB-based suspension technique developed in the Keller Lab (Figure 2A) has proven highly effective for analyzing growth factor variables and timing during cardiac differentiation (Kattman et al., 2011; Yang et al., 2008). In this protocol, EBs are formed in the presence of a low level of BMP4 and treated with optimized levels of BMP4, fibroblast growth factor 2 (FGF2), and activin A between d1 and d4, VEGFA and DKK1 during d4 to d8, and VEGF and FGF2 from d8 onward. EBs are maintained under hypoxic (5% oxygen) conditions for 12 days. Using this method, approximately 70% of HES2 EBs spontaneously contract (Yang et al., 2008). Inhibition of TGFβ/activin/NODAL and BMP signaling using the small molecules SB431542 and dorsomorphin during d3–d5 (Kattman et al., 2011) or the WNT inhibitor IWR-1 (d4 to 10) (Willems et al., 2011) has been shown to enhance this system. To overcome the limitation of EB size heterogeneity, a “forced aggregation” protocol using V- or U-shaped 96-well plates has been employed (Figure 2B). Initial versions of this system using activin A and FGF2 produced ~23% beating EBs from four different hESC lines (BurrIDGE et al., 2007). Further developments of this system, using BMP4 and FGF2 as mesoderm inducers (BurrIDGE et al., 2011) under hypoxia, increased differentiation efficiency to ~94% of EB beating in a wide range of hESC and hiPSC lines. Elliott and colleagues (Elliott et al., 2011) demonstrated a similarly high differentiation efficiency by employing a forced aggregation protocol that used BMP4, activin A, WNT3A, KITLG (SCF), and VEGFA for the first 3 days.

The suspension EB and forced aggregation methods are mature and commonly produce >70% cardiomyocytes, but these methods are technically complex and time consuming, which has led to the development of a monolayer-based method (Figure 2C) (Laflamme et al., 2007), where hPSCs are cultured to

a high density, then treated with a high dose of activin A, followed by 4 days of treatment with BMP4; crucially, the media is not changed during these 4 days to allow the secreted factors to accumulate. Contracting cells can be seen at day 12, with a purity of approximately 30% cardiomyocytes. Improvements to this system, such as the addition of WNT3A at d0–d1 and DKK1 from d5 to d11, have been demonstrated (Paige et al., 2010). Further improvements have also been made using a Matrigel overlay 1 day prior to differentiation to enhance EMT, along with the removal of insulin and addition of FGF2 from d1 to d5 and DKK1 from d5 to d7 (Uosaki et al., 2011). As with the EB system, the use of the WNT signaling inhibitors IWR-1 or IWP-4 has proven effective, in this case when applied from d3 (Hudson et al., 2011).

One important aspect of these protocols is the required identification of markers to monitor the efficiency and progress of differentiation. Although assays that measure the number of beating EBs and expression of cardiac troponin T (TNNT2) have proven effective for optimizing differentiation efficiency, more rigorous definition of the growth factor regime by analyzing cardiac mesoderm populations, such as the KDR+/PDGFRA– population (Kattman et al., 2011) and a genetically modified NKX2-5-GFP cell line, to identify the CPC population (Elliott et al., 2011) has proven effective for optimizing the ratio of activin A and BMP4 during early differentiation. The further identification of specific markers expressed during cardiac differentiation is a high priority.

Collectively, these studies demonstrate that the importance of exposure of hPSCs to various growth factors at specific times and in specific doses is essential for directing differentiation from early mesendoderm via mesoderm toward a more specific cardiac fate (Figure 3). Data collected from experiments using

both the EB and monolayer methodologies show that four major signaling pathways are involved in early cardiac differentiation of hPSCs—BMP, TGFB/activin/NODAL, WNT, and FGF—with highly specific temporal windows for effectiveness. There is also significant evidence that synergistic relationships exist between growth factors such as BMP4 and FGF2 (Yu et al., 2011). As with cardiac development in vivo, it is now clear that inhibition of WNT as well as BMP and TGFB/activin/NODAL during the mid-differentiation stage has a major role in hPSC cardiomyogenesis (Ren et al., 2011; Willems et al., 2011; Kattman et al., 2011; Zhang et al., 2011). Although less well studied, Notch inhibition using gamma-secretase inhibitor II from d0 to d8 has also been demonstrated to affect hPSC cardiac mesoderm induction (Jang et al., 2008), indicating possible involvement of a fifth signaling pathway.

The Developmental Status of hPSC-Derived Cardiomyocytes

The cardiomyocytes produced to date from hPSCs are largely immature and most analogous to fetal stages of development; these hPSC-derived cardiomyocytes exhibit automaticity (spontaneous contraction), fetal-type ion channel expression (Beqqali et al., 2006), fetal-type electrophysiological signals (Davis et al., 2011), fetal-type gene expression patterns, and fetal-type physical phenotypes (Cao et al., 2008). Conflicting data exist regarding the maturity of Ca²⁺ handling and sarcoplasmic reticulum status for hPSC-derived cardiomyocytes, although there is evidence that hPSC-derived cardiomyocytes demonstrate some mature properties (Itzhaki et al., 2011b). The lack of maturity of hPSC-derived cardiomyocytes may reduce the suitability of drug testing but may also have benefits in regards to regenerative medicine. For example, rodent fetal cardiomyocytes have been demonstrated to have enhanced cell survival over adult cardiomyocytes after engraftment into the rat heart (Reinecke et al., 1999), although the issue of automaticity still remains. Three major subtypes of hPSC-derived cardiomyocytes can be derived that have atrial-, ventricular-, or nodal-like phenotypes as determined by electrophysiological analysis of action potentials (APs). Common hPSC differentiation methodologies create a mixture of these cell types. An enriched population of nodal-like cells could potentially be used in the formation of a biological pacemaker, whereas ventricular types may be used for recovery from myocardial infarction. Zhu and colleagues demonstrated that GATA6-GFP can mark nodal-like cells (Zhu et al., 2010) and that inhibition of NRG1β/ERBB signaling can enhance the population of nodal-like cardiomyocytes. Zhang and colleagues have similarly demonstrated that retinoic acid can increase the proportion of atrial-like cardiomyocytes and that retinoic acid inhibition can increase the proportion of ventricular-like cells (Zhang et al., 2011).

Overcoming Interline Variability in Cardiac Differentiation of hiPSCs

The diversity in cardiac differentiation capacities between hESC lines has now been well established (Burridge et al., 2007; Kattman et al., 2011), and it is possible that hiPSCs may have more differentiation variability than hESCs. Many factors likely contribute to this line-to-line variation, including the conditions used to establish and maintain lines, the efficiency of reprogramming, the cell source used for reprogramming, and

levels of expression of endogenous growth factors. Maintenance of high-quality hPSCs is essential for efficient and reproducible differentiation; the level of expression of pluripotency factors such as *POU5F1* (*OCT4*) and *NANOG* has been demonstrated to have a role in mesoderm induction (Yu et al., 2011). The progress toward a defined culture of hPSCs, although initially focused on eliminating feeder layers and animal serum products to allow future clinical translation, has resulted in higher levels of control over the pluripotent state, which in turn enhances the reproducibility of differentiation and reduces costs for a future large-scale culture. hESCs were first derived on a feeder layer of mouse embryonic fibroblasts (MEFs) in media containing FBS (Thomson et al., 1998). Substantial progress has since been made in controlling hPSC culture, through second-generation media containing the commercial product knockout serum replacement (KSR) and FGF2 (Amit et al., 2000). In addition, KSR/FGF2 media conditioned on MEFs have been shown to support hPSC pluripotency on a Matrigel layer (Xu et al., 2001). Progress in elucidating a role for high levels of FGF2 along with TGF-β1 to stimulate the PI3K, MEK/ERK, and SMAD2/3 pathways has resulted in a third-generation “defined” media in which KSR is replaced with BSA or human serum albumin (HSA). Despite some difficulties in implementation (Akopian et al., 2010), several formulas have achieved widespread use, including XVIVO10 supplemented with FGF2 and TGF-β1 (Li et al., 2005) and the commercially available mTeSR1 (Ludwig et al., 2006) and Stem-Pro hESC (Wang et al., 2007). Most recently, Chen and colleagues achieved a breakthrough in hESC and hiPSC culture by formulating a fourth-generation, chemically defined medium (referred to as E8) that contains the growth factors FGF2 and TGF-β1 while lacking HSA and the antioxidant 2-mercaptoethanol (Chen et al., 2011). This medium, when used in combination with a recombinant truncated variant of vitronectin, has proved successful in deriving new hiPSC lines. Although progress toward a chemically defined culture of hPSCs has been slow and gradual, this simple and elegant solution has improved the reproducibility of pluripotent cell culture and overcomes a significant hurdle in the development of a reproducible cardiac differentiation system.

In addition to culture conditions, the considerable transcript heterogeneity within hiPSCs of a single cell line (Narsinh et al., 2011) and the hot spots of aberrant epigenomic reprogramming (Lister et al., 2011) likely contribute to variability in differentiation potential. Another source of variability for hiPSCs is cell type of origin (Martinez-Fernandez et al., 2011). It has been well documented that certain cell types, such as neural stem cells (Kim et al., 2009), can be more easily reprogrammed, but whether this tendency is due to a more malleable epigenetic state, the existing expression of a subset of pluripotency-associated genes, or the activity of epigenetic-state-modifying genes is still unclear.

The retention of epigenetic memory of cell origin (Kim et al., 2010b) was initially thought to be restricted to low passage (<15) hiPSCs (Polo et al., 2010), but recent data suggest that this effect is maintained in later passages as well (Ohi et al., 2011). This issue becomes more complex given that hPSCs progressively acquire epigenetic heterogeneity after prolonged culture, which affects subsequent differentiation (Tanasijevic et al., 2009), and that different passages of the same cell line have different cardiac potentials (Paige et al., 2010). Another

question that has not been addressed in human cells is whether the method used to induce pluripotency influences subsequent differentiation. It has been demonstrated in mouse iPSCs that *c-Myc*-independent (OSK) conditions favor cardiogenic potential of iPSCs (Martinez-Fernandez et al., 2010), and it has also been shown that hiPSCs generated under different conditions have differing cell cycles and DNA replication gene profiles (Chung et al., 2011). The effects of generating hiPSCs using defined media, matrices, and small-molecule epigenetic modifiers on subsequent differentiation have yet to be fully explored. Differences in levels of expression of endogenous signaling cytokines such as *NODAL*, *BMP4*, and *WNT3A* (Kattman et al., 2011; Paige et al., 2010) clearly contribute to the variability of cardiovascular development. Sustained TGFB/activin/*NODAL*, *BMP*, or *WNT* signaling beyond the cardiac mesoderm step can have profound inhibitory effects on the generation of contracting cardiomyocytes within the culture (Kattman et al., 2011). These differences are not necessarily related to reprogramming, as they are also observed between different hESC lines. While the underlying causes of these differences are not known, the development of quantitative assays to rapidly measure the levels of endogenous signaling would enable one to appropriately modify the induction protocols for each cell line.

Direct Reprogramming and Partial Reprogramming

It is well established that exogenous expression of *Myod* can convert murine fibroblasts directly to skeletal muscle (Davis et al., 1987). This capability was also demonstrated in a variety of other nonfibroblast cell types (Choi et al., 1990). This process is comparatively simple because *Myod* acts as a master regulator of skeletal muscle formation (Wang et al., 2003), whereas it would appear that no such single gene exists for the direct reprogramming of cardiomyocytes. It has been demonstrated that exogenous expression of *Gata4*, *Tbx5*, and *Baf60c* (a subunit of the Swi/Snf-like BAF chromatin remodeling complex) was sufficient to reprogram noncardiogenic mesoderm into beating cardiomyocytes (Takeuchi and Bruneau, 2009). In this case, *Gata4* and *Baf60c* induce *Nkx2-5* expression, which acts with *Gata4* to initiate the cardiac program, with *Tbx5* shown to be required for full cardiomyocyte differentiation. Surprisingly, none of the expected master regulators of cardiac differentiation, including *Mesp1*, *Nkx2-5*, and *Isl1*, were required for this reprogramming.

Work on reprogramming to pluripotency has provided new insight into direct reprogramming, combining high-expression retroviral vectors with the subtractive assessment strategy (Takahashi and Yamanaka, 2006). Implementing these advances, Ieda and colleagues successfully reprogrammed cardiac fibroblasts into cardiomyocyte-like cells (referred to as iCMs) via the exogenous expression of *Gata4*, *Mef2c*, and *Tbx5* (Ieda et al., 2010) (Figure 4). These investigators began with a selection of 14 key genes related to cardiac development, including transcription factors and epigenetic remodeling factors, and expressed them in cardiac fibroblasts isolated from neonatal hearts obtained from α MHC-GFP transgenic mice. Seven days following transduction, 1.7% of cells expressed GFP. By serial reduction of the 14 factors, it was found that the optimal combination of *Gata4*, *Mef2c*, and *Tbx5* resulted in GFP expression in ~20% of the cells. Approximately 6%

of cells were positive for both GFP and TNNT2 (a marker of the sarcomere structure of cardiomyocytes), although the percentage of TNNT2+ cells increased over an additional 3 weeks of culture. Ieda and colleagues went on to demonstrate that this reprogramming was not achieved via an ISL1+ CPC or pluripotent intermediate, but was the result of gradual loss of fibroblast identity and progressive upregulation of cardiomyocyte-specific genes. iCMs were induced in vitro within 3 days, although the reprogramming factors were shown to be required for 2 weeks for stable reprogramming. The iCMs produced electrophysiological and gene expression profiles similar to those of fetal cardiomyocytes, although only 30% of the iCMs (approximately 6% of the total cell number) exhibited spontaneous contraction. The GMT cocktail was also shown to be able to reprogram adult tail-tip dermal fibroblasts to iCMs to a similar efficiency, albeit with lower expression of TNNT2 and no demonstration of spontaneous contraction or electrophysiological properties. Transplantation of neonatal cardiac fibroblasts into mouse hearts 1 day after reprogramming showed that reprogramming can be achieved in vivo.

However, the main drawbacks of direct reprogramming are the low efficiency rates (~1 in 20 fibroblasts are successfully reprogrammed) at present, the use of randomly integrating retroviruses (which preclude future clinical application), the experimental complexity (which will make the GMP production process challenging), and the potential for contamination by endogenous cardiomyocytes; most importantly, use of this methodology has not yet been independently verified or demonstrated in human cells to date. Reprogramming cardiac fibroblasts into cardiac cells based on isolation of GFP-/Thy1+ cardiac fibroblasts from the heart may not necessarily be stringent enough; fluorescence-activated cell sorting (FACS) does not remove 100% of cardiac cells, and the retroviruses may still transfect even nondividing cardiac cells. The relationship between the properties of iCMs and genuine cardiomyocytes has also not been fully established using more stringent criteria. Finally, definitive proof of this methodology lies in reprogramming skin fibroblasts (and various other cell types) into cardiac cells, as can be done reproducibly for the reprogramming of various adult somatic cell types to pluripotency (Yu et al., 2007; Takahashi et al., 2007).

Partial Reprogramming

An alternative method to generate differentiated cell types from fibroblasts is to create partially reprogrammed cells that can be differentiated to the desired cell type. The potential advantage of this approach over first establishing iPSCs and then differentiating them into particular cell lineages is one of speed. This approach has been used successfully by Efe and colleagues, who showed that murine fibroblasts can be reprogrammed to cardiomyocytes using the established OSK combination (with or without *c-Myc*) in 11–12 days (Efe et al., 2011) (Figure 4). Using Nebulette-LacZ (a cardiac-muscle-specific marker) in transgenic MEFs, they carried out reprogramming under tissue culture conditions designed for cardiac differentiation rather than iPSC generation. To prevent the cells from reaching a pluripotent state, an inhibitor of Janus kinase/signal transducer and activator of transcription (JAK/STAT) was included in cultures. Cardiomyocytes were efficiently

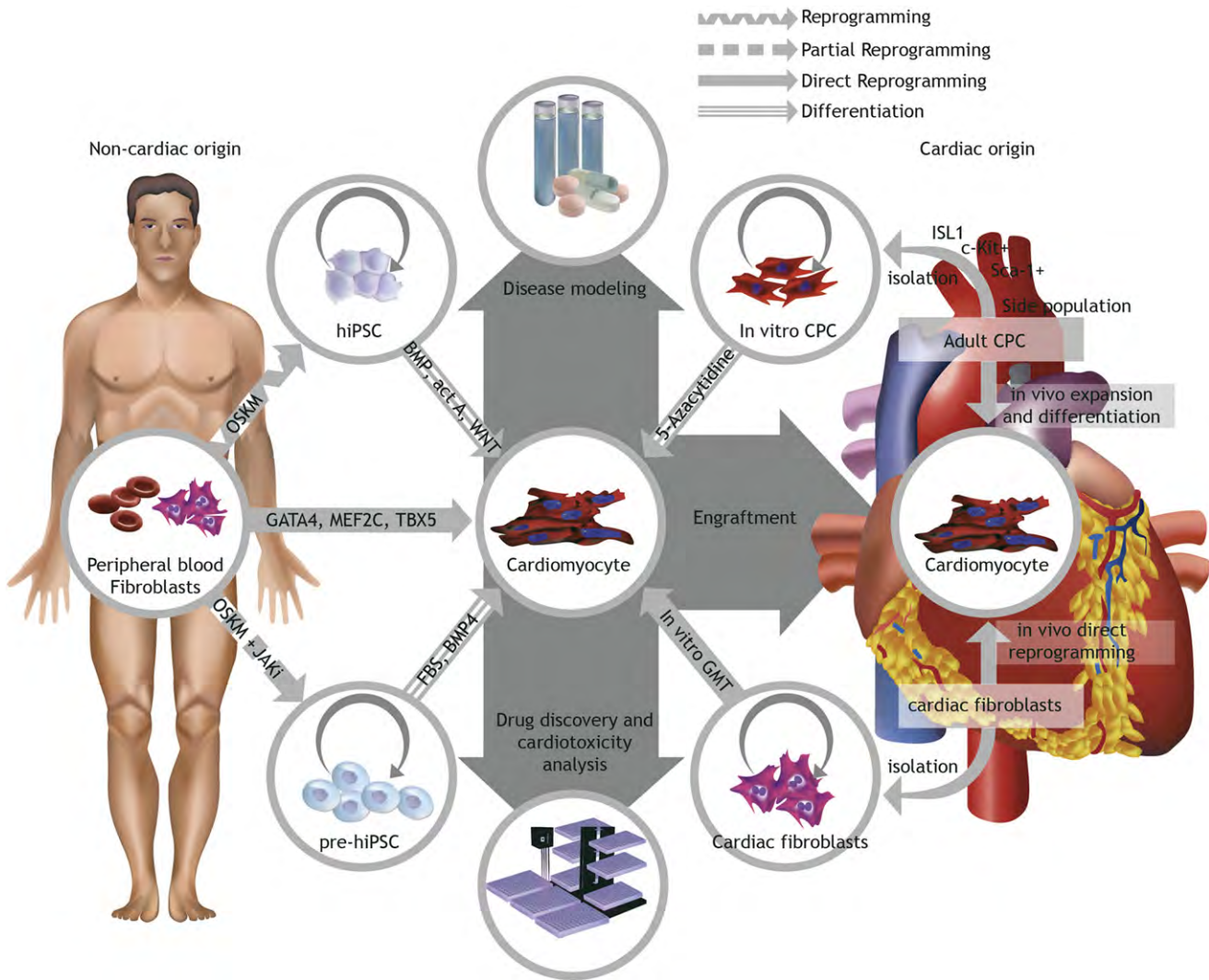


Figure 4. Schematic of De Novo Cardiomyocyte Generation and Applications

Demonstrating peripheral blood and skin fibroblasts as a cell source, hiPSC derivation and differentiation, direct reprogramming using GMT, and partial reprogramming using OSKM and a full reprogramming inhibitor. Applications in disease modeling, engraftment into the heart, drug discovery, and cardiotoxicity analysis are shown. Adult heart sources of cardiomyocytes from cardiac progenitor cells and application for in vivo expansion and differentiation are also shown, as is direct reprogramming of adult cardiac fibroblasts in vivo. Abbreviations: OSKM, *Oct4*, *Sox2*, *Klf4*, *c-Myc*; JAKi, JAK/STAT inhibitor; GMT, *Gata4*, *Mef2c*, *Tbx5*; CPC, cardiac progenitor cell.

produced and demonstrated typical gene expression profiles and calcium transients. A similar partial reprogramming protocol has been used by the same lab to generate neural progenitors (Kim et al., 2011). This technique still needs to be demonstrated in human cells, wherein an inhibitor suitable for replacing the pluripotency prevention function of JAK inhibitor would need to be used due to the differences in mechanisms that control murine and human pluripotency.

Both the direct reprogramming and partial reprogramming approaches offer exciting prospects for the future. Both methods may be entirely suitable to provide the small numbers of cells required for individualized drug screening and disease modeling. The direct reprogramming method could potentially be applied in vivo and would avoid the danger of residual pluripotent cells, thereby eliminating the risk of teratoma formation.

However, direct in vivo reprogramming could also lead to ectopic myocardium in scar tissue, which is a known risk factor for arrhythmias after infarction in humans and requires electrical ablation by a cardiac electrophysiologist. Another major concern is what the in vivo consequence would be for cells that do not receive the required dose of the GMT cocktail and are only partially reprogrammed to iCMs. Finally, in both systems, the problem of viral genomic integration and genomic mutations due to the use of retroviruses, along with issues of epigenetic status and scalability, must be overcome.

Cardiomyocyte Purification

Despite considerable progress in improving the efficiency of cardiac differentiation, the isolation of cardiomyocytes or the removal of unwanted cell populations may be required.

Traditional methods for the purification of cardiomyocytes that involve density gradient centrifugation (Xu et al., 2002) are unsuitable for large-scale practice and routinely result in only a 5- to 10-fold enrichment in cardiomyocyte populations (Murry and Keller, 2008). Genetic selection of cells based on the expression of a selectable marker driven by a lineage-restricted promoter such as NKX2-5 (Elliott et al., 2011), MYH6 (Anderson et al., 2007), MLC2V (Huber et al., 2007), and ISL1 (Bu et al., 2009) offers the possibility of isolating myocardial precursors at high purity; when this process is coupled with antibiotic selection, it is possible to generate cells at greater than 99% purity. Indeed, it has been demonstrated that an MYH6-blastocidin hiPSC line can be differentiated using simple suspension culture and 10% FBS in DMEM to produce hiPSC-derived cardiomyocytes in commercial quantities (Ma et al., 2011). The principal drawback of genetic selection is the necessity (at present) of inserting a selection cassette into the host genome, which may increase the risk of tumorigenesis and is therefore unsuitable for clinical practice.

Antibodies to cell surface markers have the advantage of not requiring genetic modification of stem cell populations, and therefore may be applicable to all hPSC lines. FACS has the ability to analyze multiple surface markers simultaneously, and it has been used to isolate cardiac progenitor populations based on the expression of the receptor tyrosine kinases KDR (FLK1/VEGFR2) and PDGFRA (Kattman et al., 2011). Because this population also contains endothelial and smooth muscle populations, the differentiated progeny represents a mixture of cardiomyocyte and vascular lineages. The recent identification of markers expressed specifically on cardiomyocytes, including EMILIN2 (Van Hoof et al., 2010), SIRPA (Dubois et al., 2011; Elliott et al., 2011), and VCAM (Elliott et al., 2011; Uosaki et al., 2011), has made it possible to isolate highly enriched populations of these cells from hESCs or hiPSCs by FACS or magnetic bead sorting.

Another nongenetic method for isolating hPSC-derived cardiomyocytes is based on the use of the mitochondrial dye tetramethylrhodamine methyl ester perchlorate (TMRM) (Hattori et al., 2010). Because this dye only functions in cardiomyocytes with high mitochondrial density, it does not detect the most immature cells that develop in the cultures (Dubois et al., 2011). Collectively, these recent discoveries have provided tools and reagents for the generation of highly enriched populations of hPSC-derived cardiomyocytes.

Future Directions and Challenges

The field is advancing quickly and we are already witnessing early applications of hPSCs in modeling human disease, understanding human biology, and developing platforms for drug discovery and predictive toxicology. Challenges remain before these applications can be widely used to develop regenerative medicine strategies for treating cardiovascular disease. The following section highlights some of these challenges.

Currently, monolayer differentiation protocols using hESCs can produce 2×10^8 enriched cardiomyocytes per T225 flask (J.D.G., unpublished data). Greater scalability could be achieved using multilayer tissue culture flasks or culture on suspended microcarriers. The EB methodology can also be adapted to high-throughput using flasks with rotating paddles (Amit

et al., 2011), or by using a simple suspension culture to produce hiPSC-derived cardiomyocytes in commercial quantities (Ma et al., 2011). The elimination of growth factors from the monolayer protocol will permit a substantial reduction in costs, and along with the application of other low-molecular-weight compounds (Willems et al., 2009), will continue to improve scalability.

Although development toward maturity in the electrophysiological properties of hPSC-cardiomyocytes has been demonstrated over time in culture, better strategies for full development of mature intracellular Ca^{2+} handling machinery and ion channels within cells in vitro must be developed. Techniques that can induce maturation include the use of small molecules such as triiodothyronine binding thyroid hormone receptors (Lee et al., 2010), overexpression or downregulation of specific genes (Yamanaka et al., 2008), and extrinsic cues from other noncardiac cells, including endothelial cells (Kim et al., 2010a). Exercise via mechanical force (Tulloch et al., 2011) (which will likely be required for tissue constructs) has also been demonstrated to enhance maturation. Other important considerations, such as classifying the type of cardiomyocyte (ventricular, atrial, or nodal) based on AP (Ma et al., 2011) and selecting relevant populations (either through development of modified differentiation protocols specific for one cell type [Zhu et al., 2010] or via selection based on electrophysiology) will need to be addressed. Indeed, the timing of when subtype fate specification occurs in development, how it is regulated, and by what signals is still largely unknown.

Finally, issues surrounding donor cell death (Nguyen et al., 2011), tumorigenesis (Cao et al., 2009; Lee et al., 2009), and prevention of immune rejection (Pearl et al., 2011; Swijnenburg et al., 2008) remain to be addressed. When engrafted, the new cardiomyocytes must both induce vascularization to keep the graft alive as well as electrically couple with the existing cardiomyocytes and function in synchronization with the existing heart muscle, or risk generating potentially fatal arrhythmias (Boudoulas and Hatzopoulos, 2009). In addition, pluripotent cells remaining in the population might be tumorigenic, and the ability to detect the presence of these cells with suitable markers (Ramirez et al., 2011), such as the recently discovered SSEA5 (Tang et al., 2011), will be crucial. To date, engraftment of hESC-derived cardiomyocytes in a single-cell format has yielded similar engraftment results to those from other cell types, with only a 5%–10% improvement in LVEF (Cao et al., 2008; van Laake et al., 2007), even with the use of “prosurvival cocktails” (Laflamme et al., 2007). Optimal engraftment may require applying cells as a 3D engineered tissue. For the successful use of these engineered tissues, the cardiomyocytes are supplemented with endothelial cells and fibroblasts, as insufficient graft vascularization can result in poor graft survival (Caspi et al., 2007). Scaffold-free hESC-derived cardiac patches have been demonstrated to be successful, along with synthetic polymers, hydrogels, and natural polymers such as fibrin and type I collagen (Tulloch et al., 2011). All work thus far with hPSC-derived cardiomyocytes has been performed in animal models (mostly in rodents), and it may be difficult for human cardiomyocytes to keep pace with the ~ 450 beats/min required in the rodent heart (Murry and Keller, 2008). This high beating rate may also mask arrhythmias. The guinea pig model has a more suitable heart rate (230–380 beats/min), but at present

immunocompromised strains suitable for cardiac engraftment do not exist, requiring the use of immunosuppressants such as cyclosporin A. Finally, assessment of human-sized cardiac patches in larger animals such as pigs will be required before transfer to the clinic is feasible.

In conclusion, our ability to generate cardiomyocytes de novo has progressed rapidly in the past 5 years, and an increasing number of potential sources of cells have become viable. The three major applications of these cardiomyocytes (in regenerative medicine, drug testing, and disease modeling) each have their own specific requirements for number of cells, speed of derivation, characterization, and similarity to adult cardiomyocytes. It is likely that a variety of approaches for making cardiomyocytes will be used in the future, depending on the specific parameters set by the application. In addition to the plethora of differentiation methods that have been devised, the core pathways controlling cardiomyocyte differentiation (BMP, TGF β /activin/NODAL, and WNT) have been identified. There are complex questions regarding the application of newly generated cardiomyocytes still to be answered, including the best methods for cell delivery and how to ensure a normal physiological response of the cells. In addition, many technical issues remain unresolved. For example, what if cells successfully engraft and electrically couple, yet do not respond normally to endogenous stimuli? In addition, the choice of hESCs versus hiPSCs as the starting material depends on the ultimate use for which the cells are intended; where ethical considerations are preeminent, hiPSCs may be the less controversial source. Certainly, for applications in genetic disease modeling or personalized therapies, hiPSCs are the obvious choice. By contrast, the extensive characterization and epigenetic stability of hESCs may make them more suitable for the large-scale generation of allogeneic cells for regenerative medicine. Such questions should not obscure the exciting prospect that, for the first time, it is possible to envisage overcoming the significant hurdles that have blocked the path to successful clinical application of de novo generated cardiomyocytes.

ACKNOWLEDGMENTS

We would like to acknowledge funding support from the NIH New Innovator Award DP2OD004437, RC1AG036142, R33HL089027, the California Institute of Regenerative Medicine RB3-05219, and the Burroughs Wellcome Foundation Career Award for Medical Scientists (J.C.W.). Due to space limitations, we are unable to include all of the important papers relevant to induced pluripotent stem cell derivation and application; we apologize to those investigators whose work was omitted here. Joseph Gold is an employee of Geron Corporation.

REFERENCES

- Akopian, V., Andrews, P.W., Beil, S., Benvenisty, N., Brehm, J., Christie, M., Ford, A., Fox, V., Gokhale, P.J., Healy, L., et al; International Stem Cell Initiative Consortium. (2010). Comparison of defined culture systems for feeder cell free propagation of human embryonic stem cells. *In Vitro Cell. Dev. Biol. Anim.* **46**, 247–258.
- Amit, M., Carpenter, M.K., Inokuma, M.S., Chiu, C.P., Harris, C.P., Waknitz, M.A., Itskovitz-Eldor, J., and Thomson, J.A. (2000). Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev. Biol.* **227**, 271–278.
- Amit, M., Laevsky, I., Miropolsky, Y., Shariki, K., Peri, M., and Itskovitz-Eldor, J. (2011). Dynamic suspension culture for scalable expansion of undifferentiated human pluripotent stem cells. *Nat. Protoc.* **6**, 572–579.
- Anderson, D., Self, T., Mellor, I.R., Goh, G., Hill, S.J., and Denning, C. (2007). Transgenic enrichment of cardiomyocytes from human embryonic stem cells. *Mol. Ther.* **15**, 2027–2036.
- Arnold, S.J., and Robertson, E.J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* **10**, 91–103.
- Beqqali, A., Kloots, J., Ward-van Oostwaard, D., Mummery, C., and Passier, R. (2006). Genome-wide transcriptional profiling of human embryonic stem cells differentiating to cardiomyocytes. *Stem Cells* **24**, 1956–1967.
- Bergmann, O., Bhardwaj, R.D., Bernard, S., Zdunek, S., Barnabé-Heider, F., Walsh, S., Zupicich, J., Alkass, K., Buchholz, B.A., Druid, H., et al. (2009). Evidence for cardiomyocyte renewal in humans. *Science* **324**, 98–102.
- Bersell, K., Arab, S., Haring, B., and Kühn, B. (2009). Neuregulin1/ErbB4 signaling induces cardiomyocyte proliferation and repair of heart injury. *Cell* **138**, 257–270.
- Bondue, A., and Blanpain, C. (2010). Mesp1: a key regulator of cardiovascular lineage commitment. *Circ. Res.* **107**, 1414–1427.
- Bondue, A., Lapouge, G., Paulissen, C., Semeraro, C., Iacovino, M., Kyba, M., and Blanpain, C. (2008). Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. *Cell Stem Cell* **3**, 69–84.
- Boudoulas, K.D., and Hatzopoulos, A.K. (2009). Cardiac repair and regeneration: the Rubik's cube of cell therapy for heart disease. *Dis Model Mech* **2**, 344–358.
- Braam, S.R., Tertoolen, L., van de Stolpe, A., Meyer, T., Passier, R., and Mummery, C.L. (2010). Prediction of drug-induced cardiotoxicity using human embryonic stem cell-derived cardiomyocytes. *Stem Cell Res. (Amst.)* **4**, 107–116.
- Bu, L., Jiang, X., Martin-Puig, S., Caron, L., Zhu, S., Shao, Y., Roberts, D.J., Huang, P.L., Domian, I.J., and Chien, K.R. (2009). Human ISL1 heart progenitors generate diverse multipotent cardiovascular cell lineages. *Nature* **460**, 113–117.
- Buckingham, M., Meilhac, S., and Zaffran, S. (2005). Building the mammalian heart from two sources of myocardial cells. *Nat. Rev. Genet.* **6**, 826–835.
- Burridge, P.W., Anderson, D., Priddle, H., Barbadillo Muñoz, M.D., Chamberlain, S., Allegrucci, C., Young, L.E., and Denning, C. (2007). Improved human embryonic stem cell embryoid body homogeneity and cardiomyocyte differentiation from a novel V-96 plate aggregation system highlights interline variability. *Stem Cells* **25**, 929–938.
- Burridge, P.W., Thompson, S., Millrod, M.A., Weinberg, S., Yuan, X., Peters, A., Mahairaki, V., Koliatsos, V.E., Tung, L., and Zambidis, E.T. (2011). A universal system for highly efficient cardiac differentiation of human induced pluripotent stem cells that eliminates interline variability. *PLoS ONE* **6**, e18293.
- Cao, F., Wagner, R.A., Wilson, K.D., Xie, X., Fu, J.D., Drukker, M., Lee, A., Li, R.A., Gambhir, S.S., Weissman, I.L., et al. (2008). Transcriptional and functional profiling of human embryonic stem cell-derived cardiomyocytes. *PLoS ONE* **3**, e3474.
- Cao, F., Li, Z., Lee, A., Liu, Z., Chen, K., Wang, H., Cai, W., Chen, X., and Wu, J.C. (2009). Noninvasive de novo imaging of human embryonic stem cell-derived teratoma formation. *Cancer Res.* **69**, 2709–2713.
- Cao, N., Liu, Z., Chen, Z., Wang, J., Chen, T., Zhao, X., Ma, Y., Qin, L., Kang, J., Wei, B., et al. (2011). Ascorbic acid enhances the cardiac differentiation of induced pluripotent stem cells promoting the proliferation of cardiac progenitor cells. *Cell Res.*, in press. Published online December 6, 2011. 10.1038/cr.2011.195.
- Carvajal-Vergara, X., Sevilla, A., D'Souza, S.L., Ang, Y.S., Schaniel, C., Lee, D.F., Yang, L., Kaplan, A.D., Adler, E.D., Rozov, R., et al. (2010). Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome. *Nature* **465**, 808–812.
- Caspi, O., Lesman, A., Basevitch, Y., Gepstein, A., Arbel, G., Habib, I.H., Gepstein, L., and Levenberg, S. (2007). Tissue engineering of vascularized cardiac muscle from human embryonic stem cells. *Circ. Res.* **100**, 263–272.
- Chen, G., Gulbranson, D.R., Hou, Z., Bolin, J.M., Ruotti, V., Probasco, M.D., Smuga-Otto, K., Howden, S.E., Diol, N.R., Propson, N.E., et al. (2011).

- Chemically defined conditions for human iPSC derivation and culture. *Nat. Methods* 8, 424–429.
- Choi, J., Costa, M.L., Mermelstein, C.S., Chagas, C., Holtzer, S., and Holtzer, H. (1990). MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. *Proc. Natl. Acad. Sci. USA* 87, 7988–7992.
- Chung, H.C., Lin, R.C., Logan, G.J., Alexander, I.E., Sachdev, P.S., and Sidhu, K.S. (2011). Human induced pluripotent stem cells derived under feeder-free conditions display unique cell cycle and DNA replication gene profiles. *Stem Cells Dev.*, in press. Published online June 1, 2011. 10.1089/scd.2010.0440.
- Costello, I., Pimeisl, I.M., Dräger, S., Bikoff, E.K., Robertson, E.J., and Arnold, S.J. (2011). The T-box transcription factor Eomesodermin acts upstream of *Mesp1* to specify cardiac mesoderm during mouse gastrulation. *Nat. Cell Biol.* 13, 1084–1091.
- David, R., Brenner, C., Stieber, J., Schwarz, F., Brunner, S., Vollmer, M., Mentle, E., Müller-Höcker, J., Kitajima, S., Lickert, H., et al. (2008). *MesP1* drives vertebrate cardiovascular differentiation through *Dkk-1*-mediated blockade of Wnt-signalling. *Nat. Cell Biol.* 10, 338–345.
- David, R., Jarsch, V.B., Schwarz, F., Nathan, P., Gegg, M., Lickert, H., and Franz, W.M. (2011). Induction of *MesP1* by *Brachyury(T)* generates the common multipotent cardiovascular stem cell. *Cardiovasc. Res.* 92, 115–122.
- Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51, 987–1000.
- Davis, R.P., van den Berg, C.W., Casini, S., Braam, S.R., and Mummery, C.L. (2011). Pluripotent stem cell models of cardiac disease and their implication for drug discovery and development. *Trends Mol. Med.* 17, 475–484.
- Denning, C., and Anderson, D. (2008). Cardiomyocytes from human embryonic stem cells as predictors of cardiotoxicity. *Drug Discov. Today Ther. Strateg.* 5, 223–232.
- Dubois, N.C., Craft, A.M., Sharma, P., Elliott, D.A., Stanley, E.G., Elefanti, A.G., Gramolini, A., and Keller, G. (2011). *SIRPA* is a specific cell-surface marker for isolating cardiomyocytes derived from human pluripotent stem cells. *Nat. Biotechnol.* 29, 1011–1018.
- Efe, J.A., Hilcove, S., Kim, J., Zhou, H., Ouyang, K., Wang, G., Chen, J., and Ding, S. (2011). Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. *Nat. Cell Biol.* 13, 215–222.
- Elliott, D.A., Braam, S.R., Koutsis, K., Ng, E.S., Jenny, R., Lagerqvist, E.L., Bibben, C., Hatzistavrou, T., Hirst, C.E., Yu, Q.C., et al. (2011). *NKX2-5(eGFP/w)* hESCs for isolation of human cardiac progenitors and cardiomyocytes. *Nat. Methods* 8, 1037–1040.
- Freund, C., Ward-van Oostwaard, D., Monshouwer-Kloots, J., van den Brink, S., van Rooijen, M., Xu, X., Zweigert, R., Mummery, C., and Passier, R. (2008). Insulin redirects differentiation from cardiogenic mesoderm and endoderm to neuroectoderm in differentiating human embryonic stem cells. *Stem Cells* 26, 724–733.
- Gaur, M., Ritner, C., Sievers, R., Pedersen, A., Prasad, M., Bernstein, H.S., and Yeghiazarians, Y. (2010). Timed inhibition of p38MAPK directs accelerated differentiation of human embryonic stem cells into cardiomyocytes. *Cytother-apy* 12, 807–817.
- Gnecchi, M., He, H., Liang, O.D., Melo, L.G., Morello, F., Mu, H., Noiseux, N., Zhang, L., Pratt, R.E., Ingwall, J.S., and Dzau, V.J. (2005). Paracrine action accounts for marked protection of ischemic heart by Akt-modified mesenchymal stem cells. *Nat. Med.* 11, 367–368.
- Grego-Bessa, J., Luna-Zurita, L., del Monte, G., Bolós, V., Melgar, P., Arandilla, A., Garratt, A.N., Zang, H., Mukoyama, Y.S., Chen, H., et al. (2007). Notch signaling is essential for ventricular chamber development. *Dev. Cell* 12, 415–429.
- Hansson, E.M., Lindsay, M.E., and Chien, K.R. (2009). Regeneration next toward heart stem cell therapeutics. *Cell Stem Cell* 5, 364–377.
- Hattori, F., Chen, H., Yamashita, H., Tohyama, S., Satoh, Y.S., Yuasa, S., Li, W., Yamakawa, H., Tanaka, T., Onitsuka, T., et al. (2010). Nongenetic method for purifying stem cell-derived cardiomyocytes. *Nat. Methods* 7, 61–66.
- Huber, I., Itzhaki, I., Caspi, O., Arbel, G., Tzukerman, M., Gepstein, A., Habib, M., Yankelson, L., Kehat, I., and Gepstein, L. (2007). Identification and selection of cardiomyocytes during human embryonic stem cell differentiation. *FASEB J.* 21, 2551–2563.
- Hudson, J., Titmarsh, D., Hidalgo, A., Wolvetang, E., and Cooper-White, J. (2011). Primitive cardiac cells from human embryonic stem cells. *Stem Cells Dev.*, in press. Published online November 9, 2011. 10.1089/scd.2011.0254.
- Ieda, M., Fu, J.D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* 142, 375–386.
- Itskovitz-Eldor, J., Schuldiner, M., Karsenti, D., Eden, A., Yanuka, O., Amit, M., Soreq, H., and Benvenisty, N. (2000). Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. *Mol. Med.* 6, 88–95.
- Itzhaki, I., Maizels, L., Huber, I., Zwi-Dantsis, L., Caspi, O., Winterstern, A., Feldman, O., Gepstein, A., Arbel, G., Hammerman, H., et al. (2011a). Modeling the long QT syndrome with induced pluripotent stem cells. *Nature* 471, 225–229.
- Itzhaki, I., Rapoport, S., Huber, I., Mizrahi, I., Zwi-Dantsis, L., Arbel, G., Schiller, J., and Gepstein, L. (2011b). Calcium handling in human induced pluripotent stem cell derived cardiomyocytes. *PLoS ONE* 6, e18037.
- Jang, J., Ku, S.Y., Kim, J.E., Choi, K., Kim, Y.Y., Kim, H.S., Oh, S.K., Lee, E.J., Cho, H.J., Song, Y.H., et al. (2008). Notch inhibition promotes human embryonic stem cell-derived cardiac mesoderm differentiation. *Stem Cells* 26, 2782–2790.
- Jopling, C., Sleep, E., Raya, M., Martí, M., Raya, A., and Izpisua Belmonte, J.C. (2010). Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature* 464, 606–609.
- Josowitz, R., Carvajal-Vergara, X., Lemischka, I.R., and Gelb, B.D. (2011). Induced pluripotent stem cell-derived cardiomyocytes as models for genetic cardiovascular disorders. *Curr. Opin. Cardiol.* 26, 223–229.
- Kattman, S.J., Witty, A.D., Gagliardi, M., Dubois, N.C., Niapour, M., Hotta, A., Ellis, J., and Keller, G. (2011). Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell* 8, 228–240.
- Kehat, I., Kenyagin-Karsenti, D., Snir, M., Segev, H., Amit, M., Gepstein, A., Livne, E., Binah, O., Itskovitz-Eldor, J., and Gepstein, L. (2001). Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. *J. Clin. Invest.* 108, 407–414.
- Kim, J.B., Greber, B., Araúzo-Bravo, M.J., Meyer, J., Park, K.I., Zaehres, H., and Schöler, H.R. (2009). Direct reprogramming of human neural stem cells by OCT4. *Nature* 461, 649.
- Kim, C., Majidi, M., Xia, P., Wei, K.A., Talantova, M., Spiering, S., Nelson, B., Mercola, M., and Chen, H.S. (2010a). Non-cardiomyocytes influence the electrophysiological maturation of human embryonic stem cell-derived cardiomyocytes during differentiation. *Stem Cells Dev.* 19, 783–795.
- Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I., et al. (2010b). Epigenetic memory in induced pluripotent stem cells. *Nature* 467, 285–290.
- Kim, J., Efe, J.A., Zhu, S., Talantova, M., Yuan, X., Wang, S., Lipton, S.A., Zhang, K., and Ding, S. (2011). Direct reprogramming of mouse fibroblasts to neural progenitors. *Proc. Natl. Acad. Sci. USA* 108, 7838–7843.
- Kühn, B., del Monte, F., Hajjar, R.J., Chang, Y.S., Lebeche, D., Arab, S., and Keating, M.T. (2007). Periostin induces proliferation of differentiated cardiomyocytes and promotes cardiac repair. *Nat. Med.* 13, 962–969.
- Kwon, C., Arnold, J., Hsiao, E.C., Taketo, M.M., Conklin, B.R., and Srivastava, D. (2007). Canonical Wnt signaling is a positive regulator of mammalian cardiac progenitors. *Proc. Natl. Acad. Sci. USA* 104, 10894–10899.
- Kwon, C., Qian, L., Cheng, P., Nigam, V., Arnold, J., and Srivastava, D. (2009). A regulatory pathway involving Notch1/beta-catenin/Is1 determines cardiac progenitor cell fate. *Nat. Cell Biol.* 11, 951–957.
- Laflamme, M.A., and Murry, C.E. (2005). Regenerating the heart. *Nat. Biotechnol.* 23, 845–856.
- Laflamme, M.A., Chen, K.Y., Naumova, A.V., Muskheli, V., Fugate, J.A., Dupras, S.K., Reinecke, H., Xu, C., Hassanipour, M., Police, S., et al. (2007). Cardiomyocytes derived from human embryonic stem cells in pro-survival

factors enhance function of infarcted rat hearts. *Nat. Biotechnol.* 25, 1015–1024.

Lavine, K.J., Yu, K., White, A.C., Zhang, X., Smith, C., Partanen, J., and Ornitz, D.M. (2005). Endocardial and epicardial derived FGF signals regulate myocardial proliferation and differentiation in vivo. *Dev. Cell* 8, 85–95.

Lee, A.S., Tang, C., Cao, F., Xie, X., van der Bogt, K., Hwang, A., Connolly, A.J., Robbins, R.C., and Wu, J.C. (2009). Effects of cell number on teratoma formation by human embryonic stem cells. *Cell Cycle* 8, 2608–2612.

Lee, Y.K., Ng, K.M., Chan, Y.C., Lai, W.H., Au, K.W., Ho, C.Y., Wong, L.Y., Lau, C.P., Tse, H.F., and Siu, C.W. (2010). Triiodothyronine promotes cardiac differentiation and maturation of embryonic stem cells via the classical genomic pathway. *Mol. Endocrinol.* 24, 1728–1736.

Li, Y., Powell, S., Brunette, E., Lebkowski, J., and Mandalam, R. (2005). Expansion of human embryonic stem cells in defined serum-free medium devoid of animal-derived products. *Biotechnol. Bioeng.* 91, 688–698.

Li, W., Wang, X., Fan, W., Zhao, P., Chan, Y.C., Chen, S., Zhang, S., Guo, X., Zhang, Y., Li, Y., et al. (2012). Modeling abnormal early development with induced pluripotent stem cells from aneuploid syndromes. *Hum. Mol. Genet.* 21, 32–45.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 471, 68–73.

Loffredo, F.S., Steinhauser, M.L., Gannon, J., and Lee, R.T. (2011). Bone marrow-derived cell therapy stimulates endogenous cardiomyocyte progenitors and promotes cardiac repair. *Cell Stem Cell* 8, 389–398.

Ludwig, T.E., Bergendahl, V., Levenstein, M.E., Yu, J., Probasco, M.D., and Thomson, J.A. (2006). Feeder-independent culture of human embryonic stem cells. *Nat. Methods* 3, 637–646.

Ma, J., Guo, L., Fiene, S.J., Anson, B.D., Thomson, J.A., Kamp, T.J., Kolaja, K.L., Swanson, B.J., and January, C.T. (2011). High purity human-induced pluripotent stem cell-derived cardiomyocytes: electrophysiological properties of action potentials and ionic currents. *Am. J. Physiol. Heart Circ. Physiol.* 301, H2006–H2017.

Martinez-Fernandez, A., Nelson, T.J., Ikeda, Y., and Terzic, A. (2010). c-MYC independent nuclear reprogramming favors cardiogenic potential of induced pluripotent stem cells. *J Cardiovasc Transl Res* 3, 13–23.

Martinez-Fernandez, A., Nelson, T.J., and Terzic, A. (2011). Nuclear reprogramming strategy modulates differentiation potential of induced pluripotent stem cells. *J Cardiovasc Transl Res* 4, 131–137.

Matsa, E., Rajamohan, D., Dick, E., Young, L., Mellor, I., Staniforth, A., and Denning, C. (2011). Drug evaluation in cardiomyocytes derived from human induced pluripotent stem cells carrying a long QT syndrome type 2 mutation. *Eur. Heart J.* 32, 952–962.

McDevitt, T.C., Laflamme, M.A., and Murry, C.E. (2005). Proliferation of cardiomyocytes derived from human embryonic stem cells is mediated via the IGF/PI 3-kinase/Akt signaling pathway. *J. Mol. Cell. Cardiol.* 39, 865–873.

Menasché, P., Alfieri, O., Janssens, S., McKenna, W., Reichenspurner, H., Trinquart, L., Vilquin, J.T., Marolleau, J.P., Seymour, B., Larghero, J., et al. (2008). The Myoblast Autologous Grafting in Ischemic Cardiomyopathy (MAGIC) trial: first randomized placebo-controlled study of myoblast transplantation. *Circulation* 117, 1189–1200.

Moretti, A., Bellin, M., Welling, A., Jung, C.B., Lam, J.T., Bott-Flügel, L., Dorn, T., Goedel, A., Höhnke, C., Hofmann, F., et al. (2010). Patient-specific induced pluripotent stem-cell models for long-QT syndrome. *N. Engl. J. Med.* 363, 1397–1409.

Mummery, C., Ward-van Oostwaard, D., Doevendans, P., Spijker, R., van den Brink, S., Hassink, R., van der Heyden, M., Ophof, T., Pera, M., de la Riviere, A.B., et al. (2003). Differentiation of human embryonic stem cells to cardiomyocytes: role of coculture with visceral endoderm-like cells. *Circulation* 107, 2733–2740.

Murry, C.E., and Keller, G. (2008). Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* 132, 661–680.

Naito, A.T., Shiojima, I., Akazawa, H., Hidaka, K., Morisaki, T., Kikuchi, A., and Komuro, I. (2006). Developmental stage-specific biphasic roles of Wnt/beta-catenin signaling in cardiomyogenesis and hematopoiesis. *Proc. Natl. Acad. Sci. USA* 103, 19812–19817.

Narsinh, K.H., Sun, N., Sanchez-Freire, V., Lee, A.S., Almeida, P., Hu, S., Jan, T., Wilson, K.D., Leong, D., Rosenberg, J., et al. (2011). Single cell transcriptional profiling reveals heterogeneity of human induced pluripotent stem cells. *J. Clin. Invest.* 121, 1217–1221.

Nguyen, P.K., Lan, F., Wang, Y., and Wu, J.C. (2011). Imaging: guiding the clinical translation of cardiac stem cell therapy. *Circ. Res.* 109, 962–979.

Ohi, Y., Qin, H., Hong, C., Blouin, L., Polo, J.M., Guo, T., Qi, Z., Downey, S.L., Manos, P.D., Rossi, D.J., et al. (2011). Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nat. Cell Biol.* 13, 541–549.

Paige, S.L., Osugi, T., Afanasiev, O.K., Pabon, L., Reinecke, H., and Murry, C.E. (2010). Endogenous Wnt/beta-catenin signaling is required for cardiac differentiation in human embryonic stem cells. *PLoS ONE* 5, e11134.

Passier, R., Oostwaard, D.W., Snapper, J., Kloots, J., Hassink, R.J., Kuijk, E., Roelen, B., de la Riviere, A.B., and Mummery, C. (2005). Increased cardiomyocyte differentiation from human embryonic stem cells in serum-free cultures. *Stem Cells* 23, 772–780.

Pasumarthi, K.B., and Field, L.J. (2002). Cardiomyocyte cell cycle regulation. *Circ. Res.* 90, 1044–1054.

Pearl, J.I., Lee, A.S., Leveson-Gower, D.B., Sun, N., Ghosh, Z., Lan, F., Ransohoff, J., Negrin, R.S., Davis, M.M., and Wu, J.C. (2011). Short-term immunosuppression promotes engraftment of embryonic and induced pluripotent stem cells. *Cell Stem Cell* 8, 309–317.

Perino, M.G., Yamanaka, S., Li, J., Wobus, A.M., and Boheler, K.R. (2008). Cardiomyogenic stem and progenitor cell plasticity and the dissection of cardiopoiesis. *J Mol Cell Cardiol* 45, 475–494.

Polo, J.M., Liu, S., Figueroa, M.E., Kulalert, W., Eminli, S., Tan, K.Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T., et al. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat. Biotechnol.* 28, 848–855.

Porrello, E.R., Mahmoud, A.I., Simpson, E., Hill, J.A., Richardson, J.A., Olson, E.N., and Sadek, H.A. (2011). Transient regenerative potential of the neonatal mouse heart. *Science* 331, 1078–1080.

Poss, K.D., Wilson, L.G., and Keating, M.T. (2002). Heart regeneration in zebrafish. *Science* 298, 2188–2190.

Qyang, Y., Martin-Puig, S., Chiravuri, M., Chen, S., Xu, H., Bu, L., Jiang, X., Lin, L., Granger, A., Moretti, A., et al. (2007). The renewal and differentiation of Isl1+ cardiovascular progenitors are controlled by a Wnt/beta-catenin pathway. *Cell Stem Cell* 1, 165–179.

Ramirez, J.M., Gerbal-Chaloin, S., Milhavet, O., Qiang, B., Becker, F., Assou, S., Lemaître, J.M., Hamamah, S., and De Vos, J. (2011). Brief report: benchmarking human pluripotent stem cell markers during differentiation into the three germ layers unveils a striking heterogeneity: all markers are not equal. *Stem Cells* 29, 1469–1474.

Reinecke, H., Zhang, M., Bartosek, T., and Murry, C.E. (1999). Survival, integration, and differentiation of cardiomyocyte grafts: a study in normal and injured rat hearts. *Circulation* 100, 193–202.

Ren, Y., Lee, M.Y., Schliffke, S., Paavola, J., Amos, P.J., Ge, X., Ye, M., Zhu, S., Senyey, G., Lum, L., et al. (2011). Small molecule Wnt inhibitors enhance the efficiency of BMP-4-directed cardiac differentiation of human pluripotent stem cells. *J. Mol. Cell. Cardiol.* 51, 280–287.

Roger, V.L., Go, A.S., Lloyd-Jones, D.M., Adams, R.J., Berry, J.D., Brown, T.M., Carnethon, M.R., Dai, S., de Simone, G., Ford, E.S., et al; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. (2011). Heart disease and stroke statistics—2011 update: a report from the American Heart Association. *Circulation* 123, e18–e209.

Shimoji, K., Yuasa, S., Onizuka, T., Hattori, F., Tanaka, T., Hara, M., Ohno, Y., Chen, H., Egasgira, T., Seki, T., et al. (2010). G-CSF promotes the proliferation of developing cardiomyocytes in vivo and in derivation from ESCs and iPSCs. *Cell Stem Cell* 6, 227–237.

- Smart, N., Bollini, S., Dubé, K.N., Vieira, J.M., Zhou, B., Davidson, S., Yellon, D., Riegler, J., Price, A.N., Lythgoe, M.F., et al. (2011). De novo cardiomyocytes from within the activated adult heart after injury. *Nature* *474*, 640–644.
- Solloway, M.J., and Harvey, R.P. (2003). Molecular pathways in myocardial development: a stem cell perspective. *Cardiovasc. Res.* *58*, 264–277.
- Swijnenburg, R.J., Schrepfer, S., Govaert, J.A., Cao, F., Ransohoff, K., Sheikh, A.Y., Haddad, M., Connolly, A.J., Davis, M.M., Robbins, R.C., and Wu, J.C. (2008). Immunosuppressive therapy mitigates immunological rejection of human embryonic stem cell xenografts. *Proc. Natl. Acad. Sci. USA* *105*, 12991–12996.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* *126*, 663–676.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* *131*, 861–872.
- Takei, S., Ichikawa, H., Johkura, K., Mogi, A., No, H., Yoshie, S., Tomotsune, D., and Sasaki, K. (2009). Bone morphogenetic protein-4 promotes induction of cardiomyocytes from human embryonic stem cells in serum-based embryoid body development. *Am. J. Physiol. Heart Circ. Physiol.* *296*, H1793–H1803.
- Takeuchi, J.K., and Bruneau, B.G. (2009). Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. *Nature* *459*, 708–711.
- Tam, P.P., and Loebel, D.A. (2007). Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* *8*, 368–381.
- Tanasijevic, B., Dai, B., Ezashi, T., Livingston, K., Roberts, R.M., and Rasmussen, T.P. (2009). Progressive accumulation of epigenetic heterogeneity during human ES cell culture. *Epigenetics* *4*, 330–338.
- Tang, C., Lee, A.S., Volkmer, J.P., Sahoo, D., Nag, D., Mosley, A.R., Inlay, M.A., Ardehali, R., Chavez, S.L., Pera, R.R., et al. (2011). An antibody against SSEA-5 glycan on human pluripotent stem cells enables removal of teratoma-forming cells. *Nat. Biotechnol.* *29*, 829–834.
- Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. *Science* *282*, 1145–1147.
- Tran, T.H., Wang, X., Browne, C., Zhang, Y., Schinke, M., Izumo, S., and Burcin, M. (2009). Wnt3a-induced mesoderm formation and cardiomyogenesis in human embryonic stem cells. *Stem Cells* *27*, 1869–1878.
- Tulloch, N.L., Muskheli, V., Razumova, M.V., Korte, F.S., Regnier, M., Hauch, K.D., Pabon, L., Reinecke, H., and Murry, C.E. (2011). Growth of engineered human myocardium with mechanical loading and vascular coculture. *Circ. Res.* *109*, 47–59.
- Ueno, S., Weidinger, G., Osugi, T., Kohn, A.D., Golob, J.L., Pabon, L., Reinecke, H., Moon, R.T., and Murry, C.E. (2007). Biphasic role for Wnt/beta-catenin signaling in cardiac specification in zebrafish and embryonic stem cells. *Proc. Natl. Acad. Sci. USA* *104*, 9685–9690.
- Uosaki, H., Fukushima, H., Takeuchi, A., Matsuoka, S., Nakatsuji, N., Yamana, S., and Yamashita, J.K. (2011). Efficient and scalable purification of cardiomyocytes from human embryonic and induced pluripotent stem cells by VCAM1 surface expression. *PLoS ONE* *6*, e23657.
- Van Hoof, D., Dormeyer, W., Braam, S.R., Passier, R., Monshouwer-Kloots, J., Ward-van Oostwaard, D., Heck, A.J., Krijgsveld, J., and Mummery, C.L. (2010). Identification of cell surface proteins for antibody-based selection of human embryonic stem cell-derived cardiomyocytes. *J. Proteome Res.* *9*, 1610–1618.
- van Laake, L.W., Passier, R., Monshouwer-Kloots, J., Verkleij, A.J., Lips, D.J., Freund, C., den Ouden, K., Ward-van Oostwaard, D., Korving, J., Tertoolen, L.G., et al. (2007). Human embryonic stem cell-derived cardiomyocytes survive and mature in the mouse heart and transiently improve function after myocardial infarction. *Stem Cell Res. (Amst.)* *1*, 9–24.
- Wang, Z., Wang, D.Z., Pipes, G.C., and Olson, E.N. (2003). Myocardin is a master regulator of smooth muscle gene expression. *Proc. Natl. Acad. Sci. USA* *100*, 7129–7134.
- Wang, L., Schulz, T.C., Sherrer, E.S., Dauphin, D.S., Shin, S., Nelson, A.M., Ware, C.B., Zhan, M., Song, C.Z., Chen, X., et al. (2007). Self-renewal of human embryonic stem cells requires insulin-like growth factor-1 receptor and ERBB2 receptor signaling. *Blood* *110*, 4111–4119.
- Willems, E., Bushway, P.J., and Mercola, M. (2009). Natural and synthetic regulators of embryonic stem cell cardiogenesis. *Pediatr. Cardiol.* *30*, 635–642.
- Willems, E., Spiering, S., Davidovics, H., Lanier, M., Xia, Z., Dawson, M., Cashman, J., and Mercola, M. (2011). Small-molecule inhibitors of the Wnt pathway potentially promote cardiomyocytes from human embryonic stem cell-derived mesoderm. *Circ. Res.* *109*, 360–364.
- Xu, C., Inokuma, M.S., Denham, J., Golds, K., Kundu, P., Gold, J.D., and Carpenter, M.K. (2001). Feeder-free growth of undifferentiated human embryonic stem cells. *Nat. Biotechnol.* *19*, 971–974.
- Xu, C., Police, S., Rao, N., and Carpenter, M.K. (2002). Characterization and enrichment of cardiomyocytes derived from human embryonic stem cells. *Circ. Res.* *91*, 501–508.
- Xu, X.Q., Graichen, R., Soo, S.Y., Balakrishnan, T., Rahmat, S.N., Sieh, S., Tham, S.C., Freund, C., Moore, J., Mummery, C., et al. (2008). Chemically defined medium supporting cardiomyocyte differentiation of human embryonic stem cells. *Differentiation* *76*, 958–970.
- Yamanaka, S., Zahanich, I., Wersto, R.P., and Boheler, K.R. (2008). Enhanced proliferation of monolayer cultures of embryonic stem (ES) cell-derived cardiomyocytes following acute loss of retinoblastoma. *PLoS ONE* *3*, e3896.
- Yang, L., Soonpaa, M.H., Adler, E.D., Roepke, T.K., Kattman, S.J., Kennedy, M., Henckaerts, E., Bonham, K., Abbott, G.W., Linden, R.M., et al. (2008). Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature* *453*, 524–528.
- Yazawa, M., Hsueh, B., Jia, X., Pasca, A.M., Bernstein, J.A., Hallmayer, J., and Dolmetsch, R.E. (2011). Using induced pluripotent stem cells to investigate cardiac phenotypes in Timothy syndrome. *Nature* *471*, 230–234.
- Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* *318*, 1917–1920.
- Yu, P., Pan, G., Yu, J., and Thomson, J.A. (2011). FGF2 sustains NANOG and switches the outcome of BMP4-induced human embryonic stem cell differentiation. *Cell Stem Cell* *8*, 326–334.
- Zhang, Y., Li, T.S., Lee, S.T., Wawrowsky, K.A., Cheng, K., Galang, G., Malliaras, K., Abraham, M.R., Wang, C., and Marbán, E. (2010). Dedifferentiation and proliferation of mammalian cardiomyocytes. *PLoS ONE* *5*, e12559.
- Zhang, Q., Jiang, J., Han, P., Yuan, Q., Zhang, J., Zhang, X., Xu, Y., Cao, H., Meng, Q., Chen, L., et al. (2011). Direct differentiation of atrial and ventricular myocytes from human embryonic stem cells by alternating retinoid signals. *Cell Res.* *21*, 579–587.
- Zhao, T., Zhang, Z.N., Rong, Z., and Xu, Y. (2011). Immunogenicity of induced pluripotent stem cells. *Nature* *474*, 212–215.
- Zhu, W.Z., Xie, Y., Moyes, K.W., Gold, J.D., Askari, B., and Laflamme, M.A. (2010). Neuregulin/ErbB signaling regulates cardiac subtype specification in differentiating human embryonic stem cells. *Circ. Res.* *107*, 776–786.
- Zwi, L., Caspi, O., Arbel, G., Huber, I., Gepstein, A., Park, I.H., and Gepstein, L. (2009). Cardiomyocyte differentiation of human induced pluripotent stem cells. *Circulation* *120*, 1513–1523.

Notes

Notes

Notes

To download available conference brochures visit: www.scripps.org/conferenceservices
Subscribe to the Scripps Conference Services & CME newsletter: http://www.scripps.org/health-education_email-sign-up
For additional information contact: Scripps Conference Services & CME, 858-652-5400, Med.edu@scrippshealth.org
Visit our Scripps Conference Services & CME Facebook Page: www.facebook.com/ScrippsCME

Cardiology

23rd Annual Coronary Interventions
October 10-12, 2012
Hilton La Jolla Torrey Pines, La Jolla, California

Dermatology

Melanoma 2013: 23rd Annual Cutaneous Malignancy Update
January 26-27, 2013
Hilton San Diego Resort, San Diego, California

Diabetes

Scripps Whittier Diabetes Professional Education and Training Programs 2012
January – October 2012
Scripps Whittier Diabetes Institute, La Jolla, California

Gastroenterology – Hepatology

27th Annual New Treatments in Chronic Liver Disease
March 31-April 1, 2012
Hyatt Regency La Jolla, San Diego, California

New Advances in Inflammatory Bowel Disease
September 8-9, 2012
San Diego Marriott La Jolla, La Jolla, California

28th Annual New Treatments in Chronic Liver Disease
March 23-24, 2013
Hilton La Jolla Torrey Pines, La Jolla, California

New Advances in Inflammatory Bowel Disease
September 7-8, 2013
Hilton San Diego Resort, San Diego, California

29th Annual New Treatments in Chronic Liver Disease
March 28-30, 2014
Hilton La Jolla Torrey Pines, La Jolla, California

Genomic Medicine

The Future of Genomic Medicine V
March 1-2, 2012
Scripps Seaside Forum
Robert Paine Scripps Forum for Science, Society and the Environment
Samuel H. Scripps Auditorium, La Jolla, California

The Future of Genomic Medicine VI
March 7-8, 2013
Scripps Seaside Forum
Robert Paine Scripps Forum for Science, Society and the Environment
Samuel H. Scripps Auditorium, La Jolla, California

Hematology – Oncology

Scripps Cancer Center's 32nd Annual Conference: Clinical Hematology and Oncology
February 18-21, 2012
Hyatt Regency La Jolla, San Diego, California

9th Annual Advanced Practice Oncology Providers Symposium: Innovation through Practice

March 2-4, 2012

Hilton San Diego Bayfront, San Diego, California

Scripps Cancer Center's 32nd Annual Oncology Nurses Symposium

October 7-10, 2012

Hilton San Diego Resort, San Diego, California

Scripps Cancer Center's 33rd Annual Conference: Clinical Hematology and Oncology

February 16-19, 2013

Westin San Diego Hotel, San Diego, California

10th Annual Advanced Practice Oncology Providers Symposium: Innovation through Practice

March 1-3, 2013

Hyatt Regency Mission Bay, San Diego, California

Scripps Cancer Center's 33rd Annual Oncology Nurses Symposium

September 29-October 2, 2013

Doubletree San Diego Mission Valley, San Diego, California

Integrative Holistic Medicine

9th Annual Natural Supplements: An Evidence-Based Update

January 19-22, 2012

Hilton San Diego Bayfront, San Diego, California

The 13th Annual Science and Clinical Application of Integrative Holistic Medicine

Jointly sponsored with the American Board of Integrative Holistic Medicine

October 28-November 1, 2012

November 2: ABIHM Board Certification Exam

Hilton San Diego Resort, San Diego, California

10th Annual Natural Supplements: An Evidence-Based Update

January 30-February 2, 2013

Hilton San Diego Resort & Spa, San Diego, California

4th Annual Integrative and Holistic Nursing Conference: Bringing Healing to You and Your Patients

April 27-28, 2013

Hilton San Diego Resort, San Diego, California

The 14th Annual Science and Clinical Application of Integrative Holistic Medicine

Jointly sponsored with the American Board of Integrative Holistic Medicine

November 3-7, 2013

November 8: ABIHM Board Certification Exam

Renaissance Vinoy Resort, St. Petersburg, Florida

Primary Care – Family Practice

Headaches: Easing the Pain

February 18, 2012

Schaetzel Center, Scripps Memorial Hospital, La Jolla, California

17th Annual Primary Care in Paradise

April 2-5, 2012

Kauai Marriott Resort & Beach Club, Lihue - Kauai, Hawaii

29th Annual Primary Care Summer Conference

Medical Specialties from the Primary Care Perspective

August 3-5, 2012

Hyatt Regency Mission Bay, San Diego, California

18th Annual Primary Care in Paradise

March 25-28, 2013

Wailea Beach Marriott Resort & Spa, Maui, Hawaii

30th Annual Primary Care Summer Conference

Medical Specialties from the Primary Care Perspective

August 2-4, 2013

Hilton San Diego Resort, San Diego, California

Quality Improvement

5th Annual Scripps Health Quality Summit

March 9, 2012

San Diego Del Mar Marriott, San Diego, California

Rehabilitation

7th Annual Brain Injury Rehabilitation Conference

March 16-17, 2012

San Diego Marriott La Jolla, La Jolla, California

8th Annual Brain Injury Rehabilitation Conference

March 22-23, 2013

Sheraton Carlsbad Resort, Carlsbad, California

Trauma – Emergency Medicine

The 2012 San Diego Day of Trauma

November 2-3, 2012

Hyatt Regency Mission Bay, San Diego, California



S A V E T H E D A T E

The Future of Genomic Medicine VI

Thursday, March 7, and Friday, March 8, 2013 • Scripps Seaside Forum • La Jolla, California

A brochure will be mailed to you in the coming months. For more information, please contact Scripps Conference Services & CME.

858-652-5400 • med.edu@scrippshealth.org • scripps.org/conferenceservices

