



**FACULTAD DE MEDICINA  
DEPARTAMENTO DE BIOQUIMICA**

**TESIS DOCTORAL**

**ANÁLISIS GENÓMICO Y FUNCIONAL DE  
PARÁMETROS DE PROTECCIÓN  
FRENTE AL VIH-1 EN PACIENTES CON  
PROGRESIÓN LENTA DE LA INFECCIÓN**

**Memoria presentada para optar al grado de  
Doctor en Biociencias Moleculares**

**Humberto Erick De La Torre Tarazona**  
**Biólogo**

**Directores de Tesis:**

José Alcamí Pertejo  
Francisco Díez Fuertes  
Instituto de Salud Carlos III

**Madrid, 2020**

**El trabajo presentado en esta Tesis Doctoral ha sido realizado en el Centro Nacional de Microbiología del Instituto de Salud Carlos III y financiado por un contrato Pre-doctoral de Formación en Investigación en Salud (PFIS-ISCIII).**



D. José Alcamí Pertejo, responsable de la Unidad de Inmunopatología del Sida y Jefe del Laboratorio de Referencia e Investigación en Retrovirus, Centro Nacional de Microbiología, Instituto de Salud Carlos III (Majadahonda, Madrid), y Director Científico de la Unidad de infección por el VIH del *Institut d'Investigacions Biomèdiques August Pi i Sunyer*, Hospital Clinic (Barcelona),

D. Francisco Díez Fuertes, investigador post-doctoral de la Unidad de Inmunopatología del Sida, Laboratorio de Referencia e Investigación en Retrovirus, Centro Nacional de Microbiología, Instituto de Salud Carlos III (Majadahonda, Madrid); *Institut d'Investigacions Biomèdiques August Pi i Sunyer*, Hospital Clinic (Barcelona),

CERTIFICAN:

Que la presente memoria titulada “ANÁLISIS GENÓMICO Y FUNCIONAL DE PARÁMETROS DE PROTECCIÓN FRENTE AL VIH-1 EN PACIENTES CON PROGRESIÓN LENTA DE LA INFECCIÓN”, que presenta el Biólogo y Máster en Virología D. Humberto Erick De La Torre Tarazona para obtener el grado de Doctor, ha sido realizada bajo nuestra dirección, autorizándola para su presentación al Tribunal Calificador.

Madrid, 23 de marzo de 2020.

Fdo. José Alcamí Pertejo  
Director de la tesis doctoral

Fdo. Francisco Díez Fuertes  
Director de la tesis doctoral

## AGRADECIMIENTOS

Durante 25 años de haber vivido en Perú y casi siete años en España, he encontrado por el camino a personas que forman parte de mi vida por haberme transmitido experiencias, alegrías, conocimientos y emociones, y a quienes les tendré una gratitud eterna y cariño infinito por haberme aportado tantísimo.

Quiero comenzar agradeciendo a Dios, porque tanto la vida y las cosas que pasan en ella, son un regalo suyo.

A la gente de mi laboratorio, la Unidad de Inmunopatología del Sida, o también conocido como comando IP. Al Dr. Pepe Alcamí, por aceptarme desde un principio para realizar el máster y después haberme dado la oportunidad de seguir con el doctorado, por transmitirme mucho conocimiento y siempre haberme apoyado en todo momento. A Olga, por su apoyo para cualquier cosa que necesité y ayudarme a buscar la solución de las cosas cuando parecen imposibles de hacerlas, eres una maestra. A Fran, por aceptar dirigirme la tesis, haberme ayudado mucho y ser ejemplo de una persona trabajadora que tiene mucha humildad en su forma de ser.

A mis amigas, Almu y Helen. Por haber compartido tantas vivencias con ustedes, dentro y fuera del lab. Muy aparte de que me han enseñado cosas del lab, con ustedes he llegado a mi top para hacer el bobo en tantas ocasiones, y vayas risas que nos hemos echado (hasta llorar en algunas ocasiones). Son personas geniales y les agradezco mucho su confianza y apoyo siempre que lo necesité. Y aunque mi amigo Raúl no es del lab, tenía que meterte en este párrafo, ya que eres un tío genial y te guardo mucha estima también.

A mi primer equipo cuando llegué a España, el de antivirales, Luismi y Manoli. Gracias por todo lo que me han enseñado con esa empatía y estilo tan desenfrenado que tienen, y también por despertar un poco de la locura que traía guardada cuando llegué a España.

A mis compis Javi, Merce, Esther, Laura, Mayte P, Rubén, etc, quienes me han enseñado mucho y también por ser tan majos en todo momento. Y cambiando de labo, mencionar a Ana V, quien con su enorme carisma se gana el cariño de la gente. Y en general, a todas las personas de ISCIII con quienes he charlado y han sido amables conmigo.

A la unidad *HIV, Inflammation et Persistence* del Instituto Pasteur de París, liderado por Michaela Müller y Asier Sáez, a quienes les agradezco por recibirme, y en especial a Caroline Passaes que me enseñó mucho con tanta amabilidad y paciencia durante la estancia que realicé en el 2018.

A todas esas personas con quienes he compartido fuera de mi trabajo. A personas que conocí en el Máster y que aún los tengo muy presentes por todos los buenos momentos compartidos: Guille, Giselle, Kathy, Aurora, Isabel, etc. A la gente que he podido conocer en la parroquia Buen Suceso y en el coro Iberoamericano de Madrid, con muchos de los cuales he pasado momentos muy buenos, en especial a mi amigo/hermano Carlitos, por tanto momento compartido.

A Norma y Richard, porque sin conocerme me recibieron en su bonita familia cuando llegué por primera vez a España, lo cual fue una bendición porque no conocía a nadie. A partir de allí formamos una bonita amistad y me han apoyado mucho durante estos años. Eso estaré eternamente agradecido.

A mi gente de Perú, a mis familiares (abuelas, tíos, primos) y todos esos grandes amigos que a pesar de la distancia y el tiempo que no nos vemos algunos, seguimos manteniendo esa amistad incondicional. Por mencionar a algunos de ellos: Lucy, Diego, Aldo, Roberto, Andy, Shinyoung, Ronald, etc. Así como a mis profes de la universidad, por haberme enseñado lo necesario para poder seguir con mi carrera profesional en España.

A quien conocí durante esta aventura que empecé en España, y quien le ha dado más felicidad a mi vida, a Ángela. Nos conocimos en situaciones únicas de nuestra vida, y hemos sabido superarlas con ayuda de Dios. Soy muy afortunado porque eres una bella persona, con buenos sentimientos y unas ganas enormes de aprender en la vida, lo cual admiro mucho. Nos casaremos pronto y no puedo estar más feliz por ello.

Y por último, y no menos importante por supuesto, a mi familia. A mis padres Humberto “Chicho” y Celi, por haberme dado todo en cuanto pudieron y haberme apoyado las decisiones que he ido tomando. Son unos padres maravillosos y soy muy feliz de tenerlos. A mis hermanos, Kevin y Fabri, con quienes tenemos tanta complicidad y somos como amigos. Son unos hermanos fabulosos, y en general me puedo regocijar de tener una familia tan buena. Les agradezco también que por primera vez vayan a salir de Perú para estar conmigo pronto y esto es un sueño para mí.

**A mis padres,  
A mis hermanos,  
A Ángela,  
A mis amigos,**

*“La ciencia sin religión está coja y la religión sin ciencia está ciega”.*

*Albert Einstein*

## RESUMEN

La infección por el virus de la inmunodeficiencia humana tipo 1 (VIH-1) continúa siendo un importante problema de salud pública a nivel global, debido al alto número de nuevas infecciones y muertes que ocasiona cada año, a pesar de los enormes esfuerzos realizados en la investigación y en la implementación de medidas preventivas y terapéuticas. Dentro de la población de personas infectadas por el VIH-1, existen grupos de individuos con fenotipos extremos de no progresión y/o control natural de la replicación viral que se denominan “no progresores a largo plazo” (LTNPs) y “controladores de élite” (ECs), respectivamente. El principal objetivo de esta tesis doctoral ha sido el estudio de estos grupos de pacientes desde diferentes abordajes tecnológicos de genotipado, así como de análisis transcriptómico. Los artículos incluidos en esta tesis quedan englobados en un amplio objetivo que persigue identificar nuevos factores asociados a los fenotipos LTNP y/o EC.

Los primeros resultados obtenidos mediante genotipado por técnicas convencionales en individuos VIH-positivos con fenotipo LTNP de una cohorte multicéntrica nacional, han identificado catorce polimorfismos genéticos asociados a una progresión lenta de la infección. El alelo “protector” HLA-B\*39 y los alelos de “riesgo” HLA-A\*24, -A\*29, -B\*08 y -B\*18 han sido asociados por primera vez con el fenotipo LTNP.

Aunque las técnicas convencionales de genotipado han permitido la identificación de variantes genéticas asociadas con diferentes fenotipos de individuos VIH-positivos, existen tecnologías de alto rendimiento que permiten el análisis de las variantes genéticas tanto a nivel del genoma completo como a nivel de exoma. Mediante la tecnología Infinium BeadChip hemos analizado los exomas de individuos LTNPs de la cohorte española LTNP-RIS, pudiéndose determinar una mayor asociación genética del SNP rs1127888 del gen *UBXN6* con el fenotipo mencionado. Además, mediante experimentos funcionales “*in vitro*” se han podido determinar que la interferencia en la expresión de *UBXN6* se asocia a un incremento en los niveles y la localización periférica de caveolina-1, y además a una disminución de la capacidad replicativa del VIH-1 en la línea celular HeLa y en cultivos primarios de células dendríticas y macrófagos.

Finalmente, evaluamos el transcriptoma de individuos LTNPs de la cohorte española LTNP-RIS, mediante la tecnología *RNA-seq*. Los resultados obtenidos demostraron una expresión génica diferencial entre los individuos vLTNPs (LTNPs virémicos), EC-LTNPs (LTNPs controladores de élite) y progresores típicos (TPs) de la infección por el VIH-1. Además, se ha identificado que la regulación positiva de genes relacionados con el

transporte y movilización del calcio puede estar asociada al control natural de la infección que presentan los individuos LTNPs que controlan la replicación viral (EC-LTNPs). Este estudio también nos ha permitido la identificación de genes y pseudogenes predictores de los fenotipos LTNP y TP mediante análisis bayesianos de clasificación supervisada, que pueden clasificar con una alta eficacia a ambos fenotipos de infección por el VIH-1. Asimismo, se ha identificado en los EC-LTNPs una correlación positiva de la expresión conjunta de los genes *CDKN1A*, *GADD45B*, *IER3* y *TNF*, en comparación con los otros grupos de pacientes incluidos en el estudio, lo cual sugiere que este mecanismo puede ser un factor importante en el control de la infección del VIH-1.

En conclusión, en esta tesis doctoral se han asociado diferentes variantes genéticas y niveles de expresión génica al fenotipo LTNP, identificando nuevos factores de protección frente a la progresión de la infección por el VIH-1.



## ABSTRACT

Human immunodeficiency virus type 1 (HIV-1) infection is still a major problem of global public health, due to the elevated number of new infections and deaths that causes every year, despite tremendous efforts in research and implementation of preventive and therapeutic approaches. In HIV-1 infected population, there are groups of individuals with extreme phenotypes of non-progression and/or natural control of infection called “long-term non progressors” (LTNPs) and “elite controllers” (ECs), respectively. The main goal of this doctoral thesis has been the study of these patients, through different genotyping techniques, as well as transcriptomic analysis. The articles included in this thesis are encompassed in a broad objective to search new factors associated LTNP and/or EC phenotypes.

The first results obtained using conventional genotyping techniques in HIV-positive individuals with LTNP phenotype from a national multicenter cohort, have identified fourteen genetic polymorphisms associated to slow disease progression. HLA-B\*39 "protective" allele and HLA-A\*24, -A\*29, -B\*08 and -B\*18 "risk" alleles, have been associated for first time to LTNP phenotype.

Although conventional genotyping techniques have allowed the identification of genetic variants associated with different phenotypes of HIV-positive individuals, there are high-throughput technologies that allow analysis of genetic variants both at complete genome and exome levels. Using Infinium BeadChip technology, we have analyzed exomes of LTNP individuals from the Spanish LTNP-RIS cohort, being able to determine a strong genetic association of the SNP rs1127888 of *UBXN6* gene to mentioned phenotype. In addition, *in vitro* functional assays have determined that interference in UBXN6 expression lead to an increasing in levels and peripheral location of caveolin-1, and also to a decrease in HIV-1 replicative capacity in HeLa cell line and primary dendritic cells and macrophages cultures.

Finally, we evaluated transcriptomes of LTNP individuals from the Spanish LTNP-RIS cohort using RNA-seq technology. The obtained results showed differential gene expression between the vLTNPs (viremic LTNPs), EC-LTNPs (elite controllers LTNPs) and typical progressors (TPs) phenotypes of HIV-1 infection. Furthermore, it has been identified that positive regulation of genes related to transport and mobilization of calcium may be associated to natural control of infection in LTNP individuals able to control viral replication (EC-LTNPs). This study has also allowed the identification of predictive genes and pseudogenes of LTNP and TP phenotypes, using supervised Bayesian classification analyzes that are able to classify both phenotypes of HIV-1 infection with high accuracy.

Likewise, a positive correlation in co-expression of *CDKN1A*, *GADD45B*, *IER3* and *TNF* genes, has been identified in EC-LTNPs, in comparison to other groups of patients included in this study, which suggest that this mechanism could be an important factor in the control of HIV-1 infection.

In conclusion, in this doctoral different genetic variants and gene expression levels have been associated to LTNP phenotype, identifying new protective factors against the progression of HIV-1 infection.

## ÍNDICE DE CONTENIDOS

AGRADECIMIENTOS.....	<i>i</i>
RESUMEN.....	<i>v</i>
ABSTRACT.....	<i>vii</i>
ABREVIATURAS.....	<i>xi</i>
1. INTRODUCCIÓN.....	1
1.1. Historia del sida y descubrimiento del VIH.....	3
1.2. Clasificación, origen y situación actual del VIH.....	3
1.3. Estructura y ciclo del virus.....	5
1.3.1. Entrada viral.....	6
1.3.2. Retrotranscripción, transporte al núcleo e integración del virus.....	7
1.3.3. Transcripción viral, síntesis de proteínas y liberación de viriones.....	7
1.4. Tratamiento antirretroviral y estrategias de curación.....	8
1.5. Etapas y progresión de la infección.....	9
1.5.1. Infección aguda.....	9
1.5.2. Infección reciente.....	10
1.5.3. Infección crónica.....	10
1.5.4. Fase avanzada.....	10
1.6. Fenotipos de progresión.....	10
1.6.1. No progresores a largo plazo (LTNPs).....	11
1.6.2. Controladores del VIH (HICs).....	12
1.6.3. Progresión inmunológica y viral en LTNPs y HICs.....	12
1.6.4. Reservorio viral en LTNPs y HICs.....	13
1.7. Factores asociados a los no progresores y controladores del VIH.....	13
1.7.1. Factores virológicos.....	14
1.7.2. Factores del hospedador.....	14
1.7.2.1. Factores de restricción.....	14
1.7.2.2. Citoquinas y receptores de citoquinas.....	15
1.7.2.3. Presentación antigénica.....	15
1.7.2.4. Funcionalidad de las células inmunitarias.....	17
1.7.2.5. Otros factores.....	18
1.8. Tecnologías “ómicas” en la investigación del VIH.....	19
1.8.1. Estudios de genómica.....	19
1.8.2. Estudios de transcriptómica.....	20
1.8.3. Otros estudios ómicos.....	21

2. OBJETIVOS .....	23
3. RESULTADOS .....	27
3.1. Artículo I: Novel association of five HLA alleles with HIV-1 progression in Spanish long-term non progressor patients .....	29
3.2. Artículo II: Association of a single nucleotide polymorphism in the <i>UBXN6</i> gene with long-term non-progression phenotype in HIV-positive individuals.....	49
3.3. Artículo III: Transcriptome sequencing of peripheral blood mononuclear cells from elite controller-long term non progressor .....	61
4. DISCUSIÓN .....	111
4.1. Determinación de polimorfismos mediante estudios de asociación genética .....	114
4.1.1. Genotipado de polimorfismos genéticos concretos en LTNPs .....	114
4.1.2. Genotipado de exoma de pacientes LTNP mediante técnicas de alto rendimiento.....	119
4.2. Identificación de nuevos patrones de expresión génica mediante análisis de transcriptoma en individuos LTNPs .....	125
4.2.1. Similitud y heterogeneidad entre LTNPs y TPs.....	126
4.2.2. Análisis de expresión diferencial entre fenotipos de infección y genes asociados al fenotipo LTNP .....	126
4.2.3. Genes predictores del fenotipo LTNP .....	127
4.2.4. Genes regulados por IFN y mecanismos celulares asociados .....	129
4.2.5. La regulación del calcio en los EC-LTNPs .....	130
4.2.6. Regulación y función de CDKN1A/p21 en el control viral.....	131
4.3. Consideraciones finales .....	132
5. CONCLUSIONES .....	135
6. REFERENCIAS .....	139
7. ANEXOS.....	155

## **LISTA DE ABREVIATURAS**

**ADN:** ácido desoxirribonucleico

**ARN:** ácido ribonucleico

**ARNm:** ARN mensajero

**CAV-1:** caveolina-1

**CDKN1A:** *cyclin dependent kinase 1a*; quinasa dependiente de ciclina 1a

**CTL(s):** *Cytolytic T Lymphocyte(s)*; linfocito(s) T citolítico(s)

**CV:** carga viral

**EC(s):** *elite controller(s)*; controlador(es) de élite

**EC-LTNP(s):** *elite controller(s)-LTNP(s)*; controladores de élite-LTNP(s)

**DC(s):** *dendritic cell(s)*; célula(s) dendrítica(s)

**DEGs:** differentially expressed genes; Genes expresados diferencialmente

**GWAS:** *genome wide association study*; estudios de asociación de genoma completo

**HLA:** *human leukocyte antigen*; antígeno leucocitario humano

**HLA-(I/II/III):** HLA de clase (I/II/III)

**HIC(s):** *HIV controller(s)*; controlador(es) del VIH

**IFN:** interferón

**ISG(s):** *interferon-stimulated gen(es)*; gen(es) estimulado(s) por interferón

**IL(s):** interleucina(s)

**KIR(s):** *Killer-cell immunoglobulin-like receptor(s)*; receptor(es) tipo inmunoglobulina de las células NKs

**LD:** *linkage disequilibrium*; desequilibrio de ligamiento

**LTNP(s):** *long-term non progressor(s)*; progresor(es) lento(s) a largo plazo

**MHC:** *major histocompatibility complex*; complejo mayor de histocompatibilidad

**NGS:** *next generation sequencing*, secuenciación de próxima generación

**NK(s):** *natural killer(s)*, asesina(s) natural(es)

**PBMCs:** *peripheral blood mononuclear cells*; células mononucleares de sangre periférica

**PCR:** *polimerase chain reaction*; reacción en cadena de la polimerasa

**qPCR:** PCR cuantitativa

**RNA-seq:** *RNA sequencing*; secuenciamiento de RNA

**RP(s):** *rapid progressor(s)*; progresor(es) rápido(s)

**Sida:** Síndrome de inmunodeficiencia adquirida

**SNP(s):** *single nucleotide polymorphism(s)*; polimorfismo(s) de nucleótido único

**TAR:** tratamiento antirretroviral

**TP(s):** *typical progressor(s)*; progresor(s) típico(s) del VIH

**TP-ART(s):** *typical progressor(s) receiving antiretroviral therapy*; progresor(es) típico(s) en tratamiento antirretroviral

**UBXN6:** *UBX domain containing protein 6*

**VC(s):** *viremic controller(s)*; controlador(es) virémico(s)

**VCP:** *valosin containing protein*; proteína que contiene valosina

**VIH:** Virus de la inmunodeficiencia humana

**vLTNP(s):** *viremic LTNP(s)*; LTNP(s) virémico(s)

**VNP(s):** *viremic non-progressor(s)*; no progresor(es) virémico(s)

# **INTRODUCCIÓN**

# 1. INTRODUCCIÓN

El VIH es un retrovirus que se ha adaptado a infectar y replicarse en linfocitos T CD4+, provocando una profunda inmunosupresión debido a la destrucción de estos linfocitos y a distintos mecanismos de interferencia con el sistema inmunitario. Como consecuencia, se producen infecciones oportunistas, desarrollo de tumores y afectación neurológica que definen el síndrome de inmunodeficiencia adquirida (sida). Desde su identificación hasta la actualidad, el VIH se ha convertido en uno de los agentes infecciosos más ampliamente estudiado. Sin embargo, aún persisten desafíos científicos como un mejor conocimiento de la biología del virus, la obtención de una vacuna eficaz o conseguir la cura de la infección.

## 1.1. Historia del sida y descubrimiento del VIH

El sida se describió por primera vez en 1981 en varones jóvenes homosexuales que padecían sarcoma de Kaposi y/o neumonía por *Pneumocystis jiroveci*. Estos pacientes presentaban un cuadro de inmunodeficiencia caracterizado por la disminución de linfocitos T CD4+ (Gottlieb *et al*, 1981). La identificación del VIH se produjo en 1983, cuando en el Instituto Pasteur de París se aisló un retrovirus a partir de nódulos linfáticos de un paciente con linfadenopatía, que fue denominado LAV (*lymphadenopathy associated virus*) (Barre-Sinoussi *et al*, 1983). Pocos meses después se aisló otros retrovirus de muestras de pacientes con sida en Estados Unidos, que se denominaron HTLV-III y ARV (*AIDS related virus*) (Levy *et al*, 1984; Popovic *et al*, 1984). Finalmente se denominó al virus que causaba dicha enfermedad como virus de la inmunodeficiencia humana (VIH). Dos años más tarde, se aisló un segundo tipo de virus (VIH-2) que también producía sida en pacientes procedentes de África Occidental (Clavel *et al*, 1986).

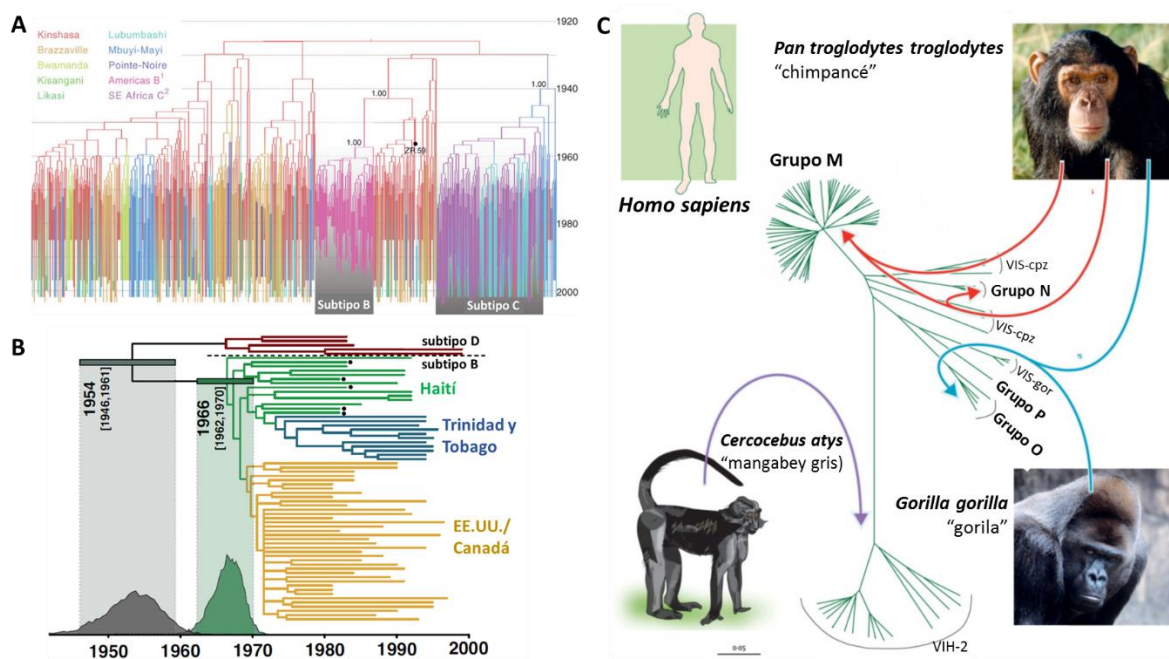
El comienzo de la expansión del VIH-1 y el VIH-2 fue a principios del siglo XX. Los estudios señalan que Leopoldville (actualmente Kinshasa, República Democrática del Congo) fue el foco de transmisión temprana del VIH, jugando un papel importante los cambios sociales y las redes de transporte para el establecimiento y diseminación del virus (Faria *et al*, 2014). Posteriormente, hacia los años 60s, el virus se expandió al continente americano causando una epidemia inicial en Haití. Luego se transmitió a Estados Unidos donde causó una epidemia silente durante aproximadamente doce años antes del comienzo de la pandemia (Gilbert *et al*, 2007) (Figura 1A, B).

## 1.2. Clasificación, origen y situación actual del VIH

Los retrovirus constituyen una familia compleja de virus ARN con siete géneros conocidos, perteneciendo el VIH al género *Lentivirus*. La principal característica de estos virus es la de infectar de una manera crónica a una gran variedad de especies de mamíferos como primates, ungulados y felinos (Weiss, 1996).



El origen del VIH se ha establecido en diversas transmisiones zoonóticas a partir de lentivirus de primates africanos no-humanos (VIS, virus de la inmunodeficiencia del simio), (Figura 1C), posiblemente como resultado de la caza y manipulación de carne de simio. Los grupos M y N del VIH-1 proceden de la transmisión de un VIS del chimpancé originario de África Occidental-Central. Los grupos O y P, proceden de la misma región y se cree que derivan de virus de procedentes de gorila o chimpancé. El VIH-1 circula en Asia, Europa, Oceanía, América y África (Gao *et al*, 1999; Revisado por Tebit y Arts, 2011). El VIH-2 tendría su origen en un virus procedente del mangabey tiznado de África Occidental, y circula en África Central y Occidental (Lemey *et al*, 2003).



**Figura 1.** Origen de la epidemia del VIH. A) Expansión del VIH a principios del siglo XX desde Kinshasa (Imagen adaptada de *Faria, Science, 2014*). B) Evolución de la epidemia del VIH en América a partir de 1960 en Haití (Imagen adaptada de *Gilbert, PNAS, 2007*). C) Procedencia de los grupos M, N, O y P del VIH-1 y del VIH-2 a partir de transmisiones zoonóticas de SIVs provenientes de primates no humanos (Imagen adaptada de *Tebit, Lancet, 2011*).

El VIH se caracteriza por presentar una elevada diversidad genética. El VIH-1 se clasifica en cuatro grandes grupos (M, N, O y P) y el VIH-2 en ocho (A, B, C, D, E, F, G y el recombinante AB). El grupo M del VIH-1 es el responsable de la pandemia y es el único que se subdivide en 11 subtipos no recombinantes (A1, A2, B, C, D, F1, F2, G, H, J y K) y en 102 formas recombinantes circulantes (CRFs) descritas hasta la fecha por recombinación entre dichos subtipos (Revisado por Tebit y Arts, 2011; Folley *et al*, 2019).

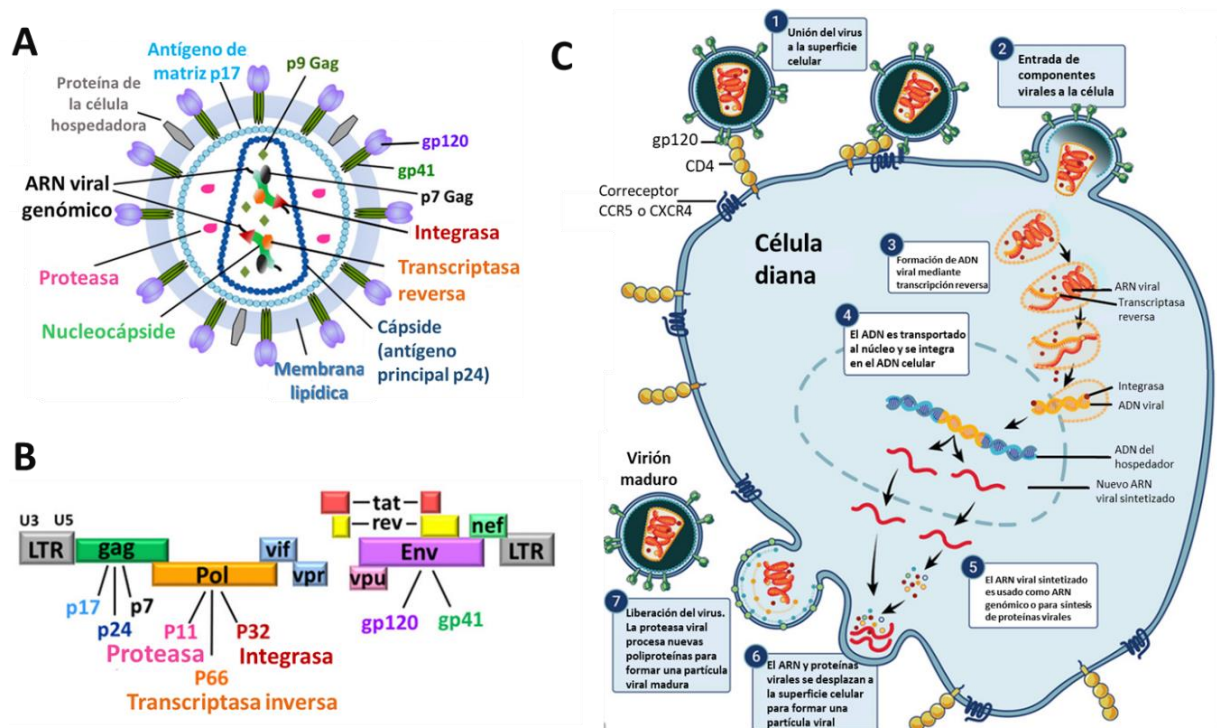
Actualmente 37,9 millones de personas están infectadas por el VIH en el mundo. Además, se han reportado 1,7 millones nuevas infecciones y 770.000 muertes causadas por enfermedades relacionadas con el sida en el año 2018. En España, se estima que

150.000 personas están infectadas por el VIH y cada año se reportan aproximadamente 3.000 nuevos casos (UNAIDS, 2019).

### 1.3. Estructura y ciclo del virus

El virión es aproximadamente esférico y mide entre 80-120 nm de diámetro. La envoltura viral está formada por una bicapa lipídica (con origen en la célula huésped) y contiene proteínas que se organizan en espículas, las cuales están formadas por tres glicoproteínas de superficie (gp120) y tres glicoproteínas transmembrana (gp41). La cápsida es cónica y está compuesta por la proteína p24, y está rodeada por una matriz compuesta de la proteína p17 que garantiza la integridad de la partícula viral. El ARN está asociado a las proteínas de la nucleocápsida p6 y p7, para protegerlo de la acción de las nucleasas, así como a los enzimas necesarios para la propagación del virus: transcriptasa inversa, proteasa e integrasa (Revisado por Li y De Clercq, 2016) (Figura 2A).

El virus posee dos copias de ARN de polaridad positiva con un tamaño aproximado de 10.000 pb (pares de bases). El ARN genómico se compone de siete elementos estructurales (LTR, TAR, RRE, PE, SLIP, CRS y INS) (Revisado por Karn y Stolzhus, 2012) y nueve genes que codifican para diecinueve proteínas en total. Los genes que codifican las proteínas estructurales (*gag*, *pol* y *env*) se encuentran en todos los retrovirus, mientras los seis restantes (*vpu*, *vpr*, *vif*, *nef*, *tat*, *rev*) codifican las proteínas reguladoras y accesorias que están implicadas en la capacidad infectiva, producción de nuevas partículas virales y/o la patogénesis del VIH (Revisado por Li y De Clercq, 2016) (Figura 2B, Tabla 1).



**Figura 2.** Características generales del VIH. A) Estructura de la partícula viral, B) Organización genómica del virus (Imágenes adaptadas de *Musumeci, Molecules, 2015*). C) Ciclo replicativo del VIH (Imagen adaptada de NIAID - NIH)

**Tabla 1:** Genes del VIH y función de las proteínas codificadas

Genes virales	Funciones de las proteínas codificadas	
<i>gag</i>	Codifica las proteínas de la cápsida. La proteína precursora p55, luego es procesada por la proteasa viral en las proteínas: p17 (matriz), p24 (core/cápside), p6 y p7 (nucleocápsida) (Revisado por Mailler <i>et al</i> , 2016).	Proteínas Estructurales
<i>pol</i>	Codifica enzimas virales: proteasa, transcriptasa inversa, RNasa e integrasa. Estas enzimas son producidas como la poliproteína precursora Gag-Pol, que es procesada por la proteasa viral (Revisado por Mailler <i>et al</i> , 2016).	
<i>env</i>	Codifica las proteínas de la envoltura viral. Es producida como un precursor (gp160) que es procesado para originar el complejo formado por la glicoproteína externa gp120 y la glicoproteína transmembrana gp41 (Revisado por Checkley <i>et al</i> , 2011).	
<i>tat</i>	Codifica la proteína Tat que está localizada primariamente en el núcleo. Se une al segmento de ARN de la región TAR ( <i>trans-activation response</i> ) y permite la elongación del mRNA a partir del promotor LTR (Revisado por Rice, 2017).	Proteínas reguladoras
<i>rev</i>	Codifica la proteína Rev que está localizada primariamente en el núcleo, aunque también actúa en el citoplasma. Se une al RRE ( <i>Rev response element</i> ) promoviendo la exportación nuclear, estabilización y utilización de ARNm viral no procesado (Revisado por Dayton, 2004).	
<i>vpu</i>	Codifica la proteína vpu. Es una proteína integral de membrana que participa en la degradación de CD4 en el retículo endoplasmático (RE) y el incremento de la liberación del virión de la membrana plasmática de las células infectadas mediante bloqueo de la proteína celular teterina (Neil <i>et al</i> , 2008).	Proteínas accesorias
<i>nef</i>	Codifica la proteína Nef que es una de las primeras proteínas del VIH producida en células infectadas. Posee una extraordinaria flexibilidad con diversas funciones y es una de las proteínas accesorias más inmunogénicas (Revisado por Basmaciogullari y Pizzato, 2014)	
<i>vpr</i>	Codifica la proteína Vpr que es incorporada en el virión y participa en la importación nuclear de los complejos de preintegración, la parada en la fase G2 del ciclo celular, la inducción de la apoptosis, la transactivación de genes celulares y la inducción de la diferenciación celular (Revisado por González, 2017).	
<i>vpx</i>	Codifica la proteína Vpx en el VIH-2, esencial para la replicación viral en los macrófagos y en las células T (Berger <i>et al</i> , 2009).	
<i>vif</i>	Codifica el factor infectivo viral (Vif), que está implicado en la protección del genoma viral impidiendo la encapsidación de dos potentes factores antivirales, las citidindeaminasas APOBEC3G y APOBEC3F, que inducen mutaciones en el genoma del virus (Revisado por Rose <i>et al</i> , 2004).	

El VIH tiene la capacidad de infectar diversos tipos celulares como macrófagos, células dendríticas (DCs), pero su diana principal son los linfocitos T CD4+ que albergan una replicación más eficiente del virus (Revisado por Wilen *et al*, 2012) (Figura 2C).

### 1.3.1. Entrada viral

La entrada del VIH en la célula se produce mediante la interacción de la proteína viral gp120 con el receptor celular CD4, que induce una serie de cambios conformacionales que exponen el dominio V3, y otras regiones adyacentes, que forman el dominio de unión de la gp120 a los receptores de quimiocinas CCR5 o CXCR4. Esta segunda interacción induce nuevos cambios en la estructura de la gp41, la cual se ancla en la membrana plasmática y produce la fusión entre las membranas viral y celular (Revisado por Wilen *et al*, 2012). La entrada del virus mediante los correceptores CCR5 y/o CXCR4 define el tropismo viral como R5, X4 o R5X4 (Berger *et al*, 1998; Revisado por Naif, 2013).

Además de estos dos receptores virales, las DCs presentan en su superficie las lectinas DC-SIGN y L-SIGN (Geijtenbeek *et al*, 2000), y otras moléculas como SIGLEC-1 (Izquierdo-Useros *et al*, 2016), que se unen de forma inespecífica al VIH. Este fenómeno

hace de la interacción entre DCs y linfocitos, denominada sinapsis inmunitaria, una zona preferente de propagación del VIH que facilita e incrementa la infección de los linfocitos T CD4+ circundantes (Revisado por Wilen *et al*, 2012). Se han reportado también que otros receptores de quimiocinas como: CCR2b, CCR3 (Doranz *et al*, 1996), CCR8 (Tiffany *et al*, 1997), CX3CR1 (Roman *et al*, 2001), CXCR6 (Limou *et al*, 2010), entre otros, participan como correceptores secundarios en la infección por el VIH-1.

### **1.3.2. Retrotranscripción, transporte al núcleo e integración del virus**

El proceso de síntesis de ADN a partir del ARN viral (o retrotranscripción) es realizado por el complejo enzimático de la transcriptasa inversa que es más efectiva tras la activación de la célula infectada, debido a que se inducen mayores niveles de nucleótidos y la acción de factores celulares (Revisado por Hu y Hughes, 2012). La nucleocápside es transportada y se desensambla a medida que progresa en el citosol, y se libera el material genético en el poro nuclear (Fernandez *et al*, 2019).

En el núcleo, el ADN viral sintetizado se acopla a una serie de factores celulares y virales (Vpr, LEDGF, CPSF6) para formar el complejo de preintegración. Luego se integra en el genoma del hospedador, constituyendo la forma proviral del VIH (Revisado por Craigie y Bushman, 2012). Después de la integración, el VIH puede permanecer latente, replicarse de forma controlada o replicarse masivamente, en función de la actividad de múltiples factores moleculares (Revisado por Coiras *et al*, 2009) y/o de la programación metabólica de las células linfocitarias (Valle-Casuso *et al*, 2019).

Las células latentemente infectadas en los diferentes compartimentos anatómicos, conforman los llamados reservorios virales. Los linfocitos T CD4+ con memoria latente son el compartimento celular predominante responsable de la persistencia viral, pero algunos estudios sugieren que las células mieloides y los progenitores hematopoyéticos, también pueden servir como reservorios virales a largo plazo. Además, se han descrito varios mecanismos que participan en el mantenimiento y perpetuación de los reservorios virales, como la replicación a bajo nivel, la expansión clonal y la proliferación homeostática (Chomont *et al*, 2009; Revisado por Kuo y Litcherfeld, 2019).

### **1.3.3. Transcripción viral, síntesis de proteínas y liberación de viriones**

La iniciación de la transcripción del genoma viral depende de factores celulares (principalmente NF- $\kappa$ B, NFAT, SP1) que producen una primera expresión del ARN viral no codificante, en concreto de la región TAR, a la que se une la proteína viral Tat que permite la elongación completa del ARN viral (Buonaguro *et al*, 1994). El ARNm del VIH se sintetiza en forma de un único transcrito que debe ser transportado al citosol y procesado en ARN de distinto tamaño. El procesamiento y transporte son regulados por Rev, que participa además en el acoplamiento de los ARNm a los ribosomas para la síntesis de las proteínas virales, que son procesadas antes de ensamblarse para constituir las partículas virales

(Revisado por Dayton, 2004). La proteína teterina secuestra los viriones en la membrana celular durante la gemación del virus, pero la proteína Vpu del VIH-1 disminuye los niveles de teterina en la superficie celular y permite la liberación de los viriones al exterior de la célula (Neil *et al*, 2008). La maduración de los viriones y el ensamblaje correcto de las proteínas de la cápside se producen durante la gemación a través de la membrana celular, mediante la acción de la proteasa viral que procesa las poliproteínas Gag y Gag-Pol y permite la formación de partículas virales maduras (Revisado por Freed, 2015).

#### **1.4. Tratamiento antirretroviral y estrategias de curación**

Los diversos fármacos antirretrovirales disponibles en la actualidad van dirigidos a interferir las diferentes etapas del ciclo viral. Se clasifican en: inhibidores de la entrada viral, inhibidores de la fusión, inhibidores de la retrotranscripción análogos/ no análogos de nucleósido/nucleótido, inhibidores de la integrasa e inhibidores de la proteasa (Zhang, 2018). Asimismo, se encuentran en fase clínica nuevos fármacos dirigidos contra la cápsida viral (Cevik y Orkin, 2019) o la proteína Tat (Mousseau y Valente, 2017), así como en investigación diversos compuestos con potente capacidad antiviral contra el VIH-1 (De La Torre-Tarazona *et al*, 2020). Desde 1996, la terapia antirretroviral (TAR) se basa en la administración de 3 fármacos dirigidos contra diferentes etapas del ciclo viral (Carpenter *et al*, 1996), la cual ha ido reduciendo la tasa mortalidad asociada a la infección por el VIH en el transcurso de los años (Crum *et al*, 2006). Actualmente, las guías clínicas recomiendan iniciar la TAR al detectarse la infección por el VIH, ya que en ensayos clínicos del inicio precoz del tratamiento se ha observado un mayor beneficio para los pacientes, como la reducción de complicaciones clínicas, apariciones de cáncer y muertes asociadas al sida (Saag *et al*, 2018).

Sin embargo, aunque la TAR controla la replicación viral de manera eficaz, no es posible curar la enfermedad mediante su administración. El establecimiento y perpetuación de los reservorios virales constituyen los principales obstáculos para lograr la cura del VIH-1 (Revisado por Chun *et al*, 2015), por lo que se han propuesto diversas estrategias dirigidas a disminuir los reservorios virales y/o curar la infección.

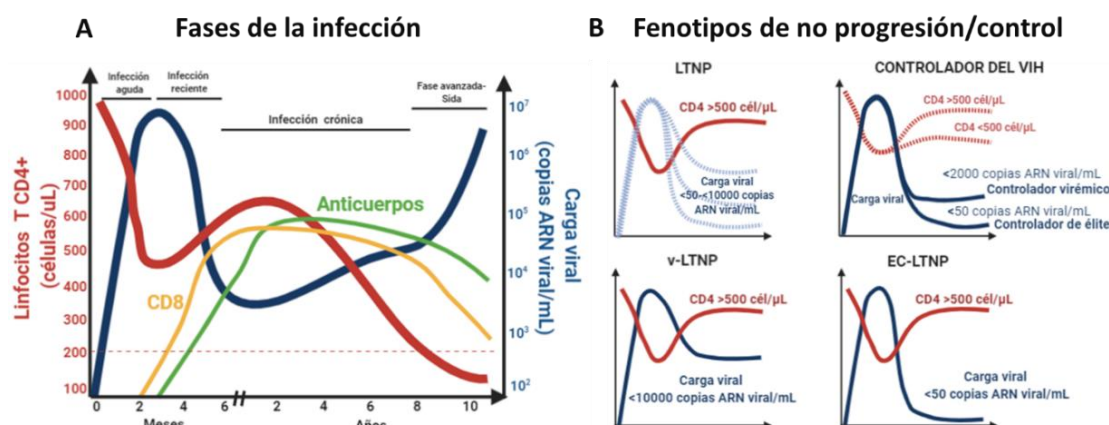
La estrategia de cura esterilizante tiene por objetivo eliminar el reservorio como en los pacientes de Berlín (Hütter *et al*, 2009) y de Londres (Gupta *et al*, 2019), quienes luego de haber recibido un trasplante de médula ósea proveniente de pacientes con la mutación *CCR5-Δ32* homocigótica, han logrado mantener cargas virales (CVs) indetectables tras la interrupción de la TAR. Se ha descrito que los polimorfismos genéticos *CCR5-Δ32* homocigótico, que consiste en la delección de 32 pb en *CCR5*, y la delección heterocigótica del codón de parada de *TNPO3* (transportina-3), confieren protección a la infección por el VIH-1 (Liu *et al*, 1996; Rodríguez-Mora *et al*, 2019).



La estrategia de cura funcional tiene por objetivo reducir los reservorios y lograr un control inmunológico, mediante propuestas como “activar y matar” (*shock and kill*) (Deeks, 2012), vacunas terapéuticas (Sneller *et al*, 2019), o una combinación entre ambas (Leth *et al*, 2016), con resultados aún poco prometedores en los ensayos clínicos realizados. También se ha propuesto que la caracterización de los factores involucrados en el control natural de la infección que presentan un grupo reducido de individuos VIH-positivos, puede ser útil para mimetizar un modelo de cura funcional (Autran *et al*, 2011).

### 1.5. Etapas y progresión de la infección

Las infecciones por el VIH-1 se transmiten mayoritariamente por vía sexual (vaginal o anal), aunque también puede transmitirse por contacto directo con sangre o por vía materno-fetal (Patel *et al*, 2014). El curso convencional de la infección consiste en una primoinfección seguida generalmente de un largo periodo clínicamente silente o solo con complicaciones menores, hasta que aparece alguna de las infecciones o neoplasias oportunistas definitorias de sida (Revisado por Naif, 2013; Figura 3A).



**Figura 3:** Fases y progresión de la infección por el VIH. A) Fases de la infección en una progresión típica. B) Fenotipos de no progresión/control del VIH. LTNP: no progresor a largo plazo, vLTNP: LTNP virémico, EC-LTNP: LTNP controlador de élite.

#### 1.5.1. Infección aguda

Esta fase puede durar de 1 a 3 semanas, pudiendo ser asintomática o presentar síntomas que se asemejan a un síndrome gripal o a una mononucleosis infecciosa aguda con adenopatías y eritema cutáneo. La mayoría de las veces, la infección de las primeras dianas se produce por una cepa única o por unas pocas (virus fundadores). Durante la infección aguda, la concentración de virus libre circulante es muy elevada y la CV suele oscilar entre las 100.000 y millones de copias de ARN del VIH/mL de plasma. Además, se produce de manera muy precoz, en menos de una semana, la diseminación del virus a los órganos linfoides centrales (Revisado por Cohen *et al*, 2011; Robb *et al*, 2016).

La respuesta inmunitaria con más actividad durante esta fase está mediada por la inmunidad innata, sin embargo es rápidamente sobrepasada por el virus. La elevada

replicación del VIH en esta fase causa una depleción muy rápida, masiva y difícilmente reversible de los linfocitos T CD4+ de memoria efectora en las mucosas, principalmente la mucosa intestinal (GALT), por un mecanismo citolítico directo del virus o por apoptosis de los linfocitos. Además, se produce la diseminación y formación de los reservorios virales, tanto en diferentes tipos celulares como en compartimentos o santuarios anatómicos (Revisado por Cohen *et al*, 2011, Robb *et al*, 2016).

### **1.5.2. Infección reciente**

Esta fase puede durar desde la cuarta semana hasta los 6 meses. La viremia disminuye debido a la generación de respuestas inmunes específicas pero aumenta la variabilidad del virus. La inmunidad adaptativa que se genera, está mediada por generación de anticuerpos por parte de los linfocitos B y la actividad citotóxica de los linfocitos T CD8+, la cual causa la destrucción de linfocitos T CD4+. Sin embargo, este control es incompleto en la mayoría de los pacientes (Revisado por Cohen *et al*, 2011).

### **1.5.3. Infección crónica**

Esta fase puede extenderse desde los 6 meses hasta los 8 años. Se da una viremia controlada y persistente, generándose variantes virales de escape a las respuestas celulares y humorales. Se produce una hiperactivación e inflamación crónicas durante la progresión de la enfermedad, que alteran el sistema inmunológico y contribuyen a la disminución progresiva de los linfocitos T CD4+ (promedio de 50 células/  $\mu$ L de sangre por año). La proliferación del virus en el tejido linfático es más focal y se produce esencialmente en los centros germinales ganglionares (Alcamí y Coiras, 2011).

### **1.5.4. Fase avanzada**

Los estadios finales de la enfermedad se caracterizan por el incremento progresivo de la CV con la emergencia de variantes con un tropismo X4 o dual R5X4, que conllevan al descenso del número de linfocitos T CD4+ por debajo de un nivel que impide una respuesta rápida y eficaz a los antígenos de memoria (< 200 células/  $\mu$ L de sangre periférica), así como al deterioro de la respuesta humoral y celular debido al agotamiento del sistema inmunológico. Clínicamente se caracteriza por la aparición de infecciones oportunistas que definen la fase sida (Alcamí y Coiras, 2011)

## **1.6. Fenotipos de progresión en la infección por el VIH**

Las fases de la infección por el VIH-1 (descritas anteriormente) pueden desarrollarse en diferentes tiempos y conducir a la clasificación de fenotipos, basados en la progresión viral (niveles de CV) y/o inmunológica (recuento de linfocitos T CD4+).

Los progresores típicos (TPs) (Figura 3A), que representan el 70 a 80% de las personas infectadas, experimentan una progresión intermedia de la enfermedad en la que tienen un

aumento de la CV, disminución de las células T CD4+ y el desarrollo de enfermedades relacionadas con el sida dentro de los 6 a 10 años posteriores a la adquisición del VIH (Revisado por Langford *et al*, 2007). Los progresores rápidos (RPs) comprenden aproximadamente el 15% de las personas infectadas y presentan una progresión acelerada de la infección, y se ha descrito que la determinación de niveles muy bajos de linfocitos T CD4+ durante el primer año de seroconversión es indicador de este grupo de individuos (Revisado por Langford *et al*, 2007; Audige *et al*, 2010; Olson *et al*, 2014). Por otra parte, también se ha descrito un grupo de pacientes capaces de controlar la infección después de la interrupción de la TAR, a los cuales se les denomina controladores post-tratamiento (PTCs, *post-treatment controllers*) (Van Gulck *et al*, 2011).

Los individuos no progresores a largo plazo (LTNPs) y controladores del VIH-1 (HICs) pueden controlar la infección por el VIH, mediante la capacidad de mantener altos niveles de linfocitos T CD4+ y controlar la replicación viral, respectivamente, sin recibir TAR y en ausencia de eventos relacionados al sida, generalmente durante varios años después de la infección (Figura 3B). Están representados por un grupo muy heterogéneo de pacientes y forman poblaciones poco frecuentes (Gurdasani *et al*, 2014). Además, ni el género ni las vías de transmisión se han asociado a este grupo de individuos VIH-positivos (Okulicz *et al*, 2009). De manera general, las definiciones LTNPs y HICs engloban a estos pacientes con fenotipos extremos del control de la infección por el VIH, pero se encuentran en la literatura hasta 600 definiciones y 26 términos para denominar a los individuos con fenotipos extremos de no progresión y/o control viral, los cuales se definen por diferentes criterios de CV, recuentos de células T CD4+ y tiempo de seguimiento (Gurdasani *et al*, 2014). La comprensión de los mecanismos involucrados puede ser esencial para lograr la remisión del VIH-1 a largo plazo o nuevas estrategias de cura del VIH.

### **1.6.1. No progresores a largo plazo (LTNPs)**

La identificación de estos pacientes data desde mediados de la década de 1990 (Klein *et al*, 1995; Buchbinder *et al*, 1999). Se considera que entre el 1-5% del total de pacientes infectados por el VIH-1 son LTNPs (Revisado por Sabin y Lundgren, 2013). Estos individuos permanecen asintomáticos y su cifra de linfocitos T CD4+ en sangre periférica es mayor de 500 células/ $\mu$ L, aunque otros estudios consideran un recuento por encima de 350 o 600 células/ $\mu$ L. El tiempo de seguimiento se ha establecido entre 7 a 10 años desde la detección de la infección (Revisado por Sabin y Lundgren, 2013).

La CV que se considera para la clasificación de estos individuos es generalmente de menos de 10.000 copias de ARN viral/mL en plasma (Revisado por Poropatich y Sullivan, 2010). La cohorte española LTNP-RIS clasifica a los LTNPs en virémicos (vLTNPs) con los criterios de un recuento de células CD4+ mayor de 500 / $\mu$ L y una CV detectable pero menor de 10.000 copias/mL durante 10 años de seguimiento (García-Merino *et al*, 2009). Por lo



general, el control de la replicación viral no es completa en estos individuos, por lo que podrían poseer mecanismos que los protejan contra la pérdida de células T CD4+ y la inmunodeficiencia inducida por el VIH (Sáez Cirión *et al*, 2014).

### 1.6.2. Controladores del VIH (HICs)

La identificación de los HICs se dio después de la introducción de la prueba de CV, que evidenció que sólo algunos de los pacientes considerados LTNPs eran capaces de controlar la infección (Rodés *et al*, 2004). Los HICs pueden mantener una CV baja o indetectable durante periodos prolongados sin TAR, y se clasifican en controladores de élite (ECs) cuando poseen CV indetectable, o controladores virémicos (VCs) si poseen entre 50-2.000 copias de ARN viral/mL de plasma (Revisado por Hunt, 2009; Dominguez-Molina *et al*, 2016). La mayoría de los HICs mantienen un recuento de células T CD4+ > 500/ $\mu$ L, pero una pequeña proporción de estos individuos pierden progresivamente estos linfocitos (Pereyra *et al*, 2009), quizás consecuencia de la activación y senescencia inmunológica derivada de la infección (Hatano *et al*, 2009). La mayoría de las definiciones consideran que alrededor del 1% de las personas infectadas por el VIH son ECs (Olson *et al*, 2013). El Consorcio Internacional de Controladores del VIH ha establecido para la clasificación de ECs: VIH-positivos durante más de un año y mínimo tres determinaciones consecutivas de ARN del VIH <75 copias/mL durante este período (Deeks y Walker, 2007). Sin embargo, en estos pacientes es posible detectar viremia con técnicas ultrasensibles y además aislar virus con capacidad replicativa (Buckheit *et al*, 2012).

La cohorte española ECRIS define como ECs a los individuos con una CV < 50 copias/mL durante al menos 12 meses de seguimiento (Dominguez-Molina *et al*, 2016). Sin embargo, la cohorte española LTNP-RIS clasifica como controladores de élite LTNPs (EC-LTNPs) a individuos con >500 células CD4/ $\mu$ L y una CV < 50 copias/mL, durante al menos 10 años de seguimiento (Casado *et al*, 2010). La coexistencia de las condiciones EC y LTNP observada en estos últimos representa el fenotipo más beneficioso contra la infección por el VIH-1 (Madec *et al*, 2005).

### 1.6.3. Progresión inmunológica y viral en LTNPs y HICs

Por lo general, se ha asociado una mejor función tímica en LTNPs y HICs en comparación con los progresores, que mejora la capacidad de producir células y mantener un recuento elevado de linfocitos T CD4+ (Revisado por Gaardbo *et al*, 2012; Yang *et al*, 2012). Por otra parte, la activación, apoptosis y senescencia de células del sistema inmunológico se expresa de manera distinta entre HICs y LTNPs, por lo que diferentes mecanismos podrían ser responsables de los recuentos conservados de células T CD4+ que presentan estos grupos de individuos (Gaardbo *et al*, 2013).

Se han reportado LTNPs y ECs que pueden controlar la infección durante más de 25 años (Mikhail *et al*, 2003; Casado *et al*, 2020). Sin embargo, hasta el 71% de los HICs

pueden perder el control y desarrollar una progresión inmuno/virológica en 10 años de seguimiento (van der Helm *et al*, 2014). La progresión inmunológica ha sido determinada principalmente en aquellos HICs que experimentan repuntes de la CV (Boufassa *et al*, 2011), aunque también pueden descender los recuentos de células CD4+ sin un incremento de la CV (Leon *et al*, 2016). Además, la viremia de bajo nivel en HICs se asocia con una disminución lenta de los recuentos de células T CD4+ durante el curso de la infección (Hatano *et al*, 2009).

Por lo general, se ha observado un menor riesgo de progresión al sida en los ECs comparado con los LTNPs y VCs (Okulicz *et al*, 2009). Sin embargo, en la cohorte CASCADE se observó que en ECs con más de 16 años de seguimiento, el 15% tuvieron recuentos de linfocitos T CD4+ <350 células/ $\mu$ L y el 7% desarrollaron sida (Madec *et al*, 2005). Por lo tanto, la progresión inmuno/virológica en HICs es frecuente y los factores involucrados serían: el corto tiempo de seguimiento, riesgo de transmisión sexual del VIH, bajo nadir (mínimo recuento de células T CD4+) u otras infecciones virales (León *et al*, 2016). Además, antes de la pérdida del control se observa una pérdida de funcionalidad, disminución de la capacidad de supresión viral y elevación de los niveles de marcadores de activación/agotamiento de las células T, y además aumentan los niveles de citoquinas proinflamatorias y la diversidad viral (Pernas *et al*, 2018; Rosas-Umbert *et al*, 2019). La TAR está recomendada en LTNP/ ECs que presentan pérdida de linfocitos T CD4+, elevada activación inmunológica, pérdida del control viral y/o aparición de comorbilidades asociadas al sida (Okulicz *et al*, 2010; Bansal *et al*, 2015; Noël *et al*, 2019).

#### **1.6.4. Reservorio viral en LTNPs y HICs**

En PBMCs de LTNPs y ECs se han encontrado niveles significativamente más bajos de ADN proviral, en comparación con otros grupos de pacientes infectados (Graf *et al*, 2011; Mendoza *et al*, 2012). También se ha determinado que las células CD4+ de ECs presentan un reservorio viral más reducido, específicamente en células de memoria en reposo y Tfh (*follicular helper*) (García *et al*, 2017). Además, un estudio reciente muestra una carga proviral 50 veces menor en células CD4+ de un grupo especial de ECs (más de 30 años de seguimiento) comparado con individuos en TAR (Casado *et al*, 2020). La baja carga proviral de estos individuos puede ser el resultado de variantes génicas del hospedador que influyen en la generación del ADN viral durante la entrada, replicación, tráfico interno viral e integración del VIH (Nissen *et al* 2018).

#### **1.7. Factores asociados a los no progresores y controladores del VIH**

Los factores que pueden reducir la gravedad de las etapas primarias de infección en los LTNPs y HICs incluyen factores del huésped (inmunidad innata y adaptativa), así como factores virales (cepas infectantes de virulencia reducida).

### 1.7.1. Factores virológicos

Los primeros reportes sobre factores virológicos asociados a fenotipos no progresores fueron descritos en la cohorte de Sidney, donde los pacientes habían sido infectados por una cepa mutante altamente atenuada con una delección en la región *nef*/LTR (Deacon *et al*, 1995), por lo que se asoció en un principio el fenotipo LTNP a la infección por virus defectivos y atenuados (Michael *et al*, 1995). Posteriormente, se fueron reportando en otras cohortes diferentes mutaciones en los genes *nef* (Mariani *et al*, 1996; Casartelli *et al*, 2003), *vpr* (Mologni *et al*, 2006) y *rev* (Papathanasopoulos *et al*, 2003) del VIH-1. Además, se ha reportado en ECs una menor actividad anti-APOBEC3G de Vif comparada con la de individuos no controladores (Kikuchi *et al*, 2015).

Las proteínas de la envoltura viral derivadas de HICs muestran una disminución significativa de la eficiencia para la entrada del virus (Lassen *et al*, 2009). Asimismo, las envolturas virales de un grupo EC-LTNPs mostraron una unión ineficaz a CD4, lo cual altera la conformación de actina/tubulina del citoesqueleto en comparación con las envolturas virales de individuos con infección crónica (Casado *et al*, 2018). Sin embargo, las envolturas virales de individuos virémicos no progresores (VNPs) (> 400 células CD4+/ $\mu$ L; CV >10 000 copias, promedio de 6 años de seguimiento) poseen la misma funcionalidad que las derivadas de individuos RPs (Cabrera-Rodríguez *et al*, 2019).

Por otra parte, se ha descrito que los LTNPs y HICs presentan virus con una reducida capacidad replicativa o *fitness*, que puede contribuir a la supresión viral en estos individuos (Tolstrup *et al*, 2006; Miura *et al*, 2010). Además, se han descrito pacientes ECs que presentan todas las secuencias provirales defectivas (Liang *et al*, 2019) y provirus integrados en zonas silentes del genoma con escasa actividad transcripcional, lo que explicaría la replicación defectiva del virus en estos individuos (Lichterfeld, 2019).

Aunque se ha demostrado que los virus defectivos o atenuados pueden ser un factor determinante del fenotipo no progresor y/o controlador en una fracción de pacientes, se ha descrito que la mayoría de éstos están infectados por virus con genes funcionales, replicación competente y capacidad patogénica (Blankson *et al*, 2007; Lamine *et al*, 2007).

### 1.7.2. Factores del hospedador

#### 1.7.2.1. Factores de restricción

Se han descrito factores celulares asociados con el control viral en las células diana de los fenotipos no progresor y/o controlador. Se ha reportado una elevada expresión de APOBEC3G asociada inversamente con la CV en individuos LTNPs, comparado con TPs sin TAR (Jin *et al*, 2005). Además, niveles superiores de APOBEC3G/F en LTNPs, se asocian a mayor índice de hipermutación viral y menor carga proviral (Kourteva *et al*, 2012). Por otra parte, se ha reportado la variante R136Q de Trim5 $\alpha$  en todos los pacientes de una cohorte africana de LTNPs (Dambaya *et al*, 2019).

La expresión de *Schlafen 11*, que actúa a nivel traduccional, es mayor en linfocitos T CD4+ de ECs comparado con no controladores (Abdel-Mohsen *et al*, 2013). Además, una mayor expresión de *CDKN1A/p21* en células T CD4+ de ECs controla la replicación viral, mediante la inhibición de las ciclinas CDK2 y CDK9 que participan en la retrotranscripción y transcripción viral, respectivamente (Chen *et al*, 2011; Leng *et al*, 2014).

Sin embargo, estos factores de restricción que pueden reducir la capacidad replicativa del VIH-1 no han sido observados en todas las cohortes estudiadas.

### 1.7.2.2. Citoquinas y receptores de citoquinas

Estudios realizados en diferentes cohortes de pacientes no progresores y HICs, han mostrado el efecto “protector” del alelo *CCR5Δ32* en el control de la CV y la progresión a sida (Fellay *et al*, 2009; Pereyra *et al*, 2010). Además, el polimorfismo *CCR5-59029 G/G* también es más frecuente en los LTNPs (Dambaya *et al*, 2019). Sin embargo, se ha reportado una mayor frecuencia de *CCR5-59353-C* en pacientes con progresión rápida a sida comparado con LTNPs (Clegg *et al*, 2000). Por otra parte, se ha observado una menor expresión de *CCR2* y *CCR5* en un grupo de HICs (Gonzalo-Gil *et al*, 2018).

Se ha asociado también el efecto “protector” del polimorfismo rs2234358 de *CXCR6* a LTNPs de la cohorte francesa GRIV, el cual reduce los niveles de expresión de este co-receptor secundario del VIH (Limou *et al*, 2010). Además, se ha descrito una mayor frecuencia de *CX3CR1-V249I* en LTNPs españoles (más de 15 años de seguimiento), comparado con no controladores e individuos no infectados (Vidal *et al*, 2005).

Respecto a las citoquinas, las células T CD4+ de LTNPs producen altos niveles de MIP-1 $\alpha$  y MIP-1 $\beta$  (ligandos de *CCR5*), en comparación con los sujetos con sida que producen cantidades extremadamente bajas de estas quimiocinas (Saha *et al*, 1998). Los homocigotos para *SDF-3'A* (ligando de *CXCR4*) han sido asociados con el fenotipo no progresor y un desarrollo más lento a sida (Hendel *et al*, 1998; Winkler *et al*, 1998).

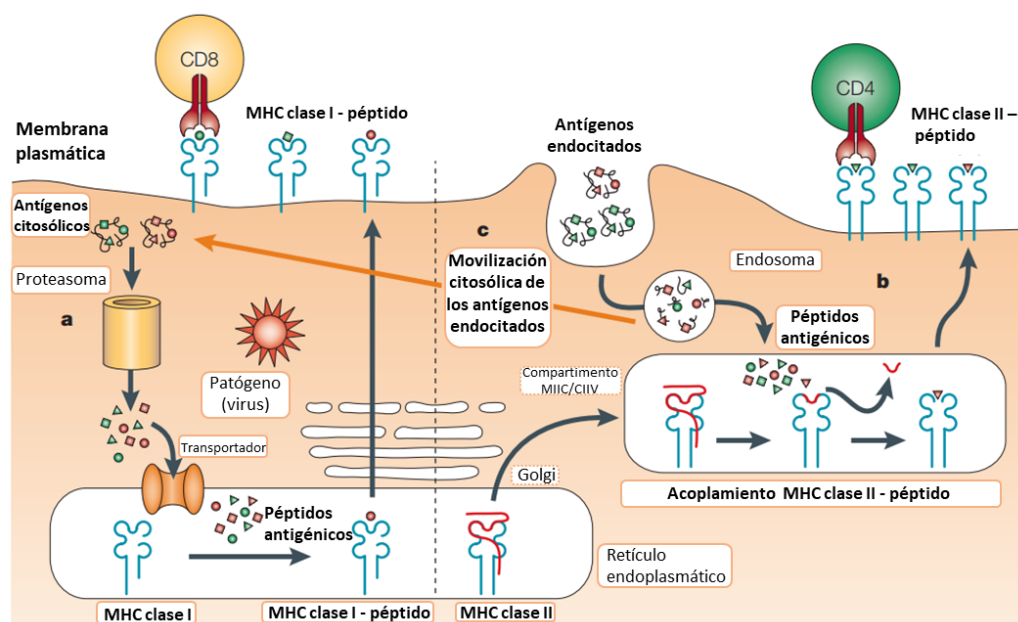
### 1.7.2.3. Presentación antigénica

La región codificante del complejo de histocompatibilidad (MHC), conocida también como la región HLA en humanos, modula la presentación de los antígenos en la superficie de todas las células (Figura 4).

El perfil genético de esta región, principalmente HLA de clase I (HLA-I), juega un papel importante en la respuesta inmune efectiva y progresión de la infección por el VIH. Las moléculas de HLA-I son reconocidas por las células CD8+ y *natural killers* (NKs), para desencadenar una respuesta citotóxica. Los alelos HLA-A, -B y -C corresponden a esta clase, y la mayoría de alelos relacionados con una progresión rápida o lenta de la infección por el VIH son HLA-B (Revisado por Goulder y Walker, 2012). Los alelos HLA-B\*27 y -B\*57 se relacionaron por primera vez con una progresión más lenta de la infección (Kaslow *et al*, 1996). Asimismo, se han descrito diversos alelos asociados positivamente con los

LTNPs y/o HICs, como son: HLA-A\*25, -B\*14/C\*0802, -B\*2705, -B\*5201, -B\*5701, -B\*5703, -B\*5801, -B\*8101 (Pereyra *et al*, 2010; Migueles *et al*, 2010; Navis *et al*, 2020), polimorfismos en la región HLA-C (Fellay *et al*, 2007), entre otros, en diversas cohortes alrededor del mundo. La frecuencia de diversos alelos HLA-B “protectores” (HLA-B\*27, -B\*57, entre otros), se expresan al menos en el 65% de los pacientes controladores y/o no progresores del VIH (Revisado por Migueles y Connors, 2010).

Los alelos HLA de clase II (HLA-II) son reconocidos principalmente por células T CD4+ (Potter *et al*, 2007). Los alelos HLA-DRB\*13, HLA-DQB\*06 y HLA-DRB\*15:02 se han asociado con fuertes respuestas específicas al VIH y/o con baja viremia en cohortes de HICs (Ferre *et al*, 2010; Ranasinghe S, 2013). Además, los alelos HLA-DR1, -DR11 y -DR15 también promueven el control del VIH en HICs (Galperin *et al*, 2018).



**Figura 4.** Proceso de la presentación antigénica a través de las moléculas HLA de clase I y II (Imagen adaptada de Heath y Carbone, Nat Rev Immunol, 2001)

La presentación de antígenos del VIH-1 a las células citotóxicas es más efectiva a través de haplotipos HLA “protectores”, debido a que exponen epítopos de regiones conservadas del virus. Por ejemplo, HLA-B\*27 presenta el epítipo KK10 y HLA-B\*57 los epítopos KF11 y TW10 de la proteína Gag, que están altamente conservados porque son críticos en la conformación de la cápside viral (Revisado por Zaunders y van Bockel, 2013). Además, los LTNP/HICs portadores de estos alelos “protectores” tienen mayor probabilidad de mantener recuentos estables de células T CD4+ (Dominguez-Molina *et al*, 2017).

Los virus son capaces de escapar de la presión de los CTLs (linfocitos T citolíticos), pero sufren una pérdida de su capacidad replicativa (Martínez-Picado *et al*, 2006). A pesar del control viral que ocurre en presencia de estos alelos “protectores”, se ha reportado replicación del virus a bajo nivel en ECs que portan el alelo HLA-B\*57, ya que se detectan mutaciones de los genes virales *gag* y *nef* (Salgado *et al*, 2010).

Por otra parte, los niveles de expresión de epítomos del VIH en las células infectadas también pueden afectar la eficacia de las respuestas CTL. La regulación de la expresión de HLA-C por miR-148 se asocia con el control viral (Kulkarni *et al*, 2011).

#### 1.7.2.4. Funcionalidad de las células inmunitarias

La función de las células del sistema inmune de individuos LTNP/HICs puede estar relacionada con la presencia de alelos HLA y polimorfismos “protectores”, o poseer características funcionales independientes a la presencia de estos alelos.

##### - Células CD8

Se ha descrito que las células T CD8+ específicas del VIH de los LTNP/HICs están presentes en mayor frecuencia, poseen mejor capacidad proliferativa, polifuncionalidad y/o respuesta citotóxica, que les confieren una mayor capacidad de eliminar las células infectadas por el VIH (Gea-Banacloche *et al*, 2000; Migueles *et al*, 2002; Betts *et al*, 2006; Migueles *et al*, 2008). Además, se ha demostrado que la funcionalidad incrementada de las células T CD8+ puede estar asociada con la expresión de distintos factores de transcripción, en particular de T-bet (Hersperger *et al*, 2011).

Las células CD8+ de HICs también pueden poseer la característica de selección de clonotipos de TCR de alta avidéz con una capacidad más amplia para reconocer variantes de epítomo y para orquestar una mejor función citolítica (Ladell *et al*, 2013). Los HICs presentan células T CD8+ con memoria quiescente que son altamente reactivas a bajas concentraciones de antígeno y pueden generar rápidamente capacidades efectoras en respuesta a recaídas del control viral (Sáez-Cirión *et al*, 2009).

Por otra parte, células T CD8+ específicas del VIH de HICs presentan una mayor expresión de genes relacionados con la función efectora y plasticidad metabólica, que mejoran su supervivencia y el desarrollo de funciones efectoras (Angin *et al*, 2019). Asimismo, la señalización de mTOR y eIF2 pueden desempeñar un papel notable para regular la función de las células T CD8+ de HICs (Chowdhury *et al*, 2018). Por último, los niveles de metabolitos presentes en HICs persistentes (que no pierden el control), están asociados a una mayor polifuncionalidad de células T CD8+ (Tarancón-Díez *et al*, 2019).

##### - Células CD4

Las células T CD4+ LTNP/ECs secretan mayores niveles de IL-2 y/o IL-21 que se asocia con una mayor polifuncionalidad y mejor capacidad citotóxica de las células T CD8+, comparado con sujetos que presentan viremia persistente (Migueles *et al*, 2000; Lichterfeld *et al*, 2004; Chevalier *et al*, 2011). Además, en los HICs las células T CD4+ con memoria específica del VIH mantienen su funcionalidad (Potter *et al*, 2007), y se han asociado con una expresión reducida de CTLA-4 (Kaufmann *et al*, 2007) y FoxO3a que median la actividad transcripcional proapoptótica (van Grevenynghe *et al*, 2008). Además, las células T CD4+ de HICs también exhiben una gran avidéz por péptidos inmunodominantes del

virus, que pueden permitirles reaccionar a bajos niveles de antígenos (Vingert *et al*, 2010). Las células T CD4+ y CD8+ polifuncionales específicas de Gag son más abundantes en la mucosa de los HICs que en individuos en TAR, lo que sugiere que células T CD4+ *helper* pueden ser clave para mantener una respuesta de CD8+ en el intestino de estos individuos (Ferre *et al*, 2010).

- Células NK

Las células NKs de LTNP/HICs contribuyen a establecer el control del VIH mediante un aumentado potencial citolítico y secretorio que pueden estar asociados con receptores particulares, o favoreciendo la inducción de una respuesta eficaz de células CD8+ a través de una coordinación óptima con las DCs (Vieillard *et al*, 2010; Sáez-Cirión *et al*, 2014). Además, se ha observado una respuesta mayor y más amplia de ADCCs (antibody dependent cell cytotoxicity) que median la actividad de NKs en LTNPs (Kulkarni *et al*, 2017).

- Células dendríticas y macrófagos

Las DCs de los HICs pueden producir específicamente IFN- $\alpha$  e inducir la apoptosis de las células infectadas o promover la expresión de factores de restricción del VIH-1 (Barblu *et al*, 2012; Pillai *et al*, 2012). Las DCs plasmacitoides mantienen una funcionalidad preservada para reducir la replicación viral en ECs (Barblu *et al*, 2012). Además, las DCs mieloides de los HICs tienen mejor capacidad de presentación de antígenos, pero producen niveles más bajos de citoquinas proinflamatorias (Huang *et al*, 2010). Por otra parte, se ha descrito también que los macrófagos de HICs tienen una baja susceptibilidad a la infección por el VIH-1 (Sáez-Cirión *et al*, 2011).

#### 1.7.2.5. Otros factores:

Los microARNs pueden influir en la transcripción de genes por mecanismos epigenéticos. La infección por el VIH-1 desencadena cambios en la expresión de miARNs, ya sea porque el virus influye en las actividades celulares o por el resultado de la respuesta a la infección (Balasubramaniam *et al*, 2018). Los miARNs pueden dirigirse a los mRNA del virus o regular la expresión de proteínas del huésped implicadas en la replicación viral (Swaminathan *et al*, 2012). La expresión de miR-31, miR-155, miR-221, miR-27a, miR-27b y miR-29b difieren significativamente entre HICs y otros grupos de individuos (Witwer *et al*, 2012; Egaña-Gorroño *et al*, 2014). Asimismo, los individuos ECs presentan niveles más altos de miR-29b-3p y miR-33a-5p en plasma en comparación con pacientes crónicamente infectados (Reynoso *et al*, 2014). Por otra parte, los niveles de miR-382-5p son más bajos en LTNPs y muestran una correlación positiva con la CV (Dey *et al*, 2016).

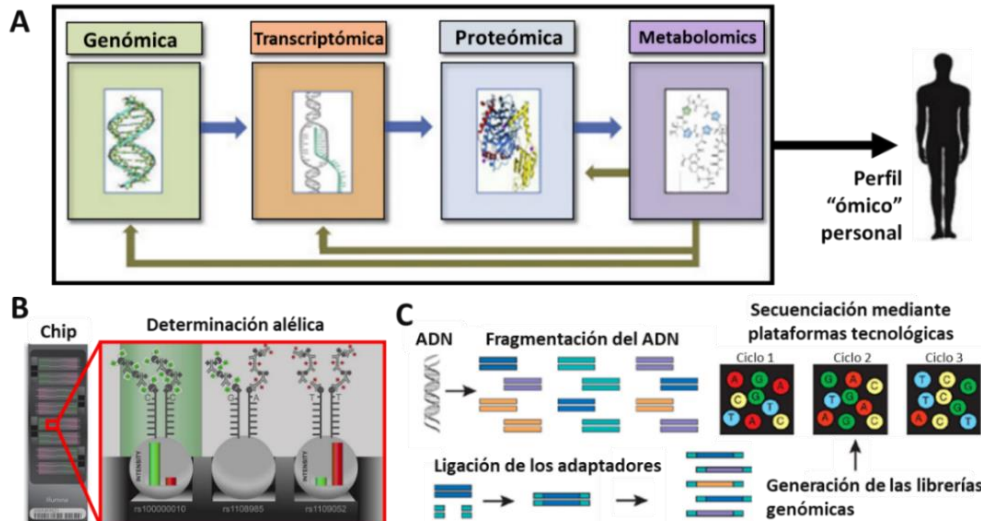
Los LTNPs y HICs presentan múltiples factores, por lo que es posible que se requiera una combinación de los mismos para alcanzarlo. Una potencial combinación favorable para alcanzar el estado LTNP/EC sería un menor número de células que alberguen una infección productiva del VIH asociado a una respuesta inmune adecuada.

## 1.8. Tecnologías “ómicas” en la investigación del VIH

Las tecnologías “ómicas” permiten estudiar un gran número de moléculas implicadas en el funcionamiento de un organismo. Las principales “ómicas” abordan el estudio del genoma (secuenciación o genotipado de genoma/exoma completo), transcriptoma, proteoma, epigenoma, microbioma y metaboloma. En el campo del VIH se han realizado principalmente estudios genómicos y transcriptómicos mediante las tecnologías de *microarrays* y *NGS* (*next-generation sequencing*, o secuenciación masiva) (Figura 5).

### 1.8.1. Estudios de genómica

El estudio de variantes genéticas es importante en la investigación de enfermedades infecciosas y puede permitir el hallazgo de un amplio conjunto de variantes de susceptibilidad o resistencia contra las infecciones virales. Los polimorfismos genéticos se pueden clasificar en inserciones/deleciones (*indels*), polimorfismos de secuencia repetida, SNPs (polimorfismos de nucleótido único), entre otros. Se han caracterizado 87,4 millones de SNPs en diferentes poblaciones, constituyendo la variación genética más común del genoma humano (The 1000 Genomes Project Consortium, 2015). Los progresos tecnológicos y científicos permiten la exploración de casi todo el genoma humano (Revisado por Jiang *et al*, 2016), con el fin de descubrir nuevas asociaciones de estos polimorfismos con diversas patologías.



**Figura 5.** Tecnologías ómicas usadas en la investigación del VIH. A) Diferentes abordajes de las “ómicas” aplicadas en pacientes (Imagen adaptada de Tosto, *Handb Clin Neurol*, 2016). B) Tecnología de *microarrays* para genotipado o análisis de expresión génica (Imagen adaptada de los protocolos de *Illumina*). C) Tecnología de genotipado de ADN mediante NGS (*next-generation sequencing*) (Imagen adaptada de Shendure, *Nat Biotech*, 2008)

El primer GWAS (*Genome wide association study*) enfocado al campo del VIH se publicó en 2007 por la cohorte Euro-CHAVI, identificándose que los polimorfismos rs2395029 (gen *HCP5*) y rs9264942 (cerca del gen *HLA-C*) tenían relación con el control de la CV (Fellay *et al*, 2007), que se confirmaron en un estudio posterior que analizaba las cohortes Euro-CHAVI y MACS. Además se identificaron otros polimorfismos asociados a



la progresión del VIH (Fellay *et al*, 2009). Posteriormente, se han desarrollado diversos estudios de GWAS para buscar nuevos marcadores del control viral y progresión a sida en pacientes con diferentes fenotipos de infección por el VIH-1 (incluidos los no progresores/controladores), en diversas cohortes como: GRIV (Limou *et al*, 2009), MACS (Herbeck *et al*, 2010), GISHEAL (Guergnon *et al*, 2012), así como el Consorcio Internacional de Controladores del VIH que incluyó individuos europeos, afroamericanos e hispanos (Pereyra *et al*, 2010). Asimismo, se han realizado metanálisis de GWAS de diferentes cohortes europeas y americanas (Limou *et al*, 2009; Troyer *et al*, 2011; Le Clerc *et al*, 2011).

Los GWAS en las diversas cohortes (la mayoría en población con ascendencia europea) han demostrado que la región HLA está fuertemente asociada tanto con la progresión al sida y con individuos LTNP/HICs. Se ha determinado que rs2395029 *HCP5* está fuertemente asociado con el fenotipo LTNP en varias cohortes. Además, otros polimorfismos asociados a los LTNPs/HICs en determinadas cohortes se encuentran en genes dentro de la región HLA (*ZNRD1*, *RNF39*, *C6orf48*, *HLA-B* y *MICA*) o fuera de la región HLA (*PROX1* y *PARD3B*) (Revisado por Limou y Zagury, 2013).

Además se han realizado estudios de GWAS en personas seronegativas con alto riesgo de infección por el VIH (HESN, *high exposure seronegative*) (Revisado por Limou y Zagury 2013; Revisado por Le Clerc *et al*, 2019), que se consideran como un modelo potencial para el desarrollo de una vacuna preventiva. Sin embargo, esta área de investigación presenta grandes desafíos en el diseño del estudio (Thèze *et al*, 2011).

Por otra parte, el exoma o secuencia que abarca todos los exones de genes que codifican proteínas, comprende entre el 1-2% del genoma (Ezkurdia *et al*, 2014). Se ha descrito además que alberga el 85% de las variantes causantes de enfermedades conocidas (Majewski *et al*, 2011). Los estudios de exoma han proporcionado un mayor conocimiento sobre los mecanismos que controlan el desarrollo, la función y la regulación de las células inmunes durante la respuesta a las infecciones (Revisado por Chou *et al*, 2012). Se han realizado estudios de exoma de pacientes infectados por el VIH-1, que han descrito variantes genéticas raras asociadas con LTNPs y ECs (Nissen *et al*, 2018).

### **1.8.2. Estudios de transcriptómica**

Desde 2003, se ha descrito que la infección por el VIH genera cambios en la expresión de genes involucrados en: división celular, transcripción, traducción, *splicing*, biosíntesis de colesterol, así como en la resistencia, permisividad o replicación viral (van't Wout *et al*, 2003; Imbeault *et al*, 2012; Xu *et al*, 2013). En estudios clínicos posteriores, se han descrito: la expresión de genes marcadores de la infección (*C5R1*, *PLAUR*, *SGK*) o del descenso de linfocitos T CD4+ (*IL12B*, *ISG20*) en pacientes VIH-positivos (Ockenhouse *et al*, 2005), así como genes que confieren resistencia a la infección por el VIH-1 (*ABL1*, *CBL*, *ICAM1*, *MAP3K5*, entre otros) (Huang *et al*, 2011).

En HICs se ha descrito que PD-1 inhibe la función de las células CD8+ específicas del VIH debido a la regulación positiva de BATF, y además presentan una expresión génica diferencial comparado con progresores (Quigley *et al*, 2010). En HICs también se han identificado genes regulados negativamente que están involucrados en la apoptosis o en las interacciones citoquina-receptor (Salgado *et al*, 2011), así como DEGs regulados por IFN en VNP comparado con ECs y RPs (Rotger *et al*, 2011). Asimismo, en LTNP se han descrito DEGs involucrados en la activación, proliferación, apoptosis y supervivencia celular (Wu *et al*, 2011; Luque *et al*, 2014). Además, metanálisis de transcriptoma han identificado diversos DEGs asociados a individuos LTNP/HICs, comparados con pacientes sanos o progresores típicos de la infección (Zhang *et al*, 2017; Ding *et al*, 2019)

Actualmente, la secuenciación del transcriptoma de célula única permite conocer la heterogeneidad transcripcional de células infectadas latentemente y tras la reactivación viral (Golumbeanu *et al*, 2018), así como la influencia del programa transcripcional de la célula huésped en la expresión proviral del VIH-1 en el reservorio (Bradley *et al*, 2018).

### 1.8.3. Otros estudios ómicos

Los diferentes estudios de proteómica han descrito que proteínas involucradas en: transporte intracelular, redireccionamiento metabólico, señalización apoptótica, homeostasis redox, ubiquitinación, dinámica de citoesqueleto (actina, beta-tubulina y anexina II), cambian sus niveles de expresión en el contexto de la infección por el VIH (Coiras *et al*, 2006; Molina *et al*, 2007; Chan *et al*, 2007; Ringrose *et al*, 2008). En ECs se han identificado proteínas involucradas en mecanismos proinflamatorios que juegan un rol en la replicación y patogénesis del VIH, así como la proteína de unión a galectina-3 como un posible biomarcador de la pérdida del control (Rodríguez-Gallego *et al*, 2019).

Estudios de epigenómica en el campo del VIH han descrito que mediante la modificación de histonas durante la latencia, se ven alterados los genes *NFIX*, *TRAF4*, *CDKN1A* y *CCND2* (Park *et al*, 2014; Kim *et al*, 2017). Además, se ha descrito una baja metilación en el promotor de *NCL5* en sujetos infectados por el VIH (Zhang *et al*, 2016).

Estudios de metabolómica indican que la expresión de moléculas proinflamatorias y la glutaminólisis están relacionados con una recuperación inmunológica tardía después del inicio de la TAR (Rosado-Sánchez *et al*, 2019). Asimismo, se ha identificado cambios en la expresión de aminoácidos, fosfolípidos y lípidos complejos que están relacionados con alteraciones metabólicas en pacientes que reciben TAR (Babu *et al*, 2019).

Con los antecedentes mencionados, en esta tesis doctoral hemos analizado polimorfismos genéticos y expresión génica de células del sistema inmune de individuos LTNP, con el fin de describir nuevos factores asociados a la progresión lenta y/o el control natural de la infección por el VIH-1, lo cual proporcionaría nuevas pistas que contribuyan a encontrar mejores estrategias de tratamiento o la cura funcional del VIH.

# **OBJETIVOS**

## 2. OBJETIVOS

El objetivo principal de esta tesis doctoral ha sido la búsqueda de nuevos factores genéticos asociados al fenotipo LTNP de individuos infectados por el VIH-1 en comparación con pacientes que tienen progresión típica de la infección (TP).

Este objetivo principal se divide en los siguientes objetivos específicos:

- 1) Caracterización genética de individuos LTNP para identificar nuevas asociaciones relacionadas con el control de la enfermedad.
  - a) Genotipado convencional de polimorfismos genéticos y de la región HLA-I (A y B) en una cohorte multicéntrica española de LTNPs.
  - b) Genotipado de exoma mediante técnicas de alto rendimiento de individuos LTNPs de la cohorte española LTNP-RIS.
  - c) Análisis funcionales *in vitro* de los marcadores asociados con el fenotipo LTNP en los análisis de exoma.
  
- 2) Análisis del transcriptoma de individuos infectados por el VIH-1 con distintos fenotipos de progresión: LNTP (vLTNP y EC-LTNP) y TP (preART-TP y postART-TP), mediante técnicas de *RNA-seq*.
  - a) Identificación y asociación de genes diferencialmente expresados al fenotipo LTNP.
  - b) Identificación de genes predictores del fenotipo LTNP mediante técnicas de clasificación supervisada.
  - c) Determinación de rutas biológicas significativamente asociadas al fenotipo LTNP.

# RESULTADOS

### 3. RESULTADOS

#### 3.1. Artículo I: Novel association of five HLA alleles with HIV-1 progression in Spanish long-term non progressor patients

El primer artículo presentado, muestra el análisis de polimorfismos genéticos y variantes de HLA en un grupo de individuos no progresores y controladores del VIH de una cohorte multicéntrica nacional de LTNPs de España.

Los primeros resultados obtenidos por genotipado mediante técnicas convencionales (qPCR, Tipado de HLA por RELI-SSO), han identificado hasta catorce factores genéticos asociados al fenotipo LTNP en la cohorte española, cinco de los cuales han sido relacionados por primera vez con este fenotipo: el alelo “protector” HLA-B\*39 y los alelos de “riesgo” HLA-A\*24, -A\*29, -B\*08 y -B\*18. Además, se ha podido determinar una mayor asociación de varios de los polimorfismos y alelos “protectores” en las subcategorías de LTNPs que presentan menor carga viral.

Este estudio fue un trabajo colaborativo entre diferentes centros, y durante el desarrollo de esta investigación nuestro grupo contribuyó en los análisis de genotipado de los polimorfismos genéticos. La mayor contribución ha sido en los análisis estadísticos, incluyendo los tests de Fisher realizados para identificar asociaciones con cada una de las categorías y subcategorías de pacientes, así como en su corrección posterior por comparaciones múltiples mediante *false discovery rate* (FDR). Estos análisis permitieron identificar diferentes genotipos y alelos “protectores” asociados al fenotipo LTNP (y sus subcategorías) comparado con diferentes poblaciones de control. Además, se contribuyó al análisis posterior y discusión de los resultados obtenidos.

## RESEARCH ARTICLE

# Novel association of five HLA alleles with HIV-1 progression in Spanish long-term non progressor patients

Eva Ramírez de Arellano<sup>1\*</sup>, Francisco Díez-Fuertes<sup>1,2</sup>, Francisco Aguilar<sup>1</sup>, Humberto Erick de la Torre Tarazona<sup>1</sup>, Susana Sánchez-Lara<sup>1,3</sup>, Yolanda Lao<sup>1</sup>, José Luis Vicario<sup>4</sup>, Felipe García<sup>2</sup>, Juan González-García<sup>5</sup>, Federico Pulido<sup>6</sup>, Félix Gutiérrez-Rodero<sup>7</sup>, Santiago Moreno<sup>8</sup>, Jose Antonio Iribarren<sup>9</sup>, Pompeyo Viciano<sup>10</sup>, Carlos Vilches<sup>11</sup>, Manuel Ramos<sup>1,3</sup>, Laura Capa<sup>1</sup>, José Alcamí<sup>1,2</sup>, Margarita Del Val<sup>1,3</sup>

**1** National Center for Microbiology, Instituto de Salud Carlos III, Majadahonda, Madrid, Spain, **2** Infectious Diseases Unit, IBIDAPS, HIVACAT, Hospital Clínic, University of Barcelona, Barcelona, Spain, **3** Viral Immunology, Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Madrid, Spain, **4** Departamento de Histocompatibilidad, Centro de Transfusión de Madrid, Madrid, Spain, **5** Hospital Universitario La Paz, Madrid, Spain, **6** HIV Unit, Instituto de Investigación Hospital 12 de Octubre (i+12), Madrid, Spain, **7** Servicio de Medicina Interna, Unidad de Enfermedades Infecciosas, Hospital General Universitario de Elche, Alicante, Spain, **8** Infectious Diseases, Hospital Ramón y Cajal, Madrid, Spain, **9** Hospital Universitario de Donostia, San Sebastián, Spain, **10** Laboratory of Immunovirology, Biomedicine Institute of Sevilla, Virgen del Rocío University Hospital, Clinic Unit of Infectious Diseases, Microbiology and Preventive Medicine, IBIS/CSIC/SAS/University of Sevilla, Sevilla, Spain, **11** Inmunogenética e Histocompatibilidad, Instituto de Investigación Sanitaria Puerta de Hierro, Majadahonda, Madrid, Spain

\* [erarellano75@gmail.com](mailto:erarellano75@gmail.com)



## OPEN ACCESS

**Citation:** Ramírez de Arellano E, Díez-Fuertes F, Aguilar F, de la Torre Tarazona HE, Sánchez-Lara S, Lao Y, et al. (2019) Novel association of five HLA alleles with HIV-1 progression in Spanish long-term non progressor patients. PLoS ONE 14(8): e0220459. <https://doi.org/10.1371/journal.pone.0220459>

**Editor:** Srinivas Mummidi, University of Texas Rio Grande Valley, UNITED STATES

**Received:** February 16, 2019

**Accepted:** July 16, 2019

**Published:** August 8, 2019

**Copyright:** © 2019 Ramírez de Arellano et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files.

**Funding:** This work was supported by Instituto de Salud Carlos III from the Spanish Ministry of Health, through Red Temática de Investigación Cooperativa en SIDA (RIS) [G03/173(2), PI050528 and RD06/0006/0033 to MDV, RD06/0006/0035 and RD12/0017/0037 to the HIV BioBank, and C03/

## Abstract

Certain host genetic variants, especially in the human leucocyte antigen (HLA) region, are associated with different progression of HIV-1-induced diseases and AIDS. Long term non progressors (LTNP) represent only the 2% of infected patients but are especially relevant because of their efficient HIV control. In this work we present a global analysis of genetic data in the large national multicenter cohort of Spanish LTNP, which is compared with seronegative individuals and HIV-positive patients. We have analyzed whether several single-nucleotide polymorphisms (SNPs) including in key genes and certain HLA-A and B alleles could be associated with a specific HIV phenotype. A total of 846 individuals, 398 HIV-1-positive patients (213 typical progressors, 55 AIDS patients, and 130 LTNPs) and 448 HIV-negative controls, were genotyped for 15 polymorphisms and HLA-A and B alleles. Significant differences in the allele frequencies among the studied populations identified 16 LTNP-associated genetic factors, 5 of which were defined for the first time as related to LTNP phenotype: the protective effect of HLA-B39, and the detrimental impact of HLA-B18, -A24, -B08 and -A29. The remaining eleven polymorphisms confirmed previous publications, including the protective alleles HLA-B57, rs2395029 (HCP5), HLA bw4 homozygosity, HLA-B52, HLA-B27, CCR2 V64I, rs9264942 (HLA-C) and HLA-A03; and the risk allele HLA bw6 homozygosity. Notably, individual Spanish HIV-negative individuals had an average of 0.12 protective HLA alleles and SNPs, compared with an average of 1.43 protective alleles per LTNP patient, strongly suggesting positive selection of LTNP. Finally, stratification of LTNP according to viral load showed a proportional relationship between the frequency of

173 and RD12/0017/0018 to CoRIS], cofinanced by ISCIII-Subdirección General de Evaluación and Fondo Europeo de Desarrollo Regional (FEDER), and also financed by Plan Nacional I+D+I from the Spanish Ministry of Science and Technology [SAF2007-60934, SAF2010-18917 and SAF2013-48754-C2-1-R to MDV], by Instituto de Salud Carlos III [Intrasalud P112/0056 to JA] and by Fundación para la Investigación y Prevención del SIDA en España (FIPSE) [to HIV Biobank]. Institutional grants from the Fundación Ramón Areces and Banco de Santander to the CBMSO are also acknowledged. FDF is supported by the Spanish Government's "Sara Borrell" postdoctoral program CD12/00515. The funders had no role in study design, decision to publish, analysis and preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** AIDS, HIV-1 infected patients that have developed AIDS; CTL, cytotoxic T lymphocytes; EC-LTNP, Elite controller LTNP; ExLTNP, former LTNP; HD, healthy donors; LTNP, HIV-1 infected long-term non progressors; LTNP-C, controller LTNP; LTNP-N, viremic non controller LTNP; OR, odds ratio; SNP, single nucleotide polymorphisms; SSO, sequence-specific oligonucleotide; TP, HIV-1 infected typical progressors; VL, viral loads.

protective alleles with control of viral load. Interestingly, no differences in the frequency of protection/risk polymorphisms were found between elite controllers and LTNPs maintaining viral loads <2.000 copies/mL throughout the follow-up.

## Introduction

The host genetic determinants influencing progression of HIV infection to disease and acquired immunodeficiency syndrome (AIDS) have been extensively studied in several cohorts of LTNP individuals of Caucasian ancestry. This is the case of several allelic variants in genes encoding the HIV-1 co-receptors and their ligands, such as CCR2 and CCR5, certain cytokines such as IL10, co-factors and interferon-induced proteins [1–9]. Among these host factors, the human major histocompatibility HLA class I complex has the strongest influence on HIV-1 progression. Thus, the HLA-B\*57 and HLA-B\*27 alleles are strongly associated with delayed HIV disease progression [10, 11] whereas HLA-B\*35 is associated with accelerated progression to AIDS [12, 13]. In addition, control of viremia and protection from AIDS is associated with HLA bw4 allelic grouping homozygosity [14]. More recent studies identified allelic variants associated with control of HIV-1 replication in *HLA-C* and HLA complex P5 (*HCP5*) [15, 16], which in turn is in tight linkage disequilibrium with the HLA-B\*5701 allele [17]. Studies based on genome-wide association strategies identified novel genetic variants associated with delayed disease progression [18–25], most of them within the HLA complex [19, 23, 24].

These data suggest that disease progression and HIV-1 replication is controlled by several loci of the human genome. However, known genes affecting disease progression and their variants do not fully explain the highly variable course of HIV-1 infection or its pathogenic mechanisms. The aim of the present study is to characterize genetically the large Spanish HIV LTNP cohort and to identify novel associations with disease control, employing a multicenter cohort of 398 Spanish HIV-1 positive patients compared with a control population of 448 healthy Spaniards. By comparing the genotype distribution of several SNPs as well as the frequency of HLA-A and HLA-B alleles, the present work proposes 5 novel HLA class I alleles related to maintenance of the LTNP status [defined as HIV-infected patients that maintain CD4-lymphocytic counts above 500 cells/uL for at least ten years in the absence of antiretroviral treatment (ART). Viral load is usually low in this group of patients (<10.000 RNA copies/ml, as defined in the Spanish LTNP-Cohort)] and confirms the role of known genetic markers associated with control of HIV-1 replication. The analysis of these genetic traits stratified by different phenotypes within LTNP patients, showed a differential effect according to the LTNP subcategory, evidencing the necessity to clearly define the LTNP condition in case/control association studies. In addition to supporting the category of EC with undetectable viral load (VL), we propose the use of a regularly maintained VL below the limit of 2,000 copies/mL as a new marker of profound and stable LTNP status.

## Materials and methods

### Patient samples

A total of 448 healthy bone marrow donors (HD), as well as 398 HIV-1 infected patients, comprising 55 AIDS patients, 213 typical progressors (TP) and 130 LTNP, were included in the study. The uninfected individuals were healthy Spanish donors from the Blood Transfusion



Centre of the Community of Madrid, Spain, and are representative of the Spanish population [26]. All HIV-1 infected patients belonged to different cohorts of patients with samples stored at the HIV BioBank (Gregorio Marañón University Hospital, Madrid, Spain), which is integrated in the Spanish AIDS research network (RIS) [27]. All the samples were collected from 2004 to 2007. CoRIS, the RIS cohort of adults with HIV infection, was launched in 2004 [28]. CoRIS is an open multicenter cohort of patients that are over 13 years of age and newly diagnosed with HIV infection in the participating hospital or treatment center they attend for the first time, and that are naïve to antiretroviral treatment. This study was reviewed and approved by the institutional Ethics committee for research and clinical trials" (CEIC) from Instituto de Salud Carlos III. All patients signed and informed consent to include their blood samples for scientific research including genetic studies in the Biobank of the Spanish AIDS Research Network. The information is subject to internal quality controls; once every 2 years, information on 10% of the cohort is audited by an external agency.

A total of 55 AIDS and 213 TP patients come from CoRIS. The AIDS group includes naïve patients late diagnosed after attending a participating center for the first time; the TPs are HIV-1 infected patients with CD4<sup>+</sup> cell loss between 50–100 cells/μl per year. The 130 LTNP patients belong to the Spanish Cohort of LTNP (LTNP-RIS), a cohort similarly managed as above, and were naïve patients who have CD4<sup>+</sup> T cell counts over 500/μl and VL < 10,000 copies/mL without antiretroviral treatment for at least 10 years after HIV diagnosis. The prototypical recruited HIV-1 infected individuals were male intravenous drug users of Spanish origin (Table 1).

Based on specific clinical data, including VL and time after seroconversion, we defined several LTNP subcategories. Thus, three mutually exclusive subcategories of LTNP have been analyzed, including ExLTNP, who are patients that lost LTNP status after at least 10 years after HIV-1 diagnosis; viremic non-controller LTNP (LTNP-N), who are LTNP maintaining detectable VL > 50 up to 10,000 copies/mL throughout the follow-up; and EC, defined as HIV-1 infected individuals with undetectable VL during follow-up. In addition, LTNP-C controllers includes a subgroup of LTNP-N maintaining VL < 2,000 copies/mL throughout the follow-up; this subcategory includes all EC but also those LTNP-N with low VL. Blood samples were processed following standard procedures [29] and frozen immediately after their processing. Peripheral blood mononuclear cells were obtained from blood of all subjects included in the study and DNA was extracted.

### Sample genotyping

Genomic DNA was used for genotyping. Most SNP tested were typed using TaqMan SNP genotyping assay following manufacturer's procedures and standardized protocols (Applied Biosystems), except for rs333 (*CCR5-Δ32*) and rs1801157 (*SDF-1*), which were determined by real time PCR employing the primers and probes described in S1 Table. The TaqMan Universal PCR Master Mix and standard thermocycling conditions were employed for all polymorphisms on an ABI PRISM 7000 system, and allele calling was performed using AutoCaller SDS Software v 1.2.3. (Applied Biosystems).

### HLA typing

Two-digit HLA-A and HLA-B typing was carried out using sequence-specific oligonucleotide (SSO) hybridization following manufacturer's procedure and standardized protocol (RELI SSO HLA Typing Kit, Invitrogen). Genomic DNA was amplified using locus-specific primers flanking exons 2 and 3 of the HLA class I genes. The PCR products were hybridized to an array of immobilized sequence-specific oligonucleotide probes. The probe-bound amplified

**Table 1. Summary of epidemiological characteristics of HIV-1 infected patients included in the analysis.**

Characteristic		TP (n = 213)	AIDS (n = 55)	LTNP (n = 130)
Age upon admission, mean (min-max)		39 (17–69)	42 (22–61)	48 (30–76)
	Unknown, n			27
Country of origin, %	Spain	100	100	77.0
	Unknown			23.0
Sex, %	Male	69.0	74.5	53.8
	Female	31.0	25.5	26.2
	Unknown			20.0
Risk group, %	Intravenous drug user	68.1	60.0	64.6
	Homosexual/bisexual	9.4	14.5	5.4
	Heterosexual	16.4	16.4	13.1
	Others (transfusion, etc.)	3.3	7.3	2.3
	Unknown	2.8	1.8	14.6

<https://doi.org/10.1371/journal.pone.0220459.t001>

product was detected by a color formation assay. All assays were automated using the Auto-RELI 48 Instrument (Dynal Biotech). The HLA-B alleles were grouped into HLA bw4 and HLA bw6 epitopes according to the official page of HLA nomenclature [30].

### Statistical analysis

Genotype frequency comparisons between groups were performed by two-tailed Fisher’s exact test in R package for each SNP (p-values of 2x3 tables). The frequency of HLA alleles was also analyzed by two-tailed Fisher’s exact test in R package (p-values of 2x2 tables). The results were corrected for multiple hypothesis testing to control the Benjamini–Hochberg false discovery rate (FDR) at a significant threshold of 0.1 to compare LTNP with different control populations (q-value). A similar correction was made to compare different subcategories of LTNP individuals with control populations, using a significant threshold of 0.05 (q-value).

## Results

### SNP and polymorphisms associated with the Spanish long term non progressors cohort phenotype

The individuals included in the analysis were genotyped for 14 different SNP and the CCR5-Δ32 polymorphism. Eleven out of 14 SNP did not differ significantly between LTNP and groups of healthy donors, AIDS patients and typical progressors (Table 2).

However, a significant difference in the genotype distribution was identified in 3 SNP (*HCP5*, *CCR2* and *5’HLA-C*) (Table 2). In the case of *HCP5*, a clearly higher frequency of the genotype TG was found in LTNP compared with HD and TP groups, and less significant with AIDS patients (Table 2). The differences in the GA/AA genotype distribution of the SNP causing the V64I mutation in *CCR2* (HIV-1 co-receptor that is associated with protection [3]) were highly significant when comparing LTNP with HD. Regarding the Δ32 deletion of the *CCR5* HIV-1 co-receptor locus that is associated with delayed HIV disease progression [1, 2, 5], a higher frequency of the protective WT/Δ32 genotype was observed in LTNP than in the AIDS group, but these differences did not reach statistical significance after FDR correction. The variant -35C/T located 35 kb upstream of the *HLA-C* locus has been associated with delayed HIV disease progression in infected patients [16]. Accordingly, a significantly higher frequency of the CC and CT genotypes was found in Spanish LTNP compared with TP (Table 2). Therefore, our data confirm the association of *HCP5*, *CCR2* and *5’HLA-C* SNPs to LTNP phenotype.

Table 2. Genotype distribution of different single nucleotide polymorphisms in distinct groups of HIV patients and in healthy donors.

SNP	Group (n)	Genotype distribution						p-value <sup>a</sup>	FDR <sup>b</sup>
		n	%	n	%	n	%		
CCR5-2459 (G/A) rs1799987		GG		GA		AA			
	HD (122)	25	20.5	53	43.4	44	36.1	0.0236	ns
	LTNP (127)	33	26.0	68	53.5	26	20.5	-	-
SDF-1 3'UTR 801 (G/A) rs1801157		GG		GA		AA			
	HD (158)	98	62.0	49	31.0	11	7.0	ns	ns
	LTNP (117)	73	62.4	38	32.5	6	5.1	-	-
RANTES -403 (G/A) rs2107538		GG		GA		AA			
	HD (164)	114	69.5	44	26.8	6	3.7	ns	ns
	AIDS (55)	34	61.8	18	32.7	3	5.5	ns	ns
	TP (212)	145	68.4	58	27.4	9	4.2	ns	ns
	LTNP (130)	90	69.2	37	28.5	3	2.3	-	-
CD32α +494 (A/G) rs1801274		AA		AG		GG			
	HD (159)	42	26.4	86	54.1	31	19.5	ns	ns
	AIDS (55)	12	21.8	31	56.4	12	21.8	ns	ns
	TP (213)	42	19.7	111	52.1	60	28.2	ns	ns
	LTNP (124)	32	25.8	61	49.2	31	25.0	-	-
Tsg101-517 (C/T) rs1857909		CC		CT		TT			
	HD (259)	209	80.7	49	18.9	1	0.4	ns	ns
	AIDS (55)	44	80.0	10	18.2	1	1.8	ns	ns
	TP (213)	177	83.1	35	16.4	1	0.5	ns	ns
	LTNP (130)	115	88.5	15	11.5	0	0	-	-
Rab27a 3'UTR (C/T) rs1050931		CC		CT		TT			
	HD (248)	161	64.9	76	30.6	11	4.4	ns	ns
	AIDS (54)	36	66.7	16	29.6	2	3.7	ns	ns
	TP (211)	144	68.2	57	27.0	10	4.7	ns	ns
	LTNP (130)	87	66.9	39	30.0	4	3.1	-	-
Rggtα (G/A) rs729421		GG		GA		AA			
	HD (177)	68	38.4	89	50.3	20	11.3	ns	ns
	AIDS (55)	16	29.1	29	52.7	10	18.2	ns	ns
	TP (213)	80	37.6	102	47.9	31	14.6	ns	ns
	LTNP (111)	52	46.8	59	53.2	18	16.2	-	-
αCatenin 3'UTR (G/T) rs288039		GG		GT		TT			
	HD (163)	84	51.5	66	40.5	13	8.0	ns	ns
	AIDS (55)	26	47.3	22	40.0	7	12.7	ns	ns
	TP (212)	114	53.8	82	38.7	16	7.5	ns	ns
	LTNP (69)	40	58.0	23	33.3	6	8.7	-	-
αCatenin 3'UTR (A/T) rs3749663		AA		AT		TT			
	HD (260)	136	52.3	102	39.2	22	8.5	ns	ns
	AIDS (55)	25	45.5	23	41.8	7	12.7	ns	ns
	TP (212)	114	53.8	81	38.2	17	8.0	ns	ns
	LTNP (129)	71	55.0	47	36.4	11	8.5	-	-
αCatenin intron (C/T) rs700626		CC		CT		TT			
	HD (168)	92	54.8	66	39.3	10	6.0	ns	ns
	LTNP (69)	39	56.5	24	34.8	6	8.7	-	-
HCP5 3'UTR (T/G)		TT		TG		GG			

(Continued)

Table 2. (Continued)

SNP	Group (n)	Genotype distribution						p-value <sup>a</sup>	FDR <sup>b</sup>
		n	%	n	%	n	%		
rs2395029	HD (254)	245	96.5	9	3.5	0	0	3.94x10 <sup>-8</sup>	1.5x10 <sup>-6</sup>
	AIDS (55)	51	92.7	4	7.3	0	0	0.0189	ns
	TP (213)	191	89.6	22	10.4	0	0	0.0044	0.057
	LTNP (128)	100	78.2	28	21.8	0	0	-	-
CCR2-V64I +190 (G/A)		GG		GA		AA			
rs1799864	HD (262)	220	84.0	40	15.3	2	0.8	0.0097	0.092
	AIDS (55)	44	80.0	10	18.2	1	1.8	ns	ns
	TP (212)	163	76.9	46	21.7	3	1.4	ns	ns
	LTNP (129)	92	71.3	34	26.4	3	2.3	-	-
CCR5 Δ32 (WT/Δ32)		WT/WT		WT/Δ32		Δ32/Δ32			
rs333	HD (246)	204	82.9	40	16.3	2	0.8	ns	ns
	AIDS (55)	50	90.9	5	9.1	0	0	0.0245	ns
	TP (213)	173	81.2	40	18.8	0	0	ns	ns
	LTNP (129)	98	76	31	24.0	0	0	-	-
5'HLA-C (C/T)		CC		CT		TT			
rs9264942	AIDS (55)	16	29.1	22	40.0	17	31.9	ns	ns
	TP (212)	41	19.3	104	49.1	67	31.6	0.0045	0.057
	LTNP (128)	36	28.1	71	55.5	21	16.4	-	-
IL-10-592 (C/A)		CC		CA		AA			
rs1800872	HD (250)	142	56.8	87	34.8	21	8.4	ns	ns
	AIDS (54)	30	55.6	17	31.5	7	13.0	ns	ns
	TP (208)	113	54.3	85	40.9	10	4.8	ns	ns
	LTNP (125)	71	56.8	49	39.2	5	4.0	-	-

HD: healthy donors, TP: typical progressors, AIDS: HIV patients with AIDS, LTNP: long term non progressors.

<sup>a</sup>p-values were calculated for each SNP comparing the genotypes of LTNP population with the other groups using Fisher's exact test ( $p < 0.05$  were considered significant, ns, not significant).

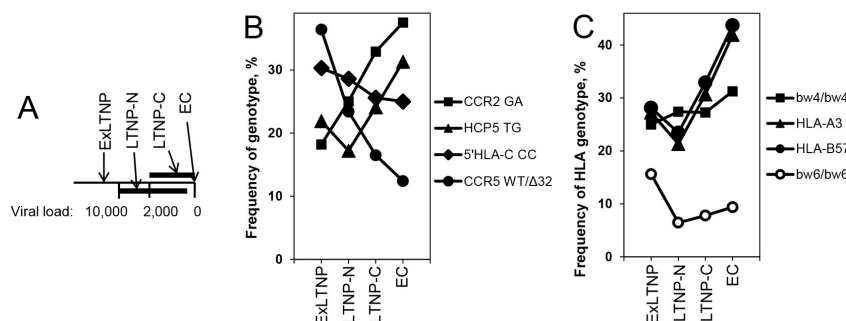
<sup>b</sup>False discovery rate (FDR) correction for multiple testing ( $\alpha = 0.1$  were considered significant, ns, not significant).

<https://doi.org/10.1371/journal.pone.0220459.t002>

### Genotype distribution of significant SNP and CCR5-Δ32 polymorphism in distinct subcategories of the Spanish LTNP cohort

As described in Methods section LTNP were stratified according to VL into 4 subcategories, ExLTNP, viremic non controllers LTNP-N, controllers LTNP-C and elite controllers EC (Fig 1A), and the genotype frequencies of the relevant genetic factors were determined (i.e. *HCP5*, *CCR2* and 5'*HLA-C* SNPs). The results confirmed the protective nature of the *HCP5* and *CCR2* genotypes, as they were more frequent in most subcategories of LTNP, especially in those subcategories with the lowest VL, the LTNP-C and the EC, than in the other HIV-infected or HD populations (Table 3). For a summary and statistics see Table 4. Actually, *HCP5* and *CCR2* SNP frequencies were gradually increased within LTNP subcategories in an inverse correlation with VL (framed data in Table 3), with percentages of *HCP5* and *CCR2* favorable genotypes peaking at the EC population with undetectable VL (Fig 1B). For a summary and statistics see Table 4.

The enrichment of the *HCP5* and *CCR2* favorable genotypes in EC-LTNP with undetectable VL was somehow expected. However, it is very noticeable that the LTNP-C controllers, whose VL are always maintained below 2,000 copies/mL, are also very significantly endowed with



**Fig 1. Genotype frequencies among subpopulations of LTNP as a function of viral load.** The frequencies of the indicated SNP or CCR5-Δ32 genotypes (B) or HLA genotypes (C) relative to the total number of individuals are plotted for the LTNP subcategories that are graphically depicted (A): ExLTNP, patients who were LTNP for 10 years but thereafter failed to fulfill any of the inclusion criteria; LTNP-N, viremic non controller LTNP, VL>50–10,000; LTNP-C, controllers, VL<2,000 copies/mL; EC, elite controllers, undetectable VL. Displayed are relevant SNP and HLA genotypes with a frequency in the LTNP subcategories above 15%, indicating those that are more frequent (filled symbols) or less frequent (open symbols) in the overall LTNP population than in HD, according to Table 3 and Fig 2.

<https://doi.org/10.1371/journal.pone.0220459.g001>

these protective genotypes (p-values in Table 3). This suggests that viral replication limited to this threshold value for many years may also be a marker of a profound and stable LTNP status.

### Allelic frequencies of HLA-A and -B in the Spanish LTNP cohort

LTNP and HD were typed for HLA class I. Most HLA alleles were not significantly different between the LTNP and the control group. From those with significant differences, several alleles seemed to favor the LTNP condition, as their allelic frequencies were significantly higher in LTNP than in HD (Fig 2); these included HLA-B57, followed by HLA-B27, -B52, -A03 and -B39. In contrast, HLA-B18 was markedly less frequent in the LTNP population, as well as HLA-A24, -B08 and -A29, and thus appeared to be detrimental for LTNP status. Stratification of the LTNP into subcategories was undertaken for most relevant alleles. Given the high number of alleles for these two HLA loci, a very low number of patients was left in most subcategories and precluded statistical analysis. Still, the strongest favorable factor HLA-B57, together with -B52, -B27 and -A03, as well as the strongest unfavorable factor HLA-B18, together with -A24 and -B08, were significantly enriched in LTNP subcategories (Table 4). Interestingly, when the frequency of the HLA allele in the LTNP subcategories was above 10% and amenable to analysis, it showed again an inverse correlation of HLA-B57 and HLA-A03 protective alleles with VL (Fig 1C), as was the case for the favorable HCP5 and CCR2 SNP.

When the HLA-B alleles were classified according to their mutually exclusive bw4 or bw6 public epitopes [30], a highly significantly greater percentage of bw4 in homozygosity was observed in the LTNP compared with HD (Table 5), confirming these alleles as protective factors for the LTNP status. The converse association of bw6/bw6 homozygosity with risk for the LTNP condition was also as strong, and both extended to most LTNP subcategories (Tables 4 and 5). As before, favorable bw4/bw4 showed a mild inverse correlation with VL while unfavorable bw6/bw6 genotype showed a mild direct correlation with VL within Spanish LTNP subcategories (Fig 1C).

**Table 3. Genotype distribution of selected SNP, which have specific alleles associated with protection or with disease progression, in distinct subcategories of HIV LTNP patients and in healthy donors.**

SNP	Group (n)	Genotype distribution						p-value <sup>a</sup>			FDR <sup>b</sup>			
		n		%		n		%		HD	AIDS	TP	HD	AIDS
<b>HCP5 3'UTR (T/G)</b>		TT		TG		GG								
<b>rs2395029</b>	HD (254)	245	96.5	9	3.5	0	0	-				-		
	AIDS (55)	51	92.7	4	7.3	0	0	ns	-			ns	-	
	TP (213)	191	89.6	22	10.4	0	0	<b>0.0045</b>	ns	-		<b>0.022</b>	ns	-
	ExLTNP (32)	25	78.1	7	21.9	0	0	<10 <sup>-3</sup>	ns	ns		<b>0.005</b>	ns	ns
	LTNP-N (64)	53	82.8	11	17.2	0	0	<10 <sup>-3</sup>	ns	ns		<b>0.005</b>	ns	ns
	LTNP-C (79)	60	75.9	19	24.1	0	0	<10 <sup>-6</sup>	<b>0.0112</b>	<b>0.004</b>		<10 <sup>-4</sup>	<b>0.044</b>	<b>0.022</b>
	EC (32)	22	68.8	10	31.3	0	0	<10 <sup>-5</sup>	<10 <sup>-2</sup>	<b>0.003</b>		<10 <sup>-4</sup>	<b>0.024</b>	<b>0.02</b>
<b>CCR2-V64I (G/A)</b>		GG		GA		AA								
<b>rs1799864</b>	HD (262)	220	84.0	40	15.3	2	0.8	-				-		
	AIDS (55)	44	80.0	10	18.2	1	1.8	ns	-			ns	-	
	TP (212)	163	76.9	46	21.7	3	1.4	ns	ns	-		ns	ns	-
	ExLTNP (33)	26	78.8	6	18.2	1	3.0	ns	ns	ns		ns	ns	ns
	LTNP-N (64)	46	71.9	16	25.0	2	3.1	<b>0.0471</b>	ns	ns		ns	ns	ns
	LTNP-C (79)	51	64.6	26	32.9	2	2.5	<10 <sup>-3</sup>	ns	ns		<b>0.005</b>	ns	ns
	EC (32)	20	62.5	12	37.5	0	0	<b>0.0203</b>	ns	ns		0.066	ns	ns
<b>5'HLA-C (C/T)</b>		CC		CT		TT								
<b>rs9264942</b>	AIDS (55)	16	29.1	22	40.0	17	31.9	-	-			-		
	TP (212)	41	19.3	104	49.1	67	31.6	-	ns	-		-	ns	
	ExLTNP (33)	10	30.3	18	54.5	5	15.2	-	ns	ns		-	ns	ns
	LTNP-N (63)	18	28.6	36	57.1	9	14.3	-	ns	<b>0.015</b>		-	ns	<b>0.054</b>
	LTNP-C (78)	20	25.6	45	57.7	13	16.7	-	ns	<b>0.033</b>		-	ns	0.099
	EC (32)	8	25.0	17	53.1	7	21.9	-	ns	ns		-	ns	ns

HD: healthy donors; TP: typical progressors; AIDS: HIV patients with AIDS; LTNP: long term non progressors; ExLTNP: patients who were LTNP for 10 years but thereafter failed to fulfill any of the inclusion criteria; LTNP-N: viremic LTNP with VL >10,000 copies/ml; LTNP-C: LTNP with VL <2,000 copies/ml; EC: elite controllers with undetectable VL.

<sup>a</sup>p-values were calculated for each SNP comparing the genotypes of LTNP population with the other groups using Fisher's exact test (p<0.05 were considered significant, ns, not significant).

<sup>b</sup>False discovery rate (FDR) correction for multiple testing (alpha = 0.1 were considered significant, ns, not significant).

<https://doi.org/10.1371/journal.pone.0220459.t003>

### Overview of genetics and LTNP status in the Spanish HIV cohorts

The 9 genotypes and alleles that are associated with LTNP status as well as those 5 unfavorable ones are listed in Table 4 and roughly ranked according to the intensity of the effect and the statistical significance. Interestingly, when analyzed as individuals concerning protective and risk factors, Spanish LTNP patients clearly stood up in comparison with HD. Almost 70% of LTNP patients had at least one HLA protective allele, and this rose to 87% when protective SNP were also considered. In contrast, only 22% of the LTNP had a detrimental allele (Fig 3). Fractions of HD controls having protective or risk alleles were very similar, for reference.

Notably, the mean number of protective minus risk HLA alleles and SNPs in individual Spanish healthy donors was balanced (0.74 protective– 0.62 risk to give a 0.12 balance per person, or an average of 0.12 protective HLA alleles and SNPs per healthy person). In sharp contrast, the mean was 12 times more marked for individual Spanish LTNP patients (1.66 protective– 0.23 risk to give an average of 1.43 protective HLA alleles and SNPs per LTNP patient),

**Table 4. Summary and statistics for differences in frequencies of the 16 genotypes and alleles associated with protection or disease progression described in this report.** Statistics apply to comparisons among distinct subcategories of Spanish LTNP patients, and with other groups of Spanish HIV patients and healthy donors.

	SNP / HLA allele	Group <sup>a</sup>	p-value <sup>a</sup>			FDR <sup>b</sup>		
			HD	AIDS	TP	HD	AIDS	TP
P1 <sup>a</sup>	HLA-B57	LTNP (128)	<10 <sup>-7</sup>	-	-	<2.3x10 <sup>-6</sup>	-	-
		ExLTNP (32)	1.0x10 <sup>-4</sup>	-	-	<5.3x10 <sup>-4</sup>	-	-
		LTNP-N (64)	1.0x10 <sup>-4</sup>	-	-	<5.3x10 <sup>-4</sup>	-	-
		LTNP-C (79)	1.0x10 <sup>-4</sup>	-	-	<5.3x10 <sup>-4</sup>	-	-
P2	HCP5 3'UTR rs2395029 (TG)	LTNP (128)	<10 <sup>-7</sup>	2.0x10 <sup>-2</sup>	4.0x10 <sup>-3</sup>	<2.3x10 <sup>-6</sup>	3.0x10 <sup>-2</sup>	8.9x10 <sup>-3</sup>
		ExLTNP (32)	6.0x10 <sup>-4</sup>	ns	ns	2.3x10 <sup>-3</sup>	ns	ns
		LTNP-N (64)	4.0x10 <sup>-4</sup>	ns	ns	1.6x10 <sup>-3</sup>	ns	ns
		LTNP-C (79)	<10 <sup>-7</sup>	1.0x10 <sup>-2</sup>	4.0x10 <sup>-3</sup>	<2.3x10 <sup>-6</sup>	1.7x10 <sup>-2</sup>	8.9x10 <sup>-3</sup>
P3	HLA bw4/bw4	LTNP (128)	1.0x10 <sup>-4</sup>	-	-	<5.3x10 <sup>-4</sup>	-	-
		ExLTNP (32)	4.9x10 <sup>-2</sup>	-	-	ns	-	-
		LTNP-N (64)	1.6x10 <sup>-3</sup>	-	-	5.5x10 <sup>-3</sup>	-	-
		LTNP-C (79)	1.0x10 <sup>-4</sup>	-	-	<5.3x10 <sup>-4</sup>	-	-
P4	HLA-B52	LTNP (128)	2.0x10 <sup>-3</sup>	-	-	5.8x10 <sup>-3</sup>	-	-
		ExLTNP (32)	1.0x10 <sup>-2</sup>	-	-	1.7x10 <sup>-2</sup>	-	-
		LTNP-N (64)	2.9x10 <sup>-2</sup>	-	-	4.1x10 <sup>-2</sup>	-	-
		LTNP-C (79)	2.1x10 <sup>-2</sup>	-	-	3.1x10 <sup>-2</sup>	-	-
P5	HLA-B27	LTNP (128)	2.0x10 <sup>-3</sup>	-	-	5.8x10 <sup>-3</sup>	-	-
		LTNP-N (64)	2.0x10 <sup>-4</sup>	-	-	9.9x10 <sup>-4</sup>	-	-
		LTNP-C (79)	1.5x10 <sup>-2</sup>	-	-	2.4x10 <sup>-2</sup>	-	-
P6	CCR2-V64I rs1799864 (GA/AA)	LTNP (129)	5.0x10 <sup>-3</sup>	ns	ns	1.1x10 <sup>-2</sup>	ns	ns
		LTNP-N (64)	3.0x10 <sup>-2</sup>	ns	ns	4.1x10 <sup>-2</sup>	ns	ns
		LTNP-C (79)	4.0x10 <sup>-4</sup>	ns	4.0x10 <sup>-2</sup>	1.6x10 <sup>-3</sup>	ns	5.0x10 <sup>-2</sup>
		EC (32)	6.0x10 <sup>-3</sup>	ns	ns	1.2x10 <sup>-2</sup>	ns	ns
P7	5'HLA-C rs9264942 (CC/CT)	LTNP (128)	-	3.0x10 <sup>-2</sup>	2.0x10 <sup>-3</sup>	-	4.1x10 <sup>-2</sup>	5.8x10 <sup>-3</sup>
		LTNP-N (63)	-	4.0x10 <sup>-2</sup>	6.0x10 <sup>-3</sup>	-	5.0x10 <sup>-2</sup>	1.2x10 <sup>-2</sup>
		LTNP-C (78)	-	ns	1.0x10 <sup>-2</sup>	-	ns	1.7x10 <sup>-2</sup>
P8	HLA-A03	LTNP (125)	1.0x10 <sup>-2</sup>	-	-	1.7x10 <sup>-2</sup>	-	-
		LTNP-C (75)	1.0x10 <sup>-2</sup>	-	-	1.7x10 <sup>-2</sup>	-	-
		EC (31)	3.0x10 <sup>-3</sup>	-	-	7.4x10 <sup>-3</sup>	-	-
P9	HLA-B39 <sup>c</sup>	LTNP (128)	2.0x10 <sup>-2</sup>	-	-	3.0x10 <sup>-2</sup>	-	-
		ExLTNP (32)	1.0x10 <sup>-2</sup>	-	-	1.7x10 <sup>-2</sup>	-	-
R1 <sup>a</sup>	HLA bw6/bw6	LTNP (128)	1.0x10 <sup>-4</sup>	-	-	<5.3x10 <sup>-4</sup>	-	-
		LTNP-N (64)	1.0x10 <sup>-4</sup>	-	-	5.3x10 <sup>-4</sup>	-	-
		LTNP-C (79)	1.0x10 <sup>-4</sup>	-	-	<5.3x10 <sup>-4</sup>	-	-
		EC (32)	3.4x10 <sup>-2</sup>	-	-	4.5x10 <sup>-2</sup>	-	-
R2	HLA-B18	LTNP (128)	9.0x10 <sup>-4</sup>	-	-	3.3x10 <sup>-3</sup>	-	-
		LTNP-N (64)	4.7x10 <sup>-2</sup>	-	-	ns	-	-
		LTNP-C (79)	3.0x10 <sup>-4</sup>	-	-	1.4x10 <sup>-3</sup>	-	-
		EC (32)	3.7x10 <sup>-2</sup>	-	-	4.8x10 <sup>-2</sup>	-	-
R3	HLA-A24	LTNP (125)	2.5x10 <sup>-3</sup>	-	-	6.9x10 <sup>-3</sup>	-	-
		LTNP-N (61)	1.3x10 <sup>-2</sup>	-	-	2.1x10 <sup>-2</sup>	-	-
		LTNP-C (75)	2.6x10 <sup>-3</sup>	-	-	6.9x10 <sup>-3</sup>	-	-

(Continued)

Table 4. (Continued)

	SNP / HLA allele	Group <sup>a</sup>	p-value <sup>a</sup>			FDR <sup>b</sup>		
			HD	AIDS	TP	HD	AIDS	TP
R4	HLA-B08	LTNP (128)	3.6x10 <sup>-3</sup>	-	-	8.6x10 <sup>-3</sup>	-	-
		LTNP-C (79)	2.4x10 <sup>-2</sup>	-	-	3.5x10 <sup>-2</sup>	-	-
R5	HLA-A29	LTNP (125)	1.1x10 <sup>-2</sup>	-	-	1.8x10 <sup>-2</sup>	-	-

<sup>a</sup>Genotypes or alleles are labelled P for 'protection' or R for 'risk' depending on whether their frequency is higher or lower in the indicated LTNP population, respectively, and roughly numerically ordered from most protective and with the highest statistical power. P1, and from most risky and with the highest statistical significance. R1.

<sup>b</sup> Only statistics for significant differences are listed (p<0.05); ns, not significant.

<sup>c</sup> Novel genetic factors described in this report in association with LTNP condition are framed.

<https://doi.org/10.1371/journal.pone.0220459.t004>

clearly indicating that LTNP is a population that has successfully undergone selection under the selective pressure of the HIV epidemics.

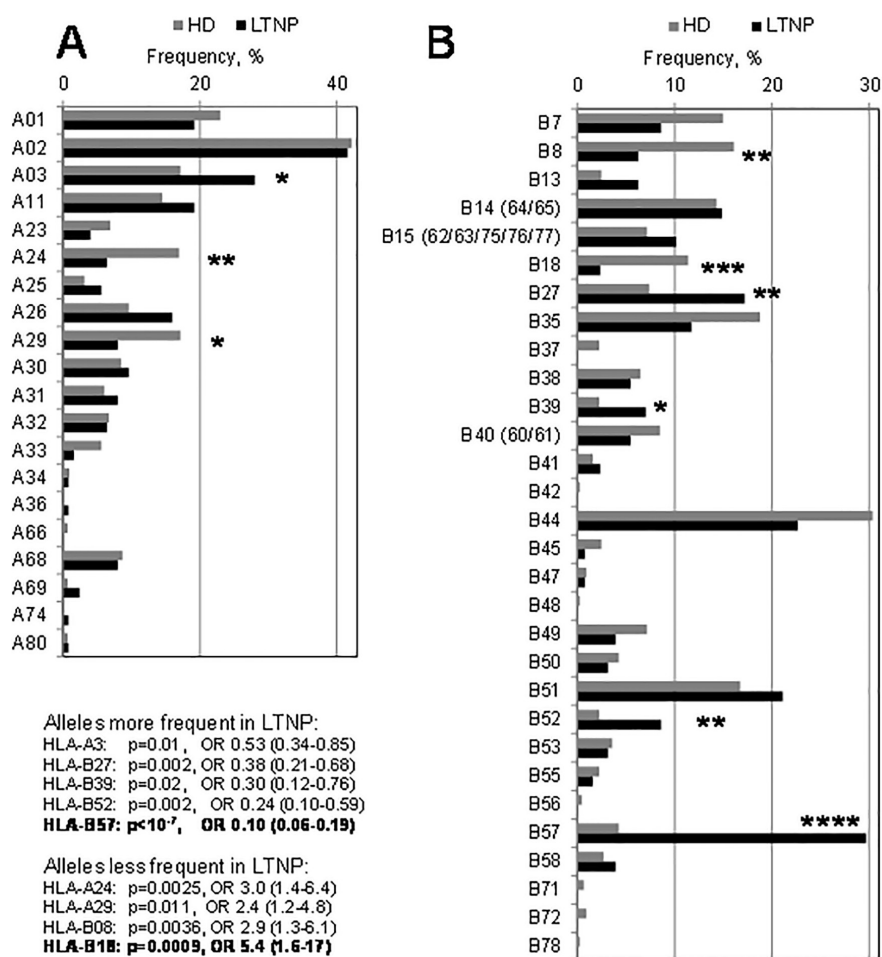
## Discussion

Several host genetic factors have been associated with HIV-1 disease progression in different cohorts of LTNP, typical progressors or rapid progressors, when compared with HIV seronegative individuals [1–25]. The present study aims to investigate the role of genetic factors in a large (n = 130) Spanish cohort of LTNP. However, the LTNP are a heterogeneous population consisting of HIV-1 infected individuals showing different phenotypes regarding their capacity to control viral replication. In this regard, the analysis has been extended to a conscientious stratification of LTNP, according to their VL, into elite controllers (EC), controllers (LTNP-C), viremic non-controller LTNP (LTNP-N) and individuals losing the LTNP status over time (ExLTNP).

Our analysis of the Spanish HIV-1 LTNP cohort and control healthy and infected populations, altogether representing 846 individuals, reveals 14 significant genetic factors. Nine of them are more frequent in the LTNP population, and thus qualify as factors that contribute to disease control and to LTNP status; in rough order of decreasing protective potency and statistical power these are the following alleles or genotypes: HLA-B57, HCP5 TG rs2395029 SNP, HLA bw4/bw4 (p<0.0001, see individual details and summary in Table 4), HLA-B52, HLA-B27, CCR2 GA/AA rs1799864 SNP (p<0.01.), 5'HLA-C CC/CT rs9264942 SNP, HLA-A03 and HLA-B39 (0.01<p<0.05). Protective alleles/genotypes range each in frequency among the LTNP population from 7% to 30%, supporting the notion that a large proportion of the LTNP phenotype may be determined by accumulation of favorable genetic traits, rather by a single strongly protective factor. Conversely, 5 genetic factors are less frequently found in LTNP and appear to represent factors favoring disease progression; in rough order of decreasing risk and statistical power these are the following alleles or genotypes: HLA bw6/bw6, HLA-B18 (p<0.001), HLA-A24, HLA-B08, (p<0.01) and HLA-A29(p<0.05).

Two out of the 14 factors reported in Table 4 are described for the first time to our knowledge in firm association with any type of HIV susceptibility to infection or disease progression and, specifically, in association with the LTNP condition. Both are unfavorable HLA alleles, HLA-B08 and -A29. In addition, another 3 factors that have been found associated with other HIV conditions are described here for the first time in association with LTNP, including the protective HLA-B39 and the risk factors HLA-B18 and -A24. Furthermore, the positive association with LTNP of two more alleles, HLA-A03 and HLA-B52, for which very limited evidence is published, is confirmed with the Spanish LTNP cohort. In the natural history of HIV





**Fig 2. Comparison of the frequencies of HLA-A and HLA-B individuals between LTNP and healthy donor populations.** HD, n = 448 donors for HLA-A and HLA-B; LTNP, n = 125 patients for HLA-A and 128 patients for HLA-B. Statistically significant differences are marked with asterisks (\*,  $p<0.05$ ; \*\*,  $p<0.01$ ; \*\*\*,  $p<0.0001$ ) and detailed in the text inset. Relevant frequencies for HD vs. LTNP are: HLA-A03 (17 vs. 28%), HLA-A24 (17 vs. 6.4%), HLA-A29 (17 vs. 8.0%), HLA-B8 (16 vs. 6.3%), HLA-B18 (11 vs. 2.3%), HLA-B27 (7.4 vs. 17%), HLA-B39 (2.2 vs. 7.0%), HLA-B52 (2.2 vs. 8.6%) and HLA-B57 (4.2 vs. 30%).

<https://doi.org/10.1371/journal.pone.0220459.g002>

infection, several HLA class I alleles have been associated consistently with HIV progression, especially HLA-B alleles [10, 12, 14, 22, 31–34], and notably, we identify here novel HLA-B as well as HLA-A alleles. Identification of several new genetic associations when compared with studies on a geographically close population as the French cohort [35] stresses the importance of assembling and studying such cohorts of patients that control HIV infection, in spite of the scarcity of such patients. It also reveals the importance of thorough studies on novel cohorts such as the Spanish one reported here. Among other factors, one possible reason for the

Table 5. Frequency comparison of HLA bw4 and bw6 allele groups between HD and LTNP subcategories.

HLA-B	Group (n)	Genotype distribution						p-values <sup>a</sup>		FDR <sup>b</sup>	
		n		%		n		%		bw4/bw4	bw6/bw6
		bw4/bw4	bw4/bw6	bw4/bw4	bw4/bw6	bw4/bw4	bw6/bw6				
HD (421)	69	16.4	239	56.8	113	26.8	-	-	-	-	
LTNP (128)	45	35.2	71	55.5	12	9.4	1.1x10 <sup>-5</sup>	2.0x10 <sup>-5</sup>	1.2x10 <sup>-4</sup>	1.0x10 <sup>-4</sup>	
ExLTNP (32)	10	31.3	17	53.1	5	15.6	0.049	ns	ns	ns	
LTNP-N (64)	22	34.4	38	59.4	4	6.3	0.002	1.2x10 <sup>-4</sup>	0.003	2.4x10 <sup>-4</sup>	
LTNP-C (79)	29	36.7	44	55.7	6	7.6	8.8x10 <sup>-5</sup>	8.4x10 <sup>-5</sup>	2.1x10 <sup>-4</sup>	2.2x10 <sup>-4</sup>	
EC (32)	13	40.6	16	50	3	9.4	0.002	0.034	0.003	0.042	

HD: healthy donors; TP: typical progressors; AIDS: HIV patients with AIDS; LTNP: long term non progressors; ExLTNP: patients who were LTNP for 10 years but thereafter failed to fulfill any of the inclusion criteria; LTNP-N: viremic LTNP with VL >10.000 copies/ml; LTNP-C: LTNP with VL <2.000 copies/ml; EC: elite controllers with undetectable VL.

<sup>a</sup>The presence of bw4/bw4 or bw6/bw6 genotypes in different categories of LTNP subcategories compared with healthy donors (HD) was calculated using Fisher's exact test (p<0.05 were considered statistically significant; ns. not significant)

<sup>b</sup>False discovery rate (FDR) correction for multiple testing (alpha = 0.1 were considered significant. ns. not significant).

<https://doi.org/10.1371/journal.pone.0220459.t005>

novelty of our data may relate to the high proportion of intravenous drug users among the Spanish LTNP compared with the majority of men who have sex with men (MSM) included in the French LTNP cohort.

The novel detrimental associations of HLA-B08 and HLA-A29 with maintenance of the LTNP status have little precedent in the literature. HLA-B08 within a common Western haplotype is frequently associated with fast progression of HIV disease, rapid CD4 T lymphocyte decline in adults and with increased mother to infant transmission [32,36,37], but the association has rarely been individually ascribed to HLA-B08. For HLA-A29 only a non-significant trend has been reported [38]. This negative association may interestingly be related to the poor recognition by A29-restricted T lymphocyte clones of viral sequence variants [39]. The large Spanish LTNP cohort data thus presents solid evidence for the first time on the negative association of these two HLA alleles with the LTNP status.

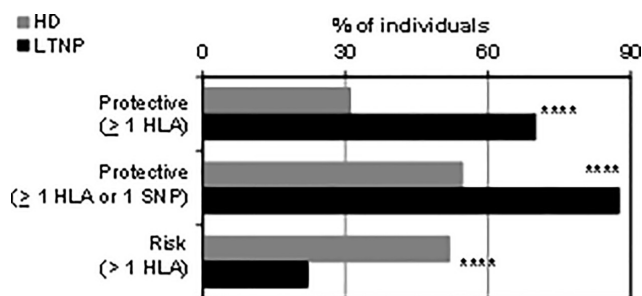


Fig 3. Individuals with a minimum of one protective or one risk factor for maintaining LTNP status: Frequency comparison between HD and LTNP. From top to bottom, the percentage of healthy controls and LTNP patients is indicated that have at least one protective HLA allele, at least one protective HLA allele or SNP factor, or at least one HLA risk allele, as listed in Table 4. \*\*\*\*, p< 0.0001. OR (95% confidence interval): 0.19 (0.12–0.30); 0.17 (0.09–0.31); 3.8 (2.4–6.1), respectively, from top to bottom. Stratification into LTNP subcategories did not provide additional information, as they were very homogeneous; p value was also <0.0001 for all comparisons between HD and each subcategory.

<https://doi.org/10.1371/journal.pone.0220459.g003>

Concerning the three HLA alleles previously reported only in HIV patients other than LTNP, the effect of HLA-B39, which is identified here as a protective allele for the Spanish LTNP cohort, to our knowledge for the first time in association with LTNP, appears to depend on the study, the geographical area or the HIV-infected population. HLA-B39 was described as a risk allele in smaller populations of Argentinian HIV<sup>+</sup> subjects [40] and of Indian serodiscordant couples [41], while, more in line with our results, as an allele associated with lower VL in Zambian HIV infected patients [42]. The second allele associated in this cohort for the first time with LTNP, HLA-A24, is described here as an unfavorable allele for the LTNP condition, and it was also early associated with rapid CD4 T lymphocyte decline [36] and with susceptibility in adults [43], promoting selection of cytotoxic T-lymphocyte escape variants in Japan [44, 45]. Whether the detrimental role of HLA-A24 for LTNP described here in the Spanish population is related to T-lymphocyte escape also in LTNP patients is currently unknown and warrants investigation. Finally, HLA-B18 was the strongest and most significant detrimental factor for the Spanish LTNP population. This HLA allele has been widely studied in HIV infected populations other than LTNP, and its favorable [38, 41] or risk [40, 46] contribution to diverse aspects of HIV disease is variable and at least seems to depend on the virus clade.

Further, HLA-A03 has been described in one report in association with French LTNP [35]. This early observation of positive association of HLA-A03 with LTNP is now confirmed with our larger and stricter Spanish LTNP cohort. Otherwise, A03 has also occasionally been associated with populations of HIV-infected patients other than LTNP [37, 47]. As for HLA-A03, we also describe a significant association of the HLA-B52 allele with delayed disease progression in the Spanish cohort of LTNP patients, confirming an international HIV controllers study [22] and a single earlier report weakly associating HLA-B52 with non-progression in a small Brazilian cohort of HIV-1 infected individuals [48].

Out of the 14 factors identified here in positive or negative association with Spanish LTNP, the remaining 7 factors were previously established, and our data are confirmatory. Previous studies have associated low HIV-1 viremia and prolonged survival with HLA-B57 [7, 10] and HLA-B27 [35] in HIV LTNP patients, and it is assumed that this is due to the antigen presentation by these alleles of conserved viral epitopes contributing to viral fitness. LTNP are also characterized by the SNP rs2395029 located at *HCP5* [18–21, 23], which is in tight linkage disequilibrium with the HLA-B\*5701 allele [13]. The fact that these HLA-B alleles display the public HLA epitope bw4 is thought to underlie the previously described and here confirmed positive role of bw4/bw4 homozygosity [14] and the converse negative role of the bw6/bw6 genotype. Interestingly, when considering HLA supertypes [49], the LTNP-associated protective HLA alleles described here clustered together in some HLA supertypes (A03, B7, B27, B58 and B62 supertypes), and segregated away from the supertypes of risk alleles (A1, A24 and B44 supertypes). As the supertypes are based on HLA antigen presentation function to cytotoxic CD8<sup>+</sup> T lymphocytes, this could possibly underlie the functional mechanism for their selective association in HIV-1 infection.

The present study confirms the strong protective effect for Spanish LTNP of *HCP5* 3'UTR TG rs2395029, *CCR2* GA/AA rs1799864 and 5'*HLA-C* CC/CT rs9264942 SNPs.

When LTNP were stratified, gradual increases of the frequencies of favorable *HCP5*, HLA-B57, HLA-A03, *CCR2* and bw4/bw4 alleles and genotypes were concomitantly observed with increasing HIV-1 control capacity, peaking at LTNP-C and EC populations, confirming a trend previously assumed for some of them in other studies that analyzed a very limited number of LTNP patients [50]. Conversely, the strongly unfavorable bw6/bw6 genotype shows a mild inverse correlation with control of VL. However, this study shows that there is no such correlation of low VL with protective 5'*HLA-C*, as published [51], nor with *CCR5* Δ32 deletion, and questions including these two SNP as markers for reduced VL [50]. While the *CCR5* Δ32

deletion has extensively been confirmed to contribute to preventing initial HIV infection [1], these data may suggest that, once infection is established in patients, it does not contribute to maintaining a profound LTNP status as strongly as HCP5, HLA-B57, -A03, CCR2, or bw4/bw4 genotypes may do.

The classification of HIV-1 infected patients based on clinical data includes LTNP, typical progressors and rapid progressors. However, this classification can be enriched incorporating the VL measurement to define a more realistic description of the LTNP status with the subcategories included in the present study, i.e. EC-LTNP, LTNP-C, LTNP-N and ExLTNP. The genetic factors influencing the LTNP status have widely been studied, even from a genome-wide perspective [18–21]. However, the control of HIV-1 replication and the delayed disease progression simultaneously observed in EC-LTNP and LTNP-C have been poorly characterized. In this regard, the present study provides new clues about the effect of known factors influencing control and resistance to HIV-1 such as *HCP5*, *CCR2*, HLA-B57 and -A03 in EC-LTNP and LTNP-C compared with the rest of LTNP. On the other hand, well-documented genetic factors associated to LTNP status such as *CCR5* rs333 or 5'*HLA-C* do not seem to have any additive effect in the EC-LTNP or LTNP-C condition with respect to the rest of LTNP. Further studies are required to discern whether the EC-LTNP and LTNP-C statuses can be considered as an accumulation of several factors previously associated with EC or LTNP or as the presence of specific unknown associations with the simultaneous observation of both phenotypes.

The fact that with new cohorts like the large multicentric and stratified Spanish ones it is still possible to identify significant associations of the LTNP with 5 new HLA alleles (one protective and 4 detrimental for the LTNP condition) underscores the strong influence of HLA on viral control. It is still open whether especially the most significant unfavorable HLA-B18 allele could play a direct functional effect on control of HIV and in long-term stability of infected LTNP patients.

## Supporting information

**S1 Table. Primers and probes employed in the determination of rs333 and rs1801157.**  
(DOCX)

## Acknowledgments

We particularly want to acknowledge the patients in this study for their participation and the HIV BioBank integrated in the Spanish AIDS Research Network and collaborating centers for the generous donation of clinical samples used in this work. This study would not have been possible without the collaboration of all the patients, medical and nursery staff and data managers who have taken part in the project. The authors would like to thank the Laboratory of Molecular Biology staff from Hospital Gregorio Marañón (Madrid, Spain), for providing patient's specimens, and the Blood Transfusion Centre of the Community of Madrid (Spain), for providing HLA typed healthy bone marrow donor samples.

## Author Contributions

**Conceptualization:** Eva Ramírez de Arellano, José Luis Vicario, Margarita Del Val.

**Data curation:** Carlos Vilches, Laura Capa.

**Formal analysis:** Eva Ramírez de Arellano, Humberto Erick de la Torre Tarazona, Margarita Del Val.

**Investigation:** Eva Ramírez de Arellano, Francisco Aguilar, Margarita Del Val.

**Methodology:** Eva Ramírez de Arellano, Francisco Aguilar, Susana Sánchez-Lara, Yolanda Lao, Manuel Ramos, Margarita Del Val.

**Project administration:** Margarita Del Val.

**Resources:** José Luis Vicario, Felipe García, Juan González-García, Federico Pulido, Félix Gutierrez-Rodero, Santiago Moreno, Jose Antonio Iribarren, Pompeyo Viciana.

**Supervision:** Eva Ramírez de Arellano, José Alcamí, Margarita Del Val.

**Validation:** Eva Ramírez de Arellano, Margarita Del Val.

**Visualization:** Eva Ramírez de Arellano, Margarita Del Val.

**Writing – original draft:** Eva Ramírez de Arellano, Francisco Díez-Fuertes, José Alcamí, Margarita Del Val.

**Writing – review & editing:** Eva Ramírez de Arellano, Francisco Díez-Fuertes, Santiago Moreno, José Alcamí, Margarita Del Val.

## References

1. Samson M, Libert F, Doranz BJ, et al. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 1996; 382:722–5. <https://doi.org/10.1038/382722a0> PMID: 8751444
2. Dean M, Carrington M, Winkler C, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* 1996; 273:1856–62. <https://doi.org/10.1126/science.273.5283.1856> PMID: 8791590
3. Smith MW, Dean M, Carrington M, et al. Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC), ALIVE Study. *Science* 1997; 277:959–65. <https://doi.org/10.1126/science.277.5328.959> PMID: 9252328
4. Winkler C, Modi W, Smith MW, et al. Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. *Science* 1998; 279:389–93. <https://doi.org/10.1126/science.279.5349.389> PMID: 9430590
5. Liu H, Chao D, Nakayama EE, et al. Polymorphism in RANTES chemokine promoter affects HIV-1 disease progression. *Proc Natl Acad Sci U S A* 1999; 96:4581–5. <https://doi.org/10.1073/pnas.96.8.4581> PMID: 10200305
6. Shin HD, Winkler C, Stephens JC, et al. Genetic restriction of HIV-1 pathogenesis to AIDS by promoter alleles of IL10. *Proc Natl Acad Sci U S A* 2000; 97:14467–72. <https://doi.org/10.1073/pnas.97.26.14467> PMID: 11121048
7. Carrington M, O'Brien SJ. The influence of HLA genotype on AIDS. *Annu Rev Med* 2003; 54:535–51. <https://doi.org/10.1146/annurev.med.54.101601.152346> PMID: 12525683
8. Machmach K, Abad-Molina C, Romero-Sanchez MC, et al. IL28B single-nucleotide polymorphism rs12979860 is associated with spontaneous HIV control in white subjects. *J Infect Dis* 2013; 207:651–5. <https://doi.org/10.1093/infdis/jis717> PMID: 23225905
9. Ghezzi S, Galli L, Kajaste-Rudnitski A, et al. Identification of TRIM22 single nucleotide polymorphisms associated with loss of inhibition of HIV-1 transcription and advanced HIV-1 disease. *AIDS* 2013; 27:2335–44. <https://doi.org/10.1097/01.aids.0000432474.76873.5f> PMID: 23921607
10. Migueles SA, Sabbaghian MS, Shupert WL, et al. HLA B\*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc Natl Acad Sci U S A* 2000; 97:2709–14. <https://doi.org/10.1073/pnas.050567397> PMID: 10694578
11. Navis M, Schellens I, van BD, et al. Viral replication capacity as a correlate of HLA B57/B5801-associated nonprogressive HIV-1 infection. *J Immunol* 2007; 179:3133–43. <https://doi.org/10.4049/jimmunol.179.5.3133> PMID: 17709528

12. Carrington M, Nelson GW, Martin MP, et al. HLA and HIV-1: Heterozygote advantage and B\*35-Cw\*04 disadvantage. *Science* 1999; 283:1748–52. <https://doi.org/10.1126/science.283.5408.1748> PMID: [10073943](https://pubmed.ncbi.nlm.nih.gov/10073943/)
13. Gao X, Bashirova A, Iversen AK, et al. AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. *Nat Med* 2005; 11:1290–2. <https://doi.org/10.1038/nm1333> PMID: [16288280](https://pubmed.ncbi.nlm.nih.gov/16288280/)
14. Flores-Villanueva PO, Yunis EJ, Delgado JC, et al. Control of HIV-1 viremia and protection from AIDS are associated with HLA-Bw4 homozygosity. *Proc Natl Acad Sci U S A* 2001; 98:5140–5. <https://doi.org/10.1073/pnas.071548198> PMID: [11309482](https://pubmed.ncbi.nlm.nih.gov/11309482/)
15. Van Manen D, Kootstra NA, Boeser-Nunnink B, et al. Association of HLA-C and HCP5 gene regions with the clinical course of HIV-1 infection. *AIDS* 2009; 23:19–28. <https://doi.org/10.1097/QAD.0b013e32831db247> PMID: [19050382](https://pubmed.ncbi.nlm.nih.gov/19050382/)
16. Thomas R, Apps R, Qi Y, et al. HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. *Nat Genet* 2009; 41:1290–4. <https://doi.org/10.1038/ng.486> PMID: [19935663](https://pubmed.ncbi.nlm.nih.gov/19935663/)
17. de Bakker PI, McVean G, Sabeti PC, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* 2006; 38:1166–72. <https://doi.org/10.1038/ng1885> PMID: [16998491](https://pubmed.ncbi.nlm.nih.gov/16998491/)
18. Dalmasso C, Carpentier W, Meyer L, et al. Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. *PLoS ONE* 2008; 3: e3907. <https://doi.org/10.1371/journal.pone.0003907> PMID: [19107206](https://pubmed.ncbi.nlm.nih.gov/19107206/)
19. Le Clerc S, Limou S, Coulonges C, et al. Genomewide association study of a rapid progression cohort identifies new susceptibility alleles for AIDS (ANRS Genomewide Association Study 03). *J Infect Dis* 2009; 200:1194–201. <https://doi.org/10.1086/605892> PMID: [19754311](https://pubmed.ncbi.nlm.nih.gov/19754311/)
20. Fellay J, Shianna KV, Ge D, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science* 2007; 317:944–7. <https://doi.org/10.1126/science.1143767> PMID: [17641165](https://pubmed.ncbi.nlm.nih.gov/17641165/)
21. Limou S, Le CS, Coulonges C, et al. Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *J Infect Dis* 2009; 199:419–26. <https://doi.org/10.1086/596067> PMID: [19115949](https://pubmed.ncbi.nlm.nih.gov/19115949/)
22. Pereyra F, Jia X, McLaren PJ, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* 2010; 330:1551–7. <https://doi.org/10.1126/science.1195271> PMID: [21051598](https://pubmed.ncbi.nlm.nih.gov/21051598/)
23. Guernon J, Dalmasso C, Broet P, et al. Single-nucleotide polymorphism-defined class I and class III major histocompatibility complex genetic subregions contribute to natural long-term nonprogression in HIV infection. *J Infect Dis* 2012; 205:718–24. <https://doi.org/10.1093/infdis/jir833> PMID: [22238471](https://pubmed.ncbi.nlm.nih.gov/22238471/)
24. McLaren PJ, Coulonges C, Ripke S, et al. Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. *PLoS Pathog* 2013; 9:e1003515. <https://doi.org/10.1371/journal.ppat.1003515> PMID: [23935489](https://pubmed.ncbi.nlm.nih.gov/23935489/)
25. Bartha I, Carlson JM, Brumme CJ, et al. A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *Elife* 2013; 2:e01123. <https://doi.org/10.7554/eLife.01123> PMID: [24171102](https://pubmed.ncbi.nlm.nih.gov/24171102/)
26. Balas A, Garcia-Sanchez F, Vicario JL. Allelic and haplotypic HLA frequency distribution in Spanish hematopoietic patients. Implications for unrelated donor searching. *Tissue Antigens* 2011; 77:45–53. <https://doi.org/10.1111/j.1399-0039.2010.01578.x> PMID: [21155721](https://pubmed.ncbi.nlm.nih.gov/21155721/)
27. Garcia-Merino I, de Las CN, Jimenez JL, et al. The Spanish HIV BioBank: a model of cooperative HIV research. *Retrovirology* 2009; 6:27. <https://doi.org/10.1186/1742-4690-6-27> PMID: [19272145](https://pubmed.ncbi.nlm.nih.gov/19272145/)
28. Caro-Murillo AM, Castilla J, Perez-Hoyos S, et al. Spanish cohort of naive HIV-infected patients (CoRIS): rationale, organization and initial results. *Enferm Infecc Microbiol Clin* 2007; 25:23–31. PMID: [17261243](https://pubmed.ncbi.nlm.nih.gov/17261243/)
29. Ballana E, Senserrich J, Pauls E, et al. ZNRD1 (zinc ribbon domain-containing 1) is a host cellular factor that influences HIV-1 replication and disease progression. *Clin Infect Dis* 2010; 50:1022–32. <https://doi.org/10.1086/651114> PMID: [20192730](https://pubmed.ncbi.nlm.nih.gov/20192730/)
30. Robinson J, Halliwell JA, Hayhurst JD, et al. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res* 2015; 43:D423–D431. <https://doi.org/10.1093/nar/gku1161> PMID: [25414341](https://pubmed.ncbi.nlm.nih.gov/25414341/)
31. Gao X, O'Brien TR, Welzel TM, et al. HLA-B alleles associate consistently with HIV heterosexual transmission, viral load, and progression to AIDS, but not susceptibility to infection. *AIDS* 2010; 24:1835–40. <https://doi.org/10.1097/QAD.0b013e32833c3219> PMID: [20588164](https://pubmed.ncbi.nlm.nih.gov/20588164/)
32. McNeil AJ, Yap PL, Gore SM, et al. Association of HLA types A1-B8-DR3 and B27 with rapid and slow progression of HIV disease. *QJM* 1996; 89:177–85. <https://doi.org/10.1093/qjmed/89.3.177> PMID: [8731561](https://pubmed.ncbi.nlm.nih.gov/8731561/)

33. Yindom LM, Leigdowicz A, Martin MP, et al. Influence of HLA class I and HLA-KIR compound genotypes on HIV-infection and markers of disease progression in a Manjako community in West Africa. *J Virol* 2010; 84(16):8202–8. <https://doi.org/10.1128/JVI.00116-10> PMID: 20519398
34. Chikata T, Murakoshi H, Koyanagi M, et al. Control of HIV-1 by an HLA-B\*52:01-C\*12:02 Protective Haplotype. *J Infect Dis* 2017; 216(11):1415–1424. <https://doi.org/10.1093/infdis/jix483> PMID: 28968792
35. Magierowska M, Theodorou I, Debre P, et al. Combined genotypes of CCR5, CCR2, SDF1, and HLA genes can predict the long-term nonprogressor status in human immunodeficiency virus-1-infected individuals. *Blood* 1999; 93:936–41. PMID: 9920843
36. Kaslow RA, Duquesnoy R, VanRaden M, et al. A1, Cw7, B8, DR3 HLA antigen combination associated with rapid decline of T-helper lymphocytes in HIV-1 infection. A report from the Multicenter AIDS Cohort Study. *Lancet* 1990; 335:927–30. [https://doi.org/10.1016/0140-6736\(90\)90995-h](https://doi.org/10.1016/0140-6736(90)90995-h) PMID: 1970024
37. Kilpatrick DC, Hague RA, Yap PL, et al. HLA antigen frequencies in children born to HIV-infected mothers. *Dis Markers* 1991; 9:21–6. PMID: 1742942
38. Farquhar C, Rowland-Jones S, Mbori-Ngacha D, et al. Human leukocyte antigen (HLA) B\*18 and protection against mother-to-child HIV type 1 transmission. *AIDS Res Hum Retroviruses* 2004; 20:692–7. <https://doi.org/10.1089/0889222041524616> PMID: 15307911
39. Wilson CC, Kalams SA, Wilkes BM, et al. Overlapping epitopes in human immunodeficiency virus type 1 gp120 presented by HLA A, B, and C molecules—Effects of viral variation on cytotoxic T-lymphocyte recognition. *Journal of Virology* 1997; 71:1256–64. PMID: 8995649
40. de Sorrentino AH, Marinic K, Motta P, et al. HLA class I alleles associated with susceptibility or resistance to human immunodeficiency virus type 1 infection among a population in Chaco Province, Argentina. *J Infect Dis* 2000; 182:1523–6. <https://doi.org/10.1086/315854> PMID: 11010837
41. Chaudhari DV, Chavan VR, Ahir SP, et al. Human leukocyte antigen B distribution in HIV discordant cohort from India. *Immunol Lett* 2013; 156:1–6. <https://doi.org/10.1016/j.imlet.2013.09.002> PMID: 24029662
42. Tang J, Tang S, Lobashevsky E, et al. Favorable and unfavorable HLA class I alleles and haplotypes in Zambians predominantly infected with clade C human immunodeficiency virus type 1. *J Virol* 2002; 76:8276–84. <https://doi.org/10.1128/JVI.76.16.8276-8284.2002> PMID: 12134033
43. Keet IP, Tang J, Klein MR, et al. Consistent associations of HLA class I and II and transporter gene products with progression of human immunodeficiency virus type 1 infection in homosexual men. *J Infect Dis* 1999; 180:299–309. <https://doi.org/10.1086/314862> PMID: 10395843
44. Goulder PJR, Edwards A, Phillips RE, et al. Identification of a novel HLA-A24-restricted cytotoxic T-lymphocyte epitope within HIV-1 NEF. *AIDS* 1997; 11:1883–4. <https://doi.org/10.1097/00002030-199715000-00015> PMID: 9412709
45. Furutsuki T, Hosoya N, Kawana-Tachikawa A, et al. Frequent transmission of cytotoxic-T-lymphocyte escape mutants of human immunodeficiency virus type 1 in the highly HLA-A24-positive Japanese population. *J Virol* 2004; 78:8437–45. <https://doi.org/10.1128/JVI.78.16.8437-8445.2004> PMID: 15280452
46. Leslie A, Matthews PC, Listgarten J, et al. Additive contribution of HLA class I alleles in the immune control of HIV-1 infection. *J Virol* 2010; 84:9879–88. <https://doi.org/10.1128/JVI.00320-10> PMID: 20660184
47. Zhang W, Wang L, Hong K, et al. Frequency of HLA-A 03 associates with HIV-1 infection in a Chinese cohort. *Sci China Life Sci* 2013; 56:1014–9. <https://doi.org/10.1007/s11427-013-4555-4> PMID: 24114445
48. Teixeira SL, de Sa NB, Campos DP, et al. Association of the HLA-B\*52 allele with non-progression to AIDS in Brazilian HIV-1-infected individuals. *Genes Immun* 2014; 15:256–62. <https://doi.org/10.1038/gene.2014.14> PMID: 24718028
49. Sidney J, Peters B, Frahm N, et al. HLA class I supertypes: a revised and updated classification. *BMC Immunol* 2008; 9:1. <https://doi.org/10.1186/1471-2172-9-1> PMID: 18211710
50. Casado C, Colombo S, Rauch A, et al. Host and viral genetic correlates of clinical definitions of HIV-1 disease progression. *PLoS ONE* 2010; 5:e11079. <https://doi.org/10.1371/journal.pone.0011079> PMID: 20552027
51. Ballana E, Ruiz-de AA, Mothe B, et al. Differential prevalence of the HLA-C-35 CC genotype among viremic long term non-progressor and elite controller HIV+ individuals. *Immunobiology* 2012; 217:889–94. <https://doi.org/10.1016/j.imbio.2011.12.012> PMID: 22333575

## Supporting information

**SI Table. Primers and probes employed in the determination of rs333 and rs1801157.**

Gene	Polymorphism	Primer/probe	Sequence
<i>CCR5</i>	rs333	Forward	5'AAGGTCTTCATTACACCTGCAGC3'
		Reverse	5'AGCAGCGGCAGGACCA3'
		allele 1	5'FAM-ACAGTCAGTATCAATTCTGGAAGAATTCCTA3'
		allele 2	5'VIC-TCTCATTTTCCATACATTAAGATAGTCATCTTTA3'
<i>SDF-1</i>	rs1801157	Forward	5'CGATCAACCTGGGCAAAGCC3'
		Reverse	5'AGCTTTGGTCCTGAGAGTCC3'
		allele 1	5'FAM-TGGGAGCCGGTCTGCCTCT3'
		allele 2	5'VIC-ACATGGGAGCCAGGTCTGCCTCTT3'



### 3.2. **Artículo II:** Association of a single nucleotide polymorphism in the *UBXN6* gene with long-term non-progression phenotype in HIV-positive individuals

En línea con los estudios de variantes genéticas relacionadas al fenotipo LTNP (Artículo I), decidimos abordar un estudio a nivel genómico de este grupo de pacientes, con el objetivo de adquirir un mayor conocimiento de los factores implicados en el control natural de la infección. Estudios de GWAS previos han determinado que variantes en genes que codifican las regiones HLA (*HLA-B*, *HCP5*), están más representados en pacientes LTNP. Por tanto, en el Artículo II, con el fin de encontrar nuevas variantes genéticas asociadas a este fenotipo, realizamos por primera vez el genotipado de exoma en pacientes de la cohorte española LTNP-RIS.

El presente artículo muestra que la frecuencia de los genotipos CT/TT del SNP rs1127888 del gen *UBXN6* es mayor en los pacientes LTNPs al compararlos con poblaciones control de individuos sanos y progresores típicos, además de las poblaciones europeas de los proyectos 1000 Genomas y el Exome Variant Server americano. Además, se pudo determinar una alta asociación de otros polimorfismos previamente reportados en genes como *HCP5* y *HLA-B*. Experimentos posteriores determinan la relación de *UBXN6* con la distribución y expresión de CAV-1 en la célula, y su asociación con la infección por el VIH-1.

Durante el desarrollo de esta investigación llevé a cabo gran parte del trabajo experimental y contribuí al diseño de experimentos. Realicé la estandarización del silenciamiento del gen *UBXN6* mediante ARN de interferencia, tanto en la línea celular HeLa como en cultivos primarios de células dendríticas y macrófagos. También realicé la producción de virus (NL4.3-Ren, JR-Ren) y los ensayos de infectividad para determinar el efecto del silenciamiento de *UBXN6* en la infección. Realicé el diseño de los cebadores para confirmar el genotipado del gen *UBXN6* en las muestras estudiadas mediante secuenciación. Finalmente, llevé a cabo parte del análisis de los datos y contribuí a la escritura del artículo detallando y discutiendo los resultados obtenidos.



Contents lists available at ScienceDirect

## Clinical Microbiology and Infection

journal homepage: [www.clinicalmicrobiologyandinfection.com](http://www.clinicalmicrobiologyandinfection.com)

## Original article

Association of a single nucleotide polymorphism in the *ubxn6* gene with long-term non-progression phenotype in HIV-positive individuals

F. Díez-Fuertes<sup>1,2,\*</sup>, H.E. De La Torre-Tarazona<sup>1,†</sup>, E. Calonge<sup>1</sup>, M. Pernas<sup>3</sup>, M. Bermejo<sup>1</sup>, J. García-Pérez<sup>1</sup>, A. Álvarez<sup>1</sup>, L. Capa<sup>1</sup>, F. García-García<sup>4</sup>, M. Saumoy<sup>5</sup>, M. Riera<sup>6</sup>, A. Boland-Auge<sup>7</sup>, C. López-Galíndez<sup>3</sup>, M. Lathrop<sup>7</sup>, J. Dopazo<sup>8,9,10</sup>, A. Sakuntabhai<sup>11</sup>, J. Alcamí<sup>1,2,\*</sup>

<sup>1</sup> AIDS Immunopathology Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain

<sup>2</sup> Hospital Clínic- Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

<sup>3</sup> Molecular Virology Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Madrid, Spain

<sup>4</sup> Unidad de Bioinformática y Bioestadística, Centro de Investigación Príncipe Felipe (CIFF), Valencia, Spain

<sup>5</sup> HIV Unit, Infectious Disease Service, Hospital Universitari de Bellvitge, Barcelona, Spain

<sup>6</sup> Servicio de Medicina Interna-Infecciosas, Hospital Universitario "Son Espases", Palma de Mallorca, Spain

<sup>7</sup> Centre National de Recherche en Génomique Humaine (CNRGH), Institut de Biologie François Jacob, CEA, Université Paris-Saclay, Evry, France

<sup>8</sup> Clinical Bioinformatics Area, Fundación Progreso y Salud (FPS), CDCA, Hospital Virgen del Rocío, Sevilla, Spain

<sup>9</sup> Bioinformatics in Rare Diseases (BiER), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), FPS, Hospital Virgen del Rocío, Sevilla, Spain

<sup>10</sup> INB-ELIXIR-es, FPS, Hospital Virgen del Rocío, Sevilla, Spain

<sup>11</sup> Functional Genetics of Infectious Diseases, Pasteur Institute, Paris, France

## ARTICLE INFO

## Article history:

Received 14 January 2019

Received in revised form

7 May 2019

Accepted 19 May 2019

Available online 31 May 2019

Editor: G. Antonelli

## Keywords:

CAV1

Caveolae

Exome

HIV-1

Lipid rafts

Long-term non-progressor

Single nucleotide polymorphism

UBXN6

VCP

## ABSTRACT

**Objectives:** The long-term non-progressors (LTNPs) are a heterogeneous group of HIV-positive individuals characterized by their ability to maintain high CD4<sup>+</sup> T-cell counts and partially control viral replication for years in the absence of antiretroviral therapy. The present study aims to identify host single nucleotide polymorphisms (SNPs) associated with non-progression in a cohort of 352 individuals. **Methods:** DNA microarrays and exome sequencing were used for genotyping about 240 000 functional polymorphisms throughout more than 20 000 human genes. The allele frequencies of 85 LTNPs were compared with a control population. SNPs associated with LTNPs were confirmed in a population of typical progressors. Functional analyses in the affected gene were carried out through knockdown experiments in HeLa-P4, macrophages and dendritic cells.

**Results:** Several SNPs located within the major histocompatibility complex region previously related to LTNPs were confirmed in this new cohort. The SNP rs1127888 (*UBXN6*) surpassed the statistical significance of these markers after Bonferroni correction ( $q = 2.11 \times 10^{-6}$ ). An uncommon allelic frequency of rs1127888 among LTNPs was confirmed by comparison with typical progressors and other publicly available populations. *UBXN6* knockdown experiments caused an increase in CAV1 expression and its accumulation in the plasma membrane. *In vitro* infection of different cell types with HIV-1 replication-competent recombinant viruses caused a reduction of the viral replication capacity compared with their corresponding wild-type cells expressing *UBXN6*.

\* Corresponding author. F. Díez-Fuertes, AIDS Immunopathology Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Ctra. Majadahonda-Pozuelo Km 2, 28220 Majadahonda, Madrid, Spain.

\*\* Corresponding author. J. Alcamí, AIDS Immunopathology Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Ctra. Majadahonda-Pozuelo Km 2, 28220 Majadahonda, Madrid, Spain.

E-mail addresses: [fdiezf@isciii.es](mailto:fdiezf@isciii.es) (F. Díez-Fuertes), [ppalcami@isciii.es](mailto:ppalcami@isciii.es) (J. Alcamí).

<sup>†</sup> These authors contributed equally to this work.

<https://doi.org/10.1016/j.cmi.2019.05.015>

1198-743X/© 2019 Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

**Conclusions:** A higher prevalence of Ala31Thr in UBXN6 was found among LTNPs within its N-terminal region, which is crucial for UBXN6/VCP protein complex formation. UBXN6 knockdown affected CAV1 turnover and HIV-1 replication capacity. **F. Díez-Fuertes, Clin Microbiol Infect 2020;26:107**

© 2019 Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

## Introduction

Disease progression in HIV-1 infected individuals is a complex mechanism governed by viral markers and host immune genetic factors [1]. On the virus side, several authors have described the long-term non-progressor (LTNP) condition as a consequence of the infection with attenuated viruses, showing critical mutations in *vpr*, *rev* or *nef* [2–4]. On the host side, few genetic markers have been associated with the LTNP condition, showing particularly strong associations with the human leucocyte antigen (HLA) class I molecules HLA-B5701 and HLA-B27 in Caucasian individuals [5], the chemokine receptors CCR5 and CCR2 [6,7], variants of CCR5 and CXCR4 chemokines, such as RANTES and CXCL12 [8,9], the antiretroviral restriction factor APOBEC3G [10], the restriction factor TRIM5 $\alpha$  [11] or the epistasis of inhibitory killer cell immunoglobulin-like receptor KIR3DL1 with molecules of the HLA-Bw4 family [12]. However, the predictive power of all these markers is limited and can explain only a small proportion of all the genetic variability found in LTNP individuals.

Several genome-wide analyses (GWAs) have been published to identify new host determinants associated with AIDS progression, including the study of cohorts such as Euro-CHAVI [13], GRIV [14], MACS [15], ACS [16] and GISHEAL [17] as well as different meta-analysis of European and American cohorts [18–20]. The single nucleotide polymorphism (SNP) rs2395029 located in HCP5 is the only polymorphism associated with the LTNP phenotype and reaching genome-wide significance (GWS) in at least more than two of these cohorts/multicothorts analyses. This SNP is in absolute linkage disequilibrium (LD) with the HLA-B\*5701 allele [14]. Other SNPs reaching GWS in concrete studies have not been confirmed by others. These SNPs are mainly located within the major histocompatibility complex (MHC) region in genes such as *ZNRD1*, *RNF39*, *C6orf48*, *HLA-B* and *MICA*, but also in non-MHC genes (i.e. *PROX1* and *PARD3B*) [13–20].

These studies interrogate between 300 000 and 10<sup>6</sup> SNPs along the whole-genome employing different technologies from Illumina and Affymetrix [13–20]. The strategy followed in the present study was to interrogate a similar number of SNPs (>240 000) but only in functional exome variants, including non-synonymous SNPs and polymorphisms in splice sites, stop variants and in promoter regions. Thus, the aim of the present study is to sift functional variants associated with the LTNP condition by genotyping a cohort of 352 individuals, including LTNPs and a control population (CP). This analysis has allowed the confirmation of different previously published genetic markers found in other European and American cohorts of Caucasian LTNPs, specifically those located in the MHC region (i.e. HCP5 and HLA genes). The association of the SNP rs1127888 with the LTNP condition is the main discovery of the present work, validated by the comparison with a Spanish population of HIV-positive patients with a typical pattern of disease progression. Functional analyses silencing UBXN6 expression were carried out to clarify its role in HIV-1 immunopathogenesis. UBXN6 silencing results in changes in CAV1 expression, an essential protein in the organization of lipid rafts and caveolae. Through the caveolin-1 binding motif in the transmembrane gp41, HIV binds to CAV1 and penetrates plasma membranes [21].

## Methods

### Study population

Samples from patients were kindly provided by the HIV-BioBank integrated in the Spanish Research Network (RIS). Samples were processed following current procedures and frozen immediately after they were received. All patients participating in the study gave their informed consent and protocols were approved by institutional ethics committees (Instituto de Salud Carlos III. CEI PI 10\_2011v3). A total of 352 individuals were included in the analysis, 85 LTNPs from the Spanish LTNP-RIS cohort and 267 control individuals. The LTNP individuals maintain CD4 counts over 500 cells/mm<sup>3</sup> and viral loads (VL <10 000 copies/mL) for at least 10 years from infection in the absence of ART. The CP was not selected for any particular parameter other than the Spanish origin. A second group of 58 HIV-positive individuals with a typical pattern of disease progression was used to sequence the region containing rs1127888 and to compare its frequency with the LTNP group. Typical progressors (TPs) were selected by a CD4<sup>+</sup> T-cell loss between 50 and 75 cells/mm<sup>3</sup> per year for at least 3 consecutive years.

### Exome genotyping

Genomic DNA from all the HIV-positive individuals was extracted from 10<sup>7</sup> peripheral blood mononuclear cells (PBMCs) with an AllPrep DNA/RNA Mini kit (Qiagen). Exome genotyping was performed with Illumina Infinium BeadChip technology. The genotypes of the control population of 267 seronegative individuals were obtained by exome sequencing [22]. The rs1127888 (UBXN6) genotypes were confirmed by an in-house PCR amplification and sequencing employing the primers 5'-CCACGGTGTCTCATGTGACTT-3' and 5'-GTCAAGGTGCCTTACCTCAG-3' to obtain a 582-bp amplicon.

### Population-based association studies

The genotype of the 85 LTNPs was obtained for 247 722 SNPs. All the individuals were genotyped in at least 95% of the positions included in the analysis. The estimation of pairwise identity by descent was calculated for two individuals genotyped twice to confirm the reproducibility of the method. A Hardy–Weinberg equilibrium test was assayed and all the SNPs in disequilibrium ( $p < 0.005$ ) in the control population were discarded. Finally, the presence of outliers that are not of European ancestry was assessed by principal components analysis [23].

The case–control association studies were carried out by a Fisher exact test of allelic association ( $p$ -value) and corrected for multiple tests by the Bonferroni's method ( $q$ -values). The LD was calculated according to the genotypes of the 503 Europeans sequenced in the 1000 Genomes Project Phase 3. The data were obtained from the EBI ftp site, and LD calculations were inferred for a window of 200 kb with VCFtools. The statistical power of the study was calculated for the SNP rs1127888 as the probability of detecting a variant assumed to cause the LTNP phenotype. The statistical power was computed using a web browser program

(<http://osse.bii.a-star.edu.sg/index.php>) specifying the sample size and the minor allele frequencies obtained in LTNP and healthy populations with a significance level of  $2.0 \times 10^{-4}$ .

#### Cell cultures

HeLa cells engineered to express CD4 and CXCR4 (HeLa–P4) were seeded in DMEM (Gibco) supplemented with 5% (v/v) fetal bovine serum (FBS), geneticine (0.5 mg/mL, Gibco), gentamicin (50 µg/mL, Gibco) and puromycin (1 µg/mL, Invivogen). Cells were cultured at 37°C in a 5% CO<sub>2</sub> humidified atmosphere and split twice a week. We evaluated the infection in human dendritic cells (DCs) and macrophages, because CAV1 is the major coat protein responsible for caveolae assembly and is highly expressed in these cell types [24]. Primary cells were generated and cultured as previously described [25].

#### Plasmids and viral stock production

The vector pNL4.3Ren was generated by cloning a *Renilla* luciferase reporter gene in the HIV-1 proviral clone pNL4.3. The pJR-Ren plasmid was generated by cloning gp160 from the JR-FL clone (R5 tropism) in place of the NL4.3 env gene in pNL4.3Ren. Infectious viral supernatants were obtained from calcium phosphate transfection on HEK293T cells of pNL4.3Ren and pJR-Ren.

#### UBXN6 knockdown

A mix of endonuclease-prepared small interfering RNAs (esiRNA) against human UBXN6 were purchased from Sigma-Aldrich Inc. (MISSION® esiRNA) to silence UBXN6 expression. Non-targeting siRNA or esiRNA against UBXN6 was transfected into HeLa–P4 cells, macrophages or DCs using HiPerfect reactive from Qiagen Inc. following the manufacturer's recommendations. Briefly, HeLa–P4 cells, macrophages or DCs were seeded into 24-well plates and 100 ng of siRNAs and 3 µL of HiPerfect were used for each transfection. Immunoblotting, immunofluorescence, detection of HIV-1 core antigens by flow cytometry and the measurement of luciferase activity were carried out as previously described [25]. The antibodies employed for these experiments were a mouse

anti-UBXN6 antibody (5C3-1, Abcam Inc.), a rabbit anti-Caveolin-1 antibody (N20, sc-894, Santa Cruz Biotechnology) and a FITC-conjugated human anti-Gag/p24 (KC57, Beckman Coulter).

#### Statistical analyses

The case–control association analyses between groups of individuals, Hardy–Weinberg equilibrium tests, Fisher exact tests (p-values) and Bonferroni corrections for multiple testing (q-values) were assayed by PLINK software [26]. Analysis of sample stratification by EIGENSTRAT package was carried out by a collection of perl scripts in Unix environments [23]. The Shapiro–Wilk test was used to test the normality of the relative infectivity obtained and Mann–Whitney or *t*-tests to detect differences between conditions in R software.

### Results

#### The SNP rs1127888 is associated with LTNP phenotype

The main difference in the comparison of the allelic frequencies in LTNPs and CP was found in the SNP rs1127888 ( $q = 2.11 \times 10^{-6}$ ) on chromosome 19, position 4 454 086 (Table 1). The SNP rs1127888 is located in position 216 of the transcript encoding UBXN6 protein, causing the missense variant A31T in the N-terminal region of the protein. The variation effect of A31T change was predicted as a tolerated variation on UBXN6 protein function (Table 1). A statistical power of 80.3% at the  $2.0 \times 10^{-4}$  significance level was reached for rs1127888 using the sample size included in the study and the allelic frequencies observed.

The frequencies of the SNP rs1127888 located in UBXN6 (which encodes the UBX domain protein 6) were confirmed by sequencing. The comparison with the European population of 1000 Genomes Project ( $p = 2.82 \times 10^{-3}$ ) and the European–American population of NHLBI Exome Sequencing Project ( $p = 3.31 \times 10^{-5}$ ) confirmed an uncommon allelic frequency of rs1127888 in the LTNP population (Fig. 1). These frequencies were also compared with a population of 58 HIV-infected individuals with a typical pattern of disease progression. No differences were found between the genotypic frequencies in TP and the rest of the control populations (Fig. 1).

**Table 1**  
Top single nucleotide polymorphism (SNP) associations to AIDS long-term non-progression

SNP	Location	Gene	Consequence <sup>a</sup>	Position in protein	SIFT <sup>b</sup>	PolyPhen <sup>c</sup>	p	q <sup>d</sup>
rs1127888	19:4454086	UBXN6	Missense variant	A31T	0.33	Benign	$7.34 \times 10^{-11}$	$2.11 \times 10^{-6}$
rs3819299	6:31354590	HLA-B	Non-coding transcript exon variant	—	—	—	$2.89 \times 10^{-6}$	$1.22 \times 10^{-4}$
rs2395029	6:31464003	HCP5	Non-coding transcript exon variant	—	—	—	$1.98 \times 10^{-5}$	$7.37 \times 10^{-4}$
rs2301734	19:36084743	WDR62	Splice region variant	T547	—	—	$2.41 \times 10^{-5}$	$8.81 \times 10^{-4}$
rs667859	8:2963223	CSMD1	Splice region variant	L2569	—	—	$2.79 \times 10^{-5}$	$1.01 \times 10^{-3}$
rs9368699	6:31802541	C6orf48	5 prime UTR variant	—	—	—	$3.29 \times 10^{-5}$	$1.17 \times 10^{-3}$
rs2306242	4:849932	GAK	Missense variant	K1265R	0.09	Benign	$3.29 \times 10^{-5}$	$1.17 \times 10^{-3}$
rs444772	8:54626497	RP1	Missense variant	R872H	0.01	Benign	$4.69 \times 10^{-5}$	$1.62 \times 10^{-3}$
rs3796375	3:45967298	FYCO1	Missense variant	A679V	0.04	Possibly damaging	$5.01 \times 10^{-5}$	$1.70 \times 10^{-3}$
rs34822421	11:563400	RASSF7	Missense variant	P319L	0.88	Benign	$5.03 \times 10^{-5}$	$1.70 \times 10^{-3}$
rs75370284	11:78210302	USP35	Missense variant	R816H	0	Possibly damaging	$5.03 \times 10^{-5}$	$1.70 \times 10^{-3}$
rs142030651	22:41117723	EP300	Missense variant	G211S	—	Unknown	$5.03 \times 10^{-5}$	$1.70 \times 10^{-3}$
rs204900	6:32088803	TNXB	Missense variant	S921A	0.02	Benign	$6.41 \times 10^{-5}$	$2.11 \times 10^{-3}$
rs822431	1:156932489	LRRRC71	Missense variant	S503A	0.1	Possibly damaging	$7.65 \times 10^{-5}$	$2.48 \times 10^{-3}$
rs2302607	16:8635267	METTL22	Missense variant	R205H	0	Probably damaging	$8.14 \times 10^{-5}$	$2.62 \times 10^{-3}$
rs7249069	19:8333122	KANK3	Missense variant	E610K	0.58	Possibly damaging	$9.08 \times 10^{-5}$	$2.90 \times 10^{-3}$

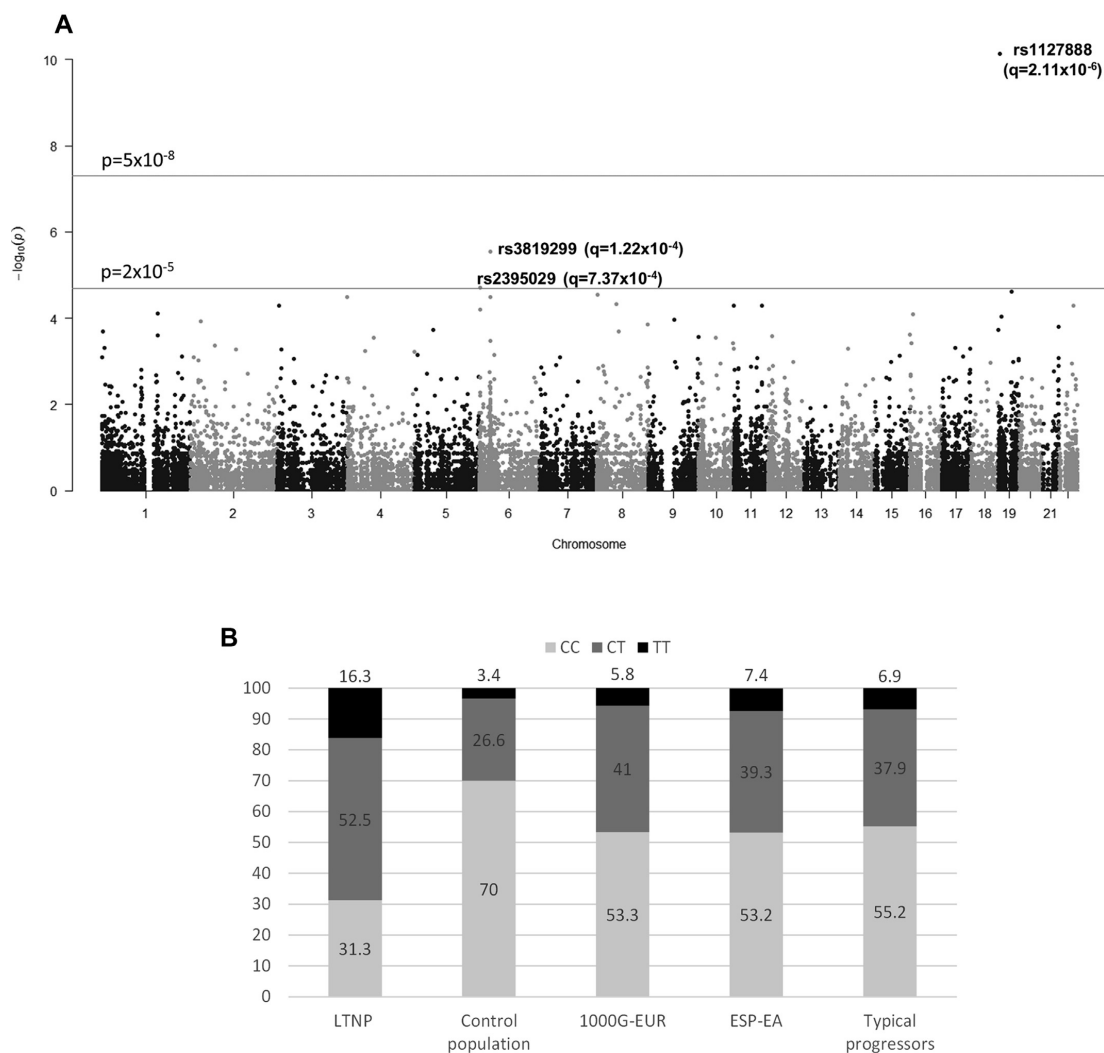
The SNPs with higher p-values in the LTNP/CP association study are shown. The table includes the location of each SNP, the gene affected and the position within the gene, the most severe consequence and the effect prediction of the amino acid substitution (SIFT and PolyPhen) and the uncorrected p-value obtained for the analysis and their corresponding p-value after Bonferroni correction (q-value).

<sup>a</sup> Most severe consequence.

<sup>b</sup> SIFT score ranges from 0 to 1. The amino acid substitution is predicted damaging if the score is  $\leq 0.05$  and tolerated if the score is  $> 0.05$ .

<sup>c</sup> PolyPhen prediction of SNPs on protein function and structure.

<sup>d</sup> Bonferroni correction for multiple testing.



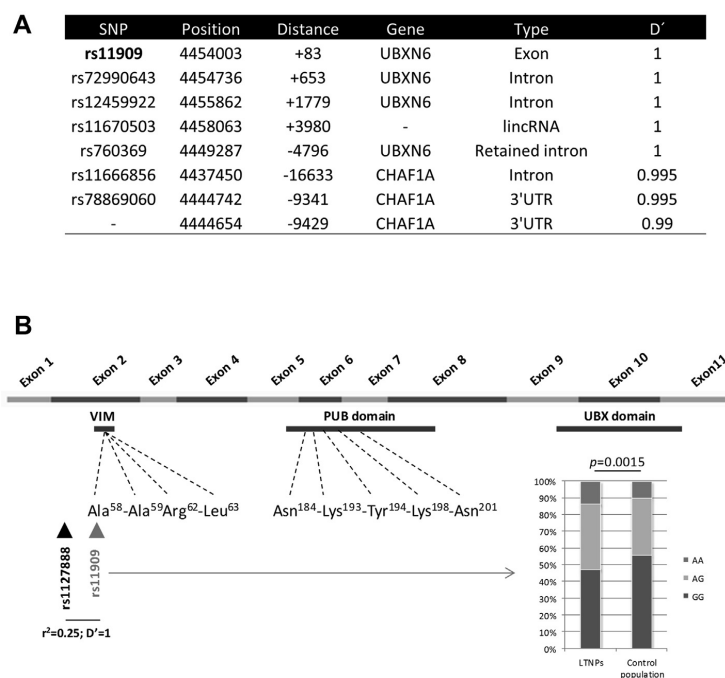
**Fig. 1.** Long-term non-progressor (LTNP)/control population (CP) association study and allelic frequencies of rs1127888 in different populations. Manhattan plot showing the p-values obtained for each SNP analyzed in the LTNP/CP association study (A). The barplot shows genotypic frequencies of rs1127888 obtained for LTNP population, CP, typical progressors (TPs), the European-American population of the NHLBI Exome Sequencing Project (ESP-EA) and the European super population of the 1000 Genomes Project Phase 3 (B).

Statistically significant differences were found comparing the frequencies obtained for LTNP and TP ( $p = 1.4 \times 10^{-2}$ ).

The genotypic frequencies of all the SNPs annotated in Ensembl in exon 2 of UBXN6 were investigated by sequencing a 582-bp amplicon (see Methods) and were compared with those of the CP. A statistically significant association was found at rs11909 ( $p = 1.5 \times 10^{-3}$ ), a SNP in partial LD with rs1127888 ( $r^2 = 0.25/D' = 1$ ), located in Ala58 of UBXN6 (Fig. 2). The mutation of this alanine to leucine completely abrogates the interaction of UBXN6 with VCP and affects CAV1 trafficking [27].

#### Other genetic markers associated with AIDS long-term non-progression

The SNP rs3819299, which is related with *HLA-B* (Table 1), was also associated with the LTNP condition ( $q = 1.22 \times 10^{-4}$ ). This SNP is in strong LD ( $r^2 \geq 0.9$ ) with several SNPs located in *HLA-B* and within a 20-kbp region located between *HLA-B* and *HCP5/MICA*, where several processed pseudogenes are annotated (please see supplementary material). The analysis also confirmed the protective role of the HCP5 rs2395029-G marker ( $q = 7.37 \times 10^{-4}$ ). Based



**Fig. 2.** Linkage disequilibrium (LD) analysis of rs1127888 and transcript structure of UBXN6 gene. A 582-bp amplicon including rs1127888 was sequenced in long-term non-progressors (LTNPs) and typical progressors (TPs). The correlation coefficient ( $D'$ ) of rs1127888 with all the single nucleotide polymorphisms (SNPs) included in this amplicon was calculated for the European super population of the 1000 Genomes Project Phase 3. The SNPs with higher values of  $D'$  are showed in the table (A). The transcript structure of UBXN6 gene is shown along with the location of the functional domains PUB, UBX and VIM (VCP interacting motif). The location of rs1127888 and rs11909 within exon 2 of UBXN6 as well as the location of several residues important for VCP/UBXN6 complex formation are indicated (according to [27,28]). Differences in rs11909 genotype frequencies obtained for LTNP and CP is also included ( $p = 0.0015$ ) (B).

on the LD analysis, rs2395029 has a strong correlation with several markers at *HCP5* and within the 20-kbp region located between *HLA-B* and *HCP5/MICA*, but not with rs3819299 (please see [supplementary material](#)). Another genetic marker located within the 5'-UTR of *C6orf48* (rs9368699) was associated with non-progression ( $q = 1.17 \times 10^{-3}$ ), supporting the results obtained by other authors [16,19]. The SNP rs9368699 is in partial LD with rs2395029 ( $r^2 = 0.615$ ;  $D' = 0.856$ ), but not with rs3819299 ( $r^2 = 0$ ;  $D' = 0.093$ ).

#### *UBXN6-knockdown modulates CAV1 expression and its cellular location*

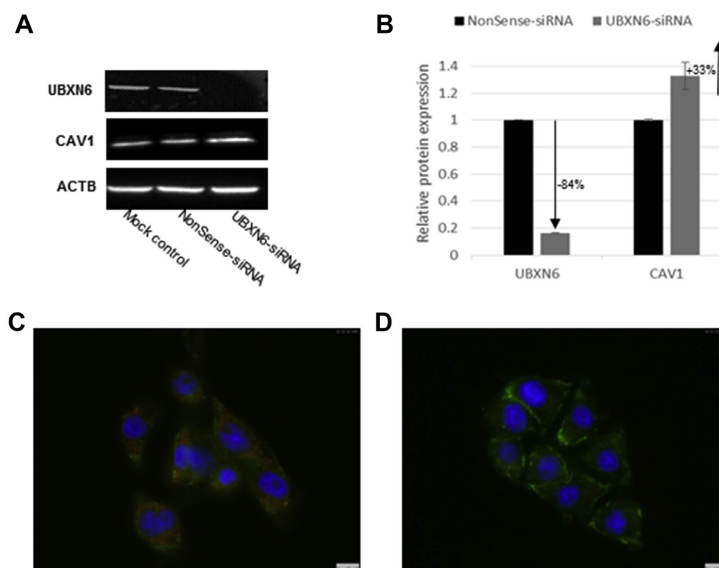
CAV1 is the major constituent of a special form of lipid rafts named caveolae and UBXN6 is involved in CAV1 expression and recycling. The effect of UBXN6-silencing in CAV1 expression and cellular distribution was assayed by Western blot and immunofluorescence. Western blot analyses confirmed that the expression of UBXN6 dropped 84% in cells treated with UBXN6-siRNA compared with cells treated with non-sense siRNA (Fig. 3). Concurrently, a 33% boost in CAV1 expression was observed in UBXN6-siRNA cells in comparison with non-sense siRNA treatment. An accumulation of CAV1 in cell periphery of cells treated with UBXN6-siRNA was observed by immunofluorescence compared with cells treated with non-sense siRNA (Fig. 3).

#### *UBXN6 knockdown reduces HIV-1 replication*

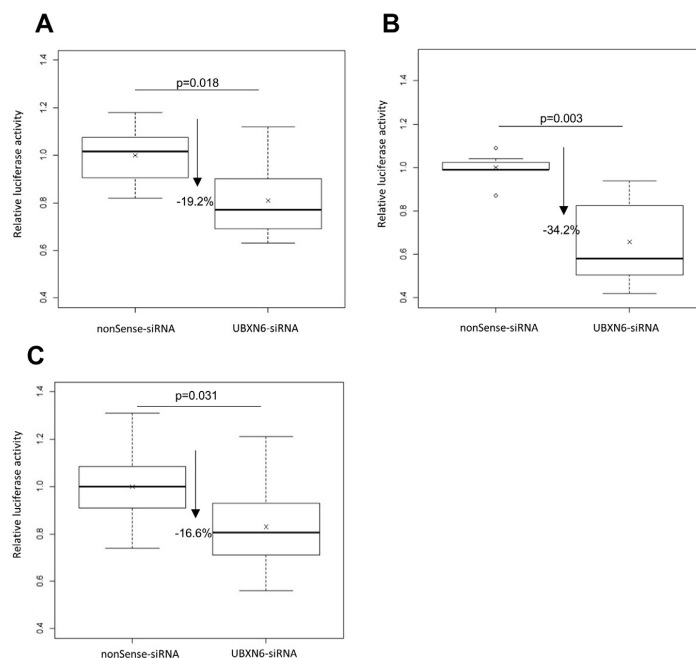
Decreased recycling and CAV1 accumulation in cell periphery could affect HIV-1 infectivity. To test this effect, different cell types treated with UBXN6-siRNA were infected with recombinant viruses expressing a reporter gene and compared with control cells. The relative luciferase activity (RLA) of HeLa-P4 cells previously treated with UBXN6-siRNA was reduced 19.2% compared with non-sense siRNA after the infection with the replication-competent NL4-3Ren virus ( $p = 0.018$ ; Fig. 4). The role of UBXN6 silencing in macrophages and DCs after infection with the R5 tropic virus JR-Ren was also explored. The RLA was reduced 34.2% in DCs treated with UBXN6-siRNA compared with non-sense-siRNA ( $p = 0.003$ ) (Fig. 4). In the case of UBXN6-knockdown macrophages, the RLA was reduced a mean of 16.6% compared with non-sense siRNA ( $p = 0.031$ ) (Fig. 4). The percentage of cells expressing p24 precursors measured by flow cytometry was lower in cells treated with UBXN6-siRNA (76.5% compared with 87% of cells treated with non-sense siRNA). This result was consistent with the lower RLA observed in cells treated with UBXN6-siRNA.

#### **Discussion**

The association of SNPs with LTNP status has been analysed extensively with genome-wide microarrays. The leading role of MHC markers in the natural delay of AIDS progression observed in LTNP has been repeatedly reported and confirmed once again in the



**Fig. 3.** Knockdown of UBXN6 and implications in CAV1 expression. The expression of UBXN6, CAV1 and ACTB proteins were detected by Western blot with specific antibodies for each treatment (A). The expression of UBXN6 dropped 84% in HeLa–P4 cells treated with the specific siRNA whereas CAV1 expression was increased 33% in these cells (B). Representative images of HeLa–P4 cells treated with non-sense siRNA (C) and UBXN6-siRNA (D) obtained by indirect immunofluorescence using specific antibodies against UBXN6 and CAV1 and secondary antibodies conjugated with Alexa Fluor (red for UBXN6 and green for CAV1). Nuclei were stained with DAPI (blue).



**Fig. 4.** Knockdown of UBXN6 and implications in HIV-1 replication capacity. Luciferase activity on HeLa–P4 (A), DCs (B) and macrophages (C) previously treated with nonSense-siRNA or UBXN6-siRNA and infected with NL4.3Ren (A) or JR-Ren (B and C) at 48 hr post infection. The relative luciferase activity obtained for each replicate is expressed as the relation with the mean obtained for non-sense siRNA treatment. Boxplots include interquartile range (boxes), median (horizontal line within boxes), mean (cross within boxes), maximum and minimum values (lines extending vertically from the boxes) and outliers (opened circles). Mann–Whitney test was used to determine statistical significance between treatments. The boxplots show the results obtained for three replicates in at least two independent experiments.



present study [13–20]. Thus, previously identified SNPs in MHC class I genes such as *HCP5*, *MICA*, *HLA-B* and in MHC class III genes such as *C6orf48* have also been found in our cohort of LTNP. Moreover, a new association in 19p13.3 related to *UBXN6* has been identified. A higher proportion of LTNPs with CT and TT genotypes for the SNP rs1127888 was found compared with CP, TP and other publicly available populations of European ancestry. The validation of this SNP in other cohorts of LTNPs of European ancestry genotyped in previously published GWAS studies [13–20] was not possible because rs1127888 and the SNPs in complete linkage disequilibrium with rs1127888 were not interrogated in the microarrays platforms employed in these studies.

The importance of *UBXN6* in HIV-1 immunopathogenesis could be explained by the specific role of *UBXN6* as a cofactor of the complex formed by *CAV1* and *VCP* [27]. The protein *CAV1* is the major constituent of caveolae, a cholesterol-rich invagination located in the plasma membrane considered as a specialized form of lipid rafts [27,28]. The *VCP* protein is an AAA<sup>+</sup>-type ATPase that binds *UBXN6* cofactor to form a protein complex that in turn targets mono-ubiquitylated *CAV1* on endosomes to ensure a correct trafficking to lysosomes where *CAV1* is degraded [28]. A *VCP*-interacting motif (*VIM*) has been identified within the N-terminal region of *UBXN6* as an independent binding site of *UBXN6* to the N-domain of *VCP* protein [27]. A higher prevalence of A31T (within the N-terminal region of *UBXN6*) among LTNPs suggests that this amino acid change could imply differences in the *UBXN6/VCP* interaction, impairing *CAV1* trafficking and leading to the accumulation of *CAV1* in plasma membrane. Furthermore, rs1127888 is in partial LD with rs11909, a SNP also showing differences in the genotypic frequencies between LTNPs and CP. This SNP is located 83 nucleotides apart from rs1127888 and is included in the Ala<sup>58</sup> codon of *VIM*. Interestingly, the mutation of this Ala<sup>58</sup> to leucine completely suppresses the interaction with *VCP* [27]. Chemical inhibition of *VCP*, mutations of crucial positions to *UBXN6/VCP* or *VCP/CAV1* interactions and siRNA-mediated depletion of *UBXN6* have previously been related to a block of *CAV1* recycling [28].

Knockdown of *UBXN6* in HeLa–P4 cells, macrophages and DCs infected with a virus that is able to produce multiple rounds of replication (NL4.3Ren/JR-Ren) caused a lower viral replication than control cells. These results could point to restricted virus trafficking in infected cells where *CAV1* internalization is reduced as a consequence of *UBXN6* depletion. These results support other studies indicating that the overexpression of *CAV1* significantly inhibits HIV replication [29]. The main discovery of the present work is the association of rs1127888-T located in *UBXN6* with LTNP phenotype along with the role of this gene regulating *CAV1* availability within cells and affecting HIV-1 replication capacity. The effect of this accumulation of *CAV1* in the plasma membrane seems to restrict virus infection by mechanisms that need to be further investigated.

### Transparency declaration

Dr Capa reports grants from European Commission, from null, during the conduct of the study. The other authors have nothing to disclose. This study was partially supported by the Spanish AIDS Research Network (RIS), funded by Instituto de Salud Carlos III and co-funded by European Regional Development Fund (ERDF) “A way to build Europe” (projects RD12/0017/0015 and RD16CIII/0002/0001, Plan Estatal de I+D+I 2013–2016), the French Government's Investissement d'Avenir programme, Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (no. ANR-10-LABX-62-IBEID), the Spanish Government's FIS-ISCI (PI16CIII/00034) and Sara Borrell Postdoctoral Programme (CD12/00515 to

F.D.F.) and the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 681137.

### Acknowledgements

We want to particularly acknowledge the patients in this study for their participation and to the HIV BioBank integrated in the Spanish AIDS Research Network and collaborating Centres for the generous gifts of clinical samples used in this work. The HIV BioBank and the AIDS Immunopathogenesis Unit are integrated in the Spanish AIDS Research Network.

### Appendix A. Supplementary data

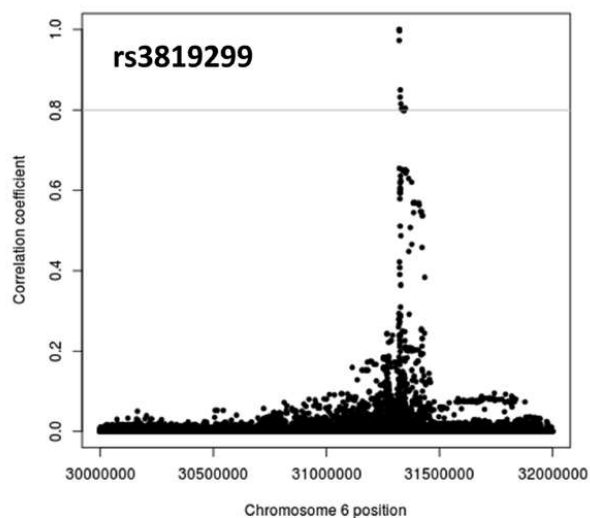
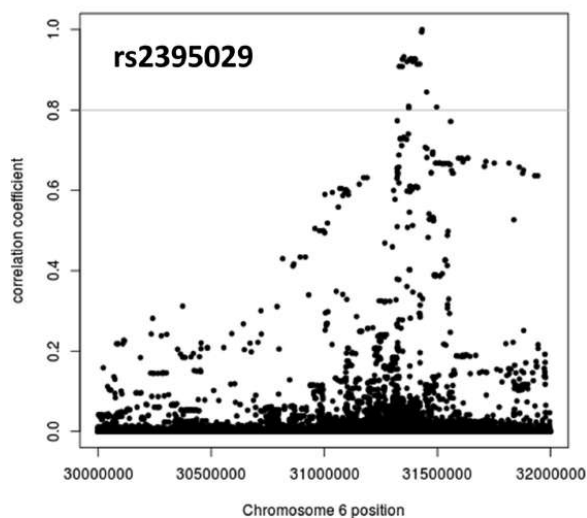
Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cmi.2019.05.015>.

### References

- Poropatich K, Sullivan Jr. Human immunodeficiency virus type 1 long-term non-progressors: the viral, genetic and immunological basis for disease non-progression. *J Gen Virol* 2011;92:247–68.
- Mologni D, Citterio P, Menzaghi B, Zanone Poma B, Riva C, Broggin V, et al. Vpr and HIV-1 disease progression: R77Q mutation is associated with long-term control of HIV-1 infection in different groups of patients. *AIDS* 2006;20:567–74.
- Iversen AK, Shpaer EG, Rodrigo AG, Hirsch MS, Walker BD, Sheppard HW, et al. Persistence of attenuated rev genes in a human immunodeficiency virus type 1-infected asymptomatic individual. *J Virol* 1995;69:5743–53.
- Casartelli N, Di Matteo G, Argentin C, Cancrini C, Bernardi S, Castelli G, et al. Structural defects and variations in the HIV-1 nef gene from rapid, slow and non-progressor children. *AIDS* 2003;17:1291–301.
- Miguel SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, Martino L, et al. HLA B\*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *J Virol* 2000;84:7581–91.
- Wang B, Palasanthiran P, Zeigler J, Cunningham A, Saksena NK. CCR5-delta 32 gene deletion in HIV-1 infected patients. *Lancet* 1997;350:741–2, 1997.
- Kostrakis LG, Huang Y, Moore JP, Wolinsky SM, Zhang L, Guo Y, et al. A chemokine receptor CCR2 allele delays HIV-1 disease progression and is associated with a CCR5 promoter mutation. *Nat Med* 1998;4:350–3.
- Scala E, D'Offizi G, Rosso R, Turriziani O, Ferrara R, Mazzone AM, et al. C-C chemokines, IL-16, and soluble antiviral factor activity are increased in cloned T cells from subjects with long-term nonprogressive HIV infection. *J Immunol* 1997;158:4485–92.
- Amara A, Gall SL, Schwartz O, Salamero J, Montes M, Loetscher P, et al. HIV coreceptor downregulation as antiviral principle: SDF-1alpha-dependent internalization of the chemokine receptor CXCR4 contributes to inhibition of HIV replication. *J Exp Med* 1997;186:139–46.
- Vázquez-Pérez JA, Ormsby CE, Hernández-Juan R, Torres KJ, Reyes-Terán G. APOBEC3G mRNA expression in exposed seronegative and early stage HIV infected individuals decreases with removal of exposure and with disease progression. *Retrovirology* 2009;6:23.
- Javanbakht H, An P, Gold B, Petersen DC, O'Huigin C, Nelson GW, et al. Effects of human TRIM5alpha polymorphisms on antiretroviral function and susceptibility to human immunodeficiency virus infection. *Virology* 2006;354:15–27.
- Alter G, Altfeld M. NK cells in HIV-1 infection: evidence for their role in the control of HIV-1 infection. *J Intern Med* 2009;265:29–42.
- Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science* 2007;317:944–7.
- Limou S, Le Clerc S, Coulonges C, Carpentier W, Dina C, Delaneau O, et al. Genome-wide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genome-wide Association Study 02). *J Infect Dis* 2009;199:419–26.
- Herbeck JT, Gottlieb GS, Winkler CA, Nelson GW, An P, Maust BS, et al. Multistage genome-wide association study identifies a locus at 1q41 associated with rate of HIV-1 disease progression to clinical AIDS. *J Infect Dis* 2010;201:618–26.
- van Manen D, Delaneau O, Kootstra NA, Boeser-Nunnink BD, Limou S, Bol SM, et al. Genome-wide association scan in HIV-1-infected individuals identifying variants influencing disease course. *PLoS One* 2011;6:e22208.
- Guernon J, Dalmasso C, Broet P, Meyer L, Westrop SJ, Imami N, et al. Single-nucleotide polymorphism-defined class I and class III major histocompatibility complex genetic subregions contribute to natural long-term non-progression in HIV infection. *J Infect Dis* 2012;205:718–24.



- [18] Troyer JL, Nelson GW, Lautenberger JA, Chinn L, McIntosh C, Johnson RC, et al. Genome-wide association study implicates PARD3B-based AIDS restriction. *J Infect Dis* 2011;203:1491–502.
- [19] Le Clerc S, Coulonges C, Delaneau O, Van Manen D, Herbeck JT, Limou S, et al. Screening low-frequency SNPs from genome-wide association study reveals a new risk allele for progression to AIDS. *J Acquir Immune Defic Syndr* 2011;56: 279–84.
- [20] McLaren PJ, Coulonges C, Ripke S, van den Berg L, Buchbinder S, Carrington M, et al. Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. *PLoS Pathog* 2013;9:e1003515.
- [21] Benferhat R, Sanchez-Martinez S, Nieva JL, Briand JP, Hovanessian AG. The immunogenic CBD1 peptide corresponding to the caveolin-1 binding domain in HIV-1 envelope gp41 has the capacity to penetrate the cell membrane and bind caveolin-1. *Mol Immunol* 2008;45:1963–75.
- [22] García-Alonso L, Jiménez-Almazán J, Carbonell-Caballero J, Vela-Boza A, Santoyo-López J, Antinolo G, et al. The role of the interactome in the maintenance of deleterious variability in human populations. *Mol Syst Biol* 2014;10:752.
- [23] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–9.
- [24] Gargalovic P, Dory L. Caveolins and macrophage lipid metabolism. *J Lipid Res* 2003;44:11–21.
- [25] Calonge E, Bermejo M, Díez-Fuertes F, Mangeot I, González N, Coiras M, et al. Different expression of interferon-stimulated genes in response to HIV-1 infection in dendritic cells based on their maturation state. *J Virol* 2017;91: e01379–16.
- [26] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
- [27] Stapf C, Cartwright E, Bycroft M, Hofmann K, Buchberger A. The general definition of the p97/valosin-containing protein (VCP)-interacting motif (VIM) delineates a new family of p97 cofactors. *J Biol Chem* 2011;286:38670–8.
- [28] Ritz D, Vuk M, Kirchner P, Bug M, Schütz S, Hayer A, et al. Endolysosomal sorting of ubiquitylated caveolin-1 is regulated by VCP and UBXD1 and impaired by VCP disease mutations. *Nat Cell Biol* 2011;3:1116–23.
- [29] Lin S, Wang XM, Nadeau PE, Mergia A. HIV infection upregulates caveolin 1 expression to restrict virus production. *J Virol* 2010;84:9487–96.



SNP	Gene	Location	R2
rs115846244	HCP5	ncRNA	0.99
rs114607072	HLA-S	pseudogen	0.93
rs41558312	MICA	Q114R	0.93
rs148792134	HCP5	ncRNA	0.93
rs140810304	HCP5	ncRNA	0.93
rs138130755	HCP5	ncRNA	0.93
rs116081995	HCP5	ncRNA	0.93
rs115841246	HCP5	ncRNA	0.93
rs114170382	ZDHHC20P2	Upstream	0.93
rs138099588	HCP5	Upstream	0.93
rs116419909	HCP5	Upstream	0.92
rs144027808	HCP5	Upstream	0.91
rs140991764	HCP5	Upstream	0.91
rs115986568	HCP5	Upstream	0.91

SNP	Gene	Location	R2
rs17192932	HLA-B	Intron	1
rs3819282	HLA-B	Intron	1
rs140769830	HLA-B	3' UTR	0.97
rs35267732	HLA-B	Upstream	0.85
rs117486637	HLA-B	Upstream	0.83
rs34955377	HLA-B	Upstream	0.81
rs13198748	ZDHHC20P2	Upstream	0.80
rs115378916	ZDHHC20P2	Upstream	0.80
rs116579023	DHFRP2	Downstream	0.80
rs4463302	ZDHHC20P2	Upstream	0.80
rs13202464	ZDHHC20P2	Upstream	0.80
rs13198903	ZDHHC20P2	Upstream	0.80

### 3.3. Artículo III: Transcriptome sequencing of peripheral blood mononuclear cells from elite controller-long term non progressor

Siguiendo en la búsqueda de factores asociados al control natural de la infección por el VIH-1, en el artículo III decidimos abordar el estudio a nivel de expresión génica de las células del sistema inmune de pacientes con fenotipos no progresor y controlador, con el objetivo de adquirir un mayor conocimiento del perfil transcriptómico asociado al control natural de la infección.

El presente artículo muestra que existe una expresión génica diferencial al comparar los EC-LTNPs con los otros fenotipos de progresión de la infección. Además, mediante algoritmos de clasificación supervisada, se pudieron identificar 20 genes y pseudogenes como predictores de la progresión, capaces de clasificar con alta eficacia entre los fenotipos LTNP y TP estudiados. Asimismo, se pudo observar que los pacientes EC-LTNP tienen una mayor expresión de genes relacionados con la movilización y transporte del calcio en la célula, así como una expresión coordinada de los genes *CDKN1A*, *GADD45B*, *IER3* y *TNF*, por lo que se proponen nuevos mecanismos asociados al control natural de la infección por el VIH.

Durante el desarrollo de esta investigación participé en la extracción de ARN de las muestras de los pacientes, análisis de la calidad del ARN, así como en el aprendizaje y la generación de las librerías genómicas. Los análisis de secuencias de datos, expresión génica, búsqueda de genes predictores, anotación funcional, los realicé conjuntamente y bajo la dirección del Dr. Francisco Díez. Finalmente, llevé a cabo parte del análisis de los datos y contribuí a la escritura del artículo detallando y discutiendo los resultados obtenidos. Además, dentro de esta línea de investigación, se presentarán datos aún no publicados sobre la expresión del gen *CDKN1A* en subpoblaciones de linfocitos T CD4+.

OPEN

# Transcriptome Sequencing of Peripheral Blood Mononuclear Cells from Elite Controller-Long Term Non Progressors

Received: 8 February 2019  
Accepted: 12 September 2019  
Published online: 03 October 2019

Francisco Díez-Fuertes<sup>1,2</sup>, Humberto Erick De La Torre-Tarazona<sup>1</sup>, Esther Calonge<sup>1</sup>, María Pernas<sup>3</sup>, María del Mar Alonso-Socas<sup>4</sup>, Laura Capa<sup>1</sup>, Javier García-Pérez<sup>1</sup>, Anavaj Sakuntabhai<sup>5</sup> & José Alcami<sup>1,2</sup>

The elite controller (EC)-long term non-progressor (LTNP) phenotype represent a spontaneous and advantageous model of HIV-1 control in the absence of therapy. The transcriptome of peripheral blood mononuclear cells (PBMCs) collected from EC-LTNPs was sequenced by RNA-Seq and compared with the transcriptomes from other phenotypes of disease progression. The transcript abundance estimation combined with the use of supervised classification algorithms allowed the selection of 20 genes and pseudogenes, mainly involved in interferon-regulated antiviral mechanisms and cell machineries of transcription and translation, as the best predictive genes of disease progression. Differential expression analyses between phenotypes showed an altered calcium homeostasis in EC-LTNPs evidenced by the upregulation of several membrane receptors implicated in calcium-signaling cascades and intracellular calcium-mobilization and by the overrepresentation of NFAT1/Elk-1-binding sites in the promoters of the genes differentially expressed in these individuals. A coordinated upregulation of host genes associated with HIV-1 reverse transcription and viral transcription was also observed in EC-LTNPs –i.e. p21/CDKN1A, TNF, IER3 and GADD45B. We also found an upregulation of ANKRD54 in EC-LTNPs and viremic LTNP in comparison with typical progressors and a clear alteration of type-I interferon signaling as a consequence of viremia in typical progressors before and after receiving antiretroviral therapy.

The chronic asymptomatic phase in HIV-1 pathogenesis is extremely variable, spanning from 2 to 25 years depending on the individual rate of disease progression defined by the interaction of host and viral factors<sup>1,2</sup>. However, a median time to AIDS since seroconversion between 8 and 11 years is generally accepted<sup>3</sup>. In order to categorize this variability, HIV-specialists have created a classification of HIV-1 infected individuals according to the disease progression, mainly measured by the loss of CD4<sup>+</sup> T cells. In this sense, an extreme phenotype observed in long-term non-progressors (LTNPs) represents about 2% of all HIV-1 infected individuals and is characterized by the preservation of CD4<sup>+</sup> T cell levels above 500 cells per µl of blood and relative low levels of viremia for at least ten years in the absence of ART<sup>3</sup>. Although some studies employ shorter periods of time to define LTNP condition, the use of 10 years of non-progression better differentiates between “true” LTNPs from those with delayed progression<sup>4</sup>. In parallel, some individuals called elite controllers (ECs) have the capacity to maintain undetectable levels of viral RNA without therapy for at least two years<sup>5</sup>. The prevalence of LTNPs and ECs have been determined in a huge military cohort of 4.586 naive HIV-1 infected individuals, representing the

<sup>1</sup>AIDS Immunopathology Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Ctra. Majadahonda-Pozuelo, Km. 2, 28220, Majadahonda, Madrid, Spain. <sup>2</sup>Hospital Clínic-Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), 08036, Barcelona, Spain. <sup>3</sup>Molecular Virology Unit, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Ctra. Majadahonda-Pozuelo, Km. 2, 28220, Majadahonda, Madrid, Spain. <sup>4</sup>Servicio de infecciones, Hospital Universitario de Canarias, 38320, Santa Cruz de Tenerife, Spain. <sup>5</sup>Functional Genetics of Infectious Diseases, Pasteur Institute, 75015, Paris, France. Francisco Díez-Fuertes and Humberto Erick De La Torre-Tarazona contributed equally. Correspondence and requests for materials should be addressed to F.D.-F. (email: [frdiez@clinic.cat](mailto:frdiez@clinic.cat)) or J.A. (email: [ppalcami@isci.es](mailto:ppalcami@isci.es))

Patient group	Gender		Population	Viral load in sample*	CD4 T cell count*
	Male	Female			
EC-LTNP	5 (62.5%)	3 (37.5%)	European	393 (Und.** – 1137)	667 (514–1081)
vLTNP	7 (87.5%)	1 (12.5%)	European	6021 (243–18900)	738 (492–1049)
TP	6 (85.7%)	1 (14.3%)	European	164900 (7620–585000)	302 (47–624)
TP-ART	6 (85.7%)	1 (14.3%)	European	Und.	554 (415–720)

**Table 1.** Main clinical characteristics of the patient groups included in the analysis. \*Mean (interval) \*\*Und. = undetectable

2.04% and 0.55%, respectively<sup>4</sup>. The coexistence of EC and LTNP conditions observed in EC-LTNPs would represent the most beneficial host phenotype against HIV-1 infection, because of the capacity of these individuals to maintain elevated levels of CD4<sup>+</sup> T cells and undetectable VLs over time<sup>6</sup>. This fact turns EC-LTNP phenotype into an interesting but infrequent group of study<sup>7,8</sup>.

EC-LTNP phenotype is considered a multifactorial phenomenon governed by viral fitness<sup>9,10</sup> and host immuno-genetic mechanisms, such as CCR5 $\Delta$ 32 heterozygosity and the presence of HLA-B57/B27 and CCR2-V64I alleles<sup>11–13</sup>. Pereyra *et al.* have described that all the SNPs in the MHC associated with EC along with the genetic variants in CCR5 and CCR2 only explain 23% of the observed variance of durable host control, evidencing that these mechanisms are far away to fully explain the EC phenotype<sup>14</sup>. The existence of additional mechanisms has been studied even at transcriptome level employing microarray technologies<sup>15,16</sup>. However, hybridization-based methods have several limitations, such as hybridization specificity, background noise, hybridization to more than one gene product and a limited quantification range owing to signal saturation<sup>17</sup>. In contrast, RNA-Seq is a relatively recent application of high throughput sequencing technologies to transcriptome profiling<sup>18</sup>. Using this technology, the transcriptome of peripheral blood mononuclear cells (PBMCs) collected from EC-LTNPs has been characterized through the comparison with viremic LTNP (vLTNP) and HIV-positive individuals with a typical pattern of disease progression before (TP) and after receiving ART (TP-ART).

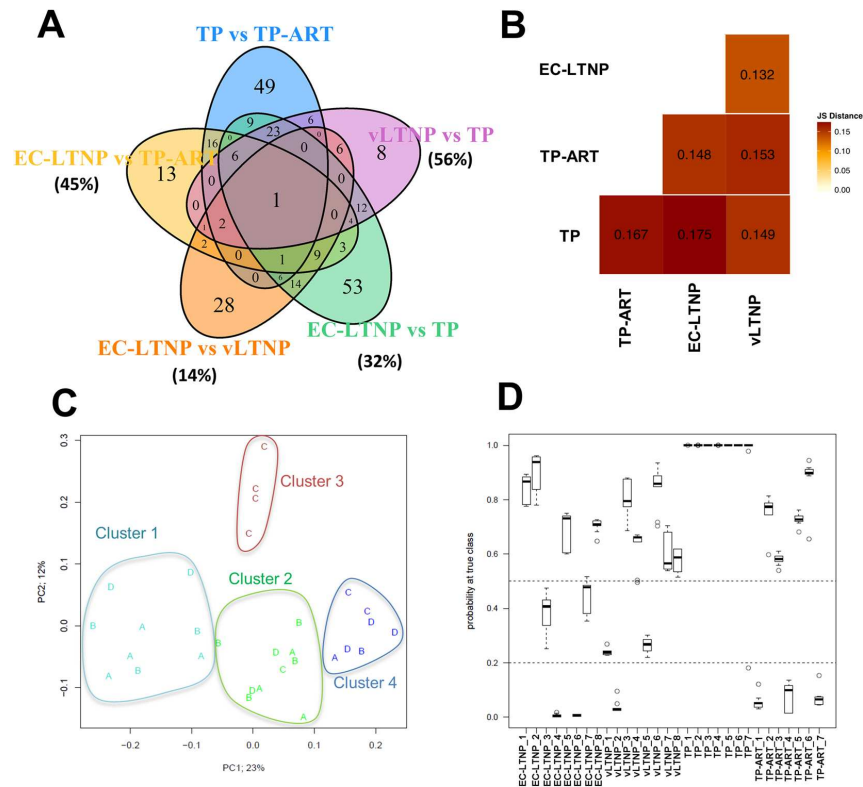
A global analysis of these transcriptomes was carried out in order to detect key transcripts to enable the classification of HIV-infected patients according to their phenotype and providing clues about the molecular mechanisms specifically associated with EC-LTNP phenotype through its comparison with vLTNP, TP and TP-ART. The understanding of viremia implications in remodeling the transcriptome machinery in patients on treatment was also investigated by the comparison of TP with TP-ART. The study of the genetic fingerprint exclusively found in EC-LTNPs would allow the molecular characterization of the most optimal immune activation against HIV-1 infection observed in nature and provides clues for the study of candidate markers for immunomodulatory drugs aiming at a functional HIV cure.

## Results

**Patients characteristics.** A total of 23 patients were included in the study, seven patients with a typical pattern of disease progression and 16 patients with a LTNP phenotype as defined in materials and methods section. Typical progressors provided two different samples for the analysis, before ART (TP) in which a CD4<sup>+</sup> T cell count depletion of 50–100 cells/mm<sup>3</sup> per year along with a detectable viral load (VL > 5,000 copies/ml) were observed; and two years after ART treatment in which VL was under the level of detection (<20 RNA copies/ml; TP-ART). LTNP were classified as EC because of their undetectable viremia or detectable viremia with VL < 2,000 copies/ml in less than 25% of all determinations during the follow-up (EC-LTNPs). All EC-LTNPs (n = 8) showed a VL < 2,000 copies/ml at sampling time and three of them had undetectable viremia. The other 8 LTNP showed detectable VLs < 10,000 copies/ml in more than 25% of all determinations during the follow-up and were considered as viremic LTNP (vLTNP). Patients' characteristics are summarized in Table 1. No differences in gender, age or origin were found between groups of individuals. All LTNP were followed for care for more than 10 years and all the individuals included in the study have a European ancestry and were diagnosed between 1988 and 1999.

**RNA-Seq quality control.** A total of 30 cDNA libraries coming from EC-LTNP, vLTNP, TP and TP-ART were analyzed. There were on average 33,689,139 single-end reads per library and a mean of 30,854,460 reads per library were aligned to the human genome (91.6%) (Supplementary Fig. S1). Approximately, half of mapped reads aligned to each strand of the genome. A median above 32 of the Phred quality score was observed across all bases at each position of the 100 bp reads (Supplementary Fig. S1). A quality score above 32 indicates that the base-calling error probability was lower than  $5.01 \times 10^{-4}$ . No statistically significant differences were identified between groups of HIV-positive individuals comparing the number of total and mapped reads.

**PBMC transcriptome profiling: global comparisons.** Different two by two comparisons were made between groups of patients. The number of differentially expressed genes (DEGs) were particularly high comparing EC-LTNPs with TP (n = 142) and TP with TP-ART (n = 119), suggesting the importance of an active viral replication in the modification of the transcriptome (Fig. 1A). According to the Jensen-Shannon distance based on the expression of these genes, the distance between EC-LTNPs and vLTNP was lower than any other comparison (Fig. 1B). TP-ART are closer to EC-LTNPs than to themselves before ART (TP), supporting the hypotheses about the significance of the viremia in altering the expression of several genes in HIV-positive individuals. The effect of the viremia was measured in all the comparisons between groups of patients as the percent of DEGs



**Figure 1.** Comparison between phenotypes. Venn diagram showing overlapped DEGs found in several comparisons analyzed in the study (TP vs TP-ART, EC-LTNP vs TP-ART, EC-LTNP vs vLTNP, EC-LTNP vs TP and vLTNP vs TP). Only *SLC37A3* were found in all the comparisons analyzed (A). Distance matrix showing similarities between phenotypes, calculated by the Jensen-Shannon divergence as implemented in the Bioconductor's package cummerbund (B). Multidimensional scaling (MDS) plot of the 30 samples based on the first two principal coordinates (PC, x and y axes). Labels A, B, C and D correspond to EC-LTNP, vLTNP, TP and TP-ART phenotypes, respectively. Color code is based on k-means clustering results with  $N = 4$ . The percentage of variability explained by each PC is indicated (C). Probabilities to be correctly classified for each individual employing the 20 best predictive genes. A total of ten independent predictions were carried out with LOOCV and the distribution of these probabilities are showed. The majority of the individuals ( $n = 22$ , 73.3%) were correctly classified and 20 of them obtained p-values  $> 0.5$  at true class after 10 repetitions (and therefore p-values  $< 0.5$  for the sum of the probabilities to be classified as any of the 3 other false classes). At the other extreme, some other individuals were repeatedly incorrectly classified with all the p-values  $< 0.2$  for the 10 models. This is the case for EC-LTNPs 4 and 6, vLTNP 2 and TP-ARTs 1, 4 and 7. EC-LTNP 4 and 6 were classified as vLTNPs for all the repetitions whereas the vLTNP 2 was classified as EC-LTNP also in all the iterations. In the case of the three TP-ARTs erroneously classified (1, 4 and 7), two of them were classified as EC-LTNPs and the other one as vLTNP. Between these two situations, 2 individuals (vLTNPs 1 and 5) were ambiguously classified with p-values at true class below 0.3 and with similar p-values to be classified as EC-LTNPs. The 10 models were able to classify correctly all the TPs with p-values close to 1 in all cases (except 1 out of the 10 models for TP 7 which was classified as vLTNP) (D).

observed in each condition which are coincident with the genes observed in TP versus TP-ART comparison. Thus, a 56%, 45% and 32% of DEGs found in vLTNP vs TP, EC-LTNP vs TP-ART and EC-LTNP vs TP comparisons, respectively, were also found in the TP versus TP-ART. However, only the 14% of DEGs found in EC-LTNP vs vLTNP were coincident with the genes found comparing TP and TP-ART (Fig. 1A).

**Selection of the most predictive genes of phenotypes.** Gene expression variability was analyzed across the 30 samples through multi-dimensional scaling (MDS) and subsequent unbiased K-means clustering into four groups. Although the clustering of groups of patients was not achieved, each phenotype was preferentially represented in one of the clusters. We found EC-LTNPs preferentially in cluster 1, vLTNPs in cluster 2, TP in

cluster 3 and TP-ART in cluster 4 (Fig. 1C). In total, 15 out of the 30 transcriptomes analyzed were correctly classified according to their corresponding phenotypes (50%). These results suggest a high level of heterogeneity within and between groups, complicating the finding of a common pattern of biomarkers associated to each phenotype.

New approximations such as the use of supervised classification techniques are necessary to describe new markers and mechanisms behind these phenotypes. Gene expression values from all the HIV-positive individuals included in the study were exposed to a gene selection process using a bias-corrected hierarchical Bayesian classification method. The expression values of a single panel of 20 genes were selected as the most accurate combination of genes to describe the phenotype variability found in the present study (Supplementary Fig. S2). The expression of most of these 20 genes is evidently different in TP compared with the other three phenotypes. However the expression of other genes such as *EIF3LP3* clearly distinguish phenotypes characterized by the control of viral replication (EC-LTNP + TP-ART) from phenotypes characterized by an active viral replication (vLTNP + TP) (Fig. 2). On his part, the expression of *XRCC6* clearly differentiates between phenotypes characterized by non-progression (EC-LTNP and vLTNP) and those characterized by a typical progression (TP and TP-ART) (Fig. 2).

Among the selected markers we found 13 genes, 6 pseudogenes with unknown function and a long intergenic non coding RNA (lincRNA) named RP11.288L9 as a negative regulator of IFI6. The functional annotation of these genes selected among the whole transcriptome identified several interferon-regulated genes (IRGs), including *HERC5*, *PARP12*, *IFI44*, *RNF213*, *MX1*, *HELZ2*, *EEF1G*, *EEF1B2*, *PARP14*, and *IFI6*. Several genes are related with RNA binding (*HERC5*, *EEF1G*, *EEF1B2*, *HELZ2*, *XRCC6*, and *PARP12*), response to virus (*IFI44*, *HERC5*, *EEF1G* and *MX1*), hydrolase activity (*HELZ2*, *XRCC6*, *RNF213* and *MX1*), eukaryotic translation elongation (*EEF1G*, *EEF1B2*, and *EEF1B2P3* as components of the eEF1 complex), NAD<sup>+</sup> ADP-ribosyltransferase activity (*PARP12* and *PARP14*), and ribosomal activity (including the ribosomal protein pseudogenes *RPL5P4*, *RPLAP5* and *RPLAP4*).

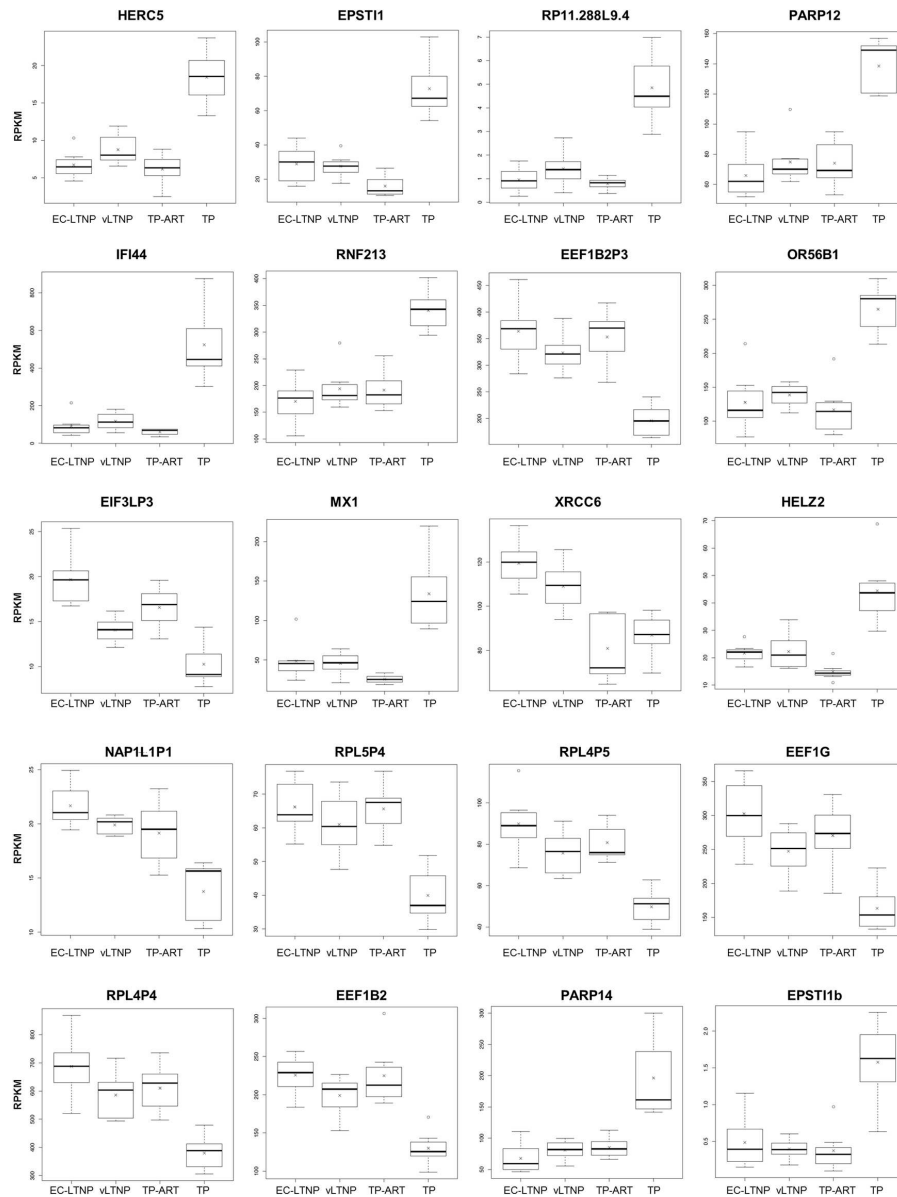
**Phenotype prediction.** A classification algorithm was created in order to evaluate the capacity of this panel of 20 genes to distinguish each phenotype. The distribution of probabilities to be correctly classified obtained in this algorithm for each individual was showed in Fig. 1D. The majority of the individuals ( $n = 22$ , 73.3%) were correctly classified (in contrast with the 50% obtained with the unbiased K-means clustering above described). Simplifying the model to only two phenotypes, LTNPs (regardless of their HIV-control capacity) and typical progressors (without considering if they are on ART or not), an accuracy of 90% ( $n = 27$ ) was achieved (compared to the 76% obtained with the clustering). In this model, only three TP-ART individuals (TP-ART-1, TP-ART-4 and TP-ART-7) were incorrectly classified as LTNPs (Fig. 1D), suggesting that some of these mechanisms associated with virus control are common between LTNPs and individuals on therapy. Interestingly, these three samples erroneously classified as LTNP (TP-ART-1, TP-ART-4 and TP-ART7) were the TP-ART samples with higher CD4<sup>+</sup> T cell counts (720, 650 and 578 cells per mm<sup>3</sup>, respectively compared with 525, 550, 439 and 415 cells per mm<sup>2</sup> found in the rest TP-ART samples). The Matthews correlation coefficient (MCC) is used in machine learning as a measure of the quality of binary classifications. MCC ranges from  $-1$  (total disagreement between prediction and observation) and  $+1$  (perfect prediction), including 0 (no better than random prediction). The MCC obtained for the classification algorithm developed in the present study was 0.81, compared with the 0.53 obtained for the K-means clustering. These results support the use of supervised data mining classification methods combined with transcript abundance estimation as a promising approximation to characterize the transcriptome profile of a heterogeneous phenotype.

**Deregulation of type I interferon signaling as a consequence of viremia.** The genes differentially expressed between any pair of phenotypes were identified using the negative binomial distribution. The expression of 119 genes was altered as a consequence of ART in TP individuals (Fig. 3). According to interferome there are evidences about the regulation of type I IFN in 92 out of these 119 genes (77.3%). The functional annotation of these 119 genes showed an enrichment of several molecular pathways related with interferon signaling and a defense response to virus (Supplementary Table S1).

Fourteen genes were included in Reactome's type I interferon signaling pathway ( $q = 7.45 \times 10^{-10}$ ). The expression of all these genes was downregulated after ART (Supplementary Fig. S3). The genes with higher differences between TP and TP-ART were the IRG *IFI27*, the antisense RNA *ARMCX3-AS1* and *ZNF275*. Five known anti-HIV IRGs were overexpressed in patients before ART including *EIF2AK2*, *ISG15*, *APOBEC3A*, *MX2* and *OAS1*, as well as other members of the OAS family, such as *OAS2*, *OAS3* and *OASL*. Five genes annotated as genes related to HIV-1 infection or resistance to AIDS were identified, including *CCL2*, *CX3CR1*, *TRIM22*, *SIGLEC1* and *TLR7*.

**EC-LTNP phenotype: activation of pathways leading to calcium release into the cytosol.** The expression of 58 genes was deregulated in EC-LTNP compared with TP-ART (Supplementary Table S2). This comparison was selected in order to avoid the viremia as a confounding factor, since in both phenotypes control of HIV-1 replication is achieved either by treatment or spontaneously. Gene Ontology analysis revealed an enrichment of several genes implicated in G-protein coupled peptide receptor activity and the positive regulation of leukocyte migration. All these receptors are overexpressed in EC-LTNP and are related with the stimulation of intracellular calcium ion mobilization (Fig. 4A).

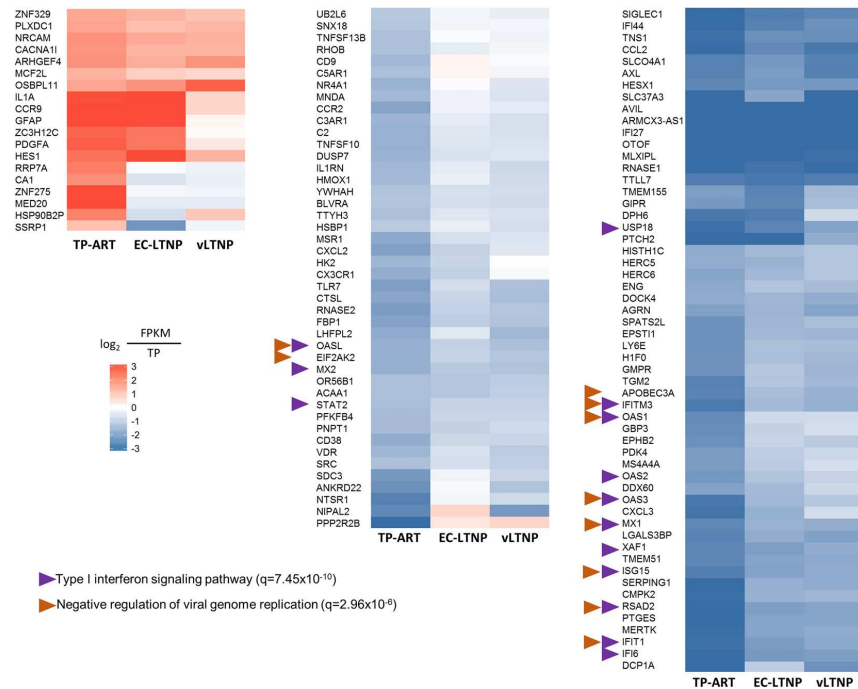
The promotor sequence of the 58 genes deregulated in EC-LTNP compared with TP-ART were analyzed to identify transcription factor binding sites (TFBS). The analysis of the TFBS showed that NFAT1 binding site was the most frequent within the promotor of these genes (50 NFAT1-binding sites) and was predicted within the promotor of 29 out of the 58 (50%) differentially expressed genes (Fig. 4B and Supplementary Table S3). Forty-three Elk-1 TFBS were also found in 29 out of the 58 genes (50%) (Fig. 4B). The overall genes with an NFAT1 or Elk-1 binding site in their promotor regions are 49 (84.5%). All these results suggest that a different regulation of the intracellular calcium signaling is observed in EC-LTNPs compared with TP-ART.



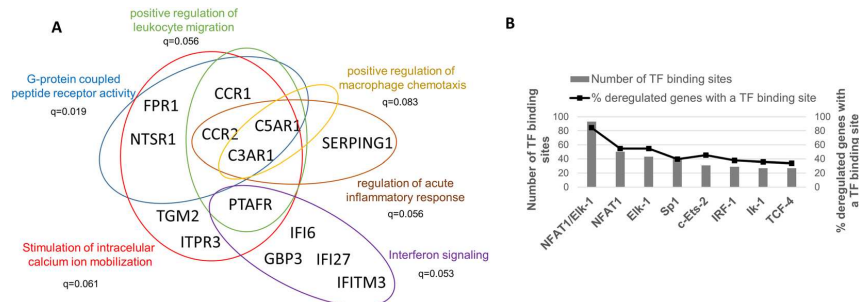
**Figure 2.** Best predictor genes of disease progression according to the hierarchical Bayesian classification model. The boxplots were generated in R and show the first and third quartile values for the RPKM distribution (upper and lower limits of the box), the median (the line splitting the box into two parts), the highest and lowest values (lines connected to the box through dashed lines), outlier values (open circles) and the mean value (crosses) for each phenotype.

**Molecular mechanisms involved in the control of HIV-1 replication in LTNP.** The expression of 70 genes was altered in EC-LTNPs in comparison with vLTNPs (Supplementary Table S4). The functional annotation of these genes showed no enrichment of any specific pathway or GO term. Some of these genes were key genes related to host mechanisms leading to modify HIV-1 replication and transcription, including the upregulation of *CDKN1A* (encoding the cyclin dependent kinase p21), TNF and the TNF-network associated gene *IER3*,





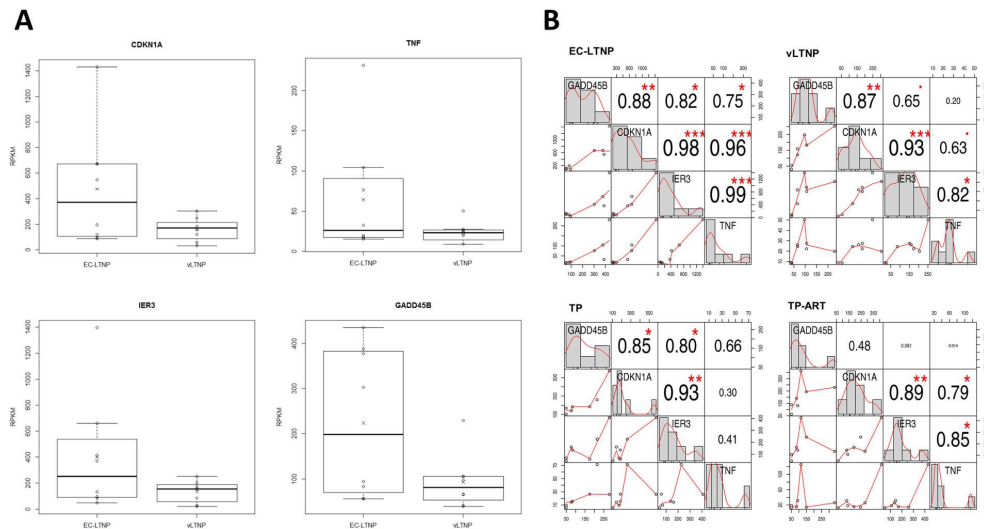
**Figure 3.** Deregulated genes in the TP/TP-ART comparison. Heatmap showing the comparison of the mean RPKM expression values for genes differentially expressed between TPs before and after receiving ART (the values for EC-LTNP and vLTNP were also showed just for the information). The RPKM expression values obtained for TPs, EC-LTNPs and vLTNPs are represented as a comparison with the values obtained for TPs.



**Figure 4.** EC-LTNP versus TP-ART comparison. Gene Ontology terms statistically significant in the comparison EC-LTNP versus TP-ART (FDR corrected p-values  $< 0.1$ ) (A). The promoter sequences of the genes differentially expressed between EC-LTNP and TP-ART were inspected to identify putative transcription factor binding sites using PROMO algorithm in ALGGEN server. The total number of transcriptional factor binding sites found and the percentage of these genes with a concrete binding site are showed (B).

and *GADD45B* (Fig. 5A). The expression of these four genes was highly correlated in EC-LTNP (statistically significant correlations), but was not in the other phenotypes (Fig. 5B).

According to the HIV-1 Human Interaction Database, there are evidences of 64 non-redundant interactions between these 4 proteins (CDKN1A, TNF, IER3 and *GADD45B*) and 13 HIV-1 proteins (Fig. 6). These results suggest that the expression of CDKN1A, TNF, IER3 and *GADD45B* are optimally coordinated in EC-LTNPs to regulate the expression of viral proteins.

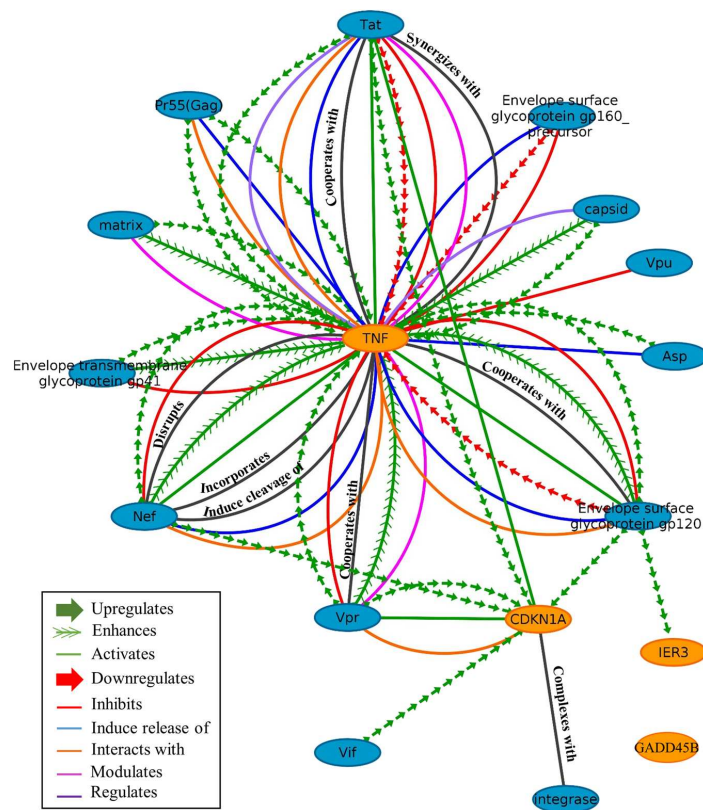


**Figure 5.** Expression of *GADD45B*, *CDKN1A*, *IER3* and *TNF* genes. RPKMs obtained for each gene in EC-LTNPs and vLTNPs (A). The correlation of these expression values between genes is shown for each group of patients. The distribution of each variable is shown on the diagonal, the bivariate scatter plots of RPKMs with a fitted line are displayed on the bottom of the diagonal and the value of the correlation plus the significance level as stars on the top of the diagonal according to Pearson parametric correlation test. This plot was generated using “PerformanceAnalytics” R package. Statistically significant p-values indicate a significant linear relationship between the expression values of two genes and are displayed as follows: \*\*\* $p < 0.001$ ; \*\* $p < 0.01$  and \* $p < 0.05$  (B).

**Genes associated with LTNP phenotype.** Common markers to all the LTNPs (EC-LTNP and vLTNP) were investigated. In order to minimize the effect of an active viral replication on the transcriptome profiles, two independent comparisons were simultaneously carried out, EC-LTNP versus TP-ART (as previously mentioned) and vLTNP versus TP, comparing two phenotypes with an active viral replication. The expression of 58 genes was dysregulated in EC-LTNPs compared with TP-ART, whereas 63 were identified in the vLTNP versus TP analysis (Fig. 7). A total of 14 genes were identified in both comparisons (EC-LTNP/TP-ART and vLTNP/TP; Fig. 7). Among these 14 genes, 9 were identified as IRGs and only *ANKRD54*, *IGHA2* and *VWA8* were not associated with the viremia in the comparison of TP versus TP-ART. A downregulation of the Von Willebrand Factor A Domain Containing 8 (*VWA8*) and the constant region of the heavy chain of IgA2 (*IGHA2*) was observed in LTNPs. Of note, only *ANKRD54* were found strongly upregulated in both EC-LTNP and vLTNP with respect typical progressors with fold changes over 25.

## Discussion

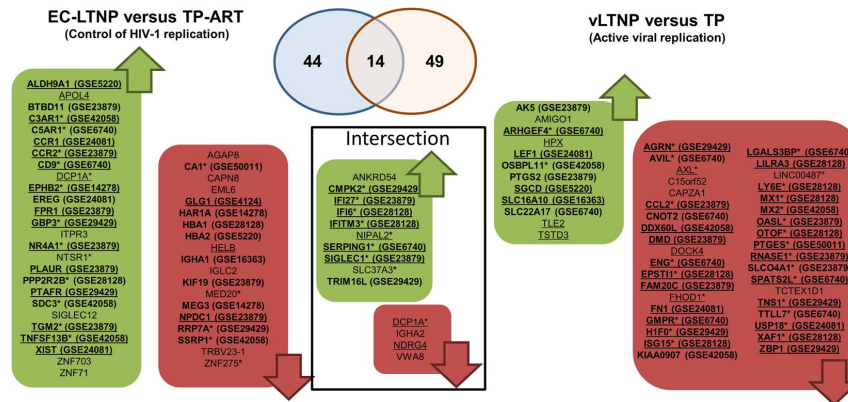
As expected from the multifactorial nature of HIV disease progression, the translation of the phenotype differences observed in EC-LTNP, vLTNP, TP and TP-ART to transcriptome differences does not seem to be obvious. A clear clustering of the individuals with different patterns of disease progression was not observed, evidencing the necessity to apply different approximations to define a group of biomarkers to better differentiate between groups of individuals. Machine learning techniques such as supervised classification are designed to analyze large amounts of data and infer a function from a training data set with several predictive variables (mRNA expression values) associated to a known output (phenotype). A mathematical model able to distinguish between LTNPs (regardless of their HIV-control capacity) and TPs (without considering if they are on ART or not) with an accuracy of 90% was obtained using hierarchical Bayesian classification algorithms combined with the selection of 20 genes as the best predictors of HIV disease progression. Several of these 20 genes were IRGs, pointing to the importance of the interferon regulation in HIV disease progression. As expected, the majority of these IRGs were downregulated in patients with low/undetectable levels of viremia. On the contrary, a particular upregulation of other IRGs implicated in reverse transcription and transcription of viral genes were especially observed in EC-LTNPs. First, an upregulation of the eukaryotic translation elongation factors of the eEF1 complex (*EEF1G*, *EEF1B2* and *EEF1B2P3*), considered critical HIV-1 reverse transcription cofactors<sup>19</sup>. Second, an upregulation of *XRCC6* which associates with Tat and TAR and repress the transcription of viral mRNAs<sup>20</sup>. Third, the downregulation of a lincRNA located within the promoter of *IFI6* gene and implicated in the negative regulation of *IFI6*<sup>21</sup>. This methodology represents an alternative way to associate differences in gene expression with a concrete phenotype, combining a solid transcript abundance estimation procedure (Tophat/Cufflinks) with machine learning approaches.



**Figure 6.** Protein-protein interaction network of CDKN1A, TNF, IER3 and GADD45B with viral proteins. The whole HIV-1 human interaction database was downloaded and all the interactions of CDKN1A, TNF, IER3 and GADD45B with viral proteins were mapped.

Deciphering the molecular mechanisms responsible for the control of HIV-1 replication for long periods of time observed in EC-LTNPs is absolutely crucial to mimic this spontaneous defense against the virus in HIV vaccinology and functional cure strategies. Different mechanisms related with reverse transcription and viral transcription have been specifically detected in EC-LTNP, which is consistent with the functional annotation obtained for the genes selected as the best predictors of disease progression. HIV-1 reverse transcription depends on the phosphorylation of viral reverse transcriptase by a host kinase named CDK2. Viral reverse transcriptase phosphorylation at a conserved threonine by CDK2 increases its efficacy and stability and enhances its viral fitness<sup>22</sup>. This mechanism is regulated by the cyclin-dependent kinase inhibitor 1A (CDKN1A)-mediated inhibition of CDK2 and has been previously described in individuals with EC phenotype<sup>22</sup>. In this study an upregulation of *CDKN1A* in EC-LTNP individuals compared with vLTNP was observed, suggesting an inactivation of CDK2 activity, avoiding the phosphorylation of viral reverse transcriptase and diminishing its efficacy and stability<sup>23</sup>. Moreover, a second mechanism related with CDKN1A has been associated with EC phenotype, describing a partial resistance of CD4<sup>+</sup> T cells from these individuals to HIV-1 infection mediated by a strong and selective upregulation of CDKN1A, also called p21<sup>24</sup>. This mechanism seems to regulate viral mRNA elongation by inactivating the enzymatic activity of CDK9, essential for the proper elongation of HIV-1 mRNA as a component of the P-TEFb (positive transcription elongation factor) complex<sup>25</sup>. This P-TEFb complex is formed by CDK9 and Cyclin T1 and is recruited after the activation of the viral LTR activity as a consequence of HIV-1 Tat protein binding to the trans-activation response (TAR) RNA structure. This mechanism is responsible for Tat-activated transcriptional elongation of viral transcripts<sup>26</sup>.

Aside from CDKN1A, other genes related with HIV-1 mRNA elongation were upregulated in EC-LTNP compared with vLTNP, including TNF, immediate early response 3 (IER3) and growth arrest and DNA damage 45 (GADD45B). GADD45B contribute to apoptosis and regulate HIV transcription<sup>27</sup>. GADD45B inhibits HIV-1 gene expression independent of CDKN1A, apparently without the need of the TAR, NF- $\kappa$ B, NRE and SP1 sites<sup>27</sup>. IER3 inhibits the most important family of Ser/Thr phosphatases, the protein phosphatase 2A (PP2A), which in



**Figure 7.** Genes associated with LTNP condition. The figure shows DEGs in the EC-LTNP versus TP-ART and vLTNP versus TP comparisons. Framed genes represent the intersection between both comparisons. Green and red boxes show upregulated and downregulated DEGs, respectively. Genes regulated by interferon are underlined whereas genes associated with an active viral replication (identified by the comparison of TP with TP-ART) are designated by an asterisk. DEGs found in other transcriptomic profiling studies of HIV-positive individuals with different degrees of disease progression were searched for a curated dataset collection<sup>34</sup>. The DEGs found in other datasets are in bold, indicating their Gene Expression Omnibus (GEO) accession numbers in parenthesis. Only datasets generated from blood cells and with a fold change of at least 2.0 were included in this figure. The included datasets and the phenotypes compared were: GSE14278 (HIV resistant vs HIV high-risk negative), GSE16363 (aviremic vs viremic), GSE23879 (elite controller vs HIV-negative), GSE24081 (controller vs progressor), GSE28128 (CD4 rapid progressors vs CD8 rapid progressors), GSE29429 (healthy vs HIV-positive), GSE4124 (HIV- vs HIV+ transmitter), GSE42058 (uninfected vs HIV infected), GSE50011 (CD4 count >500 vs CD4 count <500), GSE5220 (aviremic vs viremic), GSE6740 (CD4 uninfected vs CD4 non-progressor) and GSE6740 (CD8 non-progressor vs CD8 acute).

turn induces basal but not the Tat-activated HIV-1 transcription<sup>28</sup>. The functional implication of TNF in HIV-1 transcription is linked to its role in regulating IER3 gene expression<sup>29</sup>. One important finding in our work is the high correlation of the expression of CDKN1A, IER3, GADD45B and TNF found in EC-LTNPs compared with the other phenotypes. These results along with the analysis of the human protein interaction network with viral proteins suggest that these genes are coordinately and optimally regulated in these individuals to modulate basal and Tat-activated HIV-1 transcription.

We have also found evidences of an altered regulation of calcium-dependent signaling cascades in EC-LTNP compared with TP-ART, which is consistent with the capacity of these individuals to control viral transcription<sup>30</sup>. First, an enrichment in EC-LTNP of cell surface receptors involved in the stimulation of intracellular calcium mobilization directly (FPR1, CCR2) or by the activation of a phosphatidylinositol-calcium second messenger system (NTSR1). Second, a strong upregulation of inositol triphosphate (IP3) receptor isoform 3 (ITPR3) which mediates the mobilization of calcium ions into the cytosol in response to IP3. Third, an overrepresentation of the nuclear factor of activated T cells (NFAT1) and Elk-1 binding sites in the promoters of the genes differentially expressed. NFAT1 is present in the cytosol and are dephosphorylated by the Ca<sup>2+</sup>/calmodulin-dependent phosphatase calcineurin as a response to increased concentrations of intracellular Ca<sup>2+</sup>, causing a conformational change that results in its translocation from the cytoplasm to the nucleus and the activation of the transcription of NFAT1 target genes<sup>31</sup>. The stimulation of calcium-sensing receptors has also been associated with the activation of the transcription factor Elk-1 and the subsequent regulation of gene transcription<sup>32</sup>. Four, the increased expression in EC-LTNPs of several genes implicated in the positive regulation of the cytosolic calcium ion concentration including CXCL10, CCR9, ITPR3 and S1PR3, suggesting an additive effect on EC-LTNP over vLTNP.

Acute HIV-1 infection is characterized by a compartmentalized CD4<sup>+</sup> T cell depletion and constant viral replication, counteracted by a broad antiviral effect of the innate immune response. Type I IFNs play a leading role in this process through the activation of hundreds of IRGs. Little is known about the transcriptome changes experienced by HIV-1 infected individuals before and after receiving ART. The experimental design of the present study has allowed the identification of deregulated genes as a consequence of viremia, detecting altered mRNA levels of 119 genes, mostly regulated by interferon. The upregulation of these IRGs in patients with detectable HIV-1 replication and a typical rate of disease progression demonstrates the tight relationship between the pathogenesis of HIV infection and the chronic IFN stimulation<sup>33,34</sup>. These genes are mainly involved in the general immune response against viral infections such as OAS and MX genes<sup>35</sup>. Of note, some of these genes have been directly related with HIV pathogenesis, including *IFI27*, *IFITM3* and *TRIM22*<sup>36–38</sup>. The higher difference found comparing HIV-infected individuals before and after the treatment was the mRNA levels of the interferon alpha-inducible protein *IFI27*.

Comparing EC-LTNP vs TP-ART and vLTNP vs TP, several genes were found in both comparisons associated to both LTNP phenotypes –i.e. EC-LTNPs and vLTNPs, but only three DEGs were not affected by the viremia according to the TP versus TP-ART comparison. Thus, similar levels of these three genes (*ANKRD54*, *IGHA2* and *VWA8*) were found in all LTNPs (EC-LTNPs and vLTNPs). Neither were regulated by interferon. *ANKRD54* encodes an ankyrin repeat containing protein upregulated in LTNPs. Some proteins containing this type of ankyrin repeats interact with HIV-1 proteins such as Vpr<sup>39</sup>. Artificial ankyrins have been designed targeting the capsid domain of the HIV-1 Gag polyprotein, showing an antiviral effect at post-integration steps and inhibiting the virus assembly and egress pathway<sup>40</sup>. The role of this potential molecular marker of disease progression naturally overexpressed in LTNPs should be further investigated.

One limitation of the current study is that transcriptome was determined in total PBMCs and was not in specific lymphoid subpopulations. Nevertheless, these samples reflect different immune environments, such as possible differences in the relative proportions of cell types implicated in the immunopathogenesis of HIV-1 infection, i.e. resting CD4<sup>+</sup> T cells. This information is complementary and compatible with the study of specific subpopulations. However, the characterization of the transcriptome of a single cell can be achieved thanks to the advance in second and third generation sequencing technologies along with the relatively recent establishment of procedures to synthesize double-stranded cDNA from the extremely low quantities of mRNA present in a single cell, or even in a single nucleus<sup>41</sup>. Single cell RNA-Seq analyses will allow a deeper and definitive characterization of the immune response against HIV-1 infection observed in EC-LTNPs, identifying unequivocally which cell types are responsible for the expression of which host restriction factors. The present study exposes the changes in the transcriptome associated with different patterns of disease progression observed in HIV-positive individuals. Specifically, the analysis of EC-LTNPs as the most beneficial phenotype of immune activation against HIV-1 infection has allowed the identification of deregulated expression levels of several molecules in these patients. We propose that the coordinated upregulation of *CDKN1A*, *IER3*, *GADD45B* and *TNF* as well as the positive regulation of calcium-dependent signaling could be involved in the mechanisms leading to the slower progression to AIDS and HIV control concomitantly observed in EC-LTNPs.

## Material and Methods

**Study population.** Samples from patients were kindly provided by the HIV BioBank integrated in the Spanish Research Network (RIS). Samples were processed following current procedures and frozen at  $-80^{\circ}\text{C}$  immediately after their reception. All patients participating in the study gave their informed consent and protocols were approved by Institutional Ethical Committees (Instituto de Salud Carlos III. CEI PI 10\_2011v3). A total of 30 cDNA libraries were analyzed, including those coming from 16 HIV-positive individuals who have been classified as LTNPs within the Spanish LTNP-RIS cohort, 7 HIV-positive patients with a typical pattern of disease progression before ART (TPs) and the same 7 individuals after receiving ART (TPs-ART) from the Spanish CoRIS cohort. All the LTNP individuals maintain CD4<sup>+</sup> T cell counts over 500 cells per mm<sup>3</sup> and a VL under 10,000 copies per ml of blood in all the VL determinations during the first 10 years from infection/HIV<sup>+</sup> diagnosis. All experiments were performed in accordance with relevant guidelines and regulations.

**RNA extraction, mRNA library preparation and sequencing.** Total RNA from all the HIV-infected individuals was extracted from  $10^7$  peripheral blood mononuclear cells (PBMCs) with mRNAeasy Mini Kit (Qiagen) obtaining 5–20  $\mu\text{g}$  of total RNA. The quality of the RNA was measured in a 2100 Bioanalyzer (Agilent Technologies), obtaining a mean RNA integrity number (RIN) value of 9.2, with RIN values greater than 8 for all samples. The cDNA libraries from total RNA samples were prepared by an Illumina TruSeq RNA sample prep kit (Illumina, San Diego, CA) and were clustered onto a TruSeq single-end flow cell using a TruSeq SR Cluster Kit v3-cBot-HS (Illumina, San Diego, CA), after quantification by PicoGreen dsDNA assay kit (Life Technologies) and pooling in equimolar mixtures. Finally, the DNA sequence of each cluster on flow cells was determined employing 100 cycles of Sequencing-By-Synthesis (SBS) technology (TruSeq SBS Kit v3-HS kit) on an Illumina's HiSeq2000 Sequencing System.

**Analysis of sequencing data.** A first quality assessment was performed with FastQC and reads were trimmed with the java tool Trimmomatic with default parameters in order to remove sequences of primers and adapters employed during library preparation from the ends of sequences. The adapters and Illumina-specific primers from the reads were removed, allowing two mismatches, requiring a minimum of 30 matches in palindromic mode and a minimum of ten matches for nonpalindromic mode between the read sequence and the adapters/primers. A sliding window of 4 nucleotides were analyzed for each read and were trimmed once the average Phred quality falls below 30. The end bases below a quality score of 3 were cut. Sequences trimmed to shorter than 60 bases were removed.

Filtered reads were aligned to the human genome assembly GRCh37 using Bowtie2/TopHat2<sup>42,43</sup> and the transcript assembly was reconstructed with Cufflinks<sup>44,45</sup>. Differential expression analysis was carried out with Cuffdiff package based on the negative binomial distribution. Cuffdiff transforms these gene counts to units of reads per kilobase of transcript length per million mapped reads (RPKM) and infers probabilities before (p-values) and after FDR correction (q-values) to identify differentially expressed genes among groups of individuals (q-values < 0.05). The similarities between phenotypes were analyzed through the calculation of Jensen-Shannon distance with the Bioconductor package *cummeRbund*. The variability found between transcriptome profiles was explored in R software through K means clustering and dimensionality reduction by MDS based on the Pearson correlation using “stats” and “gplots” packages.

The genes differentially expressed between EC-LTNPs, vLTNPs, TPs and TP-ARTs were identified for each possible dual comparison between any pair of the above mentioned phenotypes (6 comparisons in total). This



type of analysis identified genes differentially expressed in individuals exhibiting an active HIV-1 replication (vLTNPs and TPs) compared with individuals with a controlled viremia (EC-LTNPs and TP-ARTs). This approximation also analyzes the distinguishing factors between EC-LTNPs and vLTNPs and the identification of the main alterations at the transcriptome level as a consequence of ART administration through the comparison of the same individuals before and after receiving anti-HIV therapy.

**Predictive genes of disease progression by data mining techniques.** The predictive model of disease progression was created combining the transcript abundance estimation obtained from Cuffdiff with a bias-corrected feature selection procedure and a hierarchical Bayesian classification<sup>46,47</sup>. The RPKMs obtained for every single gene and for all the individuals were subjected to a wrapper feature selection process and a subsequent supervised classification by a hierarchical Bayesian classification. During this process, the more informative combination of transcripts to define the progression phenotype were selected. Genes with expression values with near zero variance were eliminated prior modeling with the default parameters of nearZeroVar function included in “caret” Bioconductor package. The feature selection process of the remaining genes included the same algorithm used for classification to evaluate the importance of each gene through 10 repetitions of leave one out cross-validation (LOOCV), setting up random seeds for each repetition<sup>46</sup>. Two criteria were employed to evaluate the optimal selection of predictive genes, including ER and AMLP (which evaluate the classification models more accurately by taking into account low predictive probabilities at the true class labels<sup>47</sup>). The smaller subset of genes reaching the lower ER and AMLP in this process was selected for the final classification model. The probability for each individual to be classified as EC-LTNP, vLTNP, TP or TP-ART were calculated by the classification model using the estimated expression values of selected transcripts. The RPKMs values obtained for each individual were obtained from the output of Cuffdiff tool (genes.read\_group\_tracking file) and the Bioconductor packages “caret” and “BCBCSF” were employed for pre-processing, feature selection and classification processes<sup>47</sup>.

**Functional annotation.** The identification of enriched biological pathways, diseases or gene ontology terms associated with differentially expressed genes between groups of patients was carried out with KOBAS 2.0 tool<sup>48</sup>. This software integrates searches against the main biological databases, including Gene Ontology (GO), KEGG, PID, Reactome, PANTHER, GO, IMIM, FunDO, GAD and NHGRI GWAS Catalog. Interferome v2.01 was used to find evidences about interferon regulation of genes deregulated in any of the comparisons carried out in the present study<sup>49</sup>.

**Identification of putative transcription factor binding sites.** A 1500 base pairs sequence upstream of the start codon of the genes was retrieved from GRCh38 using Ensembl<sup>50</sup>. These sequences were inspected to identify putative transcription factor binding sites included in TRANSFAC database using PROMO algorithm in ALGGEN server<sup>51</sup>. The reliability of this methodology was calculated through the expectation of finding each binding site in a random sequence of 1000 nucleotides, considering a model with exactly the same nucleotide frequency as the query sequence (E-value). A very conservative cutoff was used to predict the transcription factor binding sites (TFBS) and only TFBS with an E-value < 0.05 and with a similarity to the matrix >95% were considered as true binding sites.

**Interactions of human proteins with HIV-1 proteins.** We downloaded the whole database of HIV-1 and human protein interactions<sup>52</sup>. The HIV-human protein interaction network was visualized in Cytoscape\_v3.1.1<sup>53</sup>. We mapped all the interactions of CDKN1A, IER3, GADD45B and TNF with viral proteins, distinguishing the type of interaction (mainly upregulation, activation, enhancement, downregulation and inhibition).

### Data Availability

The access to the raw reads for use by the scientific community can be done upon request to the authors and after approval of every single request by the Data Protection Officer of the Instituto de Salud Carlos III.

### References

- Deeks, S. G. & Walker, B. D. Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity*. **27**, 406–416 (2007).
- Poropatich, K. & Sullivan, D. J. Jr. Human immunodeficiency virus type 1 long-term non-progressors: the viral, genetic and immunological basis for disease non-progression. *J Gen Virol*. **92**, 247–268 (2010).
- Collaborative Group on AIDS Incubation and HIV Survival. Time from HIV-1 seroconversion to AIDS and death before the widespread use of highly active anti-retroviral therapy: a collaborative re-analysis. *Lancet*. **355**, 1131–1137 (2000).
- Okulicz, J. F. *et al.* Infectious Disease Clinical Research Program (IDCRP) HIV Working Group. Clinical outcomes of elite controllers, viremic controllers, and long-term nonprogressors in the US Department of Defense HIV Natural History Study. *J Infect Dis*. **200**, 1714–1723 (2009).
- Saksena, N. K., Rodes, B., Wang, B. & Soriano, V. Elite HIV controllers: myth or reality? *AIDS Rev*. **9**, 195–207 (2007).
- Maded, Y., Boufassa, F., Porter, K. & Meyer, L. Spontaneous control of viral load and CD4 cell count progression among HIV-1 seroconverters. *AIDS*. **19**, 2001–2007 (2005).
- Hunt, P. W. *et al.* Relationship between T cell activation and CD4+ T cell count in HIV-seropositive individuals with undetectable plasma HIV RNA levels in the absence of therapy. *J Infect Dis*. **197**, 126–133 (2008).
- Hunt, P. W. Natural control of HIV-1 replication and long-term nonprogression: overlapping but distinct phenotypes. *J Infect Dis*. **200**, 1636–1638 (2009).
- Mologni, D. *et al.* Vpr and HIV-1 disease progression: R77Q mutation is associated with long-term control of HIV-1 infection in different groups of patients. *AIDS*. **20**, 567–574 (2006).
- Miura, T. *et al.* Impaired replication capacity of acute/early viruses in persons who become HIV controllers. *J Virol*. **84**, 7581–7591 (2010).
- Gao, X. *et al.* AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. *Nat Med*. **11**, 1290–1292 (2005).

12. Garred, P. *et al.* Dual effect of CCR5 delta 32 gene deletion in HIV-1-infected patients. Copenhagen AIDS Study Group. *Lancet*. **349**, 1884 (1997).
13. Kostrikis, L. G. *et al.* A chemokine receptor CCR2 allele delays HIV-1 disease progression and is associated with a CCR5 promoter mutation. *Nat Med*. **4**, 350–353 (1998).
14. Pereyra, F. *et al.* The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science*. **330**, 1551–1557 (2010).
15. Giri, M. S., Nebozhyn, M., Showe, L. & Montaner, L. J. Microarray data on gene modulation by HIV-1 in immune cells: 2000–2006. *J Leukoc Biol*. **80**, 1031–1043 (2006).
16. Rotger, M. *et al.* Genome-wide mRNA expression correlates of viral control in CD4+ T-cells from HIV-1-infected individuals. *PLoS Pathog*. **6**, e1000781 (2010).
17. Okoniewski, M. J. & Miller, C. J. Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics*. **7**, 276 (2006).
18. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. **10**, 57–63 (2009).
19. Warren, K. *et al.* Eukaryotic elongation factor 1 complex subunits are critical HIV-1 reverse transcription cofactors. *PNAS* **109**, 9587–9592 (2012).
20. Jeanson, L. & Mouscadet, J. F. Ku represses the HIV-1 transcription. Identification of a putative Ku binding site homologous to the mouse mammary tumor virus NRE1 sequence in the HIV-1 long terminal repeat. *J Biol Chem*. **277**, 4918–4924 (2002).
21. Liu, X. *et al.* A novel lncRNA regulates HCV infection through IFI6. *Hepatology*. **10**, 1004–1019 (2018).
22. Leng, J. *et al.* A cell-intrinsic inhibitor of HIV-1 reverse transcription in CD4(+) T cells from elite controllers. *Cell Host Microbe*. **15**, 717–728 (2014).
23. Levkau, B. *et al.* Cleavage of p21Cip1/Waf1 and p27Kip1 mediates apoptosis in endothelial cells through activation of Cdk2: role of a caspase cascade. *Mol Cell*. **1**, 553–563 (1998).
24. Chen, H. *et al.* CD4+ T cells from elite controllers resist HIV-1 infection by selective upregulation of p21. *J Clin Invest*. **121**, 1549–1560 (2011).
25. Flores, O. *et al.* Host-cell positive transcription elongation factor b kinase activity is essential and limiting for HIV type 1 replication. *Proc Natl Acad Sci USA* **96**, 7208–7213 (1999).
26. Kim, Y. K., Bourgeois, C. F., Isel, C., Churcher, M. J. & Karn, J. Phosphorylation of the RNA polymerase II carboxyl-terminal domain by CDK9 is directly responsible for human immunodeficiency virus type 1 Tat-activated transcriptional elongation. *Mol Cell Biol*. **22**, 4622–4637 (2002).
27. Liang, Z. *et al.* GADD45 proteins inhibit HIV-1 replication through specific suppression of HIV-1 transcription. *Virology*. **493**, 1–11 (2016).
28. Faulkner, N. E., Lane, B. R., Bock, P. J. & Markovitz, D. M. Protein phosphatase 2A enhances activation of human immunodeficiency virus type 1 by phorbol myristate acetate. *J Virol*. **77**, 2276–81 (2003).
29. Wu, M. X., Ao, Z., Prasad, K. V. S., Wu, R. & Schlossman, S. F. IEX-1L, an apoptosis inhibitor involved in NF-kappa-B-mediated cell survival. *Science* **281**, 998–1001 (1998).
30. Papp, B. & Byrn, R. A. Stimulation of HIV expression by intracellular calcium pump inhibition. *J Biol Chem*. **270**, 10278–10283 (1995).
31. Okamura, H. *et al.* Concerted dephosphorylation of the transcription factor NFAT1 induces a conformational switch that regulates transcriptional activity. *Mol Cell*. **6**, 539–550 (2000).
32. Thiel, G., Lesch, A. & Keim, A. Transcriptional response to calcium-sensing receptor stimulation. *Endocrinology* **153**, 4716–4728 (2012).
33. Doyle, T., Goujon, C. & Malim, M. H. HIV-1 and interferons: who's interfering with whom? *Nat Rev Microbiol*. **13**, 403–413 (2015).
34. Hardy, G. A. *et al.* Desensitization to type I interferon in HIV-1 infection correlates with markers of immune activation and disease progression. *Blood*. **113**, 5497–5505 (2009).
35. Fagone, P. *et al.* Induction of OAS gene family in HIV monocyte infected patients with high and low viral load. *Antiviral Res*. **131**, 66–73 (2016).
36. Patro, S. C. *et al.* Shift in monocyte apoptosis with increasing viral load and change in apoptosis-related ISG/Bcl2 family gene expression in chronically HIV-1-infected subjects. *J Virol*. **89**, 799–810 (2015).
37. Winkler, M. *et al.* Rhesus macaque IFITM3 gene polymorphisms and SIV infection. *PLoS One* **12**, e0172847 (2017).
38. Singh, R. *et al.* Association of TRIM22 with the type 1 interferon response and viral control during primary HIV-1 infection. *J Virol*. **85**, 208–216 (2011).
39. Miles, M. C. *et al.* Molecular and functional characterization of a novel splice variant of ANKHD1 that lacks the KH domain and its role in cell survival and apoptosis. *FEBS J*. **272**, 4091–4102 (2005).
40. Nangola, S. *et al.* Antiviral activity of recombinant ankryrin targeted to the capsid domain of HIV-1 Gag polyprotein. *Retrovirology* **9**, 17 (2012).
41. Krishnaswami, S. R. *et al.* Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc*. **11**, 499–524 (2016).
42. Langmead, B. & Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. **9**, 357–359 (2012).
43. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. **28**, 511–515 (2010).
44. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. **14**, R36 (2013).
45. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. **7**, 562–578 (2012).
46. Diez-Fuertes *et al.* Improvement of HIV-1 coreceptor tropism prediction by employing selected nucleotide positions of env gene in a Bayesian network classifier. *J Antimicrob Chemother*. **68**, 1471–1485 (2013).
47. Longhai, L. Bias-corrected hierarchical bayesian classification with a selected subset of high-dimensional features. *J Am Stat Assoc*. **107**, 120–134 (2012).
48. Xie, C. *et al.* KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res*. **39**, W316–322 (2011).
49. Rusinova, I. *et al.* Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res*. **41**, D1040–1046 (2013).
50. Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Res*. **44**, D710–716 (2016).
51. Farré, D. *et al.* Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res*. **31**, 3651–3653 (2003).
52. Fu, W. *et al.* Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res*. **37**, D417–422 (2009).
53. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**, 2498–2504 (2003).
54. Blazkova, J. *et al.* A curated transcriptome dataset collection to investigate the immunobiology of HIV infection. *F1000Res*. **5**, 327 (2016).

### Acknowledgements

We want to particularly acknowledge the patients in this study for their participation and to the HIV BioBank integrated in the Spanish AIDS Research Network and collaborating Centres (<http://hivgmbiobank.com/donor-area/hospitals-and-centres-transferring-samples/?lang=en>) for the generous gifts of clinical samples used in this work. The HIV BioBank and the AIDS Immunopathogenesis Unit are integrated in the Spanish AIDS Research Network. The HIV BioBank is partially funded by the RD16/0025/0019 project as part of the Plan Nacional R + D + I and cofinanced by ISCIII- Subdirección General de Evaluación and el Fondo Europeo de Desarrollo Regional (FEDER)". This work was partially supported by Instituto de Salud Carlos III and co-funded by European Regional Development Fund (ERDF) "A way to build Europe" (projects RD12/0017/0015 and RD16CIII/0002/0001, Plan Estatal de I + D + I 2013–2016), the French Government's Investissement d'Avenir program, Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (n°ANR-10-LABX-62-IBED) and the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 681137. FDF was supported by the Spanish Government's Sara Borrell postdoctoral Program (CD12/00515) and HETT by the ISCIII-PFIS predoctoral program (FI14CIII/00014).

### Author Contributions

J.A. and F.D.F. conceived the study. F.D.F., H.E.T.T., E.C. and M.P. performed the experiments. F.D.F. implemented the bioinformatics and biostatistics analyses and wrote the original draft. F.D.F., H.E.T.T., M.M.A.S. and J.G.P. performed the data curation and visualization. J.A., J.G.P., L.C. and A.S. revised and edited the manuscript. J.A. acquired funding and supervised the project.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-50642-x>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

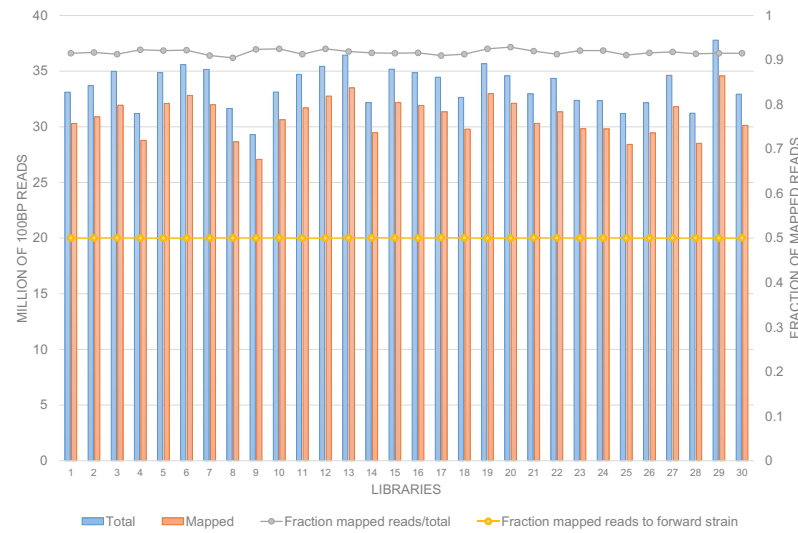


**TRANSCRIPTOME SEQUENCING OF PERIPHERAL BLOOD  
MONONUCLEAR CELLS FROM ELITE CONTROLLER-LONG TERM NON  
PROGRESSORS**

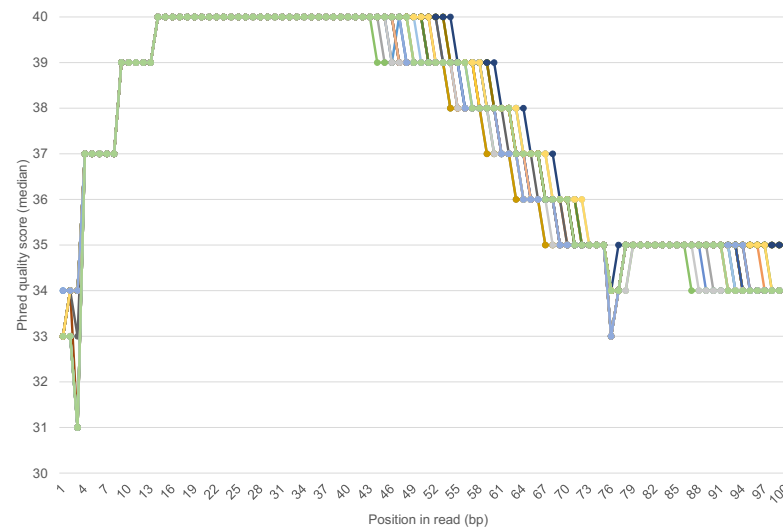
Francisco Díez-Fuertes, Humberto Erick De La Torre-Tarazona, Esther Calonge, Maria Pernas, María del Mar Alonso-Socas, Laura Capa, Javier García-Pérez, Anavaj Sakuntabhai, José Alcamí

Supplementary Figures **A**

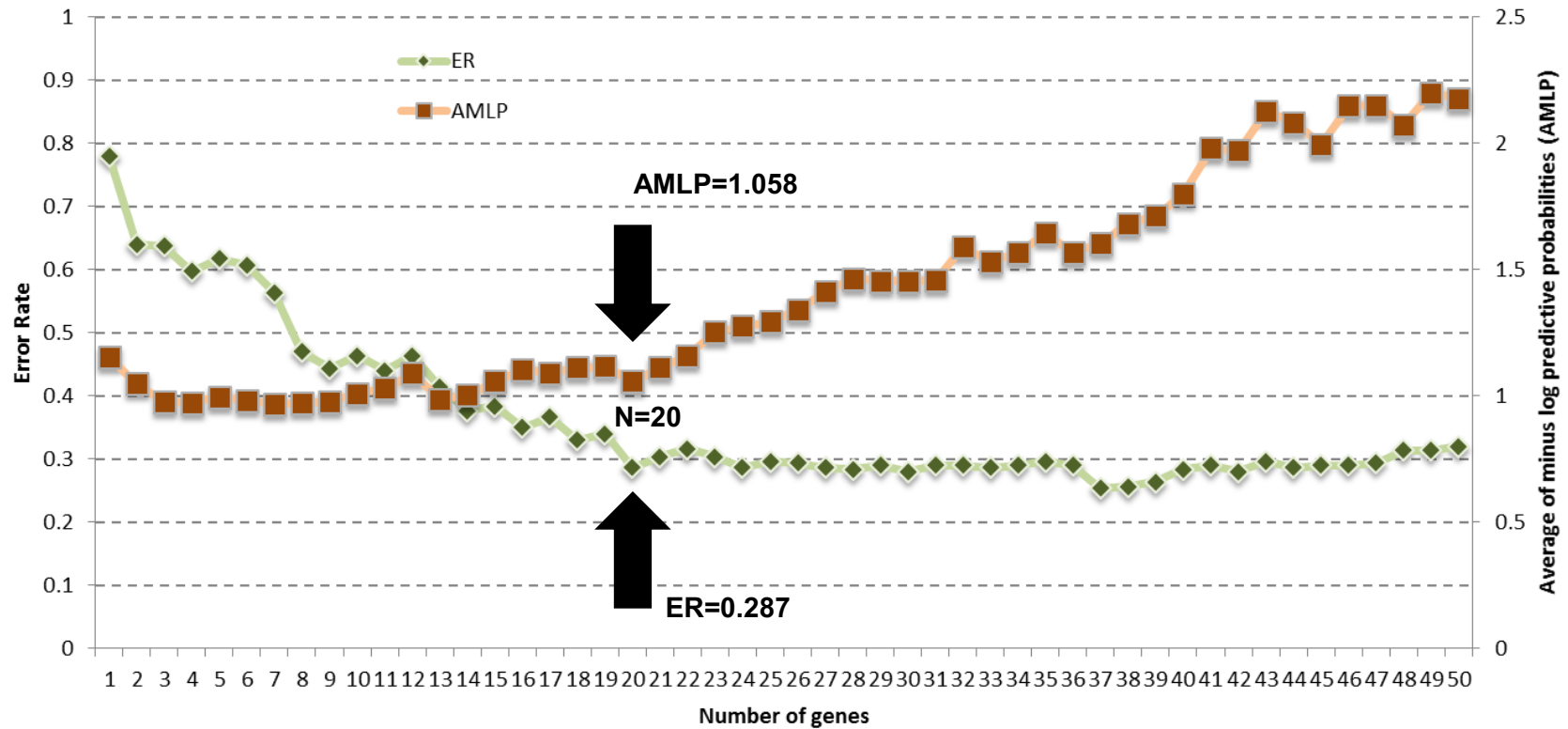
**Resultados**



**B**



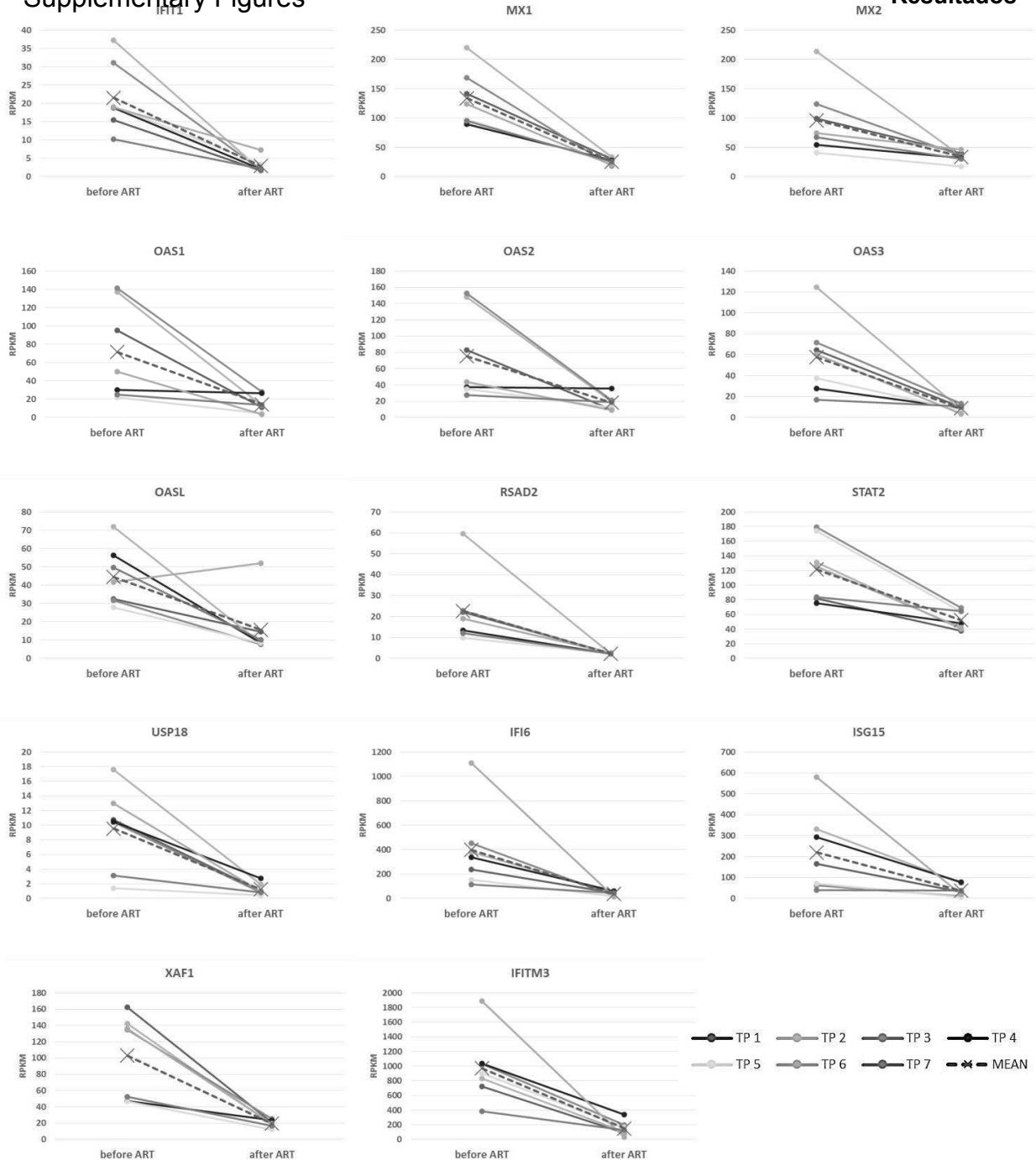
**Supplementary Figure S1. Reads per library obtained by RNA-seq and quality control.** Number of total reads obtained for each library and reads mapped to human genome, including the fraction of mapped/total reads and the fraction of mapped reads to forward strain (A). Median of the Phred quality score for the 30 libraries included in the study (B).



**Supplementary Figure S2. Identification of the best predictor genes of patient phenotype.** Error rate (ER) and average minus log predictive probabilities (AMLP) obtained in the feature selection process, evaluating the accuracy of 50 models employing the 1-50 best predictive genes. The model with 20 predictive genes was selected as the most accurate model.

## Supplementary Figures

## Resultados



**Supplementary Figure S3. Restoration of IRGs expression as a consequence of ART in typical progressors.** RPKM values obtained for the 20 IRGs differentially expressed in TPs and TPs and for each patient before and after receiving ART showing evidences about type I IFN regulation in humans.

**Supplementary Table S1. Differential expression and functional annotation of the deregulated genes identified in the TP/TP-ART comparison.**

Statistically significant pathways from Reactome and Gene Ontology were identified using KOBAS 2.0 tool [48]. Fisher's exact test (p-values) and Benjamini and Hochberg false discovery rate (corrected p-values) was employed to compare with human transcriptome. The genes associated with each pathway or term was included using Ensembl codes.

Gene	Locus	Group 1	Group2	FPKM Group 1	FPKM Group 2	log2 FC	test stat	p-value	q-value	Significant
ZNF275	X:152599612-152631799	TP	TP-ART	4,25544	142,653	5,06705	6,51531	5,00E-05	0,0121911	yes
CCR9	3:45864807-46037316	TP	TP-ART	0,514931	16,3372	4,98764	0,914852	5,00E-05	0,0121911	yes
MED20	6:41873008-41888877	TP	TP-ART	2,58391	24,5069	3,24556	4,18014	5,00E-05	0,0121911	yes
GFAP	17:42923720-42994305	TP	TP-ART	0,0754897	0,625094	3,04972	0,140327	5,00E-05	0,0121911	yes
IL1A	2:113531491-113542167	TP	TP-ART	0,673436	5,16041	2,93787	1,83364	0,00015	0,0301119	yes
PDGFA	7:536673-559933	TP	TP-ART	0,285716	1,84028	2,68727	1,38344	5,00E-05	0,0121911	yes
ZC3H12C	11:109959155-110042566	TP	TP-ART	0,509457	2,99597	2,55599	2,04269	0,00015	0,0301119	yes
HES1	3:193853038-193856579	TP	TP-ART	8,08891	41,259	2,35069	2,13076	5,00E-05	0,0121911	yes
RRP7A	22:42896584-42978044	TP	TP-ART	18,4949	82,1622	2,15135	2,15648	5,00E-05	0,0121911	yes
CA1	8:86239836-86393722	TP	TP-ART	0,141524	0,51085	1,85185	0,24374	5,00E-05	0,0121911	yes
NRCAM	7:107788067-108097161	TP	TP-ART	0,180839	0,623256	1,78512	0,767027	5,00E-05	0,0121911	yes
ARHGEF4	2:131588550-131804836	TP	TP-ART	0,151461	0,520021	1,77962	0,636149	5,00E-05	0,0121911	yes
CACNA1I	22:39966757-40085742	TP	TP-ART	0,930795	3,11828	1,74421	1,51921	0,0001	0,0217134	yes
PLXDC1	17:37213271-37310647	TP	TP-ART	2,8249	7,84085	1,47281	0,709435	0,00015	0,0301119	yes
OSBPL11	3:125244310-125314384	TP	TP-ART	12,613	34,2196	1,43991	2,59719	5,00E-05	0,0121911	yes
ZNF329	19:58637618-58666477	TP	TP-ART	1,52244	4,10288	1,43025	1,55408	0,0001	0,0217134	yes
MCF2L	13:113548691-113754053	TP	TP-ART	0,625438	1,48051	1,24316	0,646629	5,00E-05	0,0121911	yes
SSRP1	11:57093076-57103351	TP	TP-ART	96,5262	189,504	0,973239	6,95632	0,00025	0,0446698	yes
HSBP1	16:83841447-83853342	TP	TP-ART	65,7926	33,4082	-0,97772	-1,25524	0,0002	0,0380602	yes
UBE2L6	11:57318946-57335757	TP	TP-ART	135,927	66,9902	-1,02081	-1,39032	0,0001	0,0217134	yes
SNX18	5:53813588-53842415	TP	TP-ART	22,753	11,0062	-1,04774	-1,40582	0,00025	0,0446698	yes
BLVRA	7:43798278-43846943	TP	TP-ART	113,706	53,9052	-1,07681	-1,04725	0,00025	0,0446698	yes
YWHAH	22:32329506-32353693	TP	TP-ART	44,8213	20,3489	-1,13924	-1,37207	0,0001	0,0217134	yes
RHOB	2:20646834-20649200	TP	TP-ART	73,1635	32,6889	-1,16232	-1,61718	5,00E-05	0,0121911	yes
SRC	20:35973087-36034535	TP	TP-ART	59,9009	26,5698	-1,17279	-0,782	5,00E-05	0,0121911	yes

Supplementary Table S1. Differential expression and functional annotation of the deregulated genes identified in the TP/TP-ART comparison (cont.)

Gene	Locus	Group 1	Group2	FPKM Group 1	FPKM Group 2	log2 FC	test stat	p-value	q-value	Significant
TRIM22	11:5684424-5959849	TP	TP-ART	264,814	116,653	-1,18275	-1,28581	0,0001	0,0217134	yes
TNFSF13B	13:108903587-108960933	TP	TP-ART	61,3095	26,9272	-1,18705	-1,15392	5,00E-05	0,0121911	yes
C5AR1	19:47793279-47887533	TP	TP-ART	158,909	69,6735	-1,18952	-1,54027	5,00E-05	0,0121911	yes
ACAA1	3:38080695-38178733	TP	TP-ART	143,675	62,9768	-1,18992	-1,81053	0,00025	0,0446698	yes
TTYH3	7:2671584-2704491	TP	TP-ART	77,1787	33,8218	-1,19025	-1,10391	5,00E-05	0,0121911	yes
STAT2	12:56734696-56753939	TP	TP-ART	121,429	52,1285	-1,21997	-1,16445	5,00E-05	0,0121911	yes
PFKFB4	3:48554932-48601455	TP	TP-ART	24,5876	10,2783	-1,25833	-1,10596	5,00E-05	0,0121911	yes
PNPT1	2:55860871-55921045	TP	TP-ART	11,5752	4,79314	-1,272	-1,44165	0,0002	0,0380602	yes
HMOX1	22:35776353-35790207	TP	TP-ART	183,538	74,6892	-1,2971	-1,45121	0,00015	0,0301119	yes
MNDA	1:158791875-158819296	TP	TP-ART	135,851	54,198	-1,32572	-1,34679	0,00025	0,0446698	yes
IL1RN	2:113864790-113892436	TP	TP-ART	47,7947	18,9684	-1,33325	-1,5858	5,00E-05	0,0121911	yes
CD9	12:6308880-6347425	TP	TP-ART	18,2358	7,21529	-1,33764	-1,37468	5,00E-05	0,0121911	yes
VDR	12:48234207-48336831	TP	TP-ART	10,02	3,83834	-1,38432	-1,64626	0,0001	0,0217134	yes
HK2	2:75059781-75123332	TP	TP-ART	46,4385	17,666	-1,39435	-0,99475	5,00E-05	0,0121911	yes
C3AR1	12:8210897-8219067	TP	TP-ART	27,3192	10,3438	-1,40115	-1,52833	5,00E-05	0,0121911	yes
DUSP7	3:52082934-52090566	TP	TP-ART	15,056	5,60739	-1,42494	-1,72128	5,00E-05	0,0121911	yes
NR4A1	12:52416615-52453566	TP	TP-ART	96,9124	35,6112	-1,44435	-1,14579	0,00025	0,0446698	yes
TNFSF10	3:172223297-172241297	TP	TP-ART	61,1058	21,9087	-1,47981	-1,54732	5,00E-05	0,0121911	yes
C2	6:31865561-31919861	TP	TP-ART	21,0624	7,54194	-1,48166	-1,32241	5,00E-05	0,0121911	yes
OASL	12:121455493-121477045	TP	TP-ART	44,4657	15,7071	-1,50128	-1,62621	5,00E-05	0,0121911	yes
MX2	21:42733869-42782127	TP	TP-ART	96,1882	33,8948	-1,5048	-1,58348	5,00E-05	0,0121911	yes
EIF2AK2	2:37311593-37384208	TP	TP-ART	31,2239	10,7815	-1,53409	-0,68834	5,00E-05	0,0121911	yes
CX3CR1	3:39274006-39323226	TP	TP-ART	20,0874	6,88813	-1,54411	-1,42555	5,00E-05	0,0121911	yes
CD38	4:15779883-15854866	TP	TP-ART	10,7753	3,66091	-1,55745	-1,40952	5,00E-05	0,0121911	yes
HERC5	4:89378267-89427376	TP	TP-ART	18,4336	6,16551	-1,58005	-1,87639	5,00E-05	0,0121911	yes
HIST1H1C	6:26055967-26056699	TP	TP-ART	33,2718	10,8918	-1,61106	-1,61942	0,0001	0,0217134	yes
LHFPL2	5:77780985-78065844	TP	TP-ART	14,5037	4,68669	-1,62979	-1,48612	5,00E-05	0,0121911	yes

Supplementary Table S1. Differential expression and functional annotation of the deregulated genes identified in the TP/TP-ART comparison (cont.)

Gene	Locus	Group 1	Group2	FPKM Group 1	FPKM Group 2	log2 FC	test stat	p-value	q-value	Significant
DOCK4	7:111366086-111846502	TP	TP-ART	4,38906	1,41683	-1,63124	-0,80245	0,00015	0,0301119	yes
MSR1	8:15965386-16424999	TP	TP-ART	6,9104	2,19356	-1,6555	-1,39438	5,00E-05	0,0121911	yes
ENG	9:130576870-130617035	TP	TP-ART	46,5868	14,6878	-1,6653	-1,36932	5,00E-05	0,0121911	yes
CTSL	9:90340433-90462229	TP	TP-ART	47,7053	14,4428	-1,7238	-1,85796	5,00E-05	0,0121911	yes
CCR2	3:46395161-46404616	TP	TP-ART	15,7243	4,631	-1,7636	-1,6965	0,0001	0,0217134	yes
HERC6	4:89299853-89378161	TP	TP-ART	11,5359	3,36844	-1,77598	-1,12622	5,00E-05	0,0121911	yes
DDX60	4:169137443-169239958	TP	TP-ART	17,1125	4,90763	-1,80195	-1,07116	0,0002	0,0380602	yes
FBP1	9:97365414-97402720	TP	TP-ART	84,128	24,0735	-1,80514	-1,98287	5,00E-05	0,0121911	yes
AGR1	1:955502-991496	TP	TP-ART	9,17891	2,59778	-1,82104	-0,86296	5,00E-05	0,0121911	yes
CXCL2	4:74962751-74965310	TP	TP-ART	34,9619	9,83969	-1,8291	-1,76725	5,00E-05	0,0121911	yes
TLR7	X:12885156-12908687	TP	TP-ART	8,06016	2,22394	-1,85769	-0,98673	5,00E-05	0,0121911	yes
TMEM155	4:122680087-122687962	TP	TP-ART	0,710631	0,190896	-1,89632	-0,84343	5,00E-05	0,0121911	yes
RNASE2	14:21423610-21424595	TP	TP-ART	64,6505	17,0463	-1,9232	-1,99708	5,00E-05	0,0121911	yes
PDK4	7:95212810-95225803	TP	TP-ART	11,6473	3,0321	-1,94161	-1,46002	5,00E-05	0,0121911	yes
MS4A4A	11:60048013-60076445	TP	TP-ART	14,2822	3,71474	-1,94289	-1,49424	0,00015	0,0301119	yes
GIPR	19:46171458-46187181	TP	TP-ART	21,0841	5,30068	-1,99191	-1,91946	5,00E-05	0,0121911	yes
SDC3	1:31342313-31381608	TP	TP-ART	4,2978	1,05858	-2,02147	-1,12729	5,00E-05	0,0121911	yes
OAS2	12:113344581-113455556	TP	TP-ART	75,2716	18,2173	-2,04679	-1,03869	5,00E-05	0,0121911	yes
H1FO	22:38201113-38203442	TP	TP-ART	42,0384	9,62029	-2,12755	-2,47757	5,00E-05	0,0121911	yes
GMPR	6:16238810-16295780	TP	TP-ART	12,1406	2,7417	-2,14669	-2,16262	5,00E-05	0,0121911	yes
LY6E	8:144061752-144105249	TP	TP-ART	660,391	148,31	-2,15471	-2,60543	5,00E-05	0,0121911	yes
ANKRD22	10:90562486-90611796	TP	TP-ART	6,22303	1,39366	-2,15873	-1,50839	5,00E-05	0,0121911	yes
SPATS2L	2:201170603-201347088	TP	TP-ART	11,032	2,45922	-2,16542	-1,87135	5,00E-05	0,0121911	yes
EPSTI1	13:43460162-43566449	TP	TP-ART	72,8204	16,006	-2,18573	-2,25694	5,00E-05	0,0121911	yes
EPHB2	1:23037331-23241901	TP	TP-ART	2,92953	0,642543	-2,18881	-1,85583	5,00E-05	0,0121911	yes
GBP3	1:89472348-89488577	TP	TP-ART	79,8313	16,756	-2,25227	-3,8595	5,00E-05	0,0121911	yes
OAS1	12:113344581-113455556	TP	TP-ART	71,5024	14,5447	-2,29749	-0,73039	5,00E-05	0,0121911	yes

Supplementary Table S1. Differential expression and functional annotation of the deregulated genes identified in the TP/TP-ART comparison (cont.)

Gene	Locus	Group 1	Group2	FPKM Group 1	FPKM Group 2	log2 FC	test stat	p-value	q-value	Significant
NIPAL2	8:99201952-99306760	TP	TP-ART	69,9165	13,7131	-2,35008	-3,17465	5,00E-05	0,0121911	yes
MX1	21:42790833-42836215	TP	TP-ART	133,847	25,6947	-2,38104	-1,66442	5,00E-05	0,0121911	yes
HESX1	3:57231943-57261620	TP	TP-ART	2,1665	0,411142	-2,39765	-1,49649	0,00015	0,0301119	yes
TMEM51	1:14925199-15546976	TP	TP-ART	7,49694	1,4203	-2,40011	-1,98582	5,00E-05	0,0121911	yes
XAF1	17:6658765-6679144	TP	TP-ART	103,026	19,51	-2,40073	-2,05718	5,00E-05	0,0121911	yes
LGALS3BP	17:76967319-76980069	TP	TP-ART	78,8017	14,3096	-2,46124	-2,57215	5,00E-05	0,0121911	yes
APOBEC3A	22:39347594-39483760	TP	TP-ART	360,741	64,9981	-2,47249	-1,41715	5,00E-05	0,0121911	yes
AXL	19:41724873-41767929	TP	TP-ART	2,12095	0,372515	-2,50934	-1,98644	5,00E-05	0,0121911	yes
SLCO4A1	20:61272070-61317137	TP	TP-ART	6,83584	1,19093	-2,52103	-1,72201	5,00E-05	0,0121911	yes
TGM2	20:36756860-36798269	TP	TP-ART	4,78525	0,831686	-2,52449	-2,27315	5,00E-05	0,0121911	yes
NTSR1	20:61340188-61394123	TP	TP-ART	1,32674	0,230329	-2,52612	-1,59771	0,00015	0,0301119	yes
ISG15	1:948802-949920	TP	TP-ART	220,374	37,3491	-2,56081	-2,82587	5,00E-05	0,0121911	yes
TTLL7	1:84330710-84464833	TP	TP-ART	0,910299	0,150187	-2,59958	-0,697	5,00E-05	0,0121911	yes
IFITM3	11:318639-330122	TP	TP-ART	971,474	149,835	-2,6968	-3,283	5,00E-05	0,0121911	yes
OAS3	12:113344581-113455556	TP	TP-ART	57,6867	8,79402	-2,71364	-1,5105	5,00E-05	0,0121911	yes
DPH6	15:35473999-35838394	TP	TP-ART	23,424	3,49471	-2,74475	-1,9954	5,00E-05	0,0121911	yes
CXCL3	4:74902305-74904524	TP	TP-ART	7,02894	0,989164	-2,82903	-1,96655	5,00E-05	0,0121911	yes
IFIT1	10:90973254-91174314	TP	TP-ART	21,5527	2,885	-2,90123	-0,91337	0,0001	0,0217134	yes
TNS1	2:218664511-218898670	TP	TP-ART	11,8796	1,57551	-2,9146	-1,79934	5,00E-05	0,0121911	yes
USP18	22:18632598-18851707	TP	TP-ART	9,54978	1,25178	-2,93148	-2,42869	5,00E-05	0,0121911	yes
CMPK2	2:6968644-7039412	TP	TP-ART	55,6681	7,20103	-2,95058	-2,73469	5,00E-05	0,0121911	yes
MERTK	2:112656055-112787187	TP	TP-ART	20,046	2,5812	-2,9572	-1,68436	5,00E-05	0,0121911	yes
IFI44	1:79085606-79130884	TP	TP-ART	523,592	60,3014	-3,11818	-1,79993	5,00E-05	0,0121911	yes
MLXIPL	7:73007523-73038873	TP	TP-ART	2,66446	0,296714	-3,1667	-1,42167	5,00E-05	0,0121911	yes
RSAD2	2:6968644-7039412	TP	TP-ART	22,6472	2,30305	-3,29772	-1,38514	5,00E-05	0,0121911	yes
DCP1A	3:53304129-53381654	TP	TP-ART	156,795	15,7989	-3,31098	-3,29367	5,00E-05	0,0121911	yes
PTGES	9:132500609-132515326	TP	TP-ART	13,8101	1,35442	-3,34997	-2,51506	5,00E-05	0,0121911	yes



Supplementary Table S1. Differential expression and functional annotation of the deregulated genes identified in the TP/TP-ART comparison (cont.)

Gene	Locus	Group 1	Group2	FPKM Group 1	FPKM Group 2	log2 FC	test stat	p-value	q-value	Significant
LINC00487	2:6868308-6913484	TP	TP-ART	2,94714	0,280583	-3,39281	-1,52982	0,0001	0,0217134	yes
SERPING1	11:57364859-57382326	TP	TP-ART	63,7087	6,0359	-3,39985	-3,44515	5,00E-05	0,0121911	yes
IFI6	1:27992571-27998729	TP	TP-ART	397,336	36,9812	-3,4255	-3,59379	5,00E-05	0,0121911	yes
SLC37A3	7:139993492-140126050	TP	TP-ART	144,492	12,4062	-3,54185	-4,15821	5,00E-05	0,0121911	yes
RNASE1	14:21259622-21278204	TP	TP-ART	9,63642	0,705876	-3,77101	-2,05308	0,00025	0,0446698	yes
PTCH2	1:45265896-45308735	TP	TP-ART	45,8785	3,30062	-3,79701	-1,56162	5,00E-05	0,0121911	yes
CCL2	17:32581899-32584222	TP	TP-ART	54,7655	3,38053	-4,01795	-2,7271	5,00E-05	0,0121911	yes
PPP2R2B	5:145942788-146464347	TP	TP-ART	115,709	5,40316	-4,42055	-4,19933	5,00E-05	0,0121911	yes
SIGLEC1	20:3666759-3693260	TP	TP-ART	72,7761	3,26285	-4,47926	-4,91557	5,00E-05	0,0121911	yes
OTOF	2:26624783-26781566	TP	TP-ART	41,4709	1,58641	-4,70826	-3,11367	5,00E-05	0,0121911	yes
AVIL	12:58156116-58212487	TP	TP-ART	28,3522	0,921073	-4,944	-1,92349	5,00E-05	0,0121911	yes
ARMCX3-AS1	X:100877786-100895381	TP	TP-ART	35,7135	0,739008	-5,59474	-2,89421	5,00E-05	0,0121911	yes
IFI27	14:94571181-94583038	TP	TP-ART	888,536	5,78021	-7,26416	-3,54091	5,00E-05	0,0121911	yes

Supplementary Table S1. Differential expression and functional annotation of the deregulated genes identified in the TP/TP-ART comparison (cont.)

##Databases: KEGG PATHWAY, Reactome, BioCyc

##Statistical test method: hypergeometric test / Fisher's exact test

##FDR correction method: Benjamini and Hochberg

#Term	Database	ID	Input num	Backgroun	P-Value
Interferon alpha/beta signaling	Reactome	R-HSA-909733	14	67	1,13E-15
Interferon Signaling	Reactome	R-HSA-913531	19	189	1,89E-15
Cytokine Signaling in Immune system	Reactome	R-HSA-1280215	23	627	8,08E-10
ISG15 antiviral mechanism	Reactome	R-HSA-1169408	8	68	8,91E-08
Antiviral mechanism by IFN-stimulated genes	Reactome	R-HSA-1169410	8	68	8,91E-08
Immune System	Reactome	R-HSA-168256	32	1579	4,64E-07
Influenza A	KEGG PATHWAY	hsa05164	11	175	3,39E-06
Chemokine receptors bind chemokines	Reactome	R-HSA-380108	6	58	7,42E-06
Measles	KEGG PATHWAY	hsa05162	9	138	2,02E-05
Peptide ligand-binding receptors	Reactome	R-HSA-375276	9	194	2,03E-05
Interferon gamma signaling	Reactome	R-HSA-877300	6	89	7,12E-05
Cytokine-cytokine receptor interaction	KEGG PATHWAY	hsa04060	10	265	0,00053623
Hepatitis C	KEGG PATHWAY	hsa05160	7	135	0,00063326
GPCR ligand binding	Reactome	R-HSA-500792	11	451	0,00064852
Class A/1 (Rhodopsin-like receptors)	Reactome	R-HSA-373076	9	325	0,00083646
Chemokine signaling pathway	KEGG PATHWAY	hsa04062	8	189	0,00094491
Regulation of Complement cascade	Reactome	R-HSA-977606	3	27	0,00132776
Heme degradation	Reactome	R-HSA-189483	2	6	0,00142488

Supplementary Table S1. Differential expression and functional annotation of the deregulated genes identified in the TP/TP-ART comparison (cont.)

##Databases: Gene Ontology

##Statistical test method: hypergeometric test / Fisher's exact test

##FDR correction method: Benjamini and Hochberg

#Term	Database	ID	Input number	Background number	P-Value	Corrected P-Value
cellular response to type I interferon	Gene Ontology	GO:0071357	14	80	3,99E-13	7,45E-10
type I interferon signaling pathway	Gene Ontology	GO:0060337	14	80	3,99E-13	7,45E-10
response to type I interferon	Gene Ontology	GO:0034340	14	84	7,22E-13	8,98E-10
negative regulation of viral genome replication	Gene Ontology	GO:0045071	10	48	2,00E-10	1,87E-07
cytokine-mediated signaling pathway	Gene Ontology	GO:0019221	25	542	4,76E-10	3,55E-07
response to cytokine	Gene Ontology	GO:0034097	30	799	1,02E-09	6,35E-07
defense response to virus	Gene Ontology	GO:0051607	19	331	1,96E-09	1,04E-06
response to virus	Gene Ontology	GO:0009615	21	415	2,40E-09	1,12E-06
defense response	Gene Ontology	GO:0006952	42	1528	5,05E-09	2,09E-06
immune effector process	Gene Ontology	GO:0002252	28	762	5,61E-09	2,09E-06
viral genome replication	Gene Ontology	GO:0019079	11	96	7,38E-09	2,51E-06
regulation of viral genome replication	Gene Ontology	GO:0045069	10	75	9,50E-09	2,96E-06
cellular response to cytokine stimulus	Gene Ontology	GO:0071345	26	689	1,15E-08	3,31E-06
immune response	Gene Ontology	GO:0006955	40	1466	1,39E-08	3,71E-06
response to other organism	Gene Ontology	GO:0051707	30	907	1,60E-08	3,73E-06
response to external biotic stimulus	Gene Ontology	GO:0043207	30	907	1,60E-08	3,73E-06
response to biotic stimulus	Gene Ontology	GO:0009607	30	944	3,72E-08	7,89E-06
negative regulation of viral life cycle	Gene Ontology	GO:1903901	10	88	3,80E-08	7,89E-06
innate immune response	Gene Ontology	GO:0045087	26	791	1,56E-07	3,06E-05
negative regulation of viral process	Gene Ontology	GO:0048525	10	107	2,06E-07	3,84E-05
defense response to other organism	Gene Ontology	GO:0098542	21	551	2,39E-07	4,24E-05
response to external stimulus	Gene Ontology	GO:0009605	45	2107	1,47E-06	2,50E-04
response to interferon-alpha	Gene Ontology	GO:0035455	5	20	3,46E-06	5,61E-04

Supplementary Table S1. Differential expression and functional annotation of the deregulated genes identified in the TP/TP-ART comparison (cont.)

#Term	Database	ID	Input number	Background number	P-Value	Corrected P-Value
negative regulation of multi-organism process	Gene Ontology	GO:0043901	10	154	4,53E-06	7,04E-04
double-stranded RNA binding	Gene Ontology	GO:0003725	7	63	4,86E-06	7,25E-04
regulation of symbiosis, encompassing mutualism	Gene Ontology	GO:0043903	15	381	8,18E-06	1,17E-03
regulation of viral life cycle	Gene Ontology	GO:1903900	10	171	1,08E-05	1,49E-03
regulation of viral process	Gene Ontology	GO:0050792	14	358	1,73E-05	2,31E-03
regulation of vascular endothelial growth factor	Gene Ontology	GO:0010574	5	31	2,25E-05	2,89E-03
cellular response to chemical stimulus	Gene Ontology	GO:0070887	48	2597	2,65E-05	3,30E-03
vascular endothelial growth factor production	Gene Ontology	GO:0010573	5	33	2,95E-05	3,55E-03
cellular defense response	Gene Ontology	GO:0006968	6	58	3,38E-05	3,94E-03
immune system process	Gene Ontology	GO:0002376	45	2411	3,70E-05	4,06E-03
positive regulation of angiogenesis	Gene Ontology	GO:0045766	8	122	3,77E-05	4,06E-03
apoptotic cell clearance	Gene Ontology	GO:0043277	5	35	3,80E-05	4,06E-03
viral process	Gene Ontology	GO:0016032	25	1035	4,48E-05	4,64E-03
multi-organism cellular process	Gene Ontology	GO:0044764	25	1039	4,75E-05	4,79E-03
regulation of granulocyte chemotaxis	Gene Ontology	GO:0071622	5	39	6,08E-05	5,97E-03
ISG15-protein conjugation	Gene Ontology	GO:0032020	3	6	6,84E-05	6,31E-03
symbiosis, encompassing mutualism through parasitism	Gene Ontology	GO:0044403	25	1065	6,93E-05	6,31E-03
interspecies interaction between organisms	Gene Ontology	GO:0044419	25	1065	6,93E-05	6,31E-03
positive regulation of vasculature development	Gene Ontology	GO:1904018	8	135	7,41E-05	6,58E-03
angiogenesis	Gene Ontology	GO:0001525	14	417	8,43E-05	7,17E-03
regulation of immune effector process	Gene Ontology	GO:0002697	14	418	8,64E-05	7,17E-03
regulation of multi-organism process	Gene Ontology	GO:0043900	15	472	8,64E-05	7,17E-03
acute inflammatory response	Gene Ontology	GO:0002526	8	143	1,08E-04	8,67E-03
cellular response to organic substance	Gene Ontology	GO:0071310	40	2157	1,09E-04	8,67E-03
cell chemotaxis	Gene Ontology	GO:0060326	10	230	1,17E-04	9,03E-03
positive regulation of neutrophil chemotaxis	Gene Ontology	GO:0090023	4	23	1,19E-04	9,03E-03
blood vessel morphogenesis	Gene Ontology	GO:0048514	15	489	1,26E-04	9,37E-03

Supplementary Table S1. Differential expression and functional annotation of the deregulated genes identified in the TP/TP-ART comparison (cont.)

#Term	Database	ID	Input number	Background number	P-Value	Corrected P-Value
chemokine-mediated signaling pathway	Gene Ontology	GO:0070098	6	80	1,78E-04	1,30E-02
positive regulation of granulocyte chemotaxis	Gene Ontology	GO:0071624	4	26	1,81E-04	1,30E-02
positive regulation of leukocyte chemotaxis	Gene Ontology	GO:0002690	6	82	2,01E-04	1,38E-02
positive regulation of neutrophil migration	Gene Ontology	GO:1902624	4	27	2,06E-04	1,38E-02
macrophage chemotaxis	Gene Ontology	GO:0048246	4	27	2,06E-04	1,38E-02
regulation of neutrophil chemotaxis	Gene Ontology	GO:0090022	4	27	2,06E-04	1,38E-02
positive regulation of macrophage chemotaxis	Gene Ontology	GO:0010759	3	11	2,86E-04	1,87E-02
response to organic substance	Gene Ontology	GO:0010033	47	2813	3,00E-04	1,93E-02
regulation of angiogenesis	Gene Ontology	GO:0045765	9	213	3,09E-04	1,94E-02
chemotaxis	Gene Ontology	GO:0006935	15	534	3,11E-04	1,94E-02
taxis	Gene Ontology	GO:0042330	15	535	3,17E-04	1,94E-02
regulation of neutrophil migration	Gene Ontology	GO:1902622	4	31	3,33E-04	2,01E-02
regulation of complement activation	Gene Ontology	GO:0030449	4	32	3,72E-04	2,20E-02
response to lipopolysaccharide	Gene Ontology	GO:0032496	11	324	4,26E-04	2,48E-02
regulation of protein activation cascade	Gene Ontology	GO:2000257	4	34	4,59E-04	2,64E-02
regulation of cytokine production	Gene Ontology	GO:0001817	15	557	4,76E-04	2,69E-02
regulation of leukocyte chemotaxis	Gene Ontology	GO:0002688	6	98	4,96E-04	2,74E-02
positive regulation of immune system process	Gene Ontology	GO:0002684	20	878	5,00E-04	2,74E-02
regulation of immune system process	Gene Ontology	GO:0002682	27	1369	5,39E-04	2,92E-02
negative regulation of extrinsic apoptotic signaling pathway	Gene Ontology	GO:2001237	6	100	5,49E-04	2,93E-02
regulation of defense response	Gene Ontology	GO:0031347	18	756	5,69E-04	2,99E-02
response to molecule of bacterial origin	Gene Ontology	GO:0002237	11	337	5,83E-04	3,02E-02
regulation of vasculature development	Gene Ontology	GO:1901342	9	234	5,93E-04	3,03E-02
cell surface receptor signaling pathway	Gene Ontology	GO:0007166	43	2592	6,00E-04	3,03E-02
tissue remodeling	Gene Ontology	GO:0048771	7	143	6,19E-04	3,08E-02
response to lipid	Gene Ontology	GO:0033993	20	895	6,29E-04	3,09E-02
blood vessel development	Gene Ontology	GO:0001568	15	575	6,51E-04	3,16E-02

Supplementary Table S1. Differential expression and functional annotation of the deregulated genes identified in the TP/TP-ART comparison (cont.)

#Term	Database	ID	Input number	Background number	P-Value	Corrected P-Value
response to stress	Gene Ontology	GO:0006950	56	3656	6,70E-04	3,20E-02
regulation of mononuclear cell migration	Gene Ontology	GO:0071675	3	16	7,35E-04	3,47E-02
inflammatory response	Gene Ontology	GO:0006954	16	647	7,61E-04	3,55E-02
regulation of leukocyte migration	Gene Ontology	GO:0002685	7	150	8,10E-04	3,73E-02
regulation of macrophage chemotaxis	Gene Ontology	GO:0010758	3	17	8,58E-04	3,86E-02
mononuclear cell migration	Gene Ontology	GO:0071674	3	17	8,58E-04	3,86E-02
positive regulation of leukocyte migration	Gene Ontology	GO:0002687	6	110	8,81E-04	3,92E-02
regulation of acute inflammatory response	Gene Ontology	GO:0002673	5	73	9,07E-04	3,94E-02
vasculature development	Gene Ontology	GO:0001944	15	595	9,08E-04	3,94E-02
myeloid leukocyte migration	Gene Ontology	GO:0097529	7	156	1,01E-03	4,32E-02
response to interferon-gamma	Gene Ontology	GO:0034341	7	158	1,08E-03	4,59E-02

Supplementary Table S2. Differential expression identified in the EC-LTNP/TP-ART comparison

Gene	Locus	Group 1	Group2	FPKM Group 1	FPKM Group 2	log2 FC	test stat	p-value	q-value	Significant
HELB	12:66696237-67197966	EC-LTNP	TP-ART	4,83523	185,021	5,25797	4,45175	5,00E-05	0,0121911	yes
ZNF275	X:152599612-152631799	EC-LTNP	TP-ART	3,87358	142,653	5,2027	6,78297	5,00E-05	0,0121911	yes
MED20	6:41873008-41888877	EC-LTNP	TP-ART	2,03447	24,5069	3,59047	4,21793	5,00E-05	0,0121911	yes
HBA2	16:222845-223709	EC-LTNP	TP-ART	2,75	30,4794	3,47033	2,31655	5,00E-05	0,0121911	yes
HAR1A	20:61726844-61735738	EC-LTNP	TP-ART	0,590532	5,51377	3,22295	2,30359	5,00E-05	0,0121911	yes
SSRP1	11:57093076-57103351	EC-LTNP	TP-ART	22,8367	189,504	3,0528	6,12567	5,00E-05	0,0121911	yes
HBA1	16:226678-227521	EC-LTNP	TP-ART	3	18,734	2,64262	1,82352	5,00E-05	0,0121911	yes
VWA8	13:42137083-42535256	EC-LTNP	TP-ART	3,73995	22,1423	2,56572	3,29962	5,00E-05	0,0121911	yes
CA1	8:86239836-86393722	EC-LTNP	TP-ART	0,0996387	0,51085	2,35812	0,247151	5,00E-05	0,0121911	yes
RRP7A	22:42896584-42978044	EC-LTNP	TP-ART	17,7277	82,1622	2,21247	2,31365	5,00E-05	0,0121911	yes
KIF19	17:72322348-72351959	EC-LTNP	TP-ART	0,717707	3,14438	2,13131	1,65014	5,00E-05	0,0121911	yes
IGHA2	14:106053225-106054732	EC-LTNP	TP-ART	16,338	70,2289	2,10383	1,8826	5,00E-05	0,0121911	yes
MEG3	14:101245746-101327368	EC-LTNP	TP-ART	0,661139	2,81914	2,09223	0,968205	5,00E-05	0,0121911	yes
EML6	2:54950635-55199157	EC-LTNP	TP-ART	0,662611	2,7934	2,07579	0,86326	0,0002	0,0380602	yes
TRBV23-1	7:142353467-142353964	EC-LTNP	TP-ART	3,3975	13,8171	2,02391	1,40509	5,00E-05	0,0121911	yes
IGHA1	14:106173456-106175002	EC-LTNP	TP-ART	67,4389	272,391	2,01403	1,94386	5,00E-05	0,0121911	yes
GLG1	16:74481324-74641012	EC-LTNP	TP-ART	37,6327	146,576	1,9616	2,45501	5,00E-05	0,0121911	yes
CAPN8	1:223711348-223853436	EC-LTNP	TP-ART	0,210788	0,814389	1,94993	0,808696	5,00E-05	0,0121911	yes
NPDC1	9:139933774-139940655	EC-LTNP	TP-ART	4,38997	13,23	1,59153	1,04445	5,00E-05	0,0121911	yes
AGAP8	10:51224680-51371338	EC-LTNP	TP-ART	0,294849	0,865348	1,5533	0,926138	5,00E-05	0,0121911	yes
IGLC2	22:23243155-23243617	EC-LTNP	TP-ART	32,8236	88,3347	1,42825	1,4603	5,00E-05	0,0121911	yes
NDRG4	16:58496749-58547532	EC-LTNP	TP-ART	0,458486	0,790398	0,785703	0,362417	5,00E-05	0,0121911	yes
C3AR1	12:8210897-8219067	EC-LTNP	TP-ART	21,6787	10,3438	-1,06751	-1,22159	0,0002	0,0380602	yes
TNFSF13B	13:108903587-108960933	EC-LTNP	TP-ART	58,6097	26,9272	-1,12207	-1,07674	0,00015	0,0301119	yes
PTAFR	1:28473667-28520447	EC-LTNP	TP-ART	54,0195	24,2653	-1,15459	-1,50817	5,00E-05	0,0121911	yes
CCR1	3:46205095-46308197	EC-LTNP	TP-ART	92,5348	41,1043	-1,17071	-1,51545	0,0001	0,0217134	yes
ALDH9A1	1:165631452-165679205	EC-LTNP	TP-ART	72,4907	30,4516	-1,25128	-1,64597	0,00025	0,0446698	yes
PLAUR	19:44150246-44174699	EC-LTNP	TP-ART	333,639	137,553	-1,2783	-1,37601	5,00E-05	0,0121911	yes
EREG	4:75230859-75254472	EC-LTNP	TP-ART	22,3002	9,09461	-1,29397	-1,21992	0,00015	0,0301119	yes
C5AR1	19:47793279-47887533	EC-LTNP	TP-ART	178,287	69,6735	-1,35552	-1,74603	5,00E-05	0,0121911	yes

Supplementary Table S2. Differential expression identified in the EC-LTNP/TP-ART comparison (continuation)

Gene	Locus	Group 1	Group2	FPKM Group 1	FPKM Group 2	log2 FC	test stat	p-value	q-value	Significant
IFITM3	11:318639-330122	EC-LTNP	TP-ART	392,402	149,835	-1,38896	-1,70858	5,00E-05	0,0121911	yes
EPHB2	1:23037331-23241901	EC-LTNP	TP-ART	1,7001	0,642543	-1,40375	-1,20644	0,00025	0,0446698	yes
GBP3	1:89472348-89488577	EC-LTNP	TP-ART	44,4154	16,756	-1,40638	-3,23306	5,00E-05	0,0121911	yes
ZNF703	8:37553268-37557537	EC-LTNP	TP-ART	37,6481	14,1743	-1,4093	-1,56381	0,00015	0,0301119	yes
NR4A1	12:52416615-52453566	EC-LTNP	TP-ART	94,6402	35,6112	-1,41012	-1,16478	0,0001	0,0217134	yes
IFI6	1:27992571-27998729	EC-LTNP	TP-ART	98,9186	36,9812	-1,41945	-1,66949	5,00E-05	0,0121911	yes
TGM2	20:36756860-36798269	EC-LTNP	TP-ART	2,23611	0,831686	-1,42688	-1,29729	0,0001	0,0217134	yes
FPR1	19:52243790-52329442	EC-LTNP	TP-ART	172,045	62,4956	-1,46096	-1,63578	5,00E-05	0,0121911	yes
CMPK2	2:6968644-7039412	EC-LTNP	TP-ART	20,0385	7,20103	-1,4765	-1,39798	5,00E-05	0,0121911	yes
CD9	12:6308880-6347425	EC-LTNP	TP-ART	20,5748	7,21529	-1,51175	-1,56725	5,00E-05	0,0121911	yes
CCR2	3:46395161-46404616	EC-LTNP	TP-ART	13,5316	4,631	-1,54694	-1,44582	5,00E-05	0,0121911	yes
APOL4	22:36585171-36600886	EC-LTNP	TP-ART	1,33256	0,40416	-1,7212	-0,96206	0,0002	0,0380602	yes
SLC37A3	7:139993492-140126050	EC-LTNP	TP-ART	43,5736	12,4062	-1,81239	-2,08386	5,00E-05	0,0121911	yes
TRIM16L	17:18601310-18639578	EC-LTNP	TP-ART	4,45625	1,26685	-1,81459	-1,77258	5,00E-05	0,0121911	yes
SDC3	1:31342313-31381608	EC-LTNP	TP-ART	3,74052	1,05858	-1,82111	-1,04802	0,0002	0,0380602	yes
SERPING1	11:57364859-57382326	EC-LTNP	TP-ART	22,0966	6,0359	-1,87219	-1,97052	5,00E-05	0,0121911	yes
SIGLEC1	20:3666759-3693260	EC-LTNP	TP-ART	12,1492	3,26285	-1,89665	-1,8666	5,00E-05	0,0121911	yes
IFI27	14:94571181-94583038	EC-LTNP	TP-ART	25,0601	5,78021	-2,1162	-1,45337	0,0002	0,0380602	yes
NTSR1	20:61340188-61394123	EC-LTNP	TP-ART	1,1942	0,230329	-2,37428	-1,45991	0,00015	0,0301119	yes
DCP1A	3:53304129-53381654	EC-LTNP	TP-ART	83,4389	15,7989	-2,40089	-2,33641	5,00E-05	0,0121911	yes
SIGLEC12	19:51994496-52005043	EC-LTNP	TP-ART	0,926278	0,135099	-2,77743	-1,28593	0,00025	0,0446698	yes
ITPR3	6:33588029-33679504	EC-LTNP	TP-ART	65,1372	8,99487	-2,85631	-4,93623	5,00E-05	0,0121911	yes
ZNF71	19:57106631-57138118	EC-LTNP	TP-ART	7,31587	0,943736	-2,95457	-2,42014	5,00E-05	0,0121911	yes
NIPAL2	8:99201952-99306760	EC-LTNP	TP-ART	110,646	13,7131	-3,01232	-4,11377	5,00E-05	0,0121911	yes
BTBD11	12:107712105-108053419	EC-LTNP	TP-ART	82,45	7,23071	-3,51131	-5,10682	5,00E-05	0,0121911	yes
PPP2R2B	5:145942788-146464347	EC-LTNP	TP-ART	154,693	5,40316	-4,83946	-4,60005	5,00E-05	0,0121911	yes
ANKRD54	22:38226861-38285414	EC-LTNP	TP-ART	231,062	7,50059	-4,94513	-1,74397	5,00E-05	0,0121911	yes
XIST	X:73012039-73072588	EC-LTNP	TP-ART	39,7674	0,0505588	-9,61941	-1,59416	0,00015	0,0301119	yes



## Resultados

**Supplementary Table S3. Transcription binding sites predicted in the promoters of the 58 differentially expressed genes comparing EC-LTNP and TP-ART.** Each sheet includes the analysis for each gene found in the comparison. For each transcription factor site, several predicted parameters are reported [51].

The transcription factor name with the database accession number in brackets; the start and end positions of the putative binding sequences; Dissimilarity (%), which corresponds to the rate of dissimilarity between the putative and consensus sequences for a given transcription factor; String, the nucleotide sequence of potential binding site; Random Expectation (RE) indicating the expected occurrences of the match in a random sequence of the same length as the query sequence according to the dissimilarity index, presented the RE equally (equi-probability for the four nucleotides) and RE query (nucleotide frequencies as in the query sequence).

Sequence name	Factor name	Start position	End position	Dissimilarity	String	RE equally	RE query
ALDH9A1	RAR-beta:RXR-alpha [T	830	841	0	GGGCTCAGGGGA	0,00027	0,00015
ALDH9A1	Ik-1 [T02702]	1017	1029	2,374299	CAAGTAGCTGGGA	0,00047	0,00032
ALDH9A1	RelA [T00594]	996	1006	2,485523	GATTTTCCAC	0,00429	0,00427
ALDH9A1	USF2 [T00878]	1126	1135	4,003951	CAGGTGATCT	0,00858	0,00652
ALDH9A1	NF-kappaB1 [T00593]	837	847	3,680943	GGGGATCCTCC	0,01359	0,00736
ALDH9A1	NF-AT1 [T00550]	645	653	1,437145	ACTTTTCC	0,00572	0,00816
ALDH9A1	NF-AT2 [T01945]	644	653	3,814941	CACTTTTCC	0,00858	0,01068
ALDH9A1	STAT1beta [T01573]	149	158	0	TCTTGAAAT	0,01287	0,01509
ALDH9A1	Sp1 [T00759]	853	862	4,907768	TAGCCGCCCA	0,03433	0,02192
ALDH9A1	HOXD9 [T01424]	1194	1203	0	AATTTTTATT	0,00858	0,02229
ALDH9A1	HOXD9 [T01424]	1203	1212	0	TATTTTTATT	0,00858	0,02229
ALDH9A1	HOXD10 [T01425]	1194	1203	0	AATTTTTATT	0,00858	0,02229
ALDH9A1	HOXD10 [T01425]	1203	1212	0	TATTTTTATT	0,00858	0,02229
ALDH9A1	AR [T00040]	479	487	1,871872	GAAATGTCC	0,02861	0,02831
ALDH9A1	GCF [T00320]	404	412	2,140539	GCGCTGGTG	0,06866	0,03691
ALDH9A1	T3R-beta1 [T00851]	405	413	1,110682	CGCTGGTGA	0,05722	0,04741
ANKRD54	PITX2 [T02413]	194	213	1,535113	CTCACGCCTGTAATCCC	0	0
ANKRD54	STAT3 [T01493]	1457	1470	2,086875	CTGCCGGGAAACCG	0,0004	0,0004
ANKRD54	Ik-1 [T02702]	207	219	2,374299	TCCCAGCACTTTG	0,00047	0,00053
ANKRD54	NF-kappaB [T00590]	450	461	3,864667	AGGGAAACTCCA	0,00456	0,00459
ANKRD54	RelA [T00594]	449	459	3,503248	GAGGGAAACTC	0,00465	0,00464
ANKRD54	TCF-4 [T02918]	924	933	4,865142	CCTTTGAGGA	0,01431	0,01466
ANKRD54	NF-AT1 [T00550]	1473	1481	3,098758	AGCCTTCC	0,02289	0,02309
ANKRD54	TCF-4 [T02918]	653	662	2,548903	CCATCAAAGA	0,02575	0,02314
ANKRD54	AR [T00040]	956	964	1,86979	GGACAGTAA	0,02861	0,02752
ANKRD54	Sp1 [T00759]	227	236	1,566059	GGGCGGGTG	0,02432	0,02963
ANKRD54	Sp1 [T00759]	1156	1165	2,203247	TGGGCGGGCA	0,02718	0,03255
ANKRD54	Sp1 [T00759]	1214	1223	2,418403	CGGGCGGGAG	0,03004	0,03559
ANKRD54	c-Fos [T00123]	633	642	3,579011	TGACTGACTC	0,04005	0,04001
BTBD11	Sp1 [T00759]	1363	1372	0	GGGGCGGGGC	0,00143	0,00177
BTBD11	NF-kappaB [T00590]	106	117	3,935637	TGGGAACCTCCA	0,00456	0,00469
BTBD11	POU2F1 [T00641]	307	317	0,290044	ATTTGCATATA	0,00536	0,00489
BTBD11	HOXD9 [T01424]	94	103	0	TATTTTTATT	0,00858	0,0072
BTBD11	HOXD10 [T01425]	94	103	0	TATTTTTATT	0,00858	0,0072
BTBD11	POU2F1 [T00641]	271	281	0,801279	ATTTGCATCTC	0,01037	0,00959
BTBD11	PPAR-alpha:RXR-alpha	563	573	4,88658	ATCTGGGCCAG	0,01287	0,0136
BTBD11	HOXD9 [T01424]	34	43	4,321431	AGATTTTATT	0,02575	0,02218
BTBD11	HOXD10 [T01425]	34	43	4,321431	AGATTTTATT	0,02575	0,02218

## Resultados

Supplementary Table S3. (Continuation)

Sequence		End		Dissimilarity	String	RE equally	RE query
name	Factor name	Start position	position				
BTBD11	TCF-4 [T02918]	871	880	2,386325	ACTTTGAAGC	0,02575	0,02388
BTBD11	TCF-4 [T02918]	767	776	4,639022	CCTTTGAGCA	0,03147	0,03125
BTBD11	Sp1 [T00759]	1446	1455	2,491373	CGCCGCCCC	0,03004	0,03451
BTBD11	TCF-4 [T02918]	337	346	0,925008	CATTCAAAGC	0,04292	0,04125
BTBD11	NF-AT1 [T01948]	808	817	2,756277	TGGAAAGATG	0,04435	0,04257
BTBD11	c-Ets-2 [T00113]	904	912	2,945838	TGAAAGGAA	0,04578	0,04308
BTBD11	Elk-1 [T00250]	1325	1333	0,134348	CTTCCTGCA	0,04578	0,04856
C3AR1	RAR-beta [T00721]	1357	1366	0	GTTGAACCT	0,00572	0,0046
C3AR1	TCF-4 [T02918]	900	909	4,639022	CCTTTGACTA	0,03147	0,02787
C3AR1	USF2 [T00878]	68	77	4,528187	GGATCACCTG	0,0515	0,02832
C3AR1	Elk-1 [T00250]	333	341	3,381796	ATCAGGAAG	0,03433	0,03223
C3AR1	Elk-1 [T00250]	801	809	3,381796	CTTCCTGAT	0,03433	0,03223
C3AR1	SRY [T00997]	452	460	0	AGAACAAAG	0,02289	0,03894
C3AR1	c-Ets-2 [T00113]	1277	1285	2,715313	TTCTCCCC	0,05722	0,04504
C3AR1	LEF-1 [T02905]	453	460	0,641865	GAACAAAG	0,04578	0,04999
C5AR1	PITX2 [T02413]	666	685	3,070225	GCTGGGATTATAGGTG	0	0
C5AR1	PITX2 [T02413]	1069	1088	3,070225	CACACACGTGAATCCC	0	0
C5AR1	Ik-1 [T02702]	525	537	2,374299	CAAGTAGCTGGGA	0,00047	0,00046
C5AR1	Ik-1 [T02702]	660	672	2,374299	CAAAGTGCTGGGA	0,00047	0,00046
C5AR1	Ik-1 [T02702]	1082	1094	4,748597	TCCCAGTACTCG	0,00235	0,00231
C5AR1	c-Ets-2 [T00113]	893	901	0,572986	TAGGAGGAA	0,00572	0,00563
C5AR1	EBF [T05427]	1343	1353	4,04219	TCCCTGGGTG	0,0186	0,01834
C5AR1	Elk-1 [T00250]	703	711	0,957025	AGAAGGAAG	0,02289	0,02268
C5AR1	T3R-beta1 [T00851]	1060	1068	2,259951	GGGTGGTGA	0,02289	0,02311
C5AR1	IRF-1 [T00423]	1241	1249	1,616539	ATAGGGAAA	0,03433	0,03544
C5AR1	USF1 [T00874]	1069	1078	3,246902	CACACACGTG	0,04721	0,04654
C5AR1	SRY [T00997]	433	441	0,999172	CTTTGTTGC	0,04578	0,04855
CCR1	MAZ [T00490]	975	987	1,858283	GGTGGGGAGGGAC	0,00036	0,00024
CCR1	ELF-1 [T01113]	1377	1389	0,750936	TTCCAGGAAGTGG	0,00045	0,00043
CCR1	EBF [T05427]	1314	1324	3,135622	CTCCAGGGCC	0,00572	0,00406
CCR1	NF-kappaB1 [T00593]	591	601	2,871556	GGGGTGTCCCC	0,01252	0,00974
CCR1	PPAR-alpha:RXR-alpha	946	956	4,727619	TCCTGGGGCAA	0,02432	0,02037
CCR1	c-Myb [T00137]	642	649	0	GGCAGTTG	0,02289	0,02045
CCR1	c-Myb [T00137]	1362	1369	4,270092	CGCAGTTA	0,04578	0,0425
CCR1	TCF-4 [T02918]	523	532	1,235198	TTTTCAAAGC	0,04292	0,04711
CCR1	PXR-1:RXR-alpha [T05427]	499	506	4,90845	TGAACCAT	0,04578	0,04982
CCR2	USF2 [T00878]	265	274	4,003951	AGGACACCTG	0,00858	0,00533
CCR2	USF1 [T00874]	819	828	1,908112	ATACCACGTG	0,02718	0,01437
CCR2	T3R-beta1 [T00851]	1090	1098	2,259951	AGGAGGTGA	0,02289	0,01443
CCR2	HNF-1C [T01951]	1188	1196	0,90144	GTTAATAAT	0,00572	0,01612
CCR2	HNF-1B [T01950]	1187	1195	1,651022	AGTTAATAA	0,00572	0,01658
CCR2	USF1 [T00874]	823	832	3,033037	CACGTGATCT	0,04721	0,02797
CCR2	Elk-1 [T00250]	1486	1494	0,134348	CTTCCTCCA	0,04578	0,03078
CCR2	POU2F2 (Oct-2,1) [T00250]	1372	1382	4,120113	TGAATTAATA	0,01287	0,0329
CCR2	PEA3 [T00685]	781	789	0,597316	AGGATGTTA	0,03433	0,03983
CCR2	c-Ets-2 [T00113]	1077	1085	2,715313	GGGGAGGAA	0,05722	0,04628
CCR2	c-Ets-2 [T00113]	1487	1495	2,715313	TTCTCCAC	0,05722	0,04628

## Resultados

Supplementary Table S3. (Continuation)

Sequence		End		Dissimilarity	String	RE equally	RE query
name	Factor name	Start position	position				
CD9	MAZ [T00490]	1396	1408	4,524062	AGGAGGGAGGGTG	0,00141	0,00245
CD9	MAZ [T00490]	114	126	2,378507	CTCCCTCCCCTTC	0,00304	0,00599
CD9	Elk-1 [T00250]	395	403	1,645354	CTTCCTACC	0,00572	0,00701
CD9	NF-kappaB1 [T00593]	29	39	4,847696	GGGGAGCCTCT	0,01001	0,01215
CD9	PPAR-alpha:RXR-alpha	1458	1468	3,872523	CTCTGGGGCAA	0,01931	0,02171
CD9	PPAR-alpha:RXR-alpha	608	618	4,727619	CTGCCCCAGCC	0,02432	0,02746
CD9	PPAR-alpha:RXR-alpha	679	689	4,727619	GTCTGGGTCAG	0,02432	0,02746
CD9	AR [T00040]	1031	1039	1,86979	GGACAGTAA	0,02861	0,02837
CD9	NF-kappaB1 [T00593]	412	422	4,258318	CGGCCCTCCCC	0,02432	0,0291
CD9	AR [T00040]	12	20	0,865816	GCAATGTCC	0,02861	0,02969
CD9	IRF-1 [T00423]	648	656	1,616539	TTTCCCTGT	0,03433	0,0307
CD9	IRF-1 [T00423]	1453	1461	1,616539	TTTCCCTCT	0,03433	0,0307
CD9	c-Ets-2 [T00113]	965	973	2,945838	TTCTTTTCA	0,04578	0,03857
CD9	PXR-1:RXR-alpha [T00593]	490	497	4,90845	TGAACCTT	0,04578	0,04039
CD9	CREB [T00163]	51	59	4,185436	TGACGTGGG	0,04578	0,04445
CD9	AR [T00040]	599	607	4,995624	GGACAGCCC	0,04005	0,04568
CD9	ATF-2 [T00167]	1160	1169	3,160734	GTGACGTTGG	0,04578	0,04696
CD9	IRF-1 [T00423]	1037	1045	3,692688	TAAAGGAAA	0,0515	0,04819
CD9	HNF-3alpha [T02512]	1054	1061	0	TGAAAATA	0,06866	0,04899
CMPK2	STAT5B [T04684]	739	754	4,926759	AATTTCAAGAAGAGGG	0,00035	0,00026
CMPK2	FOXO4 [T03403]	452	465	4,339504	TCTTTTTGTTTGT	0,00063	0,0004
CMPK2	EBF [T05427]	1329	1339	1,477196	GCCCCAGGGCT	0,00286	0,00474
CMPK2	RelA [T00594]	431	441	3,767912	GAGTTTCCCTG	0,00608	0,00603
CMPK2	MAZ [T00490]	930	942	3,959641	GTCGGGGAGGGAG	0,0044	0,00656
CMPK2	HNF-1B [T01950]	500	508	0,825511	AGTTAATGA	0,01144	0,008
CMPK2	c-Fos [T00123]	717	726	0,556493	GAGTCAGCAA	0,00858	0,00866
CMPK2	Sp1 [T00759]	895	904	0,574521	GGGGCGGGGA	0,00572	0,01051
CMPK2	Sp1 [T00759]	1054	1063	0,336788	CCCCCGCCCC	0,00572	0,01051
CMPK2	HOXD9 [T01424]	192	201	4,321431	AATAAAAACCT	0,02575	0,01509
CMPK2	HOXD9 [T01424]	519	528	4,321431	AATAAAATAT	0,02575	0,01509
CMPK2	HOXD10 [T01425]	192	201	4,321431	AATAAAAACCT	0,02575	0,01509
CMPK2	HOXD10 [T01425]	519	528	4,321431	AATAAAATAT	0,02575	0,01509
CMPK2	NF-AT1 [T00550]	49	57	3,407861	GGAAAAGGT	0,02289	0,02007
CMPK2	HNF-1C [T01951]	501	509	2,503438	GTTAATGAG	0,03433	0,02461
CMPK2	E2F-1 [T01542]	1392	1399	0	TTTCCCGC	0,02289	0,02636
CMPK2	E2F-1 [T01542]	1426	1433	0	TTTCCCGC	0,02289	0,02636
CMPK2	IRF-1 [T00423]	45	53	1,616539	ATAGGGAAA	0,03433	0,02907
CMPK2	IRF-1 [T00423]	749	757	1,616539	AGAGGGAAA	0,03433	0,02907
CMPK2	GCF [T00320]	1213	1221	1,26923	GCGCAGGCC	0,02289	0,03363
CMPK2	c-Ets-2 [T00113]	1090	1098	1,64415	GAGGAGGAA	0,03433	0,03581
CMPK2	HNF-3alpha [T02512]	225	232	4,842999	AATTTTAA	0,06866	0,03943
CMPK2	HNF-3alpha [T02512]	520	527	4,842999	ATAAAAATA	0,06866	0,03943
CMPK2	IRF-1 [T00423]	434	442	3,689552	TTTCCCTGG	0,0515	0,04711
CMPK2	c-Myb [T00137]	562	569	1,285398	GGCAGTTA	0,04578	0,0479
CMPK2	c-Myb [T00137]	534	541	4,270092	TAACTGCG	0,04578	0,04793
CMPK2	c-Myb [T00137]	850	857	4,270092	GAACTGCG	0,04578	0,04793
DCP1A	PITX2 [T02413]	152	171	1,535113	CTCACGCCTGTAATCCC	0	0
DCP1A	Ik-1 [T02702]	165	177	2,374299	TCCCAGCACTTG	0,00047	0,00061

## Resultados

**Supplementary Table S3. (Continuation)**  
Sequence

name	Factor name	Start position	End position	Dissimilarity	String	RE equally	RE query
DCP1A	USF2 [T00878]	194	203	4,003951	AGATCACCTG	0,00858	0,0096
DCP1A	TBP [T00794]	733	742	3,743085	CTACTATAAA	0,02289	0,01346
DCP1A	RAR-beta [T00721]	87	96	1,08151	ATTAACCCCT	0,02289	0,02293
DCP1A	STAT1beta [T01573]	1317	1326	4,01053	TCAGGGAAAG	0,02575	0,02303
DCP1A	PPAR-alpha:RXR-alpha	1030	1040	3,872523	TTGCCCCAGAG	0,01931	0,02404
DCP1A	EBF [T05427]	107	117	4,016439	TTCCAGGGCT	0,0186	0,02753
DCP1A	NF-AT1 [T01948]	1260	1269	2,756277	TGGAAAGAGC	0,04435	0,03586
DCP1A	PXR-1:RXR-alpha [T056	818	825	4,90845	TGAACCTT	0,04578	0,03705
DCP1A	NF-AT1 [T00550]	1181	1189	1,970716	ATGTTTTCC	0,0515	0,03855
DCP1A	c-Myb [T00137]	1491	1498	2,687937	GGAAGTTG	0,04578	0,04392
DCP1A	NF-AT1 [T01948]	1159	1168	4,823485	CGCATTTC	0,05722	0,04961
EML6	FOXO4 [T03403]	69	82	2,169752	AGGAAACAACTTA	0,00063	0,00015
EML6	CD28RC [T00102]	147	162	4,678585	AGAAATTCAGTGGTG	0,0003	0,00016
EML6	CDX2 [T03246]	1162	1176	4,665913	GTTCTTTATGGCTTT	0,00067	0,00035
EML6	Sp1 [T00759]	567	576	0	GCCCCGCC	0,00143	0,00667
EML6	MAZ [T00490]	654	666	2,378507	GGAGGGGAGGGCA	0,00304	0,00782
EML6	MAZ [T00490]	649	661	3,162267	GGCGGGGAGGGGA	0,00355	0,00886
EML6	AhR:Arnt [T05394]	217	226	0	GCACGCCAGC	0,00429	0,01232
EML6	RAR-beta:RXR-alpha [T	942	953	4,98533	GGGCTCAGGGCA	0,00724	0,01355
EML6	IRF-1 [T00423]	52	60	1,274173	AAAAGGAAA	0,03433	0,01976
EML6	HNF-3alpha [T02512]	1403	1410	0	TAAAAATA	0,06866	0,02134
EML6	c-Ets-2 [T00113]	51	59	2,945838	GAAAAGGAA	0,04578	0,02267
EML6	NF-AT1 [T00550]	56	64	1,970716	GGAAAACCT	0,0515	0,02625
EML6	c-Myb [T00137]	1316	1323	0	CAACTGCC	0,02289	0,02871
EML6	PXR-1:RXR-alpha [T056	1189	1196	4,90845	AAGTTCA	0,04578	0,02975
EML6	c-Fos [T00123]	88	97	3,579011	GAGTCAGTTA	0,04005	0,0339
EML6	NF-AT1 [T01948]	501	510	3,445347	GCTGTTTCCA	0,05722	0,03803
EML6	AR [T00040]	1393	1401	3,382886	GGACAATTT	0,0515	0,03946
EML6	Elk-1 [T00250]	1479	1487	2,164966	CTTCCTCGC	0,04005	0,04409
EML6	PEA3 [T00685]	108	116	1,194633	AGGATGTCG	0,0515	0,04496
EPHB2	HOXD9 [T01424]	1171	1180	0	AATAAAAATT	0,00858	0,00045
EPHB2	HOXD10 [T01425]	1171	1180	0	AATAAAAATT	0,00858	0,00045
EPHB2	HNF-1C [T01951]	481	489	0	GTTAATGAT	0,01144	0,00156
EPHB2	Ik-1 [T02702]	1036	1048	4,748597	CGAGGGGTGGGA	0,00235	0,00434
EPHB2	MAZ [T00490]	827	839	1,594748	CACCCTCCCCCA	0,00107	0,00441
EPHB2	HNF-1B [T01950]	480	488	3,610263	TGTTAATGA	0,03433	0,00661
EPHB2	NF-kappaB1 [T00593]	624	634	2,687613	GGGCTTTCCC	0,00465	0,00688
EPHB2	NF-AT1 [T01948]	161	170	2,067208	CAACTTCCA	0,02861	0,01018
EPHB2	MAZ [T00490]	224	236	2,378507	GAAGGGGAGGGAT	0,00304	0,01149
EPHB2	MAZ [T00490]	1101	1113	2,378507	CGCCCTCCCTCC	0,00304	0,01149
EPHB2	ETF [T00270]	1391	1401	0	GCGCGGGGGC	0,00107	0,01316
EPHB2	Elk-1 [T00250]	650	658	0,957025	CTTCCTCA	0,02289	0,01765
EPHB2	HNF-3alpha [T02512]	1173	1180	3,500065	TAAAAAATT	0,20599	0,02443
EPHB2	NF-AT1 [T01948]	970	979	4,823485	TTTTTTTCCA	0,05722	0,02919
EPHB2	c-Myb [T00137]	404	411	0	CAACTGCC	0,02289	0,02926
EPHB2	AR [T00040]	242	250	3,59934	GGACATTTT	0,0515	0,03075
EPHB2	PEA3 [T00685]	113	121	3,710864	AGGATGACA	0,06866	0,03175
EPHB2	c-Myb [T00137]	300	307	2,570796	AAACTGCC	0,04578	0,0336

## Resultados

Supplementary Table S3. (Continuation)

Sequence		End		Dissimilarity	String	RE equally	RE query
name	Factor name	Start position	position				
EPHB2	PR B [T00696]	1185	1191	3,29756	AACATTT	0,18311	0,03446
EPHB2	PR A [T01661]	1185	1191	3,29756	AACATTT	0,18311	0,03446
EPHB2	IRF-1 [T00423]	29	37	4,968836	AGGGGGAAA	0,05722	0,03613
EPHB2	c-Ets-2 [T00113]	651	659	4,589988	TTCCTTCAC	0,0515	0,03675
EPHB2	GATA-2 [T00308]	1206	1214	1,111111	GCTTTATCT	0,06866	0,04265
EPHB2	HNF-1A [T00368]	481	488	0,287765	GTTAATGA	0,18311	0,0439
EREG	PITX2 [T02413]	112	131	1,535113	GCTGGGATTACAGGCG	0	0
EREG	Ik-1 [T02702]	106	118	2,374299	CAAAGTGCTGGGA	0,00047	0,0004
EREG	NF-kappaB1 [T00593]	1138	1148	3,787206	GGGGAACGCCA	0,00823	0,0067
EREG	USF2 [T00878]	1217	1226	0,524236	CAGGTGACCC	0,00858	0,00698
EREG	NF-AT2 [T01945]	726	735	3,814941	GGAAATAGTG	0,00858	0,01066
EREG	NF-AT2 [T01945]	1280	1289	4,979362	GGAAAAGTTT	0,01144	0,01297
EREG	NF-AT1 [T01948]	687	696	1,378139	GAGTTTCCA	0,01431	0,0172
EREG	NF-AT1 [T00550]	676	684	3,098758	GGAAAGGCT	0,02289	0,02301
EREG	NF-AT1 [T00550]	1008	1016	3,098758	GGAAAGGCT	0,02289	0,02301
EREG	NF-AT1 [T00550]	726	734	3,384125	GGAAATAGT	0,02289	0,02882
EREG	Elk-1 [T00250]	990	998	0,957025	AGAAGGAAG	0,02289	0,03159
EREG	c-Fos [T00123]	933	942	4,563121	ACAATGACTC	0,03433	0,03316
EREG	Sp1 [T00759]	1021	1030	3,79151	TGGCGGTGG	0,06723	0,04488
EREG	Elk-1 [T00250]	853	861	1,779702	AGTAGGAAG	0,04005	0,04493
FPR1	PITX2 [T02413]	718	737	1,535113	CTCATGCCTGTAATCCC	0	0
FPR1	HMG I(Y) [T02368]	1418	1430	2,773942	AGACTTCCTATTT	0,00058	0,00103
FPR1	SRF [T00764]	299	311	3,318604	CCATATTAGGGAT	0,00121	0,00142
FPR1	Elk-1 [T00250]	157	165	4,892803	CTTCCTAAC	0,00572	0,00605
FPR1	PPAR-alpha:RXR-alpha	621	631	4,88658	GTGGCCAGGG	0,01287	0,00822
FPR1	USF2 [T00878]	135	144	1,048473	CAGTCACCTG	0,01287	0,00959
FPR1	AR [T00040]	1342	1350	2,159336	CCACTGTCC	0,02861	0,02792
FPR1	NF-AT2 [T01945]	350	359	3,689016	GATATTTCC	0,02575	0,03755
FPR1	USF2 [T00878]	124	133	4,528187	TGCTCACCTG	0,0515	0,03905
FPR1	c-Myb [T00137]	1083	1090	2,687937	CAACTTCC	0,04578	0,04293
FPR1	AR [T00040]	1062	1070	4,241082	GCGCTGTCC	0,0515	0,04525
FPR1	AR [T00040]	1231	1239	3,814754	GGACACTTG	0,0515	0,04525
FPR1	T3R-beta1 [T00851]	1176	1184	1,110682	TCACCACAG	0,05722	0,04541
GBP3	NF-AT2 [T01945]	1062	1071	2,029228	GGAAATACTT	0,01144	0,02607
GBP3	Elk-1 [T00250]	591	599	3,381796	TTCAGGAAG	0,03433	0,02859
GBP3	POU2F2 (Oct-2,1) [T00	1088	1098	2,61506	AGTGTAATTCA	0,00751	0,03207
GBP3	RAR-beta [T00721]	1446	1455	2,16302	TTAAACCCT	0,05722	0,04101
GLG1	Sp1 [T00759]	1416	1425	0	GGGCGGGGC	0,00143	0,00129
GLG1	Smad3:Smad4 [T05271	319	332	4,215502	AGCTGTCTGTTTAC	0,00181	0,0018
GLG1	PEA3 [T00685]	775	783	3,113548	TTTCATCCT	0,01144	0,01192
GLG1	NF-AT1 [T00550]	1102	1110	3,098758	AGGCTTTCC	0,02289	0,02289
GLG1	Elk-1 [T00250]	1066	1074	0,822677	GGAAGGAAG	0,02289	0,02401
GLG1	Elk-1 [T00250]	1070	1078	0,822677	GGAAGGAAG	0,02289	0,02401
GLG1	Elk-1 [T00250]	1070	1078	0,822677	GGAAGGAAG	0,02289	0,02401
GLG1	Elk-1 [T00250]	1074	1082	0,822677	GGAAGGAAG	0,02289	0,02401
GLG1	Sp1 [T00759]	1331	1340	2,418403	CTCCGCCCG	0,03004	0,0281
GLG1	HNF-1B [T01950]	844	852	4,435774	TTATTAACA	0,04005	0,04136

## Resultados

Supplementary Table S3. (Continuation)

Sequence		End		Dissimilarity	String	RE equally	RE query
name	Factor name	Start position	position				
GLG1	IRF-1 [T00423]	1106	1114	4,462268	TTTCCGATT	0,04005	0,04237
GLG1	E2F-1 [T01542]	1330	1337	1,490375	TCTCCCGC	0,04578	0,04437
GLG1	c-Myb [T00137]	390	397	2,570796	AAACTGCC	0,04578	0,04539
GLG1	PXR-1:RXR-alpha [T056	1136	1143	1,63615	AAAGTTCA	0,04578	0,04703
GLG1	c-Ets-2 [T00113]	505	513	1,071163	TTCCTCTTG	0,04578	0,04846
HAR1A	EBF [T05427]	166	176	1,581175	GCCCCAGGGGC	0,00286	0,01657
HAR1A	MAZ [T00490]	334	346	3,162267	GACGGGGAGGGAG	0,00355	0,01681
HAR1A	NF-AT1 [T01948]	237	246	2,756277	TGGAAACTGC	0,04435	0,01937
HAR1A	c-Myb [T00137]	240	247	2,570796	AAACTGCC	0,04578	0,03251
HAR1A	PEA3 [T00685]	1151	1159	1,194633	GCACATCCT	0,0515	0,03325
HAR1A	TFIID [T00820]	604	610	3,075094	TTTTGCA	0,09155	0,03439
HAR1A	AR [T00040]	1269	1277	2,267638	TGAGTGTC	0,08583	0,04678
HAR1A	RAR-beta [T00721]	1484	1493	2,16302	AGGGTTCGAA	0,05722	0,04866
HBA1	CTF [T00174]	1426	1437	0	CAGCCAATGAGC	0,00072	0,00057
HBA1	Egr-1 [T00241]	1360	1375	4,92082	GCCCGGCCCGCGCAG	0,00013	0,00271
HBA1	SRY [T00997]	614	622	1,998343	CTTTGTTTA	0,02289	0,00675
HBA1	IRF-1 [T00423]	910	918	1,616539	AGAGGGAAA	0,03433	0,01179
HBA1	EBF [T05427]	1336	1346	0	CCCCAGGGGA	0,00286	0,01898
HBA1	Elk-1 [T00250]	276	284	0,822677	CTTCCTCC	0,02289	0,01919
HBA1	Elk-1 [T00250]	503	511	3,381796	CTTCCTCAG	0,03433	0,02305
HBA1	c-Fos [T00123]	719	728	3,579011	GAGTCAGTCA	0,04005	0,02529
HBA1	IRF-1 [T00423]	1168	1176	3,692688	GAAAGGAAA	0,0515	0,02602
HBA1	NF-AT1 [T00550]	1172	1180	4,056854	GAAAAGGGT	0,05722	0,02647
HBA1	T3R-beta1 [T00851]	437	445	2,259951	TCACCTCCT	0,02289	0,03682
HBA1	c-Ets-2 [T00113]	277	285	4,589988	TTCCTTCCT	0,0515	0,03774
HBA1	c-Ets-2 [T00113]	865	873	4,589988	TTCCTTCCC	0,0515	0,03774
HBA1	Elk-1 [T00250]	280	288	3,247448	CTTCCTCAC	0,05722	0,03882
HBA1	HNF-1A [T00368]	625	632	0,431647	TGTTTAAC	0,18311	0,0395
HBA1	c-Ets-2 [T00113]	1167	1175	4,017001	CGAAAGGAA	0,12016	0,04332
HBA1	NF-kappaB1 [T00593]	1384	1394	3,297969	CGGGACTCCCC	0,01359	0,04357
HBA1	LEF-1 [T02905]	614	621	2,004405	CTTTGTTT	0,13733	0,04511
HBA1	c-Jun [T00133]	722	728	4,441904	TCAGTCA	0,09155	0,04925
HBA2	PITX2 [T02413]	142	161	1,535113	GCTGGGATTACAGGCG	0	0
HBA2	PITX2 [T02413]	441	460	1,535113	GCTGGGATTACAGGCG	0	0
HBA2	CTF [T00174]	1455	1466	0	CAGCCAATGAGC	0,00072	0,0006
HBA2	Ik-1 [T02702]	435	447	2,374299	CAAAGTGCTGGGA	0,00047	0,00103
HBA2	Egr-1 [T00241]	1389	1404	4,92082	GCCCGGCCCGCGCAG	0,00013	0,00216
HBA2	MAZ [T00490]	529	541	0,797374	CACCCTCCCCTC	0,00045	0,00278
HBA2	Ik-1 [T02702]	136	148	4,748597	CAATGTGCTGGGA	0,00235	0,00484
HBA2	Ik-1 [T02702]	300	312	4,748597	CGAGTAGCTGGGA	0,00235	0,00484
HBA2	HNF-3alpha [T02512]	53	60	0	TATTTTTA	0,06866	0,00839
HBA2	HNF-3alpha [T02512]	349	356	0	TATTTTTA	0,06866	0,00839
HBA2	IRF-1 [T00423]	939	947	1,616539	AGAGGGAAA	0,03433	0,01397
HBA2	EBF [T05427]	1365	1375	0	CCCCAGGGGA	0,00286	0,01713
HBA2	TCF-4 [T02918]	270	279	0,76243	GGTCAAAGG	0,04292	0,01834
HBA2	Elk-1 [T00250]	555	563	0,822677	CTTCCTCC	0,02289	0,02103
HBA2	HNF-3alpha [T02512]	186	193	3,500065	TATTTTTT	0,20599	0,02514

## Resultados

Supplementary Table S3. (Continuation)

Sequence name	Factor name	Start position	End position	Dissimilarity	String	RE equally	RE query
HBA2	c-Fos [T00123]	748	757	3,579011	GAGTCAGTCA	0,04005	0,02692
HBA2	NF-AT1 [T00550]	1201	1209	4,056854	GGAAAGGGT	0,05722	0,02829
HBA2	IRF-1 [T00423]	1197	1205	3,692688	GAAAGGAAA	0,0515	0,02941
HBA2	NF-kappaB1 [T00593]	1413	1423	3,297969	CGGGACTCCCC	0,01359	0,04154
HBA2	c-Ets-2 [T00113]	556	564	4,589988	TTCCTTCTT	0,0515	0,0416
HBA2	c-Ets-2 [T00113]	894	902	4,589988	TTCCTTCCC	0,0515	0,0416
HBA2	Elk-1 [T00250]	559	567	3,247448	CTTCCTCAC	0,05722	0,04262
HBA2	c-Ets-2 [T00113]	1196	1204	4,017001	CGAAAGGAA	0,12016	0,04999
HELB	NF-YA [T01804]	1485	1498	0,799706	AACTGATTGGCTGA	0,00015	0,00013
HELB	Ik-1 [T02702]	1243	1255	2,374299	TCCCAGCACCTTC	0,00047	0,00031
HELB	Ik-1 [T02702]	1310	1322	2,374299	TCCCAGCCCCTCG	0,00047	0,00031
HELB	Sp1 [T00759]	1352	1361	0	GCCCCGCCCC	0,00143	0,00049
HELB	NF-E2 [T00558]	77	90	4,167221	GGCTGAGTCATTAG	0,00064	0,00067
HELB	Sp1 [T00759]	1331	1340	0,574521	TCCCCGCCCC	0,00572	0,00217
HELB	MAZ [T00490]	1369	1381	3,973255	TGCCCTCCCCTCG	0,0044	0,0024
HELB	Sp1 [T00759]	1298	1307	0,949391	GCCCCGCCCA	0,01001	0,00395
HELB	CTF [T00174]	1486	1497	2,746279	ACTGATTGGCTG	0,00465	0,00452
HELB	Sp1 [T00759]	1399	1408	2,154584	GGGGCGGAGC	0,02718	0,01245
HELB	c-Fos [T00123]	81	90	1,964632	GAGTCATTAG	0,01144	0,01347
HELB	c-Myb [T00137]	44	51	0	GGCAGTTG	0,02289	0,01785
HELB	IRF-1 [T00423]	825	833	2,418514	TTTCCTTC	0,01717	0,01961
HELB	AP-1 [T00029]	78	86	0	GCTGAGTCA	0,02289	0,02193
HELB	NF-AT1 [T00550]	454	462	3,075022	GGAAACACT	0,02289	0,02195
HELB	AR [T00040]	1010	1018	4,995624	GGACAGCCC	0,04005	0,02672
HELB	AR [T00040]	354	362	1,86979	GGACAGTAA	0,02861	0,0282
HELB	TCF-4 [T02918]	418	427	4,412902	ATGTCAAAGG	0,03147	0,03133
HELB	Elk-1 [T00250]	1252	1260	3,381796	CTTCCTGAT	0,03433	0,03408
HELB	Sp1 [T00759]	1263	1272	4,029243	TCACCGCCCA	0,06723	0,03531
HELB	NF-AT2 [T01945]	454	463	4,342522	GGAAACACTG	0,03147	0,03891
HELB	TCF-4 [T02918]	957	966	1,623895	ACTTTGAATG	0,03433	0,0418
HELB	HOXD9 [T01424]	159	168	4,080895	AATAACAGTA	0,02289	0,04278
HELB	HOXD10 [T01425]	159	168	4,080895	AATAACAGTA	0,02289	0,04278
HELB	c-Myb [T00137]	1459	1466	2,687937	GGAAGTTG	0,04578	0,04388
IFI27	RAR-beta [T00721]	1063	1072	1,08151	AGGGTTTGAT	0,02289	0,02255
IFI27	TCF-4 [T02918]	1484	1493	2,632973	GTTTCAAAGA	0,02575	0,02898
IFI27	IRF-1 [T00423]	1027	1035	1,529008	AAATGGAAA	0,03433	0,03774
IFI27	c-Fos [T00123]	203	212	3,154982	GAGCTGACTC	0,04578	0,04398
IFI27	Sp1 [T00759]	374	383	3,54287	TCTCCGCCCT	0,05579	0,04541
IFI27	TCF-4 [T02918]	438	447	0,76243	CCTTTGAATC	0,04292	0,04552
IFI27	USF2 [T00878]	34	43	4,528187	ACGGCACCTG	0,0515	0,04783
IFI27	NF-AT1 [T01948]	1030	1039	2,756277	TGAAAGCTC	0,04435	0,04791
IFI27	PXR-1:RXR-alpha [T056	999	1006	1,759733	AGAGTTCA	0,04578	0,04851
IFI6	EBF [T05427]	1194	1204	1,581175	GCCCCAGGGGG	0,00286	0,00174
IFI6	CTF [T00174]	664	675	2,895056	TAGCCAATCTGC	0,00429	0,00388
IFI6	EBF [T05427]	205	215	3,135622	CTCCAGGGGG	0,00572	0,00389
IFI6	PPAR-alpha:RXR-alpha	1192	1202	3,872523	GTGCCCCAGGG	0,01931	0,01546
IFI6	Sp1 [T00759]	1348	1357	3,103976	TGGGCGGAGC	0,02575	0,01741

## Resultados

Supplementary Table S3. (Continuation)  
Sequence

name	Factor name	Start position	End position	Dissimilarity	String	RE equally	RE query
IFI6	SRY [T00997]	1005	1013	0	GGAACAAAG	0,02289	0,03043
IFI6	TCF-4 [T02918]	333	342	4,639022	TGGTCAAAGG	0,03147	0,03223
IFI6	TCF-4 [T02918]	1265	1274	1,461318	GCTTTGAATA	0,03433	0,03774
IFI6	PXR-1:RXR-alpha [T056	403	410	1,759733	TGAACTCT	0,04578	0,04914
IFITM3	PITX2 [T02413]	195	214	0	CACACACCTGTAATCCC	0	0
IFITM3	PITX2 [T02413]	1267	1286	1,535113	GCTGGGATTACAGGAG	0	0
IFITM3	Ik-1 [T02702]	208	220	2,374299	TCCCAGCACTTTG	0,00047	0,00038
IFITM3	RBP-Jkappa [T01616]	628	639	4,988997	GATTCCTCAAGAT	0,00501	0,00513
IFITM3	USF2 [T00878]	237	246	4,003951	AGATCACCTG	0,00858	0,00747
IFITM3	POU2F2 (Oct-2,1) [T00	850	860	3,116744	GCTCTAATTCA	0,01931	0,02516
IFITM3	Elk-1 [T00250]	879	887	2,164966	GAGAGGAAAG	0,04005	0,03928
IFITM3	c-Myb [T00137]	1460	1467	2,152744	GAAGTGGC	0,04578	0,04227
IGHA1	STAT1beta [T01573]	874	883	4,01053	TGAAGGAAAT	0,02575	0,01372
IGHA1	c-Ets-2 [T00113]	873	881	4,589988	CTGAAGGAA	0,0515	0,03522
IGHA1	GATA-2 [T00308]	1308	1316	1,111111	ACCCTATCT	0,06866	0,04334
IGHA2	HOXD9 [T01424]	1252	1261	2,949288	AATATACTT	0,01717	0,00155
IGHA2	HOXD10 [T01425]	1252	1261	2,949288	AATATACTT	0,01717	0,00155
IGHA2	GATA-3 [T00311]	1286	1297	4,460245	GGCCATCTCTAA	0,00429	0,00237
IGHA2	GATA-2 [T00308]	1287	1295	0	GGCCATCT	0,01144	0,00761
IGHA2	STAT1beta [T01573]	853	862	4,01053	TGAAGGAAAT	0,02575	0,01262
IGHA2	EBF [T05427]	609	619	0,984797	TGCCCTGGGGT	0,00572	0,01746
IGHA2	AP-1 [T00029]	750	758	4,606866	TGACTCCAC	0,02289	0,02883
IGHA2	c-Ets-2 [T00113]	852	860	4,589988	CTGAAGGAA	0,0515	0,03215
IGHA2	PPAR-alpha:RXR-alpha	1274	1284	4,727619	GCCTGGGTCAG	0,02432	0,04061
IGHA2	PXR-1:RXR-alpha [T056	141	148	0,941658	TGAACTGG	0,09155	0,04944
IGLC2	FOXO4 [T03403]	228	241	3,616253	TGTTTTTGTGTGT	0,00074	0,00016
IGLC2	Smad3 [T04096]	410	423	1,563659	GTCTGTCTGTCTGT	0,00089	0,00062
IGLC2	Smad3 [T04096]	414	427	1,563659	GTCTGTCTGTCTGT	0,00089	0,00062
IGLC2	Smad3 [T04096]	418	431	1,563659	GTCTGTCTGTCTGT	0,00089	0,00062
IGLC2	Smad3 [T04096]	422	435	2,814586	GTCTGTCTGTCTCT	0,00107	0,00078
IGLC2	RAR-beta:RXR-alpha [T	905	916	2,492665	TCCCCTGTGCC	0,00161	0,00324
IGLC2	Ik-1 [T02702]	669	681	4,748597	CACAGAGCTGGGA	0,00235	0,00374
IGLC2	NF-kappaB1 [T00593]	1442	1452	3,343073	AGGGGCTCCCC	0,01359	0,03043
IGLC2	c-Ets-2 [T00113]	64	72	1,64415	TTCCTCCTT	0,03433	0,03207
IGLC2	NF-AT1 [T01948]	1349	1358	3,445347	TGAAAACAC	0,05722	0,03562
IGLC2	AR [T00040]	215	223	3,382886	GGACAATT	0,0515	0,03823
IGLC2	EBF [T05427]	729	739	4,04219	TGCCCTGGGCG	0,0186	0,04344
IGLC2	PEA3 [T00685]	861	869	3,710864	TGGATGTGA	0,06866	0,04464
IGLC2	GR [T05076]	115	121	1,444018	GTTTTTG	0,09155	0,0477
IGLC2	GR [T05076]	229	235	1,444018	GTTTTTG	0,09155	0,0477
ITPR3	FOXO4 [T03403]	1245	1258	1,446501	AACAACAAACAAA	0,00031	0,00035
ITPR3	FOXO4 [T03403]	1249	1262	1,446501	AACAACAAACAAA	0,00031	0,00035
ITPR3	FOXO4 [T03403]	1253	1266	1,446501	AACAACAAACAAA	0,00031	0,00035
ITPR3	FOXO4 [T03403]	1257	1270	1,446501	AACAACAAACAAA	0,00031	0,00035
ITPR3	FOXO4 [T03403]	1241	1254	4,885477	GTAAAACAAACAAA	0,00063	0,00068
ITPR3	EBF [T05427]	718	728	4,612817	GTCCCAGGGGC	0,00429	0,00383
ITPR3	MAZ [T00490]	89	101	3,175881	GCAGGGGAGGGAA	0,00355	0,00388



## Resultados

Supplementary Table S3. (Continuation)

Sequence		End		Dissimilarity	String	RE equally	RE query
name	Factor name	Start position	position				
ITPR3	MAZ [T00490]	319	331	3,175881	GATGGGGAGGGGT	0,00355	0,00388
ITPR3	POU2F2 (Oct-2,1) [T00	1173	1183	2,61506	ACTGTAAAACA	0,00751	0,00823
ITPR3	USF2 [T00878]	242	251	1,048473	CAGGTGACAG	0,01287	0,01239
ITPR3	PPAR-alpha:RXR-alpha	716	726	3,872523	CTGTCCCAGGG	0,01931	0,01812
ITPR3	IRF-1 [T00423]	547	555	2,418514	GAAGGGAAA	0,01717	0,01836
ITPR3	IRF-1 [T00423]	658	666	2,418514	TTCCCTTC	0,01717	0,01836
ITPR3	LEF-1 [T02905]	133	140	0,85582	CTTTGGTC	0,02289	0,02263
ITPR3	LEF-1 [T02905]	889	896	0,85582	CTTTGGTC	0,02289	0,02263
ITPR3	TBP [T00794]	201	210	0	GGTATATAAA	0,02289	0,02543
ITPR3	Elk-1 [T00250]	1439	1447	3,381796	CTCAGGAAG	0,03433	0,0352
ITPR3	TCF-4 [T02918]	1048	1057	0,98855	GCAATAAAGG	0,04292	0,04402
ITPR3	c-Fos [T00123]	690	699	3,154982	GGCCTGACTC	0,04578	0,04508
ITPR3	Elk-1 [T00250]	81	89	0,134348	AGCAGGAAG	0,04578	0,04701
ITPR3	Elk-1 [T00250]	443	451	0	GGGAGGAAG	0,04578	0,04701
ITPR3	Elk-1 [T00250]	1376	1384	0,134348	CTTCTGCT	0,04578	0,04701
ITPR3	SRY [T00997]	140	148	0,999172	CTTTGTTGT	0,04578	0,04761
KIF19	Ik-1 [T02702]	524	536	2,374299	TCCCAGCCCCTTC	0,00047	0,00087
KIF19	Sp3 [T02338]	1433	1448	4,668232	CGGGGGCGGGGGCGGC	0,00011	0,00135
KIF19	MAZ [T00490]	1456	1468	4,264019	GGCCCTCCCCTG	0,00141	0,00464
KIF19	HNF-1C [T01951]	388	396	3,273679	ATTAATAAC	0,03433	0,00882
KIF19	Sp1 [T00759]	1435	1444	0	GGGGCGGGGC	0,00143	0,0112
KIF19	EBF [T05427]	438	448	4,508838	GTCCCAGGGCA	0,00429	0,0139
KIF19	EBF [T05427]	498	508	0,984797	AGCCCTGGGGA	0,00572	0,01729
KIF19	NF-AT1 [T00550]	741	749	1,970716	AAGTTTTCC	0,0515	0,01906
KIF19	c-Fos [T00123]	973	982	2,598489	GAGTCAGGCT	0,02289	0,02589
KIF19	LEF-1 [T02905]	1025	1032	0,641865	GAGCAAAG	0,04578	0,0324
KIF19	NF-AT1 [T01948]	539	548	4,134416	TGAAAGTGA	0,06294	0,03575
KIF19	NF-AT1 [T00550]	770	778	2,619709	AACCTTCC	0,06866	0,03588
KIF19	HNF-3alpha [T02512]	11	18	3,500065	ACAAAATA	0,20599	0,03742
KIF19	ETF [T00270]	1429	1439	2,623453	GCGCCGGGGGC	0,00536	0,0383
KIF19	GR [T05076]	200	206	1,444018	CAAAAAC	0,09155	0,0394
KIF19	AP-1 [T00029]	970	978	0,348957	ACTGAGTCA	0,06866	0,04145
KIF19	Elk-1 [T00250]	517	525	2,987643	CTTCTTTC	0,05722	0,04475
KIF19	Elk-1 [T00250]	561	569	2,164966	CTTCTCTC	0,04005	0,04485
KIF19	GCF [T00320]	1265	1273	3,409768	GCGCAGGAG	0,02289	0,0467
KIF19	PR B [T00696]	21	27	3,29756	AAATGTT	0,18311	0,04774
KIF19	PR A [T01661]	21	27	3,29756	AAATGTT	0,18311	0,04774
MED20	STAT1beta [T01573]	345	354	1,112096	TCATGGAAAT	0,00858	0,00717
MED20	Sp1 [T00759]	715	724	0,574521	ACCCCGCCCC	0,00572	0,01183
MED20	SRY [T00997]	667	675	1,998343	CTTTGTTGA	0,02289	0,01805
MED20	E2F-1 [T01542]	911	918	0,993583	TGTCCCGC	0,02289	0,03076
MED20	USF1 [T00874]	1067	1076	1,858812	CACGTGGGGA	0,02718	0,03289
MED20	c-Ets-2 [T00113]	833	841	1,64415	TTCCTCCTG	0,03433	0,03718
MED20	c-Fos [T00123]	669	678	3,637699	TTGTTGACTC	0,04005	0,03802
MED20	Sp1 [T00759]	940	949	1,28618	CCCCGCCCCG	0,02432	0,04394
MED20	Sp1 [T00759]	1358	1367	1,706745	GGGGCGGGAA	0,02432	0,04394
MED20	NF-AT1 [T00550]	330	338	4,566689	GGAAACGAT	0,0515	0,04417

## Resultados

Supplementary Table S3. (Continuation)  
Sequence

name	Factor name	Start position	End position	Dissimilarity	String	RE equally	RE query
MED20	USF1 [T00874]	1063	1072	3,348091	ATCGCACGTG	0,03862	0,04433
MED20	Sp1 [T00759]	1181	1190	2,154584	GCGCCGCCCC	0,02718	0,04657
MED20	IRF-1 [T00423]	326	334	3,689552	CCAGGGAAA	0,0515	0,04756
MED20	NF-AT1 [T01948]	348	357	3,445347	TGGAAATATA	0,05722	0,04872
MEG3	NF-AT1 [T00550]	73	81	1,437145	ACTTTTTCC	0,00572	0,00425
MEG3	CTF [T00174]	280	291	2,746279	GAGCCAATGAGT	0,00465	0,00456
MEG3	USF1 [T00874]	1055	1064	0,477028	AACCCACGTG	0,01001	0,01288
MEG3	NF-Y [T00150]	1413	1420	0	ATTGTCA	0,02289	0,01865
MEG3	E2F [T00221]	394	403	0,549136	TGGCGCCAAA	0,01717	0,01918
MEG3	STAT1beta [T01573]	1328	1337	4,01053	TGAGGGAAAT	0,02575	0,02227
MEG3	c-Fos [T00123]	1400	1409	2,521126	TTAATGACTC	0,02289	0,02459
MEG3	NF-AT2 [T01945]	72	81	4,586038	TACTTTTTCC	0,03147	0,02465
MEG3	PPAR-alpha:RXR-alpha	1477	1487	3,872523	AGCTGGGGCAG	0,01931	0,02469
MEG3	AP-1 [T00029]	1404	1412	4,606866	TGACTCCAC	0,02289	0,02638
MEG3	TCF-4 [T02918]	552	561	1,687438	GCTTTGAAGC	0,03433	0,02722
MEG3	NF-kappaB1 [T00593]	1253	1263	4,302276	TGGTGTTCCCC	0,02432	0,03314
MEG3	USF1 [T00874]	1059	1068	2,121976	CACGTGGATG	0,02718	0,03373
MEG3	c-Ets-2 [T00113]	969	977	1,64415	TTCTCCTT	0,03433	0,03435
MEG3	NF-AT1 [T01948]	73	82	2,756277	ACTTTTTCCA	0,04435	0,03519
MEG3	PXR-1:RXR-alpha [T056	569	576	1,759733	AGAGTTCA	0,04578	0,03722
MEG3	Sp1 [T00759]	84	93	1,469012	CTCCCGCCCC	0,02432	0,04685
MEG3	c-Fos [T00123]	212	221	3,081206	CTCTTGACTC	0,04578	0,04829
NIPAL2	PITX2 [T02413]	58	77	1,535113	CTCACACCTGTAATCCC	0	0
NIPAL2	Ik-1 [T02702]	71	83	2,374299	TCCCAGCACTTTG	0,00047	0,00029
NIPAL2	PPAR-alpha:RXR-alpha	1471	1481	4,88658	CTGCCCAGCG	0,01287	0,00851
NIPAL2	PEA3 [T00685]	995	1003	3,113548	TTTCATCCT	0,01144	0,01514
NIPAL2	NF-kappaB1 [T00593]	1428	1438	4,414876	GGGGATTGCC	0,02432	0,0152
NIPAL2	GCF [T00320]	1401	1409	0	GCGCGGGGA	0,06866	0,02907
NIPAL2	c-Fos [T00123]	418	427	3,081206	GAGTCAAGAT	0,04578	0,03907
NIPAL2	USF1 [T00874]	193	202	3,775819	TGTGCACGTG	0,06008	0,04484
NDRG4	PITX2 [T02413]	358	377	1,535113	GCTGGGATTACAGGCA	0	0
NDRG4	HOXD9 [T01424]	93	102	2,085829	CACTTTTATT	0,00286	0,00291
NDRG4	HOXD10 [T01425]	93	102	2,085829	CACTTTTATT	0,00286	0,00291
NDRG4	EBF [T05427]	46	56	0,984797	AGCCCTGGGGA	0,00572	0,00544
NDRG4	POU2F2 (Oct-2,1) [T00	1088	1098	0,501684	TGATTTATACA	0,00966	0,00987
NDRG4	Sp1 [T00759]	1416	1425	0,949391	GCCCCGCCCA	0,01001	0,01058
NDRG4	HOXD9 [T01424]	691	700	2,949288	AATAACAATT	0,01717	0,01789
NDRG4	HOXD10 [T01425]	691	700	2,949288	AATAACAATT	0,01717	0,01789
NDRG4	c-Fos [T00123]	426	435	2,598489	AACCTGACTC	0,02289	0,02279
NDRG4	TBP [T00794]	1091	1100	0	TTTATACACT	0,02289	0,02342
NDRG4	SRY [T00997]	30	38	1,998343	CTTTGTTG	0,02289	0,02355
NDRG4	ATF-2 [T00167]	850	859	4,623667	ATGACGTCTC	0,02861	0,0283
NDRG4	Elk-1 [T00250]	1190	1198	3,381796	CTTCTCAG	0,03433	0,03451
NDRG4	c-Ets-2 [T00113]	1027	1035	1,64415	AAGGAGGAA	0,03433	0,0352
NDRG4	AR [T00040]	1128	1136	4,890444	GGACAAAGC	0,04005	0,03857
NDRG4	CREB [T00163]	851	859	1,217272	TGACGTCTC	0,04005	0,03897
NDRG4	Elk-1 [T00250]	893	901	2,164966	GCCAGGAAG	0,04005	0,04045

## Resultados

Supplementary Table S3. (Continuation)

Sequence name	Factor name	Start position	End position	Dissimilarity	String	RE equally	RE query
NDRG4	PXR-1:RXR-alpha [T056	459	466	1,63615	TGAACTAT	0,04578	0,04596
NPDC1	PITX2 [T02413]	609	628	1,535113	CTCACGCCTGTAATCCC	0	0
NPDC1	PITX2 [T02413]	744	763	3,070225	CACGCGCCTGTAATCCC	0	0
NPDC1	Ik-1 [T02702]	622	634	2,374299	TCCCAGCACTTTG	0,00047	0,00086
NPDC1	MAZ [T00490]	363	375	2,378507	TTCCCTCCCCTCT	0,00304	0,00974
NPDC1	MAZ [T00490]	1427	1439	3,986869	CAGGGGAGGGGG	0,0044	0,01327
NPDC1	USF2 [T00878]	921	930	1,048473	CTGTCACCTG	0,01287	0,01644
NPDC1	TCF-4 [T02918]	502	511	4,639022	TAGTCAAAGG	0,03147	0,02327
NPDC1	NF-kappaB1 [T00593]	457	467	2,871556	GGGGACACCCC	0,01252	0,02978
NPDC1	E2F-1 [T01542]	1082	1089	0	TTCCCCGC	0,02289	0,02989
NPDC1	AR [T00040]	896	904	2,808697	CCATTGTCC	0,04578	0,04995
NR4A1	PITX2 [T02413]	668	687	1,535113	CTCACGCCTGTAATCCC	0	0
NR4A1	PITX2 [T02413]	1221	1240	1,535113	GCTGGGATTACAGGTG	0	0
NR4A1	Ik-1 [T02702]	681	693	2,374299	TCCCAGCACTTTG	0,00047	0,00042
NR4A1	Ik-1 [T02702]	1215	1227	2,374299	CAAAGTGCTGGGA	0,00047	0,00042
NR4A1	POU2F1 [T00641]	266	276	0	AAAAATGCAAAT	0,00107	0,00133
NR4A1	MAZ [T00490]	591	603	3,189496	TCCCTCCCCTT	0,00355	0,00303
NR4A1	EBF [T05427]	1430	1440	4,334406	CCCCCTGAGGC	0,00429	0,00348
NR4A1	AR [T00040]	885	893	0,649362	GGCACTGC	0,01144	0,01066
NR4A1	GATA-2 [T00308]	550	558	0	AGATAGGGC	0,01144	0,01137
NR4A1	HOXD9 [T01424]	86	95	0	TATTTTTATT	0,00858	0,01154
NR4A1	HOXD10 [T01425]	86	95	0	TATTTTTATT	0,00858	0,01154
NR4A1	NF-kappaB1 [T00593]	127	137	4,641294	GGGGACCTCCC	0,02432	0,02137
NR4A1	c-Fos [T00123]	328	337	4,061728	TTAGTGACTC	0,02861	0,02904
NR4A1	HOXD9 [T01424]	62	71	4,321431	TGTTTTTATT	0,02575	0,03327
NR4A1	HOXD10 [T01425]	62	71	4,321431	TGTTTTTATT	0,02575	0,03327
NTSR1	TCF-4 [T02918]	689	698	2,386325	ACTTTGATGA	0,02575	0,00338
NTSR1	Sp3 [T02338]	1269	1284	4,841327	CCCGGGCGGGCGGGC	0,0001	0,00375
NTSR1	Egr-1 [T00241]	1270	1285	4,909847	CCGGGCGGGCGGGG	0,00013	0,00424
NTSR1	Ahr [T01795]	776	786	2,82917	CTCACGCCATC	0,00429	0,00512
NTSR1	MAZ [T00490]	628	640	1,581133	AGTGGGGAGGGTG	0,00107	0,00601
NTSR1	MAZ [T00490]	244	256	3,439416	GGCGGGAGGGCC	0,00125	0,00957
NTSR1	IRF-1 [T00423]	365	373	3,145547	ACACGGAAA	0,05722	0,011
NTSR1	LEF-1 [T02905]	690	697	1,703176	CTTTGATG	0,06866	0,0115
NTSR1	MAZ [T00490]	233	245	3,973255	GCAGGGGAGGGGG	0,0044	0,01986
NTSR1	NF-AT1 [T01948]	820	829	4,823485	ACCATTTCCA	0,05722	0,02004
NTSR1	ETF [T00270]	306	316	3,630042	GAGGGGGGGGC	0,00179	0,02033
NTSR1	RAR-beta:RXR-alpha [T	806	817	4,98533	TGCCCTGAGCCC	0,00724	0,02281
NTSR1	Elk-1 [T00250]	1106	1114	3,81032	CTTCCTAGC	0,06866	0,02467
NTSR1	EBF [T05427]	207	217	3,031642	CTCCAGGGGT	0,00572	0,02787
NTSR1	EBF [T05427]	274	284	0	TGCCCTGGGGG	0,00286	0,03036
NTSR1	AR [T00040]	430	438	2,267638	GGACTCA	0,08583	0,03282
NTSR1	PPAR-alpha:RXR-alpha	787	797	3,872523	CTGCCCCAGGT	0,01931	0,04061
NTSR1	AR [T00040]	265	273	3,165392	GGACAGTAG	0,04578	0,0425
PLAUR	PITX2 [T02413]	93	112	1,535113	CTCATGCCTGTAATCCC	0	0
PLAUR	Ik-1 [T02702]	106	118	2,374299	TCCCAGCACTTTG	0,00047	0,00023
PLAUR	FOXO4 [T03403]	989	1002	1,446501	AACAACAACAACA	0,00031	0,00114

## Resultados

**Supplementary Table S3. (Continuation)**  
Sequence

name	Factor name	Start position	End position	Dissimilarity	String	RE equally	RE query
PLAUR	FOXO4 [T03403]	993	1006	1,446501	AACAAACAACAAA	0,00031	0,00114
PLAUR	AhR:Arnt [T05394]	223	232	0	GCTGGCGTGC	0,00429	0,00117
PLAUR	NF-AT1 [T00550]	1484	1492	3,407861	ACATTTTCC	0,02289	0,02814
PLAUR	USF2 [T00878]	135	144	4,528187	GGATCACCTG	0,0515	0,03483
PLAUR	HOXD9 [T01424]	693	702	0	TAATTTTATT	0,00858	0,03493
PLAUR	HOXD10 [T01425]	693	702	0	TAATTTTATT	0,00858	0,03493
PLAUR	p53 [T00671]	228	234	1,270236	CGTGCCC	0,09155	0,03687
PLAUR	NF-AT1 [T01948]	1484	1493	2,067208	ACATTTTCCA	0,02861	0,04009
PLAUR	NF-AT2 [T01945]	1483	1492	4,70363	GACATTTTCC	0,03147	0,04172
PPP2R2B	MAZ [T00490]	471	483	1,581133	GGTGGGGAGGGAA	0,00107	0,00096
PPP2R2B	RelA [T00594]	548	558	1,736344	GAATTTCCCC	0,00358	0,0035
PPP2R2B	HNF-1B [T01950]	178	186	0	TAATTAACT	0,00572	0,0063
PPP2R2B	NF-kappaB1 [T00593]	547	557	3,924897	TGAAATTTCCC	0,00823	0,00773
PPP2R2B	NF-AT2 [T01945]	1309	1318	3,814941	GGAAAGAGTC	0,00858	0,00881
PPP2R2B	HNF-1C [T01951]	177	185	0	ATAATTAAC	0,01144	0,01252
PPP2R2B	EBF [T05427]	554	564	3,817907	CCCCCAGGGTA	0,0186	0,01686
PPP2R2B	Elk-1 [T00250]	27	35	3,381796	ATGAGGAAG	0,03433	0,03433
PPP2R2B	IRF-1 [T00423]	116	124	1,529008	TTTCATTT	0,03433	0,03827
PPP2R2B	Elk-1 [T00250]	507	515	1,779702	AGTAGGAAG	0,04005	0,03957
PPP2R2B	c-Fos [T00123]	826	835	3,293221	ATTTTGACTC	0,04005	0,04027
PPP2R2B	IRF-1 [T00423]	318	326	4,462268	TTTCCAATT	0,04005	0,04229
PPP2R2B	NFI/CTF [T00094]	1333	1340	1,455588	GCGGTTGG	0,04578	0,04365
PTAFR	HNF-1C [T01951]	225	233	2,841789	ATTTTAAAC	0,00572	0,00208
PTAFR	RAR-beta:RXR-alpha [T	968	979	2,492665	GGGCTCACTGGA	0,00161	0,00269
PTAFR	Ik-1 [T02702]	41	53	4,748597	CGAGTAGCTGGGA	0,00235	0,00331
PTAFR	HOXD9 [T01424]	218	227	1,131608	AACTTTTATT	0,01144	0,00433
PTAFR	HOXD10 [T01425]	218	227	1,131608	AACTTTTATT	0,01144	0,00433
PTAFR	Sp1 [T00759]	1349	1358	0	GCCCCGCCCC	0,00143	0,00477
PTAFR	PPAR-alpha:RXR-alpha	516	526	2,028114	ATGCCCCAGTT	0,00501	0,00589
PTAFR	TCF-4 [T02918]	1196	1205	4,102712	CAGTCAAAGG	0,00858	0,00867
PTAFR	USF2 [T00878]	845	854	4,003951	AGGGCACCTG	0,00858	0,0102
PTAFR	HNF-3alpha [T02512]	90	97	0	TATTTTAA	0,06866	0,02834
PTAFR	HNF-3alpha [T02512]	224	231	0	TATTTTAA	0,06866	0,02834
PTAFR	STAT1beta [T01573]	793	802	2,898434	CTTTCCCAGA	0,03862	0,0303
PTAFR	PPAR-alpha:RXR-alpha	280	290	4,727619	GCCTGGGACAC	0,02432	0,03338
PTAFR	c-Ets-2 [T00113]	1245	1253	1,64415	TTCTCCCA	0,03433	0,0356
PTAFR	HIF-1 [T01609]	639	647	2,845231	ACGTGCAAA	0,04005	0,03732
PTAFR	HIF-1 [T01609]	342	350	2,735708	ACGTGCCAG	0,03433	0,03795
PTAFR	Elk-1 [T00250]	1368	1376	2,164966	CTTCTGGC	0,04005	0,04536
RRP7A	PITX2 [T02413]	945	964	4,605338	GCTGGGATTACAGGCA	0	0
RRP7A	Sp3 [T02338]	1397	1412	1,34037	GGCGGGCGGGCCGGC	0,00001	0,00003
RRP7A	STAT5B [T04684]	503	518	3,722403	ATTTCCAAGAAGCAAC	0,00015	0,00009
RRP7A	Ik-1 [T02702]	939	951	4,748597	CAAATTGCTGGGA	0,00235	0,00315
RRP7A	Sp1 [T00759]	1432	1441	0	GGGGCGGGGC	0,00143	0,00377
RRP7A	NF-AT2 [T01945]	476	485	2,313293	TAATCTTTCC	0,00858	0,00588
RRP7A	STAT1beta [T01573]	503	512	0	ATTTCCAAGA	0,01287	0,01009
RRP7A	NF-AT1 [T01948]	355	364	1,378139	TGGAAAATCC	0,01431	0,01015
RRP7A	NF-AT1 [T00550]	477	485	0,648993	AATCTTTCC	0,01717	0,01214

## Resultados

Supplementary Table S3. (Continuation)

Sequence name	Factor name	Start position	End position	Dissimilarity	String	RE equally	RE query
RRP7A	Sp1 [T00759]	1399	1408	0,949391	CGGGCGGGGC	0,01001	0,02286
RRP7A	RAR-beta [T00721]	1	10	1,063044	AGGGTTCACC	0,02289	0,02346
RRP7A	AR [T00040]	333	341	1,86979	GGACAGTAA	0,02861	0,02717
RRP7A	EBF [T05427]	254	264	4,120419	TTCCAGGGGC	0,0186	0,03073
RRP7A	IRF-1 [T00423]	524	532	4,549799	ACTGGGAAA	0,04005	0,03137
RRP7A	NF-AT1 [T00550]	500	508	4,396744	AGCATTTC	0,0515	0,04175
RRP7A	NF-AT1 [T01948]	376	385	3,445347	CCAGTTCCA	0,05722	0,04501
RRP7A	NF-AT1 [T01948]	1295	1304	3,445347	TGAAAGCTG	0,05722	0,04501
SERPING1	PITX2 [T02413]	643	662	4,605338	GCTGGGATTACAGGTA	0	0
SERPING1	PITX2 [T02413]	1124	1143	1,535113	GCTGGGATTACAGGCG	0	0
SERPING1	Ik-1 [T02702]	1118	1130	2,374299	CAAAGTGCTGGGA	0,00047	0,00033
SERPING1	NF-kappaB1 [T00593]	1048	1058	0,112264	GGGGTTCCCC	0,00107	0,00063
SERPING1	RAR-beta:RXR-alpha [T	1289	1300	2,492665	GGCACAGGGGA	0,00161	0,00103
SERPING1	NF-kappaB [T00590]	103	114	3,935637	AGGAGTTCCA	0,00456	0,00374
SERPING1	RAR-beta:RXR-alpha [T	606	617	4,98533	GGGCTCAAGGGA	0,00724	0,00474
SERPING1	T3R-beta1 [T00851]	1438	1446	0	CGGAGGTGA	0,01144	0,00847
SERPING1	ATF-2 [T00167]	750	759	4,8769	CAGACGTCAA	0,01717	0,01659
SERPING1	NF-AT1 [T01948]	1447	1456	1,378139	GCAATTTCCA	0,01431	0,01799
SERPING1	TCF-4 [T02918]	2	11	4,639022	CCTTGAGCC	0,03147	0,03144
SERPING1	NF-AT2 [T01945]	254	263	3,201983	CAAATTTCC	0,02575	0,03281
SERPING1	CREB [T00163]	750	758	1,140913	CAGACGTCA	0,04005	0,03471
SERPING1	HOXD9 [T01424]	404	413	3,903508	CAATTATATT	0,02289	0,03931
SERPING1	HOXD10 [T01425]	404	413	3,903508	CAATTATATT	0,02289	0,03931
SDC3	PITX2 [T02413]	161	180	3,070225	GGTGGGATTACAGGCG	0	0
SDC3	Smad3 [T04096]	586	599	1,814042	CAAGACAGACAAA	0,00072	0,00041
SDC3	Sp3 [T02338]	1407	1422	4,915458	GCCCCCCCCCGGG	0,0001	0,0013
SDC3	Sp1 [T00759]	1411	1420	0	GCCCCGCCCC	0,00143	0,01296
SDC3	PPAR-alpha:RXR-alpha	484	494	3,498013	GAGCCCCAGTA	0,01073	0,0138
SDC3	EBF [T05427]	940	950	0,984797	TCCCTGGGGA	0,00572	0,01851
SDC3	IRF-1 [T00423]	495	503	2,93326	AATGGGAAA	0,05722	0,02193
SDC3	HNF-3alpha [T02512]	208	215	3,500065	TATTTTCT	0,20599	0,03183
SDC3	USF1 [T00874]	1103	1112	2,607899	CACGTGCGTG	0,02146	0,03223
SDC3	PR B [T00696]	471	477	1,892895	AACACTT	0,09155	0,03525
SDC3	PR A [T01661]	471	477	1,892895	AACACTT	0,09155	0,03525
SDC3	IRF-1 [T00423]	686	694	4,968836	AGGGGGAAA	0,05722	0,03844
SDC3	PPAR-alpha:RXR-alpha	237	247	4,727619	TGCTGGGACAC	0,02432	0,04162
SDC3	Sp1 [T00759]	1349	1358	0,336788	GGGGCGGGGG	0,00572	0,0421
SDC3	Sp1 [T00759]	1406	1415	0,679334	GGCCCCCCCC	0,00572	0,0421
SDC3	PR B [T00696]	440	446	3,29756	AAATGTT	0,18311	0,04288
SDC3	PR A [T01661]	440	446	3,29756	AAATGTT	0,18311	0,04288
SDC3	USF1 [T00874]	1099	1108	4,464121	GAGACACGTG	0,0515	0,04563
SIGLEC1	MAZ [T00490]	1107	1119	2,378507	GGTGGGAGGGGG	0,00304	0,00435
SIGLEC1	TCF-4 [T02918]	687	696	4,865142	ACCTCAAAGG	0,01431	0,0145
SIGLEC1	Elk-1 [T00250]	1023	1031	3,381796	CTCAGGAAG	0,03433	0,03365
SIGLEC1	c-Fos [T00123]	1306	1315	3,366997	TGTCTGACTC	0,04005	0,0387
SIGLEC1	c-Ets-2 [T00113]	304	312	2,945838	TTCCTTTTG	0,04578	0,03994
SIGLEC1	c-Myb [T00137]	817	824	1,285398	GAACTGCC	0,04578	0,04666



## Resultados

Supplementary Table S3. (Continuation)

Sequence		End		Dissimilarity	String	RE equally	RE query
name	Factor name	Start position	position				
TNFSF13B	RAR-beta [T00721]	122	131	2,16302	TGGGTTCAAT	0,05722	0,04778
TRBV23-1	ATF-2 [T00167]	1433	1442	1,17444	ATCACGTAC	0,01144	0,00647
TRBV23-1	Elk-1 [T00250]	675	683	0,134348	TGGAGGAAG	0,04578	0,01555
TRBV23-1	AR [T00040]	932	940	1,620358	GGAATGTCC	0,06866	0,02999
TRBV23-1	NF-1 [T00539]	490	497	2,067686	TGTGCCAA	0,09155	0,03197
TRBV23-1	CREB [T00163]	1483	1491	3,614755	GCAACGTCA	0,08011	0,0321
TRBV23-1	c-Ets-2 [T00113]	674	682	2,715313	ATGGAGGAA	0,05722	0,03525
TRBV23-1	CREB [T00163]	1433	1441	2,664517	ATCACGTCA	0,04578	0,03684
TRBV23-1	HNF-1B [T01950]	1062	1070	0,825511	TCATTAAC	0,01144	0,03927
TRBV23-1	RXR-alpha [T01345]	45	51	4,423008	TCCACCC	0,18311	0,04912
TRIM16L	PITX2 [T02413]	94	113	1,535113	CTCACGCCTGTAATCCC	0	0
TRIM16L	PITX2 [T02413]	1092	1111	4,605338	CTGACGCCTGTAATCCC	0	0
TRIM16L	Ik-1 [T02702]	107	119	2,374299	TCCCAGCACTTTG	0,00047	0,00034
TRIM16L	Ik-1 [T02702]	1105	1117	2,374299	TCCCAGCACTTTG	0,00047	0,00034
TRIM16L	Smad3 [T04096]	388	401	2,689888	AAAGACAGACATAA	0,0008	0,00122
TRIM16L	Ik-1 [T02702]	1243	1255	4,748597	TCCCAGTACTCG	0,00235	0,00171
TRIM16L	FOXO4 [T03403]	1406	1419	2,169752	TTTTTTTGTTTTGT	0,00063	0,00379
TRIM16L	NF-AT2 [T01945]	2	11	0,771098	AAATCTTCC	0,00572	0,0078
TRIM16L	POU2F2 (Oct-2,1) [T00	62	72	0,501684	TGTATAAATCA	0,00966	0,01309
TRIM16L	NF-AT1 [T00550]	3	11	0,648993	AATCTTCC	0,01717	0,02872
TRIM16L	POU2F2 (Oct-2,1) [T00	1009	1019	3,116744	TGAATTACATG	0,01931	0,02933
TRIM16L	Elk-1 [T00250]	702	710	3,381796	CTCAGGAAG	0,03433	0,03689
TRIM16L	USF2 [T00878]	1134	1143	4,528187	GGATCAGCTG	0,0515	0,03764
TRIM16L	NF-AT2 [T01945]	465	474	3,689016	AATAATTCC	0,02575	0,04053
TRIM16L	T3R-beta1 [T00851]	320	328	1,110682	TCACCACTG	0,05722	0,04372
TRIM16L	HNF-3alpha [T02512]	861	868	1,342935	TTAAATA	0,02289	0,04565
TRIM16L	TBP [T00794]	961	970	0	GAGATATAAA	0,02289	0,04736
VWA8	PEA3 [T00685]	254	262	0	TTACATCCT	0,00572	0,00734
VWA8	NF-AT1 [T00550]	758	766	1,437145	ACTTTTTCC	0,00572	0,00757
VWA8	NF-AT2 [T01945]	883	892	2,029228	AATTTTTCC	0,01144	0,01644
VWA8	NF-AT1 [T01948]	856	865	1,378139	GCAATTTCCA	0,01431	0,01834
VWA8	NF-AT1 [T00550]	884	892	0	ATTTTTTCC	0,01144	0,01837
VWA8	NF-AT1 [T00550]	977	985	3,075022	GGAAACACT	0,02289	0,02206
VWA8	Sp1 [T00759]	1367	1376	4,81072	TTTCCGCCCG	0,03433	0,02248
VWA8	GCF [T00320]	509	517	2,339499	GCGCAGGGT	0,04578	0,02721
VWA8	GCF [T00320]	1282	1290	0	GCCCAGCGC	0,06866	0,03241
VWA8	Elk-1 [T00250]	1472	1480	2,164966	CTTCCTGTC	0,04005	0,03578
VWA8	NF-AT2 [T01945]	757	766	4,586038	TACTTTTTCC	0,03147	0,03867
VWA8	AR [T00040]	748	756	2,808697	CCAGTGTC	0,04578	0,03888
VWA8	USF2 [T00878]	1239	1248	4,528187	CAGGTGAACA	0,0515	0,0412
VWA8	IRF-1 [T00423]	1072	1080	1,616539	ACAGGGAAA	0,03433	0,04751
XIST	PITX2 [T02413]	1	20	4,605338	GCTGGGATTACAAGCA	0	0
XIST	Elk-1 [T00250]	584	592	1,645354	CTTCTACC	0,00572	0,00349
XIST	FOXO4 [T03403]	258	271	3,616253	TTTTTTTGTTTGAT	0,00074	0,0052
XIST	Sp1 [T00759]	1482	1491	3,54287	TCTCCGCCCT	0,05579	0,00809
XIST	IRF-1 [T00423]	1389	1397	0	TTTCCCTT	0,00572	0,0119
XIST	NF-AT2 [T01945]	58	67	4,979362	GAACATTTCC	0,01144	0,01777

## Resultados

Supplementary Table S3. (Continuation)

Sequence		End		Dissimilarity	String	RE equally	RE query
name	Factor name	Start position	position				
XIST	POU2F1 [T00641]	315	325	0,929531	ATTTGCATACT	0,01037	0,02148
XIST	TCF-4 [T02918]	185	194	4,412902	TTCTCAAAGG	0,03147	0,02685
XIST	TCF-4 [T02918]	420	429	4,639022	CCTTTGACTT	0,03147	0,02685
XIST	Elk-1 [T00250]	526	534	1,779702	CTTCCTACT	0,04005	0,02689
XIST	c-Fos [T00123]	1314	1323	3,849714	GTTGTGACTC	0,02861	0,02718
XIST	HNF-1B [T01950]	708	716	4,953067	AGTTAAAAA	0,00572	0,02728
XIST	c-Fos [T00123]	701	710	4,351106	GAGTCATAGT	0,03433	0,03145
XIST	p53 [T00671]	25	31	4,645444	CGCGCCC	0,18311	0,03292
XIST	AP-1 [T00029]	698	706	4,902944	AAGGAGTCA	0,05722	0,03713
XIST	AP-1 [T00029]	1318	1326	4,815705	TGACTCCTG	0,05722	0,03713
XIST	c-Ets-2 [T00113]	669	677	2,715313	ATGGAGGAA	0,05722	0,03965
XIST	STAT1beta [T01573]	670	679	2,898434	TGGAGGAAAT	0,03862	0,04277
XIST	AR [T00040]	1270	1278	1,72955	TCAATGTCC	0,06866	0,04716
XIST	TCF-4 [T02918]	1156	1165	1,687438	ACATCAAAGC	0,03433	0,04751
ZNF275	Sp1 [T00759]	1450	1459	0	GGGGCGGGGC	0,00143	0,0006
ZNF275	RAR-alpha1 [T00719]	243	255	4,596281	GGGGTCAAGGGTT	0,00282	0,00208
ZNF275	WT1 [T00899]	1378	1386	0	CGCCCCCGC	0,00572	0,00241
ZNF275	Sp1 [T00759]	1391	1400	0,574521	ACCCCGCCCC	0,00572	0,00307
ZNF275	STAT5A [T04683]	816	828	4,540481	CAGTTTCTTATAA	0,00402	0,00517
ZNF275	MEF-2A [T01005]	701	711	3,898698	GGCTAAAAATA	0,00322	0,00538
ZNF275	Sp1 [T00759]	1380	1389	1,28618	CCCCGCCCCA	0,02432	0,01318
ZNF275	Elk-1 [T00250]	1238	1246	0,957025	CTTCCTTCG	0,02289	0,01955
ZNF275	POU2F2 (Oct-2,1) [T00	822	832	4,120113	CTTATAAAACA	0,01287	0,02371
ZNF275	Sp1 [T00759]	1374	1383	3,408439	AGTCCGCCCC	0,05579	0,02807
ZNF275	TBP [T00794]	820	829	3,743085	TTCTTATAAA	0,02289	0,04432
ZNF275	HOXD9 [T01424]	665	674	3,903508	AATAACAATG	0,02289	0,0461
ZNF275	HOXD10 [T01425]	665	674	3,903508	AATAACAATG	0,02289	0,0461
ZNF275	c-Ets-2 [T00113]	1227	1235	4,589988	TTCTTCGT	0,0515	0,04636
ZNF275	c-Ets-2 [T00113]	1235	1243	4,589988	TTCTTCCT	0,0515	0,04636
ZNF275	c-Ets-2 [T00113]	1239	1247	4,589988	TTCTTCGT	0,0515	0,04636
ZNF275	c-Myb [T00137]	814	821	2,570796	GGCAGTTT	0,04578	0,04869
ZNF703	HOXD9 [T01424]	497	506	2,949288	TATTGTTATT	0,01717	0,00124
ZNF703	HOXD10 [T01425]	497	506	2,949288	TATTGTTATT	0,01717	0,00124
ZNF703	TCF-4 [T02918]	128	137	0,53631	CCTTTGAAAT	0,01431	0,00535
ZNF703	HNF-3alpha [T02512]	503	510	0	TATTTTTA	0,06866	0,00859
ZNF703	MAZ [T00490]	1308	1320	3,189496	CGGGGGAGGGGA	0,00355	0,0138
ZNF703	NF-AT1 [T01948]	267	276	2,756277	TGAAAAAGTG	0,04435	0,01815
ZNF703	TCF-4 [T02918]	1474	1483	0,98855	CCTTTGAAGT	0,04292	0,01846
ZNF703	EBF [T05427]	1284	1294	1,088777	ACCCAGGGCG	0,00572	0,02035
ZNF703	PEA3 [T00685]	405	413	3,710864	TGGATGTCA	0,06866	0,03233
ZNF703	Elk-1 [T00250]	1292	1300	2,164966	GCGAGGAAG	0,04005	0,04644
ZNF71	POU2F2 (Oct-2,1) [T00	42	52	1,505053	TTTATAAATCA	0,00644	0,0041
ZNF71	CTF [T00174]	657	668	3,641537	GTCCATTGGCTG	0,00751	0,00533
ZNF71	TCF-4 [T02918]	304	313	1,934085	ATTTCAAAGT	0,03147	0,0229
ZNF71	MAZ [T00490]	812	824	3,973255	TGCCCTCCCAAC	0,0044	0,02632
ZNF71	NF-AT2 [T01945]	786	795	3,571424	CATTCTTCC	0,02575	0,03748
ZNF71	NF-AT1 [T00550]	787	795	0,648993	ATTCTTCC	0,01717	0,03884



Supplementary Table S4. Differential expression identified in the EC-LTNP/vLTNP comparison.

Gene	Locus	Group 1	Group2	FPKM Group 1	FPKM Group 2	log2 FC	test stat	p-value	q-value	Significant
TSTD3	6:99879318-99979794	EC-LTNP	vLTNP	0,517642	51,2409	6,6292	2,0327	5,00E-05	0,0121911	yes
HPX	11:6452278-6463847	EC-LTNP	vLTNP	0,269194	14,1032	5,71123	2,04822	5,00E-05	0,0121911	yes
HELB	12:66696237-67197966	EC-LTNP	vLTNP	4,83523	132,304	4,77413	4,03988	5,00E-05	0,0121911	yes
MIR210HG	11:565659-568457	EC-LTNP	vLTNP	0,202184	1,90285	3,23442	1,40873	0,00025	0,0446698	yes
HAR1A	20:61726844-61735738	EC-LTNP	vLTNP	0,590532	2,82873	2,26007	1,58101	5,00E-05	0,0121911	yes
MEG3	14:101245746-101327368	EC-LTNP	vLTNP	0,661139	3,12628	2,24142	1,03936	5,00E-05	0,0121911	yes
SGCD	5:155107801-156194799	EC-LTNP	vLTNP	0,134763	0,608153	2,17402	1,10833	5,00E-05	0,0121911	yes
TRIM16L	17:18601310-18639578	EC-LTNP	vLTNP	4,45625	19,9804	2,16468	5,43833	5,00E-05	0,0121911	yes
HBA2	16:222845-223709	EC-LTNP	vLTNP	2,75	12,0341	2,12963	1,43078	0,0002	0,0380602	yes
IMMP1L	11:31391383-31531192	EC-LTNP	vLTNP	3,47725	14,7559	2,08527	0,993535	0,00015	0,0301119	yes
IGHG1	14:106202679-106209408	EC-LTNP	vLTNP	44,3071	182,952	2,04586	2,15887	5,00E-05	0,0121911	yes
GLG1	16:74481324-74641012	EC-LTNP	vLTNP	37,6327	151,699	2,01115	2,4926	5,00E-05	0,0121911	yes
IGHG3	14:106235438-106242016	EC-LTNP	vLTNP	15,276	60,3584	1,98228	2,09262	5,00E-05	0,0121911	yes
DPH6	15:35473999-35838394	EC-LTNP	vLTNP	3,80149	14,6698	1,94821	1,44729	0,0001	0,0217134	yes
CALD1	7:134429002-134655479	EC-LTNP	vLTNP	0,374046	1,40481	1,90909	0,80908	5,00E-05	0,0121911	yes
SSRP1	11:57093076-57103351	EC-LTNP	vLTNP	22,8367	84,6868	1,89078	3,6553	5,00E-05	0,0121911	yes
MCOLN2	1:85391156-85462796	EC-LTNP	vLTNP	10,0595	32,6091	1,69672	0,877835	5,00E-05	0,0121911	yes
KIF19	17:72322348-72351959	EC-LTNP	vLTNP	0,717707	2,23624	1,63961	1,26081	0,00025	0,0446698	yes
IGLC3	22:23248511-23248973	EC-LTNP	vLTNP	27,0143	81,5374	1,59374	1,63291	5,00E-05	0,0121911	yes
MYO6	6:76458908-76629302	EC-LTNP	vLTNP	1,16845	3,43665	1,55641	1,27733	5,00E-05	0,0121911	yes
SLC14A1	18:42792959-43332485	EC-LTNP	vLTNP	1,95049	5,29061	1,4396	0,938721	5,00E-05	0,0121911	yes
SYTL2	11:85405266-85522184	EC-LTNP	vLTNP	9,62459	23,8976	1,31207	0,924672	5,00E-05	0,0121911	yes
CD8A	2:87009218-87035519	EC-LTNP	vLTNP	70,0512	172,254	1,29806	1,49917	5,00E-05	0,0121911	yes
IGHA1	14:106173456-106175002	EC-LTNP	vLTNP	67,4389	161,568	1,26049	1,46129	5,00E-05	0,0121911	yes
KIF21A	12:39687029-39837192	EC-LTNP	vLTNP	2,25017	5,15306	1,1954	0,851583	0,0001	0,0217134	yes
PDE4B	1:66258196-66843190	EC-LTNP	vLTNP	62,556	32,5737	-0,941443	-0,924592	0,00025	0,0446698	yes
PCBP3	21:47063607-47362368	EC-LTNP	vLTNP	0,660864	0,324454	-1,02634	-0,497038	0,00015	0,0301119	yes
MYADM	19:54357834-54379691	EC-LTNP	vLTNP	393,202	184,76	-1,08962	-1,01599	0,00015	0,0301119	yes
B3GNT5	3:182895830-183146566	EC-LTNP	vLTNP	24,2324	11,0338	-1,135	-1,08295	0,0002	0,0380602	yes
ANPEP	15:90328119-90358633	EC-LTNP	vLTNP	145,248	65,2577	-1,15429	-1,01224	0,0001	0,0217134	yes

Supplementary Table S4. Differential expression identified in the EC-LTNP/vLTNP comparison (continuation)

Gene	Locus	Group 1	Group2	FPKM Group 1	FPKM Group 2	log2 FC	test stat	p-value	q-value	Significant
LILRA3	19:54799853-54809952	EC-LTNP	vLTNP	38,0345	16,5275	-1,20244	-1,5244	0,0001	0,0217134	yes
NRIP1	21:16333554-16438084	EC-LTNP	vLTNP	26,7756	11,6297	-1,2031	-1,40274	0,00015	0,0301119	yes
RAB11FIP1	8:37715971-37757001	EC-LTNP	vLTNP	55,1097	23,5858	-1,22439	-1,22018	0,00025	0,0446698	yes
DCP1A	3:53304129-53381654	EC-LTNP	vLTNP	83,4389	35,3374	-1,23953	-3,04013	5,00E-05	0,0121911	yes
GADD45B	19:2476119-2478257	EC-LTNP	vLTNP	223,795	93,5924	-1,25771	-0,925683	5,00E-05	0,0121911	yes
SERTAD1	19:40927498-40931932	EC-LTNP	vLTNP	60,5254	24,9936	-1,27598	-1,5039	5,00E-05	0,0121911	yes
ICAM1	19:10362576-10399695	EC-LTNP	vLTNP	195,942	77,0965	-1,34569	-1,19068	0,0002	0,0380602	yes
ALDH9A1	1:165631452-165679205	EC-LTNP	vLTNP	72,4907	26,9271	-1,42873	-1,98265	5,00E-05	0,0121911	yes
PHLDA1	12:76419226-76427712	EC-LTNP	vLTNP	12,0595	4,44117	-1,44116	-1,45917	5,00E-05	0,0121911	yes
TNF	6:31543272-31546113	EC-LTNP	vLTNP	64,3611	23,5839	-1,44839	-1,59735	5,00E-05	0,0121911	yes
MB21D2	3:192514537-192635950	EC-LTNP	vLTNP	4,20613	1,45219	-1,53426	-1,37458	0,00015	0,0301119	yes
TCL1B	14:96116834-96158980	EC-LTNP	vLTNP	1,23226	0,421665	-1,54713	-0,919164	0,00025	0,0446698	yes
MME	3:154741912-154901750	EC-LTNP	vLTNP	0,556404	0,189041	-1,55743	-0,451558	5,00E-05	0,0121911	yes
CDKN1A	6:36644304-36655246	EC-LTNP	vLTNP	476,929	160,479	-1,57139	-1,52836	5,00E-05	0,0121911	yes
IER3	6:30710705-30712331	EC-LTNP	vLTNP	400,362	133,416	-1,58538	-1,68579	5,00E-05	0,0121911	yes
OSM	22:30658603-30671413	EC-LTNP	vLTNP	246,937	81,8595	-1,59292	-1,06276	0,0001	0,0217134	yes
CD200	3:112021324-112085457	EC-LTNP	vLTNP	1,65943	0,547525	-1,5997	-1,09289	0,00025	0,0446698	yes
KLF5	13:73593487-73651678	EC-LTNP	vLTNP	4,89823	1,60408	-1,61052	-1,28468	0,0001	0,0217134	yes
MN1	22:28144241-28201218	EC-LTNP	vLTNP	4,31954	1,32918	-1,70034	-1,20879	0,0002	0,0380602	yes
PHLDA2	11:2949502-2950685	EC-LTNP	vLTNP	22,4494	6,61484	-1,7629	-1,32359	0,0001	0,0217134	yes
LDLR	19:11200037-11266484	EC-LTNP	vLTNP	79,7232	23,3913	-1,76903	-1,06941	5,00E-05	0,0121911	yes
NR4A3	9:102584136-102629282	EC-LTNP	vLTNP	26,9358	7,89595	-1,77034	-1,821	5,00E-05	0,0121911	yes
CAPZA1	1:113009162-113215488	EC-LTNP	vLTNP	720,488	208,974	-1,78565	-1,38732	5,00E-05	0,0121911	yes
SLC37A3	7:139993492-140126050	EC-LTNP	vLTNP	43,5736	12,466	-1,80546	-2,08966	5,00E-05	0,0121911	yes
LGALS9C	17:18353557-18398259	EC-LTNP	vLTNP	1,35223	0,381898	-1,82408	-0,731409	0,00015	0,0301119	yes
NRIP3	11:9001547-9025598	EC-LTNP	vLTNP	7,8656	2,12508	-1,88804	-1,18848	5,00E-05	0,0121911	yes
CDCP1	3:45123671-45187914	EC-LTNP	vLTNP	0,61534	0,165151	-1,8976	-1,21957	5,00E-05	0,0121911	yes
ITPR3	6:33588029-33679504	EC-LTNP	vLTNP	65,1372	16,8691	-1,9491	-5,60276	5,00E-05	0,0121911	yes
SHQ1	3:72798427-72911065	EC-LTNP	vLTNP	11,4214	2,95231	-1,95182	-1,65566	5,00E-05	0,0121911	yes

Supplementary Table S4. Differential expression identified in the EC-LTNP/vLTNP comparison (continuation)

Gene	Locus	Group 1	Group2	FPKM Group 1	FPKM Group 2	log2 FC	test stat	p-value	q-value	Significant
HES1	3:193853038-193856579	EC-LTNP	vLTNP	76,8756	19,0514	-2,01263	-1,6406	5,00E-05	0,0121911	yes
MAGI2-AS3	7:77646392-79100524	EC-LTNP	vLTNP	1,52855	0,336005	-2,18561	-0,776815	5,00E-05	0,0121911	yes
CXCL10	4:76932336-77033955	EC-LTNP	vLTNP	11,7364	2,57317	-2,18937	-1,67724	5,00E-05	0,0121911	yes
ZC3H12C	11:109959155-110042566	EC-LTNP	vLTNP	2,45186	0,531485	-2,20578	-1,82569	5,00E-05	0,0121911	yes
APOL4	22:36585171-36600886	EC-LTNP	vLTNP	1,33256	0,27945	-2,25354	-1,01969	5,00E-05	0,0121911	yes
IL1A	2:113531491-113542167	EC-LTNP	vLTNP	6,42364	1,07256	-2,58233	-1,86049	5,00E-05	0,0121911	yes
NIPAL2	8:99201952-99306760	EC-LTNP	vLTNP	110,646	17,44	-2,66548	-4,5632	5,00E-05	0,0121911	yes
GFAP	17:42923720-42994305	EC-LTNP	vLTNP	0,733242	0,0849109	-3,11027	-0,166604	5,00E-05	0,0121911	yes
BTBD11	12:107712105-108053419	EC-LTNP	vLTNP	82,45	4,64372	-4,15017	-6,23559	5,00E-05	0,0121911	yes
CCR9	3:45864807-46037316	EC-LTNP	vLTNP	14,2968	0,804151	-4,15208	-0,61211	5,00E-05	0,0121911	yes
C15orf52	15:40623652-40633168	EC-LTNP	vLTNP	26,3543	1,03728	-4,66717	-3,04105	5,00E-05	0,0121911	yes

# DISCUSIÓN

## 4. DISCUSIÓN

Desde la aparición del VIH, aún no ha sido posible desarrollar una vacuna protectora eficaz o encontrar la cura para la infección, a pesar de los enormes esfuerzos que se han realizado. El estudio de individuos con fenotipo no progresor y/o controlador del VIH es importante porque su caracterización puede brindarnos evidencias sobre los factores virales, genéticos e inmunológicos involucrados en el control natural de la infección, que contribuirían en las estrategias de cura funcional y diseño de vacunas terapéuticas.

La heterogeneidad de estos individuos complica el hallazgo de nuevos factores asociados al control de la infección. El tiempo de seguimiento es importante en los LTNPs, ya que una mejor supervivencia se ha observado en aquellos que son definidos a los 10 años de infección estable en comparación con aquellos con 7 años de infección (Okulicz *et al*, 2009), lo que sugiere que un menor tiempo de seguimiento no distingue adecuadamente entre un verdadero LTNP y un progresor del VIH-1.

Los individuos ECs considerados en nuestros estudios son EC-LTNPs, es decir mantienen CVs indetectables y niveles de linfocitos CD4+ > 500 células/μL. El estudio de EC-LTNPs es relevante porque representan el fenotipo más beneficioso de la infección por el VIH-1, debido al control inmunológico y virológico que mantienen estos individuos durante al menos 10 años (Casado *et al*, 2010). Además, los EC-LTNPs o individuos con características similares, son considerados como un potencial modelo de cura funcional (Autran *et al*, 2011; López-Galíndez *et al*, 2019). De manera similar se definen a los HICs de la cohorte NIAID (*National Institute of Allergy and Infectious Diseases*), los cuales tienen una CV < 50 copias de ARN viral/mL y recuentos estables de T CD4+ (Migueles *et al*, 2008). Un estudio en la cohorte francesa de la ANRS (*Agence Nationale de Recherche sur le Sida et les hépatites virales*), muestra que los HICs y LTNPs representan poblaciones parcialmente superpuestas, constituyendo los *elite*-LTNPs el 0.22% del total de individuos infectados por el VIH (Grabar *et al*, 2009).

Sin embargo, otras cohortes definen a los controladores del VIH sin considerar el recuento de linfocitos T CD4+. Un estudio de una cohorte multicéntrica francesa define al grupo de HICs evaluando a los pacientes por un período de tiempo mayor de 10 años, con la diferencia de que las CVs están por debajo de 400 copias/mL en más del 90% de los valores medidos (Lambotte *et al*, 2005). Más cercano a nuestra definición, un estudio multicéntrico de HICs de Estados Unidos define a los ECs con una CV < 50 copias/mL (Pereyra *et al*, 2008). Es importante la valoración de los linfocitos T CD4+, ya que se ha demostrado que la proporción de ECs con al menos un recuento de células CD4+ < 500/μL oscila entre 45-53% (Olson *et al*, 2012) y la pérdida de células CD4+ en el fenotipo VNP es más común que en aquellos sin CV detectable (Pereyra *et al*, 2009).

El objetivo principal de la presente tesis doctoral es comprender mejor el control natural de la infección por el VIH-1 observado en LTNPs y HICs mediante una aproximación holística basada en la biología de sistemas, con el fin de tener un mayor conocimiento que contribuya a un mejor tratamiento o la cura funcional de la infección.

#### **4.1. Determinación de polimorfismos mediante estudios de asociación genética**

Los virus han ejercido una presión selectiva, constante y potente sobre los genes humanos a lo largo de la evolución, que se puede estimar mediante la frecuencia de variantes alélicas, principalmente en genes involucrados en la respuesta inmune, estructuras de receptores virales y en la interacción directa con componentes virales (Fumagalli *et al*, 2010). La ventaja de estudiar variantes genéticas en el ADN es la constancia en el tiempo, a diferencia de la expresión génica que puede ser variable en el curso de la infección, por la administración del tratamiento o el tipo celular estudiado.

En nuestros análisis hemos evaluado principalmente SNPs. Los SNPs consisten en el cambio de un nucleótido en la secuencia de ADN, generalmente con dos alelos posibles y son la forma más abundante de variación genética en el genoma humano (Bush y Moore, 2012; The 1000 Genomes Project Consortium, 2015). La mayoría de SNPs tienen un impacto mínimo en los sistemas biológicos, aunque pueden tener consecuencias funcionales causando cambios en: los aminoácidos, la estabilidad del transcrito de ARNm y la afinidad de unión al factor de transcripción (Griffith *et al*, 2008).

Se han desarrollado diversas tecnologías para el genotipado de variantes genéticas, basados en técnicas que discriminan los alelos más sistemas de detección apropiados. En general, las tecnologías de genotipado de SNPs han progresado rápidamente con la aparición de métodos novedosos, más rápidos y más baratos, así como mejoras en los métodos existentes (Revisado por: Kim *et al*, 2007; Jiang *et al*, 2016).

##### **4.1.1. Genotipado de polimorfismos genéticos concretos en LTNPs**

Estudios previos señalan que el control de la infección por el VIH se asocia con diversos alelos ubicados principalmente en el cromosoma 6, y además la mayoría de la evidencia de los estudios de asociación genética apunta a genes HLA-I (McLaren *et al*, 2015). En este primer trabajo (Artículo I) se analizó principalmente polimorfismos en genes que codifican quimiocinas, receptores de quimiocinas y la región HLA.

Hemos observado una alta frecuencia de los genotipos rs2395029-TG (*HCP5*), rs1799864-GA/AA (*CCR2*) y rs9264942-CC/CT (5' *HLA-C*) en LTNPs, en comparación con individuos sanos o TPs (Tabla 2, Artículo I). Los polimorfismos rs1799864 y rs9264942 se relacionan con la mutación V641T de *CCR2* y con los niveles de expresión de *HLA-C*, respectivamente, y se han asociado previamente con el control de la infección por el VIH (Smith *et al*, 1997; Fellay *et al*, 2009). El SNP rs2395029 de *HCP5* también ha sido

ampliamente asociado con los LTNP (Revisado por Limou y Zagury, 2013) y también encontramos su alta asociación en nuestro análisis de exoma de LTNP (Artículo II). Los polimorfismos rs9264942 y rs2395029, se han asociado previamente con la progresión lenta en individuos LTNP españoles (Rodríguez Da Silva *et al*, 2015). Por tanto, se confirma la asociación de estos tres polimorfismos con el fenotipo LTNP. Por otra parte, se evidencia una mayor frecuencia de los SNPs mencionados en *HCP5* y *CCR2* en aquellas subcategorías de LTNP con la menor CV, LTNP-Cs (LTNP con CV < 2.000 copias/mL) y ECs (Figura 1 y Tabla 3, Artículo I). Sin embargo, no evidenciamos una asociación de rs9264942 con la CV como en trabajos previos (Thomas *et al*, 2009; Fellay *et al*, 2009).

El polimorfismo *CCR5*Δ32 causa una mutación en el correceptor *CCR5*, y el alelo wt/Δ32 heterocigótico ha sido asociado con un retraso en la progresión del VIH (Dean *et al*, 1996; Rodríguez Da Silva *et al*, 2015). Además, se ha asociado previamente la alta frecuencia de este polimorfismo en otra cohorte de LTNP (Stewart *et al*, 1997). Se ha descrito previamente que el 15% de la población española presenta el genotipo *CCR5* wt/Δ32 (Solloch *et al*, 2017), coincidiendo con nuestra población control de individuos sanos. A pesar que se observa una frecuencia más alta del genotipo wt/Δ32 en LTNP (24%), la diferencia no fue significativa al compararlas con los otros grupos del estudio (Tabla 2, Artículo I). De manera similar, el 25% de individuos de la cohorte de LTNP evaluada en el estudio de exomas presentó el genotipo wt/Δ32 (Anexo 1, Tabla S2). Investigaciones previas han señalado que la frecuencia de *CCR5*-Δ32 no parece ser diferente ni estar enriquecida en cohortes de HICs (Pereyra *et al*, 2008), e incluso se han identificado LTNP/ECs que no presentan los alelos *CCR5*-Δ32 y HLA “protectores” (Nissen *et al*, 2018). La delección *CCR5*Δ32 contribuye en la prevención de la infección inicial por el VIH (Samson *et al*, 1996), pero una vez establecida la infección no contribuye con el fenotipo LTNP de igual manera que otros polimorfismos genéticos.

Ha sido reportada previamente la asociación de los otros polimorfismos evaluados con la progresión del VIH u otras enfermedades virales. Variantes en el SNP rs1799987 (*CCR5*) han sido asociadas a la susceptibilidad a encefalitis japonesa (Deval *et al*, 2019) o al daño renal por la crioglobulinemia relacionada con el VHC (Wang *et al*, 2018). El polimorfismo rs1801274 (*FCGR2A*) se ha asociado al desarrollo de neumonías severas causadas por el virus influenza AH1N1 (Zúñiga *et al*, 2012) o la progresión del dengue (Noecker *et al*, 2014). El genotipo rs1801157-TT (*CXCL12*) se asocia con una baja recuperación inmunológica en los pacientes que comienzan la TAR con recuentos bajos de linfocitos T CD4+ (Restrepo *et al*, 2019). Variantes en el polimorfismo rs2107538 (*CCL5*) están asociadas al desarrollo de arteriosclerosis en infectados por el VIH-1 (Ibañez *et al*, 2014) o a una infección severa por el enterovirus 71 (Li *et al*, 2015).

Además, tampoco se ha mostrado una asociación de los otros polimorfismos evaluados (rs1857909, rs1050931, rs729421, rs288039, rs3749663, rs700626) con el fenotipo LTNP. Otras investigaciones realizadas con técnicas de genotipado convencional en sangre o PBMCs, han reportado asociaciones de variantes alélicas en *CX3CR1*, *IL6*, *PDCD1*, *TNF* y *ZNRD1* con la condición LTNP y/o EC (Vidal *et al*, 2005; Ballana *et al*, 2010; Nasi *et al*, 2013; Loureiro Dos Reis *et al*, 2019).

Por otra parte, el estudio de la región HLA es muy importante debido a que hay trabajos que proponen que variantes o genes localizados fuera de esta región se asocian difícilmente con la progresión del VIH-1 (McLaren *et al*, 2017). La tipificación de HLA es importante en la investigación de enfermedades y en la práctica clínica de trasplante de órganos. Los métodos de tipado de HLA utilizados comprenden los serológicos, moleculares (PCR con *primers* y sondas específicos de secuencia, secuenciación directa del ADN) y los basados en NGS que permiten el análisis de la secuencia completa de la región HLA (Hosomichi *et al*, 2015; Althaf *et al*, 2017). Los métodos convencionales por PCR-SSO y tipado basado en secuenciación de PCR son de primera línea en investigación y diagnóstico de HLA (Hosomichi *et al*, 2015), y han sido empleados en nuestro trabajo.

La presencia de polimorfismos “protectores” en la región HLA están asociados con la no progresión a largo plazo, debido al control viral temprano que se produce durante la infección (Limou y Zagury, 2013). Nuestros resultados indican que los alelos HLA-A\*03, -B\*27, -B\*39, -B\*52 y -B\*57 favorecen la condición LTNP, debido a las altas frecuencias alélicas en estos individuos comparado con donantes sanos. El alelo HLA-A\*03 ha sido asociado a LTNPs de una cohorte francesa (Magierowska *et al*, 1999), así como al control de la infección por el VIH en poblaciones asiáticas (Zhang *et al*, 2013). Asimismo, el alelo HLA-B\*52 se ha asociado con el control inmunológico y/o viral de individuos VIH-positivos en varias cohortes (Pereyra *et al*, 2010; Teixeira *et al*, 2014; Chikata *et al*, 2017).

Estudios previos han demostrado la asociación de los alelos HLA-B\*27 y HLA-B\*57 con baja viremia en pacientes LTNPs (Magierowska *et al*, 1999; Migueles *et al*, 2000), que se relacionan con la presentación antigénica de epítomos virales conservados que afectan al *fitness* viral (Martínez-Picado *et al*, 2006). La protección de HLA-B\*57 ocurre en las etapas tempranas de la infección y está asociado con un descenso lento de las células T CD4+. En cambio, el efecto de HLA-B\*27 se produce en la infección tardía, cuando los niveles de células T CD4+ han descendido, y se asocia con un retraso en la aparición de la fase sida (Gao *et al*, 2005). Sin embargo, se ha descrito que una alta proporción de individuos con alelos HLA-B\*27 y/o -B\*57 progresan a la enfermedad, y otras personas que no poseen alelos protectores permanecen asintomáticos (Goulder y Walker, 2012). Además entre un 15 a 70% de HICs no presentan el alelo HLA-B\*57 en diferentes cohortes estudiadas (Migueles *et al*, 2000; Flores-Villanueva *et al*, 2001; Pereyra *et al*, 2008; Tang *et al*, 2010).



Se ha reportado previamente la asociación de HLA-B\*39 con una baja CV durante la infección por el VIH (Tang *et al*, 2002). Nuestro estudio asocia por primera vez este alelo “protector” con el fenotipo LTNP. Los alelos HLA- A\*03:01, -B\*27:05, - B\*39:02 y -B\*57:01 han sido asociados al control viral en una cohorte mexicana de individuos VIH-positivos (Valenzuela-Ponce *et al*, 2018). Además, se observa una mayor frecuencia de los alelos “protectores” HLA-B\*57 y HLA-A\*03 en los grupos EC y LTNP-C, los cuales correlacionan inversamente con la CV (Figura 1, Artículo I). La mayoría de los alelos HLA “protectores” presentan epítomos de Gag, debido a la abundancia de esta proteína y también porque está conservada para la formación de la cápside viral (Miura *et al*, 2009; Chen *et al*, 2012).

Por otra parte, los alelos HLA-A\*24, -A\*29, -B\*08 y -B\*18, son considerados de “riesgo” porque se han asociado previamente con una progresión rápida de la enfermedad. Además, nuestro trabajo los ha asociado por primera vez al fenotipo LTNP debido a su menor frecuencia en este grupo, evidenciándose además una diferencia de éstos en las subcategorías de LTNPs (Tabla 4, Artículo I). El alelo HLA-A\*24 ha sido asociado previamente con una rápida disminución de los linfocitos T CD4+ (Kaslow *et al*, 1990). La presencia del alelo HLA-A\*29 se asocia con un mayor riesgo de infección de infantes lactantes y mayor riesgo de superinfección por el VIH (Farquhar *et al*, 2004; Vesa *et al*, 2017). Por otra parte, niveles elevados de HLA-A se asocian a un control más deficiente del VIH, ya que afecta la expresión de HLA-E, ligando para el receptor de la célula NK, alterando la eliminación de células infectadas por el VIH (Ramsuran *et al*, 2018).

El alelo HLA-B\*08 ha sido asociado a una rápida disminución de los linfocitos T CD4+ y con una aumentada transmisión materno-fetal del VIH (Kilpatrick *et al*, 1991; McNeil *et al*, 1996). Por otra parte, HLA-B\*18 ha sido estudiado en otras poblaciones de individuos VIH-positivos, y puede tener un efecto favorable o de riesgo durante la infección (Farquhar *et al*, 2004; de Sorrentino *et al*, 2000). Sin embargo, no se ha encontrado una asociación negativa del alelo HLA-B\*35 con los LTNP/ECs como en estudios previos (Pereyra *et al*, 2010; Teixeira *et al*, 2012).

Los alelos HLA “de riesgo” pueden promover la selección de variantes de escape a los CTLs o un disminuido reconocimiento de los epítomos virales por los linfocitos T (Wilson *et al*, 1999; Furutsuki *et al*, 2004). Por el contrario, se ha reportado que los PTCs presentan una mayor frecuencia de los alelos de “riesgo” (HLA-B\*07 y -B\*35) y no presentan altas frecuencias de alelos HLA “protectores” como los LTNP/HICs (Sáez-Cirión *et al*, 2013).

Las moléculas HLA muestran los epítomos Bw4 o Bw6, que son mutuamente excluyentes y se identifican mediante cinco aminoácidos variables en las posiciones 77 y 80-83 (Muller *et al*, 1989). El motivo Bw4 está presente en el tercio de las moléculas HLA-B identificadas, así como también en moléculas HLA-A (Parham, 2005). HLA-B\*27 y -B\*57, generalmente con mayor frecuencia en HICs, son miembros del grupo Bw4 (Flores-

Villanueva *et al*, 2001). El motivo HLA-Bw4 se expresa en células T CD4+ y es un ligando de KIR (*Killer-cell immunoglobulin-like receptor*) (Parham, 2005). La región KIR constituye una familia poligénica y polimórfica de receptores, con motivos inhibidores y activadores (Vilches y Parham, 2002) que son expresados por las células NK (Fauriat *et al*, 2008) y las células T CD8+ (Bjorkstrom *et al*, 2015), participando tanto en la inmunidad innata como adaptativa. El epítipo HLA-Bw4-80I de células T CD4+ interacciona con KIR3DS1 de las NKs, dando como resultado una CV más baja y riesgo reducido de progresión de la enfermedad (Martin *et al*, 2002; Boulet *et al*, 2010). De manera similar, se ha observado una fuerte supresión viral por las células T CD8+ que expresan KIR3DL1 sobre las células T CD4+ infectadas (Lu *et al*, 2016). El control de la replicación viral mediada por células NK es modesta y es más activa en la etapa inicial que durante la fase crónica (Long *et al*, 2008). Por ello, es posible que en LTNPs/HICs que presentan HLA-Bw4, las células NKs contribuyan al control en las etapas iniciales de la infección, ya que estos individuos tienen niveles más bajos de viremia que los TPs durante la infección aguda (Goujard *et al*, 2009). Nuestros resultados evidencian un mayor porcentaje de Bw4/Bw4 homocigotos en el grupo LTNP comparado con los individuos sanos, mostrándose además una diferencia entre las subcategorías de LTNPs de acuerdo a su CV (Tablas 4 y 5, Artículo I). Esto confirma el rol positivo de Bw4 observado durante la infección por el VIH (Flores-Villanueva *et al*, 2001).

Aunque hemos hallado nuevos alelos HLA asociados al fenotipo LTNP mediante técnicas convencionales, el genotipado de la región HLA también es posible mediante plataformas de *Illumina* que pueden analizar hasta 576 loci de 96 individuos en una sola carrera, teniendo la capacidad de distinguir locus que son ambiguos para el genotipado realizado por otras técnicas (Ehrenberg *et al*, 2017).

En conclusión, hemos determinado nueve factores genéticos asociados con un efecto “protector” en LTNPs y cuyas frecuencias individuales oscilan de 7 a 30% en la población estudiada, lo cual sugiere que la acumulación de estas variantes determina el fenotipo LTNP. Además, a diferencia del grupo control no infectado, se puede observar que casi un 70% de los LTNPs tuvieron al menos un alelo HLA “protector” y casi un 90% al considerarse los otros SNPs “protectores”. Asimismo, se ha determinado en nuestra cohorte que un paciente LTNP presenta de media hasta 12 veces más alelos HLA y SNPs “protectores” que los individuos sanos. A pesar de ello, estudios previos indican que las variantes genéticas de la región HLA y SNPs “protectores”, sólo explican entre el 14-23% del control viral observado (Pelak *et al*, 2010; Pereyra *et al*, 2010; Mc Laren *et al*, 2015), evidenciando que estos factores no explican en su totalidad el control de la infección por el VIH-1, por lo que nuevos abordajes son necesarios en la investigación de los LTNPs y HICs.

#### 4.1.2. Genotipado de exoma de pacientes LTNP mediante técnicas de alto rendimiento

Las tecnologías de genotipado de alto rendimiento actuales incluyen sistemas basados en *microarrays* fluorescentes (Affymetrix), en microesferas fluorescentes (Luminex, Illumina, Q-dot), métodos Pergelen e Invader (Ragoussis, 2009), así como metodologías basadas en NGS (Kratz y Caminci, 2014). Los GWAS han revolucionado el campo de la genética de enfermedades complejas en la década pasada, ya que evalúan cientos de miles a millones de polimorfismos en los genomas de individuos para identificar asociaciones genotipo-fenotipo (Visscher *et al*, 2017). Los diversos trabajos de GWAS vienen desarrollándose para el estudio de diferentes patologías infecciosas causadas por diversos agentes como: virus de la hepatitis C (VHC) (Matsuura *et al*, 2017), virus de la hepatitis B (Li *et al*, 2016), virus del Epstein Barr (Yao *et al*, 2017). Uno de los éxitos más notables de GWAS en enfermedades infecciosas fue la identificación de variantes de IFNL3 asociada con la eliminación del VHC después del tratamiento (ribavirina e IFN- $\alpha$ ) (Suppiah *et al*, 2009) o con el aclaramiento espontáneo del virus (Thomas *et al*, 2009). En el campo del VIH, el estudio de genomas se viene usando en diversas áreas como la asociación del control de la CV con polimorfismos genéticos (Wei *et al*, 2015), sensibilidad a antirretrovirales (Carr *et al*, 2017) y también para caracterizar individuos no progresores y/o controladores del VIH (Limou *et al*, 2009; Pereyra *et al*, 2010; Guernon *et al*, 2012).

La asociación de los estudios genómicos con la biología no es sencilla, porque la asociación entre un polimorfismo de un locus genómico y un rasgo fenotípico, no implica siempre causalidad (Visscher *et al*, 2017). Sin embargo, desde la aplicación de GWAS para diversas enfermedades, se han podido determinar más de 10.000 asociaciones entre variantes genéticas y uno o más rasgos complejos (Welter *et al*, 2014)

Para el estudio de genomas se emplean por lo general tecnologías basadas en *microarrays* en chip o mediante la secuenciación de genomas completos (Bush y Moore, 2012; Revisado por Tam *et al*, 2019). Las plataformas principales que se han utilizado para la mayoría de los GWAS son los productos de *Illumina* y *Affymetrix*, que ofrecen diferentes enfoques para medir la variación de SNPs (Distefano *et al*, 2011). La primera generación de matrices de genotipado sólo cubre variaciones comunes del genoma (frecuencia  $\geq 5\%$ ) y no podían identificar SNPs, indels y variantes estructurales menos frecuentes que puedan estar asociados con la patogénesis. La nueva generación de matrices de genotipado ofrece el análisis simultáneo de hasta 5 millones de tagSNPs e incorpora el conocimiento del Proyecto *1000 Genomes* (variantes genéticas con una frecuencia de hasta 1%), que amplía la posibilidad de hallar nuevos factores asociados con el curso de la infección por el VIH (Limou y Zangury, 2013). Publicaciones de GWAS han empleado plataformas de *Illumina* para el análisis de factores relacionados con la no progresión al VIH en cohortes, como:

GRIV (Limou *et al*, 2009), *International HIV Controller Study* (Pereyra *et al*, 2010), ACS (van Manen *et al*, 2011), o en estudios de múltiples cohortes (Le Clerc *et al*, 2011). Adicionalmente, la tecnología *Illumina* también se emplea para realizar estudios de genotipado viral de HICs (Caetano *et al*, 2018).

El estudio de variaciones polimórficas en los genes, en particular del conjunto de exones o exoma, es de gran interés en la salud humana porque posiblemente alberga la variación más funcional (Botstein y Risch, 2003). El exoma humano contiene 233.785 exones, de los cuales aproximadamente el 80% tienen menos de 200 pb de longitud, lo que constituye un total de aproximadamente el 1-2 % del genoma completo (Venter *et al*, 2001; Sakharkara *et al*, 2004; Ezkurdia *et al*, 2014). Aunque el exoma comprende una fracción muy pequeña del genoma, se postula que alberga el 85% de las mutaciones que tienen un gran efecto sobre diferentes enfermedades (Majewski *et al*, 2011).

El estudio de exomas mediante *microarrays* en general ha tenido limitaciones para identificar variantes raras (MAF < 1%) en loci que se describieron inicialmente en GWAS previos (Visscher *et al*, 2017). En este trabajo hemos analizado el exoma de individuos LTNP mediante la tecnología de *microarrays* de *Illumina*, con la intención de asociar nuevos SNPs con la no progresión a la infección. Los *microarrays* de *Illumina* seleccionan SNPs basados en el LD (tagSNPs) de las poblaciones de HapMap, enfocándose en variaciones centrales del gen y enriqueciendo la cobertura de la región HLA altamente polimórfica (Limou y Zagury, 2013). Además, utiliza una tecnología basada en *beads* con secuencias de ADN ligeramente más largas para detectar los alelos, proporcionando una mejor especificidad (Bush y Moore, 2012). En nuestro estudio fueron analizados 242.898 SNPs, que es adecuado teniendo en cuenta que previamente se ha propuesto que entre 200.000–300.000 tagSNPs son requeridos para mapear la mayor parte de la variación en el genoma (Judson *et al*, 2002).

El umbral del valor  $p$  de significación estadística de  $5 \times 10^{-8}$  se ha convertido en un estándar en los estudios de GWAS, para frecuencias alélicas menores (MAFs) superiores a 5% en la población europea (Pe'er *et al*, 2008). Sin embargo, para estudios de exoma de individuos con ascendencia europea, los umbrales significativos del valor  $p$  que han sido propuestos son de  $1 \times 10^{-6}$  para polimorfismos con una  $MAF \geq 5\%$  (Fadista *et al*, 2016). Nuestros resultados evidencian una asociación positiva de los polimorfismos rs2395029 (*HCP5*) y rs9368699 (*HLA-B*) con los individuos LTNP, con valores de  $p$  cercanos al umbral estadístico (Tabla 1, Artículo II), de manera similar a GWAS realizados previamente en otras cohortes de LTNP y HICs.

La asociación del polimorfismo rs2395029 de *HCP5* con la progresión lenta y control viral, ha sido ampliamente demostrado en análisis individual o múltiple de diferentes cohortes como Euro-CHAVI, MACS, PRIMO, GRIV, GISHEAL (Fellay *et al*, 2007;

Dalmasso *et al*, 2008; Limou *et al*, 2009; Herbeck *et al*, 2010; Guernon *et al*, 2012). Este polimorfismo muestra una fuerte asociación con la CV y se asocia en menor medida cuando se valora sólo la progresión de la enfermedad como fenotipo de análisis (Fellay *et al*, 2007). La proporción de individuos que presentan el genotipo T/G en el polimorfismo rs2395029 en esta cohorte corresponde al 19.5% (Anexo 1, Tabla S1), parecido a la cohorte estudiada en el Artículo I que presenta un 21,8%. Nuestros análisis muestran que rs2395029 presenta un alto LD con otros polimorfismos del mismo gen y *MICA* (Material suplementario, Artículo II). El gen *MICA* codifica la secuencia A relacionada con el MHC-I, y se ha reportado polimorfismos en este gen que se encuentran en alto LD con el SNP rs2395029, lo cual sugiere un papel de *MICA* en la progresión y el control del VIH (Le Clerc *et al*, 2014). En trabajos previos, este SNP presenta un alto LD con otros SNPs de *BAT1-5*, *HLA-B*, *HLA-C*, *LTB*, *MICB*, *PSORS1C1*, *TNF* y *RDBP* (Limou *et al*, 2009; Limou y Zagury, 2013).

*HCP5* codifica un retrovirus endógeno con homología de secuencia con el gen *pol* del VIH-1, que planteó su papel como ARN antisentido que interfiere con la replicación viral (Kulski y Dawkins, 1999). Sin embargo, se ha observado que inhibiendo la expresión de las dos formas alélicas de este gen no afecta a la infectividad del VIH-1 *in vitro*, lo que sugiere que rs2395029 no es la variante causal directa del fenotipo (Yoon *et al*, 2010). El polimorfismo rs2395029 está en alto LD con HLA-B\*5701, un alelo de HLA “protector” asociado previamente con baja viremia y aparición tardía del sida (Stephens, 2005; van Manen *et al*, 2011). La función de este alelo es la presentación de epítomos específicos de VIH-1 capaces de iniciar una fuerte respuesta de CTLs y el reconocimiento del complejo HLA-péptido por el TCR (Chen *et al*, 2012; Kloverpris *et al*, 2012).

La asociación del SNP rs3819299 del gen *HLA-B* con el fenotipo LTNP, no ha sido reportado previamente. Sin embargo son evaluados mayormente los polimorfismos rs13202464 y rs4349859 del alelo “protector” HLA-B\*27. El polimorfismo rs3819299 se encuentra en alto LD con otros SNPs de *HLA-B*, *ZDHHC20P2* y *DHFRP2* (Material suplementario, Artículo II). En otras investigaciones también se han asociado otros polimorfismos en HLA-B con el control del VIH, rs9266409 (río abajo) (Fellay *et al*, 2009) y rs2523608 (intrón 5) (Pelak *et al*, 2010). El SNP rs2523608 posee un alto LD con HLA-B\*5703, que ha sido asociado al control de la CV en afroamericanos (Pelak *et al*, 2010).

La asociación del polimorfismo rs9368699 de *C6orf48* (MHC-III) con el control viral, ha sido reportada previamente en otra cohorte de individuos VIH-positivos (Euro-CHAVI – MACS) (Nititham *et al*, 2017). De igual manera, estudios en la cohorte GISHEAL (cohorte francesa ALT conjuntamente con la cohorte italiana ELVIS) revelaron que rs9368699 presentan una asociación específica con la condición LTNP (Guernon *et al*, 2012). En nuestra cohorte, un 15% de LTNPs poseen el genotipo CT en este polimorfismo, a diferencia de sólo un 3% en la población control, lo cual es significativo (Anexo I, Tabla 1).

Los valores de  $p$  para los SNPs mencionados, tienen una significación menor comparado con otros trabajos previos. De manera similar, no se observó una significación estadística de los SNPs analizados en la región HLA de 117 individuos hispanos en otra cohorte de estudio (Pereyra *et al*, 2010). Esto se debe a exigencias estadísticas relacionadas con el tamaño muestral de la cohorte (Limou y Zagury, 2013).

Algunos de los genes relacionados con los SNPs más significativos hallados en nuestro trabajo (Tabla I, Artículo II), han mostrado previamente una asociación con la infección por el VIH. Una variante del gen *CSMD1* (rs7840128), ha sido relacionado al trastorno cognitivo asociado al VIH (Jia *et al*, 2017). Por otra parte, se ha descrito que Tat del VIH se asocia a EP300 para regular la actividad transactivadora de Tat y la integración del material genético viral (Vendel y Lumb, 2003). El polimorfismo rs2234358 del receptor de quimiocinas CXCR6 (intrón 14 de *FYCO1*), ha sido asociado con la no progresión a largo plazo al sida (Limou *et al*, 2010). Además, se ha observado una regulación negativa de *FYCO1* en ECs y VCs (Gonzalo-Gil *et al*, 2018).

Por otra parte, mediante los GWAS realizados se han reportado diversos polimorfismos asociados con el control viral en diferentes cohortes, que no han sido encontrados en nuestro trabajo debido a que se encuentran en zonas intergénicas o intrónicas, como son: rs9264942 (cercano al gen *HLA-C*) (Fellay *et al*, 2007); rs8192591 (variante sin sentido de *NOTCH4*); rs9468692 (relacionado a una variante sin sentido de *TRIM10*) (Fellay *et al*, 2009); rs11725412 (entre los genes *TBC1D1* y *KLF3*) (Dalmasso *et al*, 2008); rs3131018 (polimorfismo intrónico de *PSORS1C3*) (Pereyra *et al*, 2010), entre otros.

Los estudios de exoma plantean que variantes génicas fuera de la región HLA no contribuyen significativamente al control de la infección (McLaren *et al*, 2017). Sin embargo, el hallazgo más novedoso fue la alta asociación del SNP rs1127888 del gen *UBXN6* (cromosoma 19) al fenotipo LTNP ( $p=7.34 \times 10^{-11}$ ). Polimorfismos en la región KIR, también en el cromosoma 19, han sido asociados a los HICs (Tomescu *et al*, 2012).

*UBXN6* (*UBX domain-containing protein 6*) es un cofactor de *VCP* (*valosin containing protein*) que está bien conservado en los eucariotas (Carim-Todd *et al*, 2001) y tiene una función celular poco conocida. Sin embargo, se ha demostrado que *UBXN6* es una proteína abundante y estable que se localiza en el núcleo y el citosol (Madsen *et al*, 2008). Además, *UBXN6* conjuntamente con la proteína *VCP* participan en la ubiquitinación y el tráfico endolisosomal de caveolina 1 (*CAV-1*) (Schuberth *et al*, 2008; Ritz *et al*, 2011).

No se ha asociado previamente la relación de *UBXN6* con la infección por el VIH-1 u otras infecciones virales. Sin embargo, se ha descrito que otra proteína de la misma familia (*UBXN1*) interfiere con la respuesta inmune antiviral mediada por RIG-I, al dirigirse a la proteína mitocondrial MAVS e inhibiendo la señal de activación de NF- $\kappa$ B vía TNF (Wang *et al*, 2015). Por otra parte, mutaciones en *VCP* causan un trastorno degenerativo sistémico

denominado IBMPFD (miopatía con cuerpos de inclusión asociada con la enfermedad ósea de Paget y demencia fronto-temporal) (Watts *et al*, 2004; Wehl *et al*, 2009).

Las frecuencias alélicas de rs1127888 fueron diferentes en el grupo LTNP. Los promedios de las frecuencias alélicas de los genotipos CT y TT en los diferentes grupos control, fueron de 36.2 y 5.9%, respectivamente. Sin embargo, en los individuos LTNPs las frecuencias alélicas de los genotipos CT y TT, fueron de 52.6 y 16.3%, respectivamente (Figura 1, Artículo II). Por lo tanto, proponemos que la mayor frecuencia de los alelos TT y CT de rs1127888 en los LTNPs tiene un efecto indirecto en la patogénesis del VIH, a través de mecanismos relacionados con la expresión y localización de CAV-1 en la célula, como resultado de una reducida interacción del complejo CAV-1/UBXN6/VCP (Anexo II, Figura S1). Este déficit en la interacción señalada podría ser debido a la mutación A31T de la región N-terminal de la proteína, que se origina por la presencia de timina en el polimorfismo mencionado (Figura 2, Artículo II). Además, se muestra que el polimorfismo rs1127888 está en alto LD con rs11909, que está incluido en el codón de Ala58 de la región VIM de UBXN6. Un estudio ha evidenciado que la mutación A58L en UBXN6 suprime la formación del complejo UBXN6/VCP (Stapf *et al*, 2011)

CAV-1 es una proteína de andamiaje de 21-24 kDa y componente estructural importante de las caveolas, que son pequeñas invaginaciones de la membrana plasmática compuestas también de colesterol, fosfolípidos y esfingolípidos (Rothberg *et al*, 1992). Esta proteína está altamente expresada en diferentes tipos celulares como células endoteliales, adipocitos, macrófagos y DCs (Galbiati *et al*, 2001; Harris *et al*, 2002), mientras que las células CD4+ no expresan CAV-1 (Wang *et al*, 2011). Se han descrito interacciones entre caveolas y patógenos, bacterias y virus, que podrían haber evolucionado para facilitar su entrada y evitar rutas que conduzcan a su destrucción en la célula huésped (Harris *et al*, 2002). La regulación positiva de CAV-1 durante la infección por el VIH ha sido relacionada con una infección persistente a bajo nivel, principalmente en macrófagos (Mergia, 2017).

La disminución de la expresión de UBXN6 (que evitaría la formación del complejo CAV-1/UBXN6/VCP) causa una disminución de la infección por el VIH-1 en diferentes tipos celulares (Figura 4, Artículo II). CAV-1 puede inhibir la infección por el VIH-1 a través de diferentes mecanismos que han sido descritos previamente (Anexo II, Figura S2). La infección por el VIH-1 produce una regulación positiva de CAV-1 mediada por la proteína Tat del VIH (Lin *et al*, 2010), causando una reducción significativa en la replicación viral a través de la represión transcripcional modulada por NF- $\kappa$ B, reduciendo la activación de IKK $\beta$ , IKK $\alpha$ , I $\kappa$ B $\alpha$  y p65, así como la posterior translocación de la proteína NF- $\kappa$ B p65 al núcleo (Wang *et al*, 2011). Por otra parte, se ha demostrado que CAV-1 interactúa con la gp41 en las balsas lipídicas, bloqueando la fusión de membranas y la apoptosis mediada por la envoltura de VIH (Huang *et al*, 2007; Wang *et al*, 2010). Este mecanismo podría

relacionarse con una localización periférica de CAV-1, que observamos mediante la regulación negativa de UBXN6 (Figura 3, Artículo II).

La regulación del colesterol en la célula es importante para la infectividad del VIH-1. CAV-1 es una molécula importante para el mantenimiento de la homeostasis del colesterol celular, debido a que media el transporte de colesterol recién sintetizado desde el RE a la membrana plasmática (Frank *et al*, 2006) e influye indirectamente en la transferencia a aceptores extracelulares tales como lipoproteína de alta densidad (HDL) o apolipoproteína A-I (apoA-I). La infección por el VIH afecta el flujo de salida de colesterol dependiente de ABCA-1 por apoA-I, debido a que Nef interfiere en la regulación y redistribución de ABCA-1 (Mujawar *et al*, 2006). CAV-1 interacciona con Nef para contrarrestar los efectos de éste en el flujo de colesterol por apoA-I, aumentando así el transporte de colesterol que puede conducir a su agotamiento. Además, se ha descrito que las células presentadoras de antígeno de pacientes progresores poseen más niveles de expresión de ABCA-1 comparadas con las de no progresores (Prasad y Bukrinsky, 2014). Todo este proceso puede influir en el contenido de colesterol incorporado en las partículas virales, que disminuye la infectividad de éstas en los nuevos ciclos de infección (Lin *et al*, 2012).

Además, CAV-1 puede bloquear la vía TLR4 e inhibir la producción de citoquinas proinflamatorias inducidas por LPS (Wang *et al*, 2009). Esto sugiere un papel de CAV-1 en la regulación negativa de las tormentas de citoquinas proinflamatoria en LTNPs, quienes presentan niveles más bajos de genes de las rutas de inflamación comparado con TPs, según lo observado en nuestros análisis del transcriptoma (Artículo III). Los mecanismos descritos podrían darse de manera simultánea y explicar el control virológico parcial, así como el mantenimiento de recuentos estables de linfocitos T CD4+ en los LTNPs.

Por otra parte, filamina A puede modular la expresión y actividad antiviral de teterina, y además puede interactuar directamente con CAV-1 (Dotson *et al*, 2016). Nuestros análisis no evidencian diferencias en la expresión de teterina al disminuir los niveles de UBXN6 (Anexo II, Figura S3), por lo que no sugerimos la participación de teterina en el control de la infección en nuestra cohorte de LTNPs. Sin embargo, otros estudios han asociado variantes genéticas en *RICH2* (que interactúa con teterina) con individuos LTNPs (Le Clerc *et al*, 2011; Paximadis *et al*, 2017).

El efecto de CAV-1, tanto en DCs como en macrófagos, tiene una gran relevancia en el escenario de la infección por el VIH. Aunque las células T CD4+ representan el principal reservorio de la infección persistente, investigaciones realizadas sugieren que los macrófagos también contribuyen significativamente a la persistencia del virus durante la TAR. Los macrófagos están presentes en prácticamente todos los tejidos del organismo, incluidas ubicaciones con muy baja población de células T, como el cerebro, donde la inflamación mediada por la infección puede provocar secuelas patológicas. Además, se ha



reportado que los macrófagos pueden albergar una infección productiva del VIH (Clayton *et al*, 2017; Rodrigues *et al*, 2017). Asimismo, también se ha descrito que las DCs pueden tener una infección productiva del VIH *in vivo*, y la regulación positiva de CAV-1 en estas células reduce los niveles de CD40, cuya activación está asociada a la transmisión del VIH a las células T CD4+ (Fong *et al*, 2002; Li *et al*; 2012).

Otros estudios de exoma en el campo del VIH, han permitido la identificación de: variantes en el gen *CNOT1* como potenciales candidatos para predecir la respuesta al tratamiento con vacunas terapéuticas basadas en DCs (Moura *et al*, 2014), y la variante E88T que causa la pérdida de función de SIGLEC1/CD169 (receptor de superficie de células mieloides), afectando la captura y la trans-infección del VIH-1 (Martinez-Picado *et al*, 2016). Además, la secuenciación de exoma en LTNPs y ECs ha permitido identificar variantes raras en los genes *FGD6*, *FN1*, *MAP1A*, *PIK3C2B*, *PIK3R5*, *PRKCA*, *PRKDC* y *TAB2*, que están involucrados en la detección inmune innata, infectividad dependiente de CD4, tráfico viral y transcripción del VIH, aunque se requieren más estudios para establecer su papel en el control de la infección (Nissen *et al*, 2018).

En conclusión, los estudios de genotipado mediante técnicas de alto rendimiento han generado información importante sobre factores genéticos asociados a LTNPs y HICs. Sin embargo, aún queda por identificar el papel de otros polimorfismos como las variaciones del número de copias y variantes genéticas interactuantes. Además, las restricciones estadísticas derivadas del bajo número de muestras para el análisis (en comparación con otras enfermedades) también influyen en la limitación de estos estudios (Shungin *et al*, 2015; Fuchsberger *et al*, 2016). Por otro lado, aunque actualmente la aplicación clínica del conocimiento derivado de los factores genéticos no modificables que contribuyen al control espontáneo de la infección es limitada, los hallazgos obtenidos sirven para comprender con mayor profundidad los mecanismos asociados a la no progresión y/o control del VIH, que podría contribuir a un diseño personalizado de tratamientos o vacunas en el futuro.

#### **4.2. Identificación de nuevos patrones de expresión génica mediante análisis de transcriptoma en individuos LTNPs**

El análisis del transcriptoma de individuos infectados por el VIH es muy importante, debido a que la expresión génica puede modificarse debido a la infección del virus en las células diana (Peng *et al*, 2014), o también el programa transcripcional de la célula infectada puede influir en la expresión proviral (Bradley *et al*, 2018).

Este trabajo empleó la tecnología *RNA-seq*, cuyas ventajas sobre los *microarrays* de ADNc son: el alto nivel de reproducibilidad de datos que reduce la cantidad de réplicas técnicas, identificación y cuantificación de diferentes isoformas, así como la detección de nuevos transcritos (Agarwal *et al*, 2010). Además, la creciente popularidad de tecnologías NGS ha permitido disminuir el costo de los experimentos (Kratz y Caminci, 2014).

#### 4.2.1. Similitud y heterogeneidad entre LTNPs y TPs

La divergencia Jensen-Shannon es un método estadístico para medir la similitud entre dos grupos (Antell *et al*, 2016), y se ha reportado que esta herramienta puede ser útil para el análisis de datos de expresión génica por *RNA-seq* (Park *et al*, 2016). Nuestros resultados indican una mayor similitud entre los fenotipos no progresores (EC-LTNPs y vLTNPs), y éstos a su vez tienen una distancia menor con los TP-ARTs comparado con TPs (sin TAR) (Figura 1B, Artículo III). De manera similar, se ha descrito que los perfiles transcripcionales de células CD4+ de ECs son similares a los de TP-TARs (Vigneault *et al*, 2011). Además, confirmamos que la infección por el VIH-1 modifica la expresión génica de las células infectadas (Peng *et al*, 2014) y se demuestra que los LTNPs comparten un perfil transcriptómico común a pesar de diferentes grados de control viral.

El análisis de escalamiento multidimensional (MDS) es empleado en estudios de *RNA-seq* y la búsqueda de patrones de DEGs asociados a infecciones virales (Oh *et al*, 2019). Nuestros análisis muestran una gran heterogeneidad en la expresión génica de los individuos con diferentes fenotipos de infección por el VIH, ya que una correcta agrupación fenotípica se dio sólo en el 50% de los individuos (Figura 1C, Artículo III).

#### 4.2.2. Análisis de expresión diferencial entre fenotipos de infección y genes asociados al fenotipo LTNP

En el análisis para determinar DEGs entre los fenotipos de estudio, resaltamos los datos más relevantes. La expresión de *SLC37A3* es superior en LTNP/ECs y diferencial en todas las comparaciones. Mutaciones en este gen se han asociado con el hiperinsulinismo congénito de la infancia (Cappello *et al*, 2018). De manera similar, se ha reportado que otros genes de la familia SLC (*SLC30A1*, *SLC7A5*) presentan una regulación positiva en LTNPs y HICs (Ding *et al*, 2019; Lee *et al*, 2019).

Por otra parte, se han asociado 14 DEGs con la condición LTNP (EC-LTNPs/TP-ARTs vs vLTNPs/TPs) (Figura 7, Artículo III), de los cuales *VWA8*, *ANKRD54* e *IGHA2* no están afectados por la viremia y tampoco son regulados por IFN. Estos tres genes son asociados por primera vez al fenotipo LTNP. *ANKRD54* se encuentra incrementado con mayor diferencia (25 veces) en los LTNPs. Además, se han descrito que proteínas con repeticiones de anquirina interactúan con la proteína Vpr del VIH-1 (Miles *et al*, 2005) y anquirinas artificiales dirigidas a la proteína Gag de la cápsida del VIH-1 muestran un efecto antiviral en la fase posterior a la integración viral (Nangola *et al*, 2012). Además, *ANKRD54* modula el desplazamiento nucleo-citoplasmático de BTK, una proteína tirosina quinasa con un papel fundamental en el desarrollo y activación de linfocitos B (Gustafsson *et al*, 2017). Por otra parte, una mayor expresión de *IGHA2* ha sido asociada a los linfocitos B de memoria CD27- IgA+ (Berkowska *et al*, 2011), cuyo porcentaje en sangre es significativamente mayor en pacientes infectados por el VIH-1 en comparación a individuos

sanos (Cagigi *et al*, 2009). Una menor expresión de este gen en HICs podría asociarse a una menor tasa de hipermutación somática de los anticuerpos durante la infección. La expresión de *ANKRD54* e *IGHA2*, podría ser evaluada específicamente en linfocitos B de LTNPs. Por otra parte, la expresión de *VWA8* se ha asociado con la progresión de algunos tipos de cáncer (Marcucci *et al*, 2014; Yuan *et al*, 2018).

Respecto a los otros genes asociados al fenotipo LTNP: *CMPK2* ha sido relacionado a la restricción del VIH *in vitro* (El-Diwany *et al*, 2018); genes de la familia *IFITM* presentan una actividad antiviral mediante la inhibición de la entrada del VIH (Lee *et al*, 2018); *SIGLEC-1* es un receptor de DCs que media la transinfección del VIH (Izquierdo-Useros *et al*, 2012); *SERPING1* es una proteasa que participa en la vía del complemento de la inmunidad innata y ha sido asociada a la restricción del VIH, presentando una expresión incrementada en monocitos de individuos infectados (Sanfilippo *et al*, 2017).

De manera similar a nuestros resultados, otros análisis de transcriptoma han reportado una mayor expresión de *SERPINB2* en células CD4+ y monocitos de LTNPs comparado con pacientes sanos o TP-ARTs (Wu *et al*, 2013; Zhang *et al*, 2017) y una mayor expresión de *CMPK1* en células T de LTNPs comparado con TPs (Ding *et al*, 2019). Por otra parte, se ha observado mayores niveles de *IFI27*, *IFITM3* e *IFI6* en PBMCs de individuos con mejor recuperación de células CD4+ tras el inicio de la TAR (Woelk *et al*, 2010), que tendría relación con nuestros resultados que indican que una mayor expresión de estos genes pueden contribuir a mantener niveles altos de células T CD4+ en LTNPs.

Además, se evidencia una menor expresión de *FN1* en vLTNPs comparado con TPs, así como una mayor expresión de *SDC3* y *EPHB2* en EC-LTNPs comparado con TP-ARTs (Figura 7, Artículo III). Se ha observado que las variantes R2425H y P2016L de *FN1* (fibronectina-1) pueden reducir sus niveles de expresión en LTNPs (Nissen *et al*, 2018). De manera similar, en otros análisis de transcriptoma se han observado una expresión más elevada de *SCD3* en células CD4 de ECs comparado con TPs en TAR (Vigneault *et al*, 2011), así como una mayor expresión de *EPHB2* en células CD4+ de pacientes resistentes al VIH comparado con los susceptibles a la infección (McLaren *et al*, 2010).

#### 4.2.3. Genes predictores del fenotipo LTNP

Se han implementado modelos bayesianos para la detección de biomarcadores mediante análisis de datos de transcriptoma (Huo *et al*, 2019). Además, los métodos bayesianos jerárquicos han sido aplicados al análisis de datos provenientes de *RNA-Seq* (Lee *et al*, 2015), y utilizando métodos de clasificación supervisada usando esta herramienta hemos determinado un grupo de 20 genes capaces de predecir el fenotipo LTNP con una alta fiabilidad (Figura 2, Artículo III).

En la búsqueda de genes predictores se emplea la validación cruzada LOOCV (*leave one out cross validation*) para determinar la probabilidad de clasificación correcta de un

determinado fenotipo. Esa herramienta ha permitido clasificar individuos que presentan alta o baja recuperación de células CD4+ tras iniciar la TAR, empleando datos de expresión de 40 genes obtenidos por *microarrays* de *Illumina* (Woelk *et al*, 2010). Nuestros análisis alcanzan un nivel de predicción de: 75% si se clasifican los cuatro fenotipos de estudio, 90% si se clasifican los individuos en LTNPs y progresores (con TAR o sin ella), o un 100% si se consideran sólo los LTNPs y TPs (sin TAR) (Figura 1D, Artículo III).

Algunos de estos genes predictores, como *XCCR6*, *HERC5*, *MX1* e *IFI44*, ya habían sido asociados con la infección por el VIH. *XCCR6* codifica la subunidad Ku70 de la proteína Ku, que está involucrada en la represión de la transcripción del VIH-1 (Jeanson y Mouscadet, 2002), lo cual tendría relación con una mayor expresión de *XCCR6* que observamos en el fenotipo LTNP (Figura 2, Artículo III). Por otra parte, la expresión de *HERC5* es inducida por IFN-I y codifica una ligasa celular E3 que inhibe la replicación del VIH-1 al bloquear la producción de Gag (Woods *et al*, 2011) e interferir con la exportación nuclear del ARN dependiente de Rev/RRE (Woods *et al*, 2014). De manera similar, se ha reportado en una menor expresión de *HERC5* y una mayor expresión de *XCCR2* en células T de LTNPs en metanálisis de transcriptoma (Ding *et al*, 2019; Lee *et al*, 2019)

Respecto a *MX1*, se ha publicado que Vpr del VIH induce cambios en la expresión de este gen en macrófagos (Zahoor *et al*, 2014). Asimismo, se ha observado una mayor expresión basal de Mx1 en PBMCs de modelos animales con mayor resistencia a la infección en mucosa por el SIV (Aamer *et al*, 2014). Aunque el mecanismo de la acción antiviral de *MX1* no se comprende completamente, parece interferir con el tráfico de nucleoproteínas virales en las células infectadas (Verhelst *et al*, 2012). Por otra parte, *IFI44* se asocia a la región LTR del VIH-1, suprimiendo la transcripción viral y previniendo la reactivación del VIH-1 latente (Power *et al*, 2015). De manera similar, otros análisis de transcriptoma han reportado una menor expresión de *IFI44*, *MX1*, *EPSTI1* y *PARP12* en células CD4+ y/o CD8+ de HICs comparados con progresores (Hycza *et al*, 2007; Quigley *et al*, 2010; Rotger *et al*, 2011).

Por otra parte, se ha mostrado una alta correlación de EIF3L y EIF4A con el control del VIH-1 en LTNP/ECs (Ding *et al*, 2019). En nuestros análisis, EIF3LP3 tiene una mayor expresión en fenotipos que controlan el virus (EC- LTNPs y TP-ARTs) (Figura 2, Artículo III). Además, otro estudio ha descrito que la señalización vía eIF2 puede jugar un papel importante en la regulación de la función de las células CD8+ de los ECs (Chowdhury *et al*, 2018). Por otra parte, observamos que la mayoría de los genes predictores tienen niveles similares entre EC-LTNPs y TP-ARTs, de manera similar a lo observado previamente entre células CD4+ de ECs y progresores en TAR (Vigneault *et al*, 2011).

La expresión de varios genes predictores ha sido asociada también a otras infecciones virales. *EPSTI1*, *EEF1G* y *MX1* están relacionados a la infección por influenza (Verhelst *et*

*al*, 2012; Sammaibashi *et al*, 2018; Ghobadi *et al*, 2019), *PARP12* y *PARP14* a coronavirus (Grunewald *et al*, 2019), *HELZ2* al virus del dengue (Fusco *et al*, 2017), *lncRNA RP11-288L9.4* regula la infección por el VHC por una vía independiente a JAK/STAT (Liu *et al*, 2019). Para otros genes como: *EEF1B2P3*, *OR56B1*, *RPL4P4*, *RPL5P4* y *RPL4P5*, no se les ha atribuido funciones específicas ni asociación con alguna patología. Sin embargo, componentes ribosomales (*RPS20*, *RPS28*, *RPS15A*, *RPS25*, *RPS6*, *RPS21*, *RPS3*, *RPL36*, *RPL9*, *RPL31*, *RPL23*, *RPL27*, *RPL30*, *RPL29*, *RPL35*) y factores de iniciación traduccional (*EIF2S3*, *EIF1*, *EIF2C2*, *EIF4G2*, *EIF4A3*) presentan niveles más elevados en no progresores en comparación con progresores crónicos (Ding *et al*, 2019). Los defectos en componentes del ribosoma y el bloqueo de iniciación traduccional parecen tener un papel importante en la progresión y/o control del VIH.

#### 4.2.4. Genes regulados por IFN y mecanismos celulares asociados

Los genes estimulados por interferón (ISGs), tanto predictores como DEGs entre la comparación de fenotipos, mostraron niveles bajos y similares entre EC-LTNPs, vLTNPs y TP-ARTs (Figuras 2 y 7, Artículo III). Estos datos correlacionan con metanálisis de transcriptoma que han identificado una reducida expresión de diversos ISGs en células CD4+ y CD8+ de LTNPs (Zhang *et al*, 2017; Ding *et al*, 2019; Lee *et al*, 2019).

Los ISGs: *IFI44*, *ISG15*, *LY6E*, *MX1*, *MX2*, *OAS1*, *OAS2*, *OAS3*, *OASL*, *USP18*, *XAF1* y *TNFSF10* (*TRAIL*), entre otros, descendieron sus niveles tras la administración de la TAR en TPs (Tabla S1 y Figura S3, Artículo III). Otro estudio de transcriptoma ha determinado hasta 234 DEGs por el efecto de la TAR, descendiendo los niveles de *OAS1* y *OAS3* como en nuestros resultados. Además, los genes relacionados con la activación y respuesta antiviral presentaron una alta regulación antes de la TAR (da Conceicao *et al*, 2014). Por otra parte, un estudio de *RNA-seq* ha observado una menor expresión de *TRAIL* en LTNPs y ECs comparado con progresores (Lee *et al*, 2019; Paim *et al*, 2019).

Además, se muestra también una menor expresión de los genes: *EPSTI1*, *ISG15*, *LY6E*, *MX1*, *MX2*, *OASL*, *OTOF*, *RNASE1*, *USP18* y *XAF1* en vLTNPs comparado con TPs (Figura 7, Artículo III), así como una regulación significativa de los mecanismos de respuesta a IFN-I entre ambos fenotipos (Anexo III, Figura S4). De manera similar, diferentes análisis de transcriptoma observan una regulación negativa de *ISG15*, *LY6E*, *MX1*, *OAS1/2/3*, *OTOF*, *RNASE2*, *USP18*, *XAF1*, entre otros ISGs, en células T y/o monocitos de LTNP/HICs comparado con progresores del VIH (Quigley *et al* 2010; Rotger *et al*, 2011; Wu *et al*, 2013; Zhang *et al*, 2017; Ding *et al*, 2019; Lee *et al*, 2019).

Durante la fase aguda de la infección, el IFN-I reduce la replicación del VIH a través de la respuesta innata, pero en la fase crónica contribuye a los niveles de inflamación, depleción de las células T y progresión de la enfermedad (Revisado por Cheng *et al*, 2017). Se ha descrito que LTNPs poseen respuestas proinflamatorias reducidas, debido a

variantes en las vías TLR y NOD2 (Nissen *et al*, 2018). Además, se ha observado que los HICs tienen niveles bajos de marcadores de activación como IP-10, MIG y MIP-1 $\beta$ . Además, la expresión de citoquinas inflamatorias correlacionan negativamente con los niveles de células T CD4+ en VCs (Platten *et al*, 2016). Sin embargo, los VCs tienen una menor expresión de marcadores de activación, senescencia y apoptosis comparados con LTNPs (Gaardbo *et al*, 2013), por lo que diferentes mecanismos estarían involucrados en la conservación de células T CD4+ de LTNPs y HICs.

Por otra parte, otro trabajo de transcriptoma de PBMCs ha reportado que una regulación positiva de *IL-8*, *HIG2* y *SIK2* en LTNPs, así como otros genes que se asocian al control de la activación, proliferación, inhibición de la apoptosis y supervivencia de células T (*HSH2D*, *STAT5B*). Además, los LTNPs presentan una menor expresión de *FOXO3A* y *CTSS* comparado con TP-ARTs (Luque *et al*, 2014). Otra investigación ha identificado una mayor expresión de las vías canónicas Wnt/-catenin, AKT y MAPK, involucradas en la supervivencia celular en células T de LTNPs (Wu *et al*, 2011).

#### 4.2.5. La regulación del calcio (Ca<sup>+2</sup>) celular en los EC-LTNPs

El Ca<sup>+2</sup> intracelular participa en la señalización celular, la función mitocondrial y la muerte celular (Duchen, 2000; Contreras *et al*, 2010). La absorción del Ca<sup>+2</sup> por las mitocondrias activa los enzimas del ciclo de Krebs y la fosforilación oxidativa, lo que conduce a una mayor producción de ATP (Nasr *et al*, 2003). El Ca<sup>+2</sup> también puede actuar como un segundo mensajero en células como los linfocitos, que en reposo mantienen una baja concentración de Ca<sup>+2</sup> y se incrementa tras diversos estímulos antigénicos para cumplir con sus funciones de defensa inmunológica (Vig *et al*, 2009). Además, la vía Ca<sup>+2</sup>/calcineurina puede regular el factor nuclear de células T activadas (NFAT) que a su vez modula diversas funciones de respuesta inmunológica (Revisado por Vaeth y Feske, 2018).

Se han descrito que algunos virus regulan las concentraciones del Ca<sup>+2</sup> en el citoplasma y mitocondrias de la célula huésped, lo cual permite la expresión de genes virales, la replicación del virus y el control de la viabilidad celular (Foti *et al*, 1999; Gong *et al*, 2001). La proteína Nef del VIH afecta la señalización del Ca<sup>+2</sup> intracelular al interactuar con una familia de tirosina quinasas (Foti *et al*, 1999) y también mediante la interacción Nef/IP3R1 (receptor del inositol trifosfato) (Manninen *et al*, 2002). Investigaciones recientes proponen la intervención farmacológica del canal del Ca<sup>+2</sup> o la liberación de Ca<sup>+2</sup> desde el RE como estrategias para el desarrollo de potentes antivirales (Chen *et al*, 2019).

Se ha determinado 58 DEGs entre EC-LTNPs y TP-ARTs, así como una regulación positiva de mecanismos involucrados a la movilización y transporte del Ca<sup>+2</sup> en los EC-LTNPs. Además, la mayoría de los DEGs poseen sitios de unión transcripcional a NFAT/EIk-1 en sus promotores (Figura 4, Artículo III). Nuestros resultados indican que los genes *NR4A1*, *FPR1* y *C5AR1* están sobreexpresados en el fenotipo EC-LTNPs (Figura

7, Artículo III). De manera similar, en otros estudios se han evidenciado que *FPR1* y *NR4A2* están sobreexpresados en células CD4<sup>+</sup> de LTNPs al compararlos con pacientes no infectados (Lee *et al*, 2019). Asimismo, se ha reportado una mayor expresión de *C5AR1* en células CD4 de no progresores comparado con progresores (Zhang *et al*, 2017; Lee *et al*, 2019). Por otra parte, otros estudios de transcriptoma han descrito que la vía PI3K-Akt, asociada a la señalización y transporte del Ca<sup>+2</sup>, está positivamente regulada en células T de LTNPs y ECs (Zhang *et al*, 2017; Chowdhury *et al*, 2018). Por último, se ha observado una mayor expresión de CALM-1 en PBMCs de LTNPs comparado con TPs, que expresa la calmodulina que participa en la transducción de la señal del Ca<sup>+2</sup> (Luque *et al*, 2014).

#### 4.2.6. Regulación y función de CDKN1A/p21 en el control viral

La proteína p21 (codificada por *CDKN1A*) funciona como un potente inhibidor de las quinasas dependientes de ciclinas, un grupo de enzimas del huésped requeridos para la replicación efectiva del VIH-1 y otros virus (Schang *et al*, 2002; Schang, 2006). Además, p21 es un componente celular endógeno en células madre que proporciona una barrera molecular contra el VIH-1 (Zhang *et al*, 2007). Asimismo, puede restringir el VIH-1 en DCs mediante la reducción de la biosíntesis de desoxinucleósido trifosfato, la regulación de la actividad antiviral de SAMHD1 y la regulación negativa de varios enzimas involucrados en la biosíntesis de dNTPs (Valle-Casuso *et al*, 2017).

Nuestros resultados evidencian una mayor expresión de *CDKN1A* en EC-LTNPs comparado con otros fenotipos de progresión (Figura 5, Artículo III). De manera similar, se ha observado una mayor expresión de este gen en células CD4<sup>+</sup> de HICs comparado con individuos sanos o progresores del VIH (Sáez-Cirión *et al*, 2011; Vigneault *et al*, 2011). Las células T CD4<sup>+</sup> de pacientes HICs son capaces de resistir la infección debido a una mayor expresión de p21 que inhibe la fosforilación dependiente de CDK2 de la transcriptasa inversa del VIH-1 y reduce significativamente la eficacia de la transcripción reversa (Leng *et al*, 2014). Además, p21 inactiva la actividad enzimática de CDK9, componente del complejo P-TEFb junto con la ciclina T1, regulando la elongación del ARNm viral (Chen *et al*, 2011). Por otra parte, un metanálisis reciente ha descrito que una mayor expresión de *RBM38* puede regular la estabilidad de p21 en los ECs (Lee *et al*, 2019).

La expresión de *CDKN1A* está altamente correlacionada con la de *GADD45B*, *IER3* y *TNF* en EC-LTNPs (Figura 5, Artículo III). Además, se ha determinado que estas proteínas pueden interactuar con 13 proteínas del VIH-1 (Figura 6, Artículo III). Por otra parte, se determinó una correlación positiva de la expresión de estos genes en PBMCs de ECs de la cohorte francesa CODEX (Anexo III, Figura S5), que reafirma su papel en el control viral.

Se ha descrito que la sobreexpresión de proteínas GADD45A/B reduce la producción viral a través de la supresión de la transcripción del promotor LTR viral (Liang *et al*, 2016). Además, un estudio reciente evidencia una regulación positiva de *GADD45A* en células T

de LTNP/ECs comparado con progresores del VIH (Hycza *et al*, 2007; Zhang *et al*, 2017; Ding *et al*, 2019). Por otra parte, el papel de TNF ha sido ampliamente descrito en la infección por el VIH (Pasquereau *et al*, 2017), y su implicación funcional en la transcripción viral está relacionada con la regulación de la expresión de *IER3* (Wu *et al*, 1998). *IER3* inhibe la proteína fosfatasa 2A (PP2A), que a su vez induce la transcripción basal del VIH (Faulkner *et al*, 2003). De manera similar, se ha observado que la expresión de *IER3* es superior en células CD8+ de ECs comparado con no progresores, y en monocitos de LTNP comparado con progresores en TAR (Rotger *et al*, 2011; Wu *et al*, 2013).

Profundizando en el estudio de *CDKN1A*, evaluamos su expresión en poblaciones de células T de pacientes sanos. Las células CD4+ expresan mayores niveles de *CDKN1A* comparado con las CD8+. Además, las células CD4+ *naïve* expresan menos este gen comparado con células de memoria central, memoria efectora y terminales diferenciadas (Anexo III, Figura S6). En PBMCs estimuladas con IL-7, involucrada en la proliferación homeostática de células T (Chomont *et al*, 2009), se observa un aumento de la expresión de *CDKN1A* en células T CD4+ totales y en las subpoblaciones analizadas (Anexo III, Figura S6). Asimismo, también observamos que aumenta los niveles de fosforilación de SAMHD1 (Anexo 2, Figura S7), así como la infección viral en las poblaciones celulares estudiadas (Anexo III, Figura S8). Por tanto, proponemos que la actividad antiviral de *CDKN1A/p21* en células CD4+ de LTNP/ECs sería en un contexto de reducida activación celular y que su expresión podría ser regulada por mecanismos epigenéticos. Además, la expresión de *CDKN1A* en subpoblaciones de células CD4 de individuos sanos parece no estar asociadas con los niveles de infectividad del VIH-1.

Se han descrito algunas características de las subpoblaciones de T CD4+ de LTNP y HICs. Se ha determinado que la pérdida de células T CD4+ *naïve* entre ECs y TPs es similar, a pesar de la capacidad de los ECs de mantener CVs indetectables (Yang *et al*, 2012). Además, se ha descrito que las células CD4+ de memoria central tienen una mejor preservación, y mayores niveles de CCR7 y de IL7R en HICs (Potter *et al*, 2007), así como una menor expresión de CCR5 en un grupo de LTNP (Descours *et al*, 2012). Por último, las células de memoria central y efectora de ECs presentan una menor susceptibilidad a la apoptosis mediada por Fas (van Grevenynghe *et al*, 2008). Sin embargo, el estudio del transcriptoma en subpoblaciones celulares de estos individuos ha sido poco abordado.

#### 4.3. Consideraciones finales

La investigación en el área del VIH que relaciona variantes y expresión génicas con la progresión y el control de la infección, ha proporcionado un mayor conocimiento sobre la patogénesis del VIH-1, evidenciando que los mecanismos antivirales y respuestas inmunológicas son necesarias para la defensa del hospedador. La existencia de individuos capaces de controlar y/o no progresar en la infección por el VIH-1 ha planteado la



posibilidad de utilizar vacunas terapéuticas que induzcan la generación de respuestas inmunes efectivas similares a los LTNP/ECs (Thèze *et al*, 2011). Sin embargo, el logro de un completo control de la infección por los LTNPs/HICs no puede ser explicado por la presencia de un solo factor, sino que tiene un carácter multifactorial (Thèze *et al*, 2011).

Se ha propuesto que individuos EC-LTNPs con varios alelos HLA “protectores”, niveles estables de células CD4+, prolongado control viral y ausencia de evolución del virus, pueden haber alcanzado la cura funcional del VIH (López-Galíndez *et al*, 2019). Por lo que, el estudio de los EC-LTNPs evaluados en este trabajo puede complementarse con análisis virológicos e inmunológicos en un mayor tiempo de seguimiento, con el fin de determinar una posible cura funcional de estos individuos en el futuro.

Por otra parte, los análisis de transcriptoma de PBMCs han permitido obtener información para el manejo clínico de pacientes VIH-positivos, así como la identificación de genes cuyo rol en la infección todavía no se comprende. Por lo que, es necesario realizar análisis funcionales sobre estos genes con función desconocida, así como estudios de transcriptoma en subpoblaciones celulares o de célula única de los LTNP/HICs. De esta manera se podrá relacionar la expresión génica con parámetros virológicos y/o inmunológicos para encontrar nuevas alternativas para el tratamiento de la infección por el VIH, ya que se ha propuesto que la integración de datos “ómicos” con sistemas experimentales podría ser el futuro de la medicina personalizada (Bush y Moore, 2012).

En el campo de las ómicas para el estudio del VIH pueden implementarse análisis adicionales. La integración de datos genómicos del hospedador y del virus ha asociado SNPs de la región HLA-I con 48 variantes de aminoácidos del VIH-1 (Bartha *et al*, 2013). Otro análisis de integración del perfil transcriptómico de ARNm y microARN de personas infectadas por el VIH, ha permitido determinar rutas y biomarcadores involucrados en la neurodegeneración asociada a la infección por el VIH (Zhou *et al*, 2012). Para otras enfermedades, los datos de GWAS han permitido desarrollar sistemas de análisis que permiten predecir la expresión génica y mecanismos moleculares asociados, así como correlacionar niveles de metabolitos en sangre con variantes génicas específicas de enzimas involucrados en su metabolismo (Illig *et al*, 2010; Gamazon *et al*, 2015).

En conclusión, los resultados de esta tesis doctoral muestran nuevas asociaciones de alelos HLA y el polimorfismo rs1127888 del gen *UBXN6* con el fenotipo LTNP. Además, el análisis de transcriptoma ha permitido la identificación de genes predictores que clasifican fenotipos LTNP con alta eficacia, así como la regulación positiva de genes involucrados en el metabolismo del calcio y una expresión coordinada de los genes *CDKN1A*, *GADD45B*, *IER3* y *TNF* en EC-LTNPs. Sin embargo, se necesitan más estudios para contribuir a un mayor conocimiento en la inmunopatogenia del VIH-1, una de las enfermedades infecciosas más importantes de los últimos tiempos.

# CONCLUSIONES

## 5. CONCLUSIONES

- 1) En el estudio de pacientes VIH-positivos con una progresión lenta de la infección a largo plazo (LTNPs), hemos identificado cinco nuevas asociaciones de alelos en la región HLA con este fenotipo: el alelo “protector” HLA-B\*39 y los alelos de “riesgo” HLA-A\*24, -A\*29, -B\*08 y -B\*18.
- 2) Los pacientes con fenotipo LTNP presentan de media 12 veces más alelos HLA y SNPs “protectores” comparado con la población VIH-negativa.
- 3) Se ha identificado un nuevo polimorfismo asociado a la no progresión del VIH-1. Los genotipos TT y TC en el SNP rs1127888 del gen *UBXN6* presentan mayores frecuencias alélicas en LTNPs en comparación con los distintos grupos control de pacientes sanos o infectados por el VIH-1.
- 4) En experimentos “*in vitro*”, la disminución de la expresión de *UBXN6* incrementa los niveles y la localización periférica de caveolina-1 en la célula, y además causa una disminución de los niveles de infección del VIH-1 en varios tipos celulares.
- 5) Los fenotipos LTNP virémico (vLTNP), LTNP controlador de élite (EC-LTNP) y progresor típico con o sin tratamiento antirretroviral (TP-ART y TP, respectivamente), presentan una expresión genética diferencial. Se han identificado catorce genes asociados al fenotipo LTNP, independientemente del control viral. Además, se ha confirmado la expresión diferencial de genes regulados por interferón tipo I entre los fenotipos mencionados.
- 6) La similitud en los patrones de expresión genética entre individuos vLTNPs y EC-LTNPs es mayor respecto a los TPs. A su vez, los TP-ARTs tienen una mayor similitud con los EC-LTNPs cuando se comparan frente a los TPs sin TAR.
- 7) Se han identificado veinte genes y pseudogenes predictores del fenotipo LTNP mediante análisis bayesianos de clasificación supervisada, que permite clasificar a individuos LTNPs y TPs hasta con un 90% de exactitud.
- 8) Los individuos EC-LTNPs presentan una regulación positiva de genes relacionados con las rutas de transporte y movilización del calcio en la célula, al compararlos con TP-ARTs.
- 9) Los individuos EC-LTNPs presentan una expresión coordinada de los genes *CDKN1A*, *IER3*, *GADD45B* y *TNF*, a diferencia de los otros grupos de pacientes infectados por el VIH-1 que no presentan una expresión coordinada de estos cuatro genes.

# REFERENCIAS

## 6. REFERENCIAS

- Aamer H, Rajakumar P, Nyaundi J, Murphey-Corb M. Resistance to simian immunodeficiency virus low dose rectal challenge is associated with higher constitutive TRIM5 $\alpha$  expression in PBMC. *Retrovirology*. 2014;11(1):39. doi:10.1186/1742-4690-11-39
- Abdel-Mohsen M, Raposo R, Deng X et al. Expression profile of host restriction factors in HIV-1 elite controllers. *Retrovirology*. 2013;10(1):106. doi:10.1186/1742-4690-10-106
- Agarwal A, Koppstein D, Rozowsky J et al. Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays. *BMC Genomics*. 2010;11(1):383. doi:10.1186/1471-2164-11-383
- Alcamí J, Coiras M. Inmunopatogenia de la infección por el virus de la inmunodeficiencia humana. *Enferm Infecc y Microbiol Clin*. 2011;29(3):216-226. doi:10.1016/j.eimc.2011.01.006
- Althaf M, El Kossi M, Jin J, Sharma A, Halawa A. Human leukocyte antigen typing and crossmatch: A comprehensive review. *World J Transplant*. 2017;7(6):339-348. doi:10.5500/wjt.v7.i6.339
- Anqin M, Volant S, Passaes C et al. Metabolic plasticity of HIV-specific CD8+ T cells is associated with enhanced antiviral potential and natural control of HIV-1 infection. *Nat Metab*. 2019;1(7):704-716. doi:10.1038/s42255-019-0081-4
- Antell G, Dampier W, Aiamkitsumrit B et al. Utilization of HIV-1 envelope V3 to identify X4- and R5-specific Tat and LTR sequence signatures. *Retrovirology*. 2016;13(1). doi:10.1186/s12977-016-0266-9
- Audigé A, Taffé P, Rickenbach M et al. Low Postseroconversion CD4 Count and Rapid Decrease of CD4 Density Identify HIV+ Fast Progressors. *AIDS Res Hum Retroviruses*. 2010;26(9):997-1005. doi:10.1089/aid.2009.0263
- Autran B, Descours B, Avettand-Fenoel V, Rouzioux C. Elite controllers as a model of functional cure. *Curr Opin HIV AIDS*. 2011;6(3):181-187. doi:10.1097/coh.0b013e328345a328
- Balasubramaniam M, Pandhare J, Dash C. Are microRNAs Important Players in HIV-1 Infection? An Update. *Viruses*. 2018;10(3):110. doi:10.3390/v10030110
- Ballana E, Senserrich J, Pauls E et al. ZNRD1 (zinc ribbon domain-containing 1) is a host cellular factor that influences HIV-1 replication and disease progression. *Clin Infect Dis*. 2010;50(7):1022-32. doi:10.1086/651114
- Bansal A, Sterrett S, Erdmann N et al. Normal T-cell activation in elite controllers with preserved CD4+ T-cell counts. *AIDS*. 2015;29(17):2245-2254. doi:10.1097/qad.0000000000000860
- Barblu L, Machmach K, Gras C et al. Plasmacytoid Dendritic Cells (pDCs) From HIV Controllers Produce Interferon- $\alpha$  and Differentiate Into Functional Killer pDCs Under HIV Activation. *J Infect Dis*. 2012;206(5):790-801. doi:10.1093/infdis/jis384
- Barre-Sinoussi F, Chermann J, Rey F et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*. 1983;220(4599):868-871. doi:10.1126/science.6189183
- Bartha I, Carlson J, Brumme C, et al. A genome to genome analysis of associations between human genetic variation, HIV-1 sequence diversity and viral control. *Elife*. 2013; 29;2:e01123. doi: 10.7554/eLife.01123
- Berger E, Doms R, Fenyo E et al. A new classification for HIV-1. *Nature*. 1998;391(6664):240-240. doi:10.1038/34571
- Basmaciogullari S, Pizzato M. The activity of Nef on HIV-1 infectivity. *Front Microbiol*. 2014;5. doi:10.3389/fmicb.2014.00232
- Berger G, N'guyen X, Cordeil S et al. The HIV-2/SIVSM Vpx protein and an antiviral restriction in primary myeloid cells. *Retrovirology*. 2009;6(S2). doi:10.1186/1742-4690-6-s2-p12
- Berkowska M, Driessen G, Bikos V et al. Human memory B cells originate from three distinct germinal center-dependent and -independent maturation pathways. *Blood*. 2011;118(8):2150-2158. doi:10.1182/blood-2011-04-345579
- Betts M, Nason M, West S et al. HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells. *Blood*. 2006;107(12):4781-4789. doi:10.1182/blood-2005-12-4818
- Björkström N, Béziat V, Cichocki F et al. CD8 T cells express randomly selected KIRs with distinct specificities compared with NK cells. *Blood*. 2012;120(17):3455-3465. doi:10.1182/blood-2012-03-416867
- Blankson J, Bailey J, Thayil S et al. Isolation and Characterization of Replication-Competent Human Immunodeficiency Virus Type 1 from a Subset of Elite Suppressors. *J Virol*. 2006;81(5):2508-2518. doi:10.1128/jvi.02165-06
- Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet*. 2003;33(S3):228-237. doi:10.1038/ng1090
- Boufassa F, Saez-Cirion A, Lechenade J et al. CD4 Dynamics over a 15 Year-Period among HIV Controllers Enrolled in the ANRS French Observatory. *PLoS ONE*. 2011;6(4):e18726. doi:10.1371/journal.pone.0018726
- Boulet S, Song R, Kanya P et al. HIV Protective KIR3DL1 and HLA-B Genotypes Influence NK Cell Function Following Stimulation with HLA-Devoid Cells. *J Immunol*. 2010;184(4):2057-2064. doi:10.4049/jimmunol.0902621
- Bradley T, Ferrari G, Haynes B, Margolis D, Browne E. Single-Cell Analysis of Quiescent HIV Infection Reveals Host Transcriptional Profiles that Regulate Proviral Latency. *Cell Rep*. 2018;25(1):107-117.e3. doi:10.1016/j.celrep.2018.09.020

- Buchbinder S, Vittinghoff E. HIV-infected long-term nonprogressors: epidemiology, mechanisms of delayed progression, and clinical and research implications. *Microbes Infect.* 1999;1(13):1113-1120. doi:10.1016/s1286-4579(99)00204-x
- Buckheit R, Allen T, Alme A et al. Host factors dictate control of viral replication in two HIV-1 controller/chronic progressor transmission pairs. *Nat Commun.* 2012;3(1). doi:10.1038/ncomms1697
- Buonaguro L, Buonaguro F, Giraldo G, Ensoli B. The human immunodeficiency virus type 1 Tat protein transactivates tumor necrosis factor beta gene expression through a TAR-like structure. *J Virol.* 1994;68(4):2677-2682. doi:10.1128/jvi.68.4.2677-2682.1994
- Bush W, Moore J. Chapter 11: Genome-Wide Association Studies. *PLoS Comput Biol.* 2012;8(12):e1002822. doi:10.1371/journal.pcbi.1002822
- Cabrera-Rodríguez R, Hebmann V, Marfil S et al. HIV-1 envelope glycoproteins isolated from Viremic Non-Progressor individuals are fully functional and cytopathic. *Sci Rep.* 2019;9(1). doi:10.1038/s41598-019-42075-3
- Caetano D, Côrtes F, Bello G et al. Next-generation sequencing analyses of the emergence and maintenance of mutations in CTL epitopes in HIV controllers with differential viremia control. *Retrovirology.* 2018;15(1). doi:10.1186/s12977-018-0444-z
- Cagigi A, Du L, Dang L et al. CD27- B-Cells Produce Class Switched and Somatic Hyper-Mutated Antibodies during Chronic HIV-1 Infection. *PLoS ONE.* 2009;4(5):e5427. doi:10.1371/journal.pone.0005427
- Cappello A, Curcio R, Lappano R, Maggiolini M, Dolce V. The Physiopathological Role of the Exchangers Belonging to the SLC37 Family. *Front Chem.* 2018;6. doi:10.3389/fchem.2018.00122
- Carim-Todd L, Escarceller M, Estivill X, Sumoy L. Identification and characterization of UBXD1, a novel UBX domain-containing gene on human chromosome 19p13, and its mouse ortholog. *Biochim Biophys Acta.* 2001;1517(2):298-301. doi:10.1016/s0167-4781(00)00248-7
- Carpenter C, Fischl MA, Hammer SM et al. Antiretroviral Therapy for HIV Infection in 1996. *JAMA.* 1996;276(2):146. doi:10.1001/jama.1996.03540020068031
- Carr D, Bourgeois S, Chaponda M et al. Genome-wide association study of nevirapine hypersensitivity in a sub-Saharan African HIV-infected population. *J Antimicrob Chemother.* 2017:dkw545. doi:10.1093/jac/dkw545
- Casado C, Colombo S, Rauch A et al. Host and Viral Genetic Correlates of Clinical Definitions of HIV-1 Disease Progression. *PLoS ONE.* 2010;5(6):e11079. doi:10.1371/journal.pone.0011079
- Casado C, Marrero-Hernández S, Márquez-Arce D et al. Viral Characteristics Associated with the Clinical Nonprogressor Phenotype Are Inherited by Viruses from a Cluster of HIV-1 Elite Controllers. *mBio.* 2018;9(2). doi:10.1128/mbio.02338-17
- Casado C, Galvez C, Pernas M et al. Permanent control of HIV-1 pathogenesis in exceptional elite controllers: a model of spontaneous cure. *Sci Rep.* 2020;10(1). doi:10.1038/s41598-020-58696-y
- Casartelli N, Di Matteo G, Argentini C et al. Structural defects and variations in the HIV-1 nef gene from rapid, slow and non-progressor children. *AIDS.* 2003;17(9):1291-1301. doi:10.1097/00002030-200306130-00003
- Cevik M, Orkin C. Insights into HIV-1 capsid inhibitors in preclinical and early clinical development as antiretroviral agents. *Expert Opin Investig Drugs.* 2019;28(12):1021-1024. doi:10.1080/13543784.2019.1692811
- Chan E, Qian W, Diamond D et al. Quantitative Analysis of Human Immunodeficiency Virus Type 1-Infected CD4+ Cell Proteome: Dysregulated Cell Cycle Progression and Nuclear Transport Coincide with Robust Virus Production. *J Virol.* 2007;81(14):7571-7583. doi:10.1128/jvi.00288-07
- Checkley M, Luttge B, Freed E. HIV-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation. *J Mol Biol.* 2011;410(4):582-608. doi:10.1016/j.jmb.2011.04.042
- Chen H, Li C, Huang J et al. CD4+ T cells from elite controllers resist HIV-1 infection by selective upregulation of p21. *J Clin Invest.* 2011;121(4):1549-1560. doi:10.1172/jci44539
- Chen H, Ndhlovu Z, Liu D et al. TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. *Nat Immunol.* 2012;13(7):691-700. doi:10.1038/ni.2342
- Chen D, Balamurugan A, Ng H, Cumberland W, Yang O. Epitope targeting and viral inoculum are determinants of Nef-mediated immune evasion of HIV-1 from cytotoxic T lymphocytes. *Blood.* 2012;120(1):100-111. doi:10.1182/blood-2013-04-499822
- Chen X, Cao R, Zhong W. Host Calcium Channels and Pumps in Viral Infections. *Cells.* 2019;9(1):94. doi:10.3390/cells9010094
- Cheng L, Yu H, Li G et al. Type I interferons suppress viral replication but contribute to T cell depletion and dysfunction during chronic HIV-1 infection. *JCI Insight.* 2017;2(12). doi:10.1172/jci.insight.94366
- Chevalier M, Julg B, Pyo A et al. HIV-1-Specific Interleukin-21+ CD4+ T Cell Responses Contribute to Durable Viral Control through the Modulation of HIV-Specific CD8+ T Cell Function. *J Virol.* 2010;85(2):733-741. doi:10.1128/jvi.02030-10
- Chikata T, Murakoshi H, Koyanaqi M et al. Control of HIV-1 by an HLA-B\*52:01-C\*12:02 Protective Haplotype. *J Infect Dis.* 2017;216(11):1415-1424. doi:10.1093/infdis/jix483
- Chomont N, El-Far M, Ancuta P et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat Med.* 2009;15(8):893-900. doi:10.1038/nm.1972
- Chou J, Ohsumi T, Geha R. Use of whole exome and genome sequencing in the identification of genetic causes of primary immunodeficiencies. *Curr Opin Allergy Clin Immunol.* 2012;12(6):623-628. doi:10.1097/acf.0b013e3283588ca6

- Chowdhury F, Ouyang Z, Buzon M, Walker B, Lichtenfeld M, Yu X. Metabolic pathway activation distinguishes transcriptional signatures of CD8+ T cells from HIV-1 elite controllers. *AIDS*. 2018;32(18):2669-2677. doi:10.1097/qad.0000000000002007
- Chun T, Moir S, Fauci A. HIV reservoirs as obstacles and opportunities for an HIV cure. *Nat Immunol*. 2015;16(6):584-589. doi:10.1038/ni.3152
- Clavel F, Guetard D, Brun-Vezinet F et al. Isolation of a new human retrovirus from West African patients with AIDS. *Science*. 1986;233(4761):343-346. doi:10.1126/science.2425430
- Clayton K, Garcia V, Clements J, Walker B. HIV Infection of Macrophages: Implications for Pathogenesis and Cure. *Pathogens and Immunity*. 2017;2(2):179. doi:10.20411/10.20411/pai.v2i2.204
- Cleqq A, Ashton L, Biti R et al. CCR5 promoter polymorphisms, CCR5 59029A and CCR5 59353C, are under represented in HIV-1-infected long-term non-progressors. *AIDS*. 2000;14(2):103-108. doi:10.1097/00002030-200001280-00004
- Clerc S, Coulonges C, Delaneau O et al. Screening Low-Frequency SNPs From Genome-Wide Association Study Reveals a New Risk Allele for Progression to AIDS. 2011;56(3):279-284. doi:10.1097/qai.0b013e318204982b
- Cohen M, Shaw G, McMichael A, Haynes F. Acute HIV-1 Infection. *N Engl J Med*. 2011; 364(20): 1943–1954.
- Coiras M, Camafeita E, Ureña T et al. Modifications in the human T cell proteome induced by intracellular HIV-1 Tat protein expression. *Proteomics*. 2006;6(S1):S63-S73. doi:10.1002/pmic.200500437
- Contreras L, Drago I, Zampese E, Pozzan T. Mitochondria: The calcium connection. *Biochim Biophys Acta*. 2010;1797(6-7):607-618. doi:10.1016/j.bbabi.2010.05.005
- Craigie R, Bushman F. HIV DNA Integration. *Cold Spring Harb Perspect Med*. 2012;2(7):a006890-a006890. doi:10.1101/cshperspect.a006890
- Crum N, Riffenburgh R, Wegner S et al. Comparisons of Causes of Death and Mortality Rates Among HIV-Infected Persons. *JAIDS*. 2006;41(2):194-200. doi:10.1097/01.qai.0000179459.31562.16
- da Conceicao V, Dyer W, Gandhi K, Gupta P, Saksena N. Genome-wide analysis of primary peripheral blood mononuclear cells from HIV+ patients-pre-and post-HAART show immune activation and inflammation the main drivers of host gene expression. *Mol Cell Ther*. 2014;2(1):11. doi:10.1186/2052-8426-2-11
- Dalmasso C, Carpentier W, Meyer L et al. Distinct Genetic Loci Control Plasma HIV-RNA and Cellular HIV-DNA Levels in HIV-1 Infection: The ANRS Genome Wide Association 01 Study. *PLoS ONE*. 2008;3(12):e3907. doi:10.1371/journal.pone.0003907
- Dambaya B, Nguéfeu Nkenfou C, Mekue L, et al. TRIM5 $\alpha$  136Q, CCR5 Promoter 59029G And CCR264I Alleles Impact The Progression Of HIV In Children And Adolescents. *Appl Clin Genet*. 2019; 12: 203–211. doi: 10.2147/TACG.S205335
- Dayton A. Within you, without you: HIV-1 Rev and RNA export. *Retrovirology*. 2004; 1: 35. doi: 10.1186/1742-4690-1-35
- de Sorrentino A, Marinic K, Motta P, Sorrentino A, López R, Illioich E. HLA Class I Alleles Associated with Susceptibility or Resistance to Human Immunodeficiency Virus Type 1 Infection among a Population in Chaco Province, Argentina. *J Infect Dis*. 2000;182(5):1523-1526. doi:10.1086/315854
- Deacon N, Tsykin A, Solomon A et al. Genomic Structure of an Attenuated Quasi Species of HIV-1 from a Blood Transfusion Donor and Recipients. *Science*. 1995;270(5238):988-991. doi:10.1126/science.270.5238.988
- Dean M, Carrington M, Winkler C et al. Genetic Restriction of HIV-1 Infection and Progression to AIDS by a Deletion Allele of the CCR5 Structural Gene. *Science*. 1996;273(5283):1856-1862. doi:10.1126/science.273.5283.1856
- Deeks S, Walker B. Human Immunodeficiency Virus Controllers: Mechanisms of Durable Virus Control in the Absence of Antiretroviral Therapy. *Immunity*. 2007;27(3):406-416. doi:10.1016/j.immuni.2007.08.010
- Deeks S. Shock and kill. *Nature*. 2012;487(7408):439-440. doi:10.1038/487439a
- Descours B, Avettand-Fenoel V, Blanc C et al. Immune Responses Driven by Protective Human Leukocyte Antigen Alleles From Long-term Nonprogressors Are Associated With Low HIV Reservoir in Central Memory CD4 T Cells. *Clinical Infectious Diseases*. 2012;54(10):1495-1503. doi:10.1093/cid/cis188
- Deval H, Alagarasu K, Mittal M et al. Association of single nucleotide polymorphisms in TNFA and CCR5 genes with Japanese Encephalitis: A study from an endemic region of North India. *J Neuroimmunol*. 2019;336:577043. doi:10.1016/j.jneuroim.2019.577043
- Dey R, Soni K, Saravanan S et al. Anti-HIV microRNA expression in a novel Indian cohort. *Sci Rep*. 2016;6(1). doi:10.1038/srep28279
- Ding J, Ma L, Zhao J et al. An integrative genomic analysis of transcriptional profiles identifies characteristic genes and patterns in HIV-infected long-term non-progressors and elite controllers. *J Transl Med*. 2019;17(1). doi:10.1186/s12967-019-1777-7
- Distefano J, Taverna D. Technological issues and experimental design of gene association studies. *Methods Mol Biol*. 2011;700: 3–16. doi: 10.1007/978-1-61737-954-3 1
- Dominguez-Molina B, Tarancon-Diez L, Hua S et al. HLA-B\*57 and IFNL4-related polymorphisms are associated with protection against HIV-1 disease progression in controllers. *Clin Infect Dis*. 2016:ciw833. doi:10.1093/cid/ciw833
- Dominguez-Molina B, Leon A, Rodriguez C et al. Analysis of Non-AIDS-Defining Events in HIV Controllers. *Clin Infect Dis*. 2016. doi: 10.1093/cid/ciw120.
- Doranz B, Rucker J, Yi Y et al. A Dual-Tropic Primary HIV-1 Isolate That Uses Fusin and the  $\beta$ -Chemokine Receptors CKR-5, CKR-3, and CKR-2b as Fusion Cofactors. *Cell*. 1996;85(7):1149-1158. doi:10.1016/s0092-8674(00)81314-8

- Dotson D, Woodruff E, Villalta F, Dong X. Filamin A Is Involved in HIV-1 Vpu-mediated Evasion of Host Restriction by Modulating Tetherin Expression. *J Biol Chem.* 2016;291(8):4236-4246. doi:10.1074/jbc.m115.708123
- Duchon M. Mitochondria and calcium: from cell signalling to cell death. *J Physiol.* 2000;529(1):57-68. doi:10.1111/j.1469-7793.2000.00057.x
- Egaña-Gorroño L, Escribà T, Boulanger N et al. Differential MicroRNA Expression Profile between Stimulated PBMCs from HIV-1 Infected Elite Controllers and Viremic Progressors. *PLoS ONE.* 2014;9(9):e106360. doi:10.1371/journal.pone.0106360
- Ehrenberg P, Geretz A, Sindhu R et al. High-throughput next-generation sequencing to genotype six classical HLA loci from 96 donors in a single MiSeq run. *HLA.* 2017;90(5):284-291. doi:10.1111/tan.13133
- El-Diwany R, Soliman M, Sugawara S et al. CMPK2 and BCL-G are associated with type 1 interferon-induced HIV restriction in humans. *Sci Adv.* 2018;4(8):eaat0843. doi:10.1126/sciadv.aat0843
- Ezkurdia I, Juan D, Rodriguez J et al. Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Hum Mol Genet.* 2014;23(22):5866-5878. doi:10.1093/hmg/ddu309
- Fadista J, Manning A, Florez J, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet.* 2016;24(8):1202-1205. doi:10.1038/ejhg.2015.269
- Faria N, Rambaut A, Suchard M et al. The early spread and epidemic ignition of HIV-1 in human populations. *Science.* 2014;346(6205):56-61. doi:10.1126/science.1256739
- Farquhar C, Rowland-Jones S, Mbori-Ngacha D et al. Human Leukocyte Antigen (HLA) B\*18 and Protection against Mother-to-Child HIV Type 1 Transmission. *AIDS Res Hum Retrovir.* 2004;20(7):692-697. doi:10.1089/0889222041524616
- Faulkner N, Lane B, Bock P, Markovitz D. Protein Phosphatase 2A Enhances Activation of Human Immunodeficiency Virus Type 1 by Phorbol Myristate Acetate. *J Virol.* 2003;77(3):2276-2281. doi:10.1128/jvi.77.3.2276-2281.2003
- Fauriat C, Andersson S, Björklund A et al. Estimation of the Size of the Alloreactive NK Cell Repertoire: Studies in Individuals Homozygous for the Group A KIR Haplotype. *J Immunol.* 2008;181(9):6010-6019. doi:10.4049/jimmunol.181.9.6010
- Fellay J, Ge D, Shianna K et al. Common Genetic Variation and the Control of HIV-1 in Humans. *PLoS Genet.* 2009;5(12):e1000791. doi:10.1371/journal.pgen.1000791
- Fellay J, Shianna K, Ge D et al. A Whole-Genome Association Study of Major Determinants for Host Control of HIV-1. *Science.* 2007;317(5840):944-947. doi:10.1126/science.1143767
- Fernandez J, Machado A, Lyonnais S et al. Transportin-1 binds to the HIV-1 capsid via a nuclear localization signal and triggers coating. *Nat Microbiol.* 2019;4(11):1840-1850. doi:10.1038/s41564-019-0575-6
- Ferre A, Hunt P, McConnell D et al. HIV Controllers with HLA-DRB1\*13 and HLA-DQB1\*06 Alleles Have Strong, Polyfunctional Mucosal CD4+ T-Cell Responses. *J Virol.* 2010;84(21):11020-11029. doi:10.1128/jvi.00980-10
- Flores-Villanueva P, Yunis E, Delgado J et al. Control of HIV-1 viremia and protection from AIDS are associated with HLA-Bw4 homozygosity. *PNAS.* 2001;98(9):5140-5145. doi:10.1073/pnas.071548198
- Foley B, Leitner T, Apetrei C et al. HIV Sequence Compendium 2018. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, LA-UR 18-25673.
- Fong L, Mengozzi M, Abbey N, Herndier B, Engleman E. Productive Infection of Plasmacytoid Dendritic Cells with Human Immunodeficiency Virus Type 1 Is Triggered by CD40 Ligation. *J Virol.* 2002;76(21):11033-11041. doi:10.1128/jvi.76.21.11033-11041.2002
- Foti M, Cartier L, Piquet V et al. The HIV Nef Protein Alters Ca<sup>2+</sup> Signaling in Myelomonocytic Cells through SH3-mediated Protein-Protein Interactions. *J Biol Chem.* 1999;274(49):34765-34772. doi:10.1074/jbc.274.49.34765
- Frank P, Cheung M, Pavides S, Llaverias G, Park D, Lisanti M. Caveolin-1 and regulation of cellular cholesterol homeostasis. *Am J Physiol Heart Circ Physiol.* 2006;291(2):H677-H686. doi:10.1152/ajpheart.01092.2005
- Freed E. HIV-1 assembly, release and maturation. *Nat Rev Microbiol.* 2015;13(8):484-496. doi:10.1038/nrmicro3490
- Fuchsberger C, Flannick J, Teslovich T, et al. The genetic architecture of type 2 diabetes. *Nature.* 2016; 536 (7614), 41–47. doi: 10.1038/nature18642
- Fumagalli M, Pozzoli U, Cagliani R et al. Genome-Wide Identification of Susceptibility Alleles for Viral Infections through a Population Genetics Approach. *PLoS Genet.* 2010;6(2):e1000849. doi:10.1371/journal.pgen.1000849
- Furutsuki T, Hosoya N, Kawana-Tachikawa A et al. Frequent Transmission of Cytotoxic-T-Lymphocyte Escape Mutants of Human Immunodeficiency Virus Type 1 in the Highly HLA-A24-Positive Japanese Population. *J Virol.* 2004;78(16):8437-8445. doi:10.1128/jvi.78.16.8437-8445.2004
- Fusco D, Pratt H, Kandilas S et al. HELZ2 Is an IFN Effector Mediating Suppression of Dengue Virus. *Front Microbiol.* 2017;8. doi:10.3389/fmicb.2017.00240
- Gaardbo J, Hartling H, Gerstoft J, Nielsen S. Thirty Years with HIV Infection—Nonprogression Is Still Puzzling: Lessons to Be Learned from Controllers and Long-Term Nonprogressors. *AIDS Res Treat.* 2012;2012:1-14. doi:10.1155/2012/161584
- Gaardbo J, Hartling H, Ronit A et al. Different Immunological Phenotypes Associated with Preserved CD4+ T Cell Counts in HIV-Infected Controllers and Viremic Long Term Non-Progressors. *PLoS ONE.* 2013;8(5):e63744. doi:10.1371/journal.pone.0063744



- Galbiati F, Volonte' D, Liu J et al. Caveolin-1 Expression Negatively Regulates Cell Cycle Progression by Inducing G0/G1 Arrest via a p53/p21/WAF1/Cip1-dependent Mechanism. *Mol Biol Cell*. 2001;12(8):2229-2244. doi:10.1091/mbc.12.8.2229
- Galperin M, Farenc C, Mukhopadhyay M et al. CD4 + T cell-mediated HLA class II cross-restriction in HIV controllers. *Sci Immunol*. 2018;3(24):eaat0687. doi:10.1126/sciimmunol.aat0687
- Gamazon E, Wheeler H, Shah K et al. Agene-based association method for mapping traits using reference transcriptome data. *Nat Genet*. 2015;47(9):1091-1098. doi:10.1038/ng.3367
- Gao F, Bailes E, Robertson D et al. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature*. 1999;397(6718):436-441. doi:10.1038/17130
- Gao X, Bashirova A, Iversen A et al. AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. *Nat Med*. 2005;11(12):1290-1292. doi:10.1038/nm1333
- Garcia M, Górgolas M, Cabello A et al. Peripheral T follicular helper Cells Make a Difference in HIV Reservoir Size between Elite Controllers and Patients on Successful cART. *Sci Rep*. 2017;7(1). doi:10.1038/s41598-017-17057-y
- Garcia-Merino, I., de las Cuevas, N., Jimenez, J. L., Gallego, J., Gomez, C., Prieto, C., ... (shivbb), S. H. B. (2009). *The Spanish HIV BioBank: a model of cooperative HIV research*. *Retrovirology*, 6(1), 27. doi:10.1186/1742-4690-6-27
- Gea-Banacloche J, Miqueles S, Martino L et al. Maintenance of Large Numbers of Virus-Specific CD8+ T Cells in HIV-Infected Progressors and Long-Term Nonprogressors. *J Immunol*. 2000;165(2):1082-1092. doi:10.4049/jimmunol.165.2.1082
- Geijtenbeek T, Kwon D, Torensma R et al. DC-SIGN, a Dendritic Cell-Specific HIV-1-Binding Protein that Enhances trans-Infection of T Cells. *Cell*. 2000;100(5):587-597. doi:10.1016/s0092-8674(00)80694-7
- Gilbert M, Rambaut A, Wasiuk G, Spira T, Pitchenik A, Worobey M. The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci*. 2007;104(47):18566-18570. doi:10.1073/pnas.0705329104
- Golumbeanu M, Cristinelli S, Rato S et al. Single-Cell RNA-Seq Reveals Transcriptional Heterogeneity in Latent and Reactivated HIV-Infected Cells. *Cell Rep*. 2018;23(4):942-950. doi:10.1016/j.celrep.2018.03.102
- Gong G, Waris G, Tanveer R, Siddiqui A. Human hepatitis C virus NS5A protein alters intracellular calcium levels, induces oxidative stress, and activates STAT-3 and NF- B. *PNAS*. 2001;98(17):9599-9604. doi:10.1073/pnas.171311298
- González M. The HIV-1 Vpr Protein: A Multifaceted Target for Therapeutic Intervention. *Int J Mol Sci*. 2017;18(1):126. doi:10.3390/ijms18010126
- Gonzalo-Gil E, Rapuano P, Ikediobi U et al. Transcriptional down-regulation of ccr5 in a subset of HIV+ controllers and their family members. *Elife*. 2019;8. doi:10.7554/elife.44360
- Gottlieb M, Schroff R, Schanker H et al. Pneumocystis Carinii Pneumonia and Mucosal Candidiasis in Previously Healthy Homosexual Men. Evidence of a New Acquired Cellular Immunodeficiency. *J Urol*. 1982;128(2):444-444. doi:10.1016/s0022-5347(17)52990-0
- Goujard C, Chaix M, Lambotte O et al. Spontaneous Control of Viral Replication during Primary HIV Infection: When Is "HIV Controller" Status Established?. *Clin Infect Dis*. 2009;49(6):982-986. doi:10.1086/605504
- Goulder P, Walker B. HIV and HLA Class I: An Evolving Relationship. *Immunity*. 2012;37(3):426-440. doi:10.1016/j.immuni.2012.09.005
- Grabar S, Selinger-Leneman H, Abgrall S, Pialoux G, Weiss L, Costagliola D. Prevalence and comparative characteristics of long-term nonprogressors and HIV controller patients in the French Hospital Database on HIV. *AIDS*. 2009;23(9):1163-1169. doi:10.1097/qad.0b013e32832b44c8
- Graf E, Mexas A, Yu J et al. Elite Suppressors Harbor Low Levels of Integrated HIV DNA and High Levels of 2-LTR Circular HIV DNA Compared to HIV+ Patients On and Off HAART. *PLoS Pathog*. 2011;7(2):e1001300. doi:10.1371/journal.ppat.1001300
- Griffith O, Montgomery S, Bernier B et al. ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic Acids Res*. 2007;36(Database):D107-D113. doi:10.1093/nar/gkm967
- Grunewald M, Chen Y, Kuny C et al. The coronavirus macrodomain is required to prevent PARP-mediated inhibition of virus replication and enhancement of IFN expression. *PLoS Pathog*. 2019;15(5):e1007756. doi:10.1371/journal.ppat.1007756
- Guernon J, Dalmasso C, Broet P et al. Single-Nucleotide Polymorphism-Defined Class I and Class III Major Histocompatibility Complex Genetic Subregions Contribute to Natural Long-term Nonprogression in HIV Infection. *J Infect Dis*. 2012;205(5):718-724. doi:10.1093/infdis/jir833
- Gupta R, Abdul-Jawad S, McCoy L et al. HIV-1 remission following CCR5Δ32/Δ32 haematopoietic stem-cell transplantation. *Nature*. 2019;568(7751):244-248. doi:10.1038/s41586-019-1027-4
- Gurdasani D, Iles L, Dillon D et al. A systematic review of definitions of extreme phenotypes of HIV control and progression. *AIDS*. 2014;28(2):149-162. doi:10.1097/qad.000000000000049
- Gustafsson M, Mohammad D, Ylösmäki E et al. ANKRD54 preferentially selects Bruton's Tyrosine Kinase (BTK) from a Human Src-Homology 3 (SH3) domain library. *PLoS ONE*. 2017;12(4):e0174909. doi:10.1371/journal.pone.0174909
- Harris J, Werling D, Hope J, Taylor G, Howard C. Caveolae and caveolin in immune cells: distribution and functions. *Trends Immunol*. 2002;23(3):158-164. doi:10.1016/s1471-4906(01)02161-5
- Hatano H, Delwart E, Norris P et al. Evidence for Persistent Low-Level Viremia in Individuals Who Control Human Immunodeficiency Virus in the Absence of Antiretroviral Therapy. *J Virol*. 2008;83(1):329-335. doi:10.1128/jvi.01763-08

- Hendel H, Hénon N, Lebuanec H et al. Distinctive Effects of CCR5, CCR2, and SDF1 Genetic Polymorphisms in AIDS Progression. *J Acquir Immune Defic Syndr Hum Retrovirology*. 1998;19(4):381-386. doi:10.1097/00042560-199812010-00009
- Herbeck J, Gottlieb G, Winkler C et al. Multistage Genomewide Association Study Identifies a Locus at 1q41 Associated with Rate of HIV-1 Disease Progression to Clinical AIDS. *J Infect Dis*. 2010;201(4):618-626. doi:10.1086/649842
- Hersperger A, Martin J, Shin L et al. Increased HIV-specific CD8+ T-cell cytotoxic potential in HIV elite controllers is associated with T-bet expression. *Blood*. 2011;117(14):3799-3808. doi:10.1182/blood-2010-12-322727
- Hosomichi K, Shiina T, Tajima A, Inoue I. The impact of next-generation sequencing technologies on HLA research. *J Hum Genet*. 2015;60(11):665-673. doi:10.1038/jhg.2015.102
- Hsuan H, Lichterfeld M. Recent progress in understanding HIV reservoirs. *Curr Opin HIV AIDS*. 2018; 13(2): 137–142. doi:10.1097/COH.0000000000000441.
- Hu W, Hughes S. HIV-1 Reverse Transcription. *Cold Spring Harb Perspect Med*. 2012;2(10):a006882-a006882. doi:10.1101/cshperspect.a006882
- Huang J, Burke P, Cung T et al. Leukocyte Immunoglobulin-Like Receptors Maintain Unique Antigen-Presenting Properties of Circulating Myeloid Dendritic Cells in HIV-1-Infected Elite Controllers. *J Virol*. 2010;84(18):9463-9471. doi:10.1128/jvi.01009-10
- Huang J, Lu L, Lu H, Chen X, Jiang S, Chen Y. Identification of the HIV-1 gp41 Core-binding Motif in the Scaffolding Domain of Caveolin-1. *J Biol Chem*. 2006; 282(9):6143-6152. doi:10.1074/jbc.m607701200
- Hunt P. Natural Control of HIV-1 Replication and Long-Term Nonprogression: Overlapping but Distinct Phenotypes. *J Infect Dis*. 2009;200(11):1636-1638. doi:10.1086/646610
- Huo Z, Song C, Tseng G. Bayesian latent hierarchical model for transcriptomic meta-analysis to detect biomarkers with clustered meta-patterns of differential expression signals. *Ann Appl Stat*. 2019;13(1):340-366. doi:10.1214/18-aos1188
- Hütter G, Nowak D, Mossner M et al. Long-Term Control of HIV by CCR5Delta32/Delta32 Stem-Cell Transplantation. *N Engl J Med*. 2009;360(7):692-698. doi:10.1056/nejmoa0802905
- Ibáñez L, Velli P, Font R et al. HIV-Infection, Atherosclerosis and the Inflammatory Pathway: Candidate Gene Study in a Spanish HIV-Infected Population. *PLoS ONE*. 2014;9(11):e112279. doi:10.1371/journal.pone.0112279
- Illig T, Gieger C, Zhai G et al. A genome-wide perspective of genetic variation in human metabolism. *Nat Genet*. 2010; 42 (2), 137–141. doi: 10.1038/ng.507
- Imbeault M, Giquère K, Ouellet M, Tremblay M. Exon Level Transcriptomic Profiling of HIV-1-Infected CD4+ T Cells Reveals Virus-Induced Genes and Host Environment Favorable for Viral Replication. *PLoS Pathog*. 2012;8(8):e1002861. doi:10.1371/journal.ppat.1002861
- Izquierdo-Useros N, Lorzate M, Puertas M et al. Siglec-1 Is a Novel Dendritic Cell Receptor That Mediates HIV-1 Trans-Infection Through Recognition of Viral Membrane Gangliosides. *PLoS Biol*. 2012;10(12):e1001448. doi:10.1371/journal.pbio.1001448
- Jeanson L, Mouscadet J. Ku Represses the HIV-1 Transcription. *J Biol Chem*. 2002;277(7):4918-4924. doi:10.1074/jbc.m110830200
- Jia P, Zhao Z, Hulgán T et al. Genome-wide association study of HIV-associated neurocognitive disorder (HAND): A CHARTER group study. *Am J Med Genet B Neuropsychiatr Genet*. 2017; 174(4):413-426. doi:10.1002/ajmg.b.32530
- Jiang Z, Wang H, Michal J et al. Genome Wide Sampling Sequencing for SNP Genotyping: Methods, Challenges and Future Development. *Int J Biol Sci*. 2016;12(1):100-108. doi:10.7150/ijbs.13498
- Jin X, Brooks A, Chen H, Bennett R, Reichman R, Smith H. APOBEC3G/CEM15 (hA3G) mRNA Levels Associate Inversely with Human Immunodeficiency Virus Viremia. *J Virol*. 2005;79(17):11513-11516. doi:10.1128/jvi.79.17.11513-11516.2005
- Judson R, Salisbury B, Schneider J, Windemuth A, Stephens J. How many SNPs does a genome-wide haplotype map require?. *Pharmacogenomics*. 2002;3(3):379-391. doi:10.1517/14622416.3.3.379
- Kam J, Stoltzfus C. Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression. *Cold Spring Harb Perspect Med*. 2011;2(2):a006916-a006916. doi:10.1101/cshperspect.a006916
- Kaslow R, vanRaden M, Friedman H et al. A1, Cw7, B8, DR3 HLA antigen combination associated with rapid decline of T-helper lymphocytes in HIV-1 infection. *Lancet*. 1990;335(8695):927-930. doi:10.1016/0140-6736(90)90995-h
- Kaslow R, Carrington M, Apple R et al. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat Med*. 1996;2(4):405-411. doi:10.1038/nm0496-405
- Kaufmann D, Kavanagh D, Pereyra F et al. Upregulation of CTLA-4 by HIV-specific CD4+ T cells correlates with disease progression and defines a reversible immune dysfunction. *Nat Immunol*. 2007;8(11):1246-1254. doi:10.1038/ni1515
- Kikuchi T, Iwabu Y, Tada T et al. Anti-APOBEC3G Activity of HIV-1 Vif Protein Is Attenuated in Elite Controllers. *J Virol*. 2015;89(9):4992-5001. doi:10.1128/jvi.03464-14
- Kilpatrick DC, Haque RA, Yap PL, et al. HLA antigen frequencies in children born to HIV-infected mothers. *Dis Markers* 1991; 9:21–6. PMID: 1742942
- Kim S, Misra A. SNP Genotyping: Technologies and Biomedical Applications. *Annu Rev Biomed Eng*. 2007;9(1):289-320. doi:10.1146/annurev.bioeng.9.060906.152037
- Kim K, Lee S, Son J et al. Identification of novel genes associated with HIV-1 latency by analysis of histone modifications. *Hum Genomics*. 2017;11(1). doi:10.1186/s40246-017-0105-7

- Klein M, Miedema F. Long-term survivors of HIV-1 infection. *Trends Microbiol.* 1995;3(10):386-391. doi:10.1016/s0966-842x(00)88984-2
- Kløverpris H, Harndahl M, Carlson J et al. HIV control through a single nucleotide on the HLA-I locus. *Retrovirology.* 2012;9(S2). doi:10.1186/1742-4690-9-s2-o47
- Kourteva Y, De Pasquale M, Allos T, McMunn C, D'Aquila R. APOBEC3G expression and hypermutation are inversely associated with human immunodeficiency virus type 1 (HIV-1) burden in vivo. *Virology.* 2012;430(1):1-9. doi:10.1016/j.virol.2012.03.018
- Kratz A, Carninci P. The devil in the details of RNA-seq. *Nat Biotechnol.* 2014; 32(9):882 - 884. <https://doi.org/10.1038/nbt.3015> PMID: 25203036
- Kulkarni S, Savan R, Qi Y et al. Differential microRNA regulation of HLA-C expression and its association with HIV control. *Nature.* 2011;472(7344):495-498. doi:10.1038/nature09914
- Kulkarni A, Kurle S, Shete A et al. Indian Long-term Non-Progressors Show Broad ADCC Responses with Preferential Recognition of V3 Region of Envelope and a Region from Tat Protein. *Front Immunol.* 2017;8. doi:10.3389/fimmu.2017.00005
- Kulski J, Dawkins R. The P5 multicopy gene family in the MHC is related in sequence to human endogenous retroviruses HERV-L and HERV-16. *Immunogenetics.* 1999;49(5):404-412. doi:10.1007/s002510050513
- Ladell K, Hashimoto M, Iglesias M et al. A Molecular Basis for the Control of Preimmune Escape Variants by HIV-Specific CD8+ T Cells. *Immunity.* 2013;38(3):425-436. doi: 10.1016/j.immuni.2012.11.021
- Lambotte O, Boufassa F, Madec Y et al. HIV Controllers: A Homogeneous Group of HIV-1-Infected Patients with Spontaneous Control of Viral Replication. *Clin Infect Dis.* 2005;41(7):1053-1056. doi:10.1086/433188
- Lamine A, Caumont-Sarcos A, Chaix M et al. Replication-competent HIV strains infect HIV controllers despite undetectable viremia (ANRS EP36 study). *AIDS.* 2007;21(8):1043-1045. doi:10.1097/qad.0b013e3280d5a7ac
- Langford S, Ananworanich J, Cooper D. Predictors of disease progression in HIV infection: a review. *AIDS Res Ther.* 2007;4(1):11. doi:10.1186/1742-6405-4-11
- Lassen K, Lobritz M, Bailey J et al. Elite Suppressor-Derived HIV-1 Envelope Glycoproteins Exhibit Reduced Entry Efficiency and Kinetics. *PLoS Pathog.* 2009;5(4):e1000377. doi:10.1371/journal.ppat.1000377
- Le Clerc S, Delaneau O, Coulonges C et al. Evidence After Imputation for a Role of MICA Variants in Nonprogression and Elite Control of HIV Type 1 Infection. *J Infect Dis.* 2014;210(12):1946-1950. doi:10.1093/infdis/jiu342
- Lee J, Ji Y, Liang S, Cai G, Müller P. Bayesian Hierarchical Model for Differential Gene Expression Using RNA-Seq Data. *Stat Biosci.* 2015;7(1):48-67. doi:10.1007/s12561-013-9096-7
- Lee S, Park Y, Yoon C, Kim K, Kim K. Meta-analysis of gene expression profiles in long-term non-progressors infected with HIV-1. *BMC Med Genomics.* 2019;12(1). doi:10.1186/s12920-018-0443-x
- Lee W, Fu R, Liang C, Sloan R. IFITM proteins inhibit HIV-1 protein synthesis. *Sci Rep.* 2018;8(1). doi:10.1038/s41598-018-32785-5
- Lemey P, Pybus O, Wang B, Saksena N, Salemi M, Vandamme A. Tracing the origin and history of the HIV-2 epidemic. *Proc Natl Acad Sci.* 2003;100(11):6588-6592. doi:10.1073/pnas.0936469100
- Leng J, Ho H, Buzon M et al. A Cell-Intrinsic Inhibitor of HIV-1 Reverse Transcription in CD4+ T Cells from Elite Controllers. *Cell Host Microbe.* 2014;15(6):717-728. doi:10.1016/j.chom.2014.05.011
- Leon A, Perez I, Ruiz-Mateos E et al. Rate and predictors of progression in elite and viremic HIV-1 controllers. *AIDS.* 2016;30(8):1209-1220. doi:10.1097/qad.0000000000001050
- Leth S, Schleimann M, Nissen S et al. Combined effect of Vacc-4x, recombinant human granulocyte macrophage colony-stimulating factor vaccination, and romidepsin on the HIV-1 reservoir (REDUC): a single-arm, phase 1B/2A trial. *The Lancet HIV.* 2016;3(10):e463-e472. doi:10.1016/s2352-3018(16)30055-8
- Levy J, Hoffman A, Kramer S. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science.* 1984 Aug 24;225(4664):840-2.
- Li J, Geng S, Xie X et al. Caveolin-1-Mediated Negative Signaling Plays a Critical Role in the Induction of Regulatory Dendritic Cells by DNA and Protein Coimmunization. *J Immunol.* 2012;189(6):2852-2859. doi:10.4049/jimmunol.1102828
- Li G, De Clercq E. HIV Genome-Wide Protein Associations: a Review of 30 Years of Research. *Microbiol Mol Biol Rev.* 2016;80(3):679-731. doi:10.1128/mmbr.00065-15
- Li M, Bai A, Xie G et al. Association of Chemotactic Chemokine Ligand 5 Polymorphisms with the Risk of Developing Severe Enterovirus 71 Infection. *Am J Trop Med Hyg.* 2015;93(4):709-713. doi:10.4269/ajtmh.14-0745
- Li Y, Si L, Zhai Y, et al. Genome-wide association study identifies 8p21.3 associated with persistent hepatitis B virus infection among Chinese. *Nat Commun.* 2016; 31;7:11664.
- Liang Z, Liu R, Zhang H et al. GADD45 proteins inhibit HIV-1 replication through specific suppression of HIV-1 transcription. *Virology.* 2016;493:1-11. doi:10.1016/j.virol.2016.02.014
- Lichterfeld M, Kaufmann D, Yu X et al. Loss of HIV-1-specific CD8+ T Cell Proliferation after Acute HIV-1 Infection and Restoration by Vaccine-induced HIV-1-specific CD4+ T Cells. *J Exp Med.* 2004;200(6):701-712. doi:10.1084/jem.20041270
- Lichterfeld M. Multidimensional profiling of HIV-1 reservoir cells. 2019. Comunicación oral. XI Congreso Nacional Gesida.
- Limou S, Le Clerc S, Coulonges C et al. Genomewide Association Study of an AIDS-Nonprogression Cohort Emphasizes the Role Played by HLA Genes (ANRS Genomewide Association Study 02). *J Infect Dis.* 2009;199(3):419-426. doi:10.1086/596067

- Limou S, Coulonges C, Herbeck J et al. Multiple-Cohort Genetic Association Study Reveals CXCR6 as a New Chemokine Receptor Involved in Long-Term Nonprogression to AIDS. *J Infect Dis.* 2010;202(6):908-915. doi:10.1086/655782
- Limou S, Zagury J. Immunogenetics: Genome-Wide Association of Non-Progressive HIV and Viral Load Control: HLA Genes and Beyond. *Front Immunol.* 2013;4. doi:10.3389/fimmu.2013.00118
- Lin S, Wang X, Nadeau P, Mergia A. HIV Infection Upregulates Caveolin 1 Expression To Restrict Virus Production. *J Virol.* 2010;84(18):9487-9496. doi:10.1128/jvi.00763-10
- Lin S, Nadeau P, Wang X, Mergia A. Caveolin-1 reduces HIV-1 infectivity by restoration of HIV Nef mediated impairment of cholesterol efflux by apoA-I. *Retrovirology.* 2012;9(1):85. doi:10.1186/1742-4690-9-85
- Liu R, Paxton W, Choe S et al. Homozygous Defect in HIV-1 Coreceptor Accounts for Resistance of Some Multiply-Exposed Individuals to HIV-1 Infection. *Cell.* 1996;86(3):367-377. doi:10.1016/s0092-8674(00)80110-5
- Liu X, Duan X, Holmes J et al. A Long Noncoding RNA Regulates Hepatitis C Virus Infection Through Interferon Alpha-Inducible Protein 6. *Hepatology.* 2019;69(3):1004-1019. doi: 10.1002/hep.30266
- Long B, Ndhlovu L, Oksenberg J et al. Conferral of Enhanced Natural Killer Cell Function by KIR3DS1 in Early Human Immunodeficiency Virus Type 1 Infection. *J Virol.* 2008;82(10):4785-4792. doi:10.1128/jvi.02449-07
- Loureiro Dos Reis M, Queiroz M, da Silva B et al. IL6 and FAS/FASL gene polymorphisms maybe associated with disease progression in HIV-1-positive ethnically mixed patients. *J Med Virol.* 2019. doi: 10.1002/jmv.25651.
- Lu W, Chen S, Lai C et al. Suppression of HIV Replication by CD8+ Regulatory T-Cells in Elite Controllers. *Front Immunol.* 2016;7. doi:10.3389/fimmu.2016.00134
- Luque M, Santos C, Mairena E et al. Gene expression profile in long-term non progressor HIV infected patients: In search of potential resistance factors. *Mol Immunol.* 2014;62(1):63-70. doi:10.1016/j.molimm.2014.05.016
- Madec Y, Boufassa F, Porter K, Meyer L. Spontaneous control of viral load and CD4 cell count progression among HIV-1 seroconverters. *AIDS.* 2005;19(17):2001-2007. doi:10.1097/01.aids.0000194134.28135.cd
- Madsen L, Andersen K, Prag S et al. Ubx1 is a novel co-factor of the human p97 ATPase. *Int J Biochem Cell Biol.* 2008;40(12):2927-2942. doi:10.1016/j.biocel.2008.06.008
- Magierowska M, Theodorou I, Debré P et al. Combined Genotypes of CCR5, CCR2, SDF1, and HLA Genes Can Predict the Long-Term Nonprogressor Status in Human Immunodeficiency Virus-1-Infected Individuals. *Blood.* 1999;93(3):936-941. doi:10.1182/blood.v93.3.936.403k08\_936\_941
- Mailler E, Bernacchi S, Marquet R, Paillart J, Vivet-Boudou V, Smyth R. The Life-Cycle of the HIV-1 Gag-RNA Complex. *Viruses.* 2016;8(9):248. doi:10.3390/v8090248
- Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you?. *J Med Genet.* 2011;48(9):580-589. doi:10.1136/jmedgenet-2011-100223
- Manninen A, Saksela K. HIV-1 Nef Interacts with Inositol Trisphosphate Receptor to Activate Calcium Signaling in T Cells. *J Exp Med.* 2002;195(8):1023-1032. doi:10.1084/jem.20012039
- Marcucci G, Yan P, Maharry K et al. Epigenetics Meets Genetics in Acute Myeloid Leukemia: Clinical Impact of a Novel Seven-Gene Score. *J Clin Oncol.* 2014;32(6):548-556. doi:10.1200/jco.2013.50.6337
- Marees A, de Kluiver H, Stringer S et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res.* 2018;27(2):e1608. doi:10.1002/mpr.1608
- Mariani R, Kirchoff F, Greenough T, Sullivan J, Desrosiers R, Skowronski J. High frequency of defective nef alleles in a long-term survivor with nonprogressive human immunodeficiency virus type 1 infection. *J Virol.* 1996;70(11):7752-7764. doi:10.1128/jvi.70.11.7752-7764.1996
- Martin M, Gao X, Lee J et al. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nat Genet.* 2002;31(4):429-434. doi:10.1038/ng934
- Martinez-Picado J, Prado J, Fry E et al. Fitness Cost of Escape Mutations in p24 Gag in Association with Control of Human Immunodeficiency Virus Type 1. *J Virol.* 2006;80(7):3617-3623. doi:10.1128/jvi.80.7.3617-3623.2006
- Martinez-Picado J, McLaren P, Erkizia I et al. Identification of Siglec-1 null individuals infected with HIV-1. *Nat Commun.* 2016;7(1). doi:10.1038/ncomms12412
- Matsuura K, Sawai H, Ikeo K et al. Genome-Wide Association Study Identifies TLL1 Variant Associated With Development of Hepatocellular Carcinoma After Eradication of Hepatitis C Virus Infection. *Gastroenterology.* 2017;152(6):1383-1394. doi:10.1053/j.gastro.2017.01.041
- McLaren P, Ball T, Wachihi C et al. HIV-Exposed Seronegative Commercial Sex Workers Show a Quiescent Phenotype in the CD4+T Cell Compartment and Reduced Expression of HIV-Dependent Host Factors. *J Infect Dis.* 2010;202(S3):S339-S344. doi:10.1086/655968
- McLaren P, Coulonges C, Bartha I et al. Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load. *PNAS.* 2015;112(47):14658-14663. doi:10.1073/pnas.1514867112
- McLaren P, Pulit S, Gurdasani D et al. Evaluating the Impact of Functional Genetic Variation on HIV-1 Control. *J Infect Dis.* 2017 Nov 27;216(9):1063-1069. doi: 10.1093/infdis/jix470.
- McNeil A, Yap P, Gore S et al. Association of HLA types A1-B8-DR3 and B27 with rapid and slow progression of HIV disease. *QJM.* 1996;89(3):177-186. doi:10.1093/qjmed/89.3.177
- Mendoza D, Johnson S, Peterson B et al. Comprehensive analysis of unique cases with extraordinary control over HIV replication. *Blood.* 2012;119(20):4645-4655. doi:10.1182/blood-2011-10-381996
- Mergia A. The Role of Caveolin 1 in HIV Infection and Pathogenesis. *Viruses.* 2017;9(6):129. doi:10.3390/v9060129

- Michael N, Chang G, d'Arcy L et al. Defective accessory genes in a human immunodeficiency virus type 1-infected long-term survivor lacking recoverable virus. *J Virol.* 1995;69(7):4228-4236. doi:10.1128/jvi.69.7.4228-4236.1995
- Migueles S, Sabbaghian M, Shupert W et al. HLA B\*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *PNAS.* 2000; 97(6):2709-2714. doi:10.1073/pnas.050567397
- Migueles S, Laborico A, Shupert W et al. HIV-specific CD8+ T cell proliferation is coupled to perforin expression and is maintained in nonprogressors. *Nat Immunol.* 2002;3(11):1061-1068. doi:10.1038/ni845
- Migueles S, Osborne C, Royce C et al. Lytic Granule Loading of CD8+ T Cells Is Required for HIV-Infected Cell Elimination Associated with Immune Control. *Immunity.* 2008;29(6):1009-1021. doi:10.1016/j.immuni.2008.10.010
- Migueles S, Connors M. Long-term Nonprogressive Disease Among Untreated HIV-Infected Individuals. *JAMA.* 2010;304(2):194. doi:10.1001/jama.2010.925
- Mikhail M, Wang B, Saksena N. Mechanisms involved in non-progressive HIV disease. *AIDS Rev.* 2003; 5, 230–244.
- Miles M, Janket M, Wheeler E et al. Molecular and functional characterization of a novel splice variant of ANKHD1 that lacks the KH domain and its role in cell survival and apoptosis. *FEBS Journal.* 2005;272(16):4091-4102. doi:10.1111/j.1742-4658.2005.04821.x
- Miura T, Brumme Z, Brockman M et al. Impaired Replication Capacity of Acute/Early Viruses in Persons Who Become HIV Controllers. *J Virol.* 2010;84(15):7581-7591. doi:10.1128/jvi.00286-10
- Molina L, Grimaldi M, Robert-Hebmann V et al. Proteomic analysis of the cellular responses induced in uninfected immune cells by cell-expressed X4 HIV-1 envelope. *Proteomics.* 2007;7(17):3116-3130. doi:10.1002/pmic.200700306
- Mologni D, Citterio P, Menzaghi B et al. Vpr and HIV-1 disease progression: R77Q mutation is associated with long-term control of HIV-1 infection in different groups of patients. *AIDS.* 2006;20(4):567-574. doi:10.1097/01.aids.0000210611.60459.0e
- Moura R, Pontillo A, D'Adamo P, Pirastu N, Coelho A, Crovella S. Exome analysis of HIV patients submitted to dendritic cells therapeutic vaccine reveals an association of CNOT1 gene with response to the treatment. *J Int AIDS Soc.* 2014;17(1):18938. doi:10.7448/ias.17.1.18938
- Mousseau G, Valente S. Didehydro-Cortistatin A: a new player in HIV-therapy?. *Expert Rev Anti Infect Ther.* 2015;14(2):145-148. doi:10.1586/14787210.2016.1122525
- Mujawar Z, Rose H, Morrow M et al. Human Immunodeficiency Virus Impairs Reverse Cholesterol Transport from Macrophages. *PLoS Biol.* 2006;4(11):e365. doi:10.1371/journal.pbio.0040365
- Müller C, Engler-Blum G, Gekeler V, Steiert I, Weiss E, Schmidt H. Genetic and serological heterogeneity of the supertypic HLA-B locus specificities Bw4 and Bw6. *Immunogenetics.* 1989;30(3):200-207. doi:10.1007/bf02421207
- Naif H. Pathogenesis of HIV infection. *Infect Dis Rep.* 2013;5(1S):6. doi:10.4081/idr.2013.s1.e6
- Nangola S, Urvoas A, Valerio-Lepiniec M et al. Antiviral activity of recombinant ankryrin targeted to the capsid domain of HIV-1 Gag polyprotein. *Retrovirology.* 2012;9(1):17. doi:10.1186/1742-4690-9-17
- Nasi M, Riva A, Borghi V et al. Novel genetic association of TNF- $\alpha$ -238 and PDCD1-7209 polymorphisms with long-term non-progressive HIV-1 infection. *Int J Infect Dis.* 2013; 17(10):e845-e850. doi:10.1016/j.ijid.2013.01.003
- Nasr P, Gursahani H, Pang Z et al. Influence of cytosolic and mitochondrial Ca<sup>2+</sup>, ATP, mitochondrial membrane potential, and calpain activity on the mechanism of neuron death induced by 3-nitropropionic acid. *Neurochem Int.* 2003;43(2):89-99. doi:10.1016/s0197-0186(02)00229-2
- Navis M, Schellens I, van Baarle D et al. Viral Replication Capacity as a Correlate of HLA B57/B5801-Associated Nonprogressive HIV-1 Infection. *J Immunol.* 2007;179(5):3133-3143. doi:10.4049/jimmunol.179.5.3133
- Neil S, Zang T, Bieniasz P. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature.* 2008;451(7177):425-430. doi:10.1038/nature06553
- Nissen S, Christiansen M, Helleberg M et al. Whole Exome Sequencing of HIV-1 long-term non-progressors identifies rare variants in genes encoding innate immune sensors and signaling molecules. *Sci Rep.* 2018;8(1). doi:10.1038/s41598-018-33481-0
- Nititham J, Gupta R, Zeng X, et al. Psoriasis risk SNPs and their association with HIV-1 control. *Hum Immunol.* 2017; 78(2): 179–184.
- Noecker C, Amaya-Larios I, Galeana-Hernández M, Ramos-Castañeda J, Martínez-Vega R. Contrasting associations of polymorphisms in Fc $\gamma$ RIIIa and DC-SIGN with the clinical presentation of dengue infection in a Mexican population. *Acta Trop.* 2014;138:15-22. doi:10.1016/j.actatropica.2014.05.021
- Noël N, Saez-Cirion A, Avettand-Fenoël V, Boufassa F, Lambotte O. HIV controllers: to treat or not to treat? Is that the right question?. *The Lancet HIV.* 2019;6(12):e878-e884. doi:10.1016/s2352-3018(19)30264-4
- Nyanhete T, Frisbee A, Bradley T et al. HLA class II-Restricted CD8+ T cells in HIV-1 Virus Controllers. *Sci Rep.* 2019;9(1). doi:10.1038/s41598-019-46462-8
- Ockenhouse C, Bernstein W, Wang Z, Vahey M. Functional Genomic Relationships in HIV-1 Disease Revealed by Gene-Expression Profiling of Primary Human Peripheral Blood Mononuclear Cells. *J Infect Dis.* 2005;191(12):2064-2074. doi:10.1086/430321
- Oh S, Lim S, Song M, et al. Whole Transcriptome Analyses Reveal Differential mRNA and microRNA Expression Profiles in Primary Human Dermal Fibroblasts Infected with Clinical or Vaccine Strains of Varicella Zoster Virus. *Pathogens.* 2019;8(4):183. doi:10.3390/pathogens8040183

- Okulicz J, Grandits G, Weintrob A et al. CD4 T Cell Count Reconstitution in HIV Controllers after Highly Active Antiretroviral Therapy. *Clin Infect Dis*. 2010;50(8):1187-1191. doi:10.1086/651421
- Okulicz J, Marconi V, Landrum Met al. Clinical Outcomes of Elite Controllers, Viremic Controllers, and Long-Term Nonprogressors in the US Department of Defense HIV Natural History Study. *J Infect Dis*. 2009;200(11):1714-1723. doi:10.1086/646609
- Olson A, Guiguet M, Zangerle R et al. Evaluation of Rapid Progressors in HIV Infection as an Extreme Phenotype. *JAIDS*. 2014;67(1):15-21. doi:10.1097/qai.0000000000000240
- Olson A, Meyer L, Prins Met al. An Evaluation of HIV Elite Controller Definitions within a Large Seroconverter Cohort Collaboration. *PLoS ONE*. 2014;9(1):e86719. doi:10.1371/journal.pone.0086719
- Paim A, Cummins N, Natesampillai S et al. HIV elite control is associated with reduced TRAIL short expression. *AIDS*. 2019;33(11):1757-1763. doi:10.1097/qad.0000000000002279
- Papathanasopoulos M, Patience T, Meyers T, McCutchan F, Morris L. Full-Length Genome Characterization of HIV Type 1 Subtype C Isolates from Two Slow-Progressing Perinatally Infected Siblings in South Africa. *AIDS Res Hum Retrovir*. 2003;19(11):1033-1037. doi:10.1089/088922203322588396
- Parham P. MHC class I molecules and kirs in human history, health and survival. *Nat Rev Immunol*. 2005;5(3):201-214. doi:10.1038/nri1570
- Park J, Lim C, Ham S, Kim S, Choi B, Roh T. Genome-wide analysis of histone modifications in latently HIV-1 infected T cells. *AIDS*. 2014;28(12):1719-1728. doi:10.1097/qad.0000000000000309
- Park Y, Lim S, Nam J, Kim S. Measuring intratumor heterogeneity by network entropy using RNA-seq data. *Sci Rep*. 2016;6(1). doi:10.1038/srep37767
- Pasquereau S, Kumar A, Herbein G. Targeting TNF and TNF Receptor Pathway in HIV-1 Infection: from Immune Activation to Viral Reservoirs. *Viruses*. 2017;9(4):64. doi:10.3390/v9040064
- Patel P, Borkowf C, Brooks J, Lasry A, Lansky A, Mermin J. Estimating per-act HIV transmission risk. *AIDS*. 2014;28(10):1509-1519. doi:10.1097/qad.0000000000000298
- Paximadis M, Ngqobe R, Chaisson R, Martinson N, Tiemessen C. RICH2 is implicated in viraemic control of HIV-1 in black South African individuals. *Infect Genet Evol*. 2017;49:78-87. doi:10.1016/j.meegid.2017.01.007
- Pe'er I, Yelensky R, Altshuler D, Daly M. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol*. 2008;32(4):381-385. doi:10.1002/gepi.20303
- Pelak K, Goldstein D, Wallev N et al. Host Determinants of HIV-1 Control in African Americans. *J Infect Dis*. 2010;201(8):1141-1149. doi:10.1086/651382
- Peng X, Sova P, Green R et al. Deep Sequencing of HIV-Infected Cells: Insights into Nascent Transcription and Host-Directed Therapy. *J Virol*. 2014;88(16):8768-8782. doi:10.1128/jvi.00768-14
- Pereyra F, Addo M, Kaufmann D et al. Genetic and Immunologic Heterogeneity among Persons Who Control HIV Infection in the Absence of Therapy. *J Infect Dis*. 2008;197(4):563-571. doi:10.1086/526786
- Pereyra F, Jia X, McLaren P et al. The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation. *Science*. 2010;330(6010):1551-1557. doi:10.1126/science.1195271
- Pereyra F, Palmer S, Miura T et al. Persistent Low-Level Viremia in HIV-1 Elite Controllers and Relationship to Immunologic Parameters. *J Infect Dis*. 2009;200(6):984-990. doi:10.1086/605446
- Pernas M, Tarancón-Diez L, Rodríguez-Gallego E, et al. Factors Leading to the Loss of Natural Elite Control of HIV-1 Infection. *J Virol*. 2018. doi: 10.1128/JVI.01805-17.
- Pillai SK, Abdel-Mohsen M, Guatelli J et al. Role of retroviral restriction factors in the interferon-alpha-mediated suppression of HIV-1 in vivo. *PNAS*. 2012; 109, 3035-3040. (doi:10.1073/pnas.1111573109)
- Platten M, Jung N, Trapp S et al. Cytokine and Chemokine Signature in Elite Versus Viremic Controllers Infected with HIV. *AIDS Res Hum Retrovir*. 2016;32(6):579-587. doi:10.1089/aid.2015.0226
- Popovic M, Sarngadharan M, Read E, Gallo R. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science*. 1984;224(4648):497-500. doi:10.1126/science.6200935
- Poropatich K, Sullivan D. Human immunodeficiency virus type 1 long-term non-progressors: the viral, genetic and immunological basis for disease non-progression. *J Gen Virol*. 2010;92(2):247-268. doi:10.1099/vir.0.027102-0
- Potter S, Lacabaratz C, Lambotte O et al. Preserved Central Memory and Activated Effector Memory CD4+ T-Cell Subsets in Human Immunodeficiency Virus Controllers: an ANRS EP36 Study. *J Virol*. 2007;81(24):13904-13915. doi:10.1128/jvi.01401-07
- Power D, Santoso N, Dieringer M et al. IFI44 suppresses HIV-1 LTR promoter activity and facilitates its latency. *Virology*. 2015;481:142-150. doi:10.1016/j.virol.2015.02.046
- Prasad V, Bukrinsky M. New Clues to Understanding HIV Nonprogressors: Low Cholesterol Blocks HIV Trans Infection. *mBio*. 2014;5(3). doi:10.1128/mbio.01396-14
- Quigley M, Pereyra F, Nilsson B et al. Integrative Genomic Analysis of HIV-Specific CD8+ T Cells Reveals That PD-1 Inhibits T Cell Function by Upregulating the AP-1 Transcription Factor BATF. *Blood*. 2009;114(22):916-916. doi:10.1182/blood.v114.22.916.916
- Ragoussis J. Genotyping Technologies for Genetic Research. *Annu Rev Genomics Hum Genet*. 2009;10(1):117-133. doi:10.1146/annurev-genom-082908-150116
- Ramsuran V, Naranbhai V, Horowitz A, et al. Elevated HLA-A expression impairs HIV control through inhibition of NKG2A-expressing cells. *Science*. 2019;365(6452):eaav7985. doi:10.1126/science.aav7985
- Ransinghe S, Cutler S, Davis I et al. Association of HLA-DRB1-restricted CD4+ T cell responses with HIV immune control. *Nat Med*. 2013;19(7):930-933. doi:10.1038/nm.3229

- Restrepo C, Gutierrez-Rivas M, Pacheco Y et al. Genetic variation in CCR2 and CXCL12 genes impacts on CD4 restoration in patients initiating cART with advanced immunosuppression. *PLoS ONE*. 2019;14(3):e0214421. doi:10.1371/journal.pone.0214421
- Reynoso R, Laufer N, Hackl M et al. MicroRNAs differentially present in the plasma of HIV elite controllers reduce HIV infection in vitro. *Sci Rep*. 2014;4(1). doi:10.1038/srep05915
- Rice A. The HIV-1 Tat Protein: Mechanism of Action and Target for HIV-1 Cure Strategies. *Curr Pharm Des*. 2017;23(28). doi:10.2174/1381612823666170704130635
- Ringrose J, Jeeninga R, Berkhout B, Speijer D. Proteomic Studies Reveal Coordinated Changes in T-Cell Expression Patterns upon Infection with Human Immunodeficiency Virus Type 1. *J Virol*. 2008;82(9):4320-4330. doi:10.1128/jvi.01819-07
- Ritz D, Vuk M, Kirchner P et al. Endolysosomal sorting of ubiquitylated caveolin-1 is regulated by VCP and UBXD1 and impaired by VCP disease mutations. *Nat Cell Biol*. 2011;13(9):1116-1123. doi:10.1038/ncb2301
- Robb M, Ananworanich J. Lessons from acute HIV infection. *Curr Opin HIV AIDS*. 2016;11(6):555-560. doi:10.1097/coh.0000000000000316
- Rodés B, Toro C, Paxinos E et al. Differences in disease progression in a cohort of long-term non-progressors after more than 16 years of HIV-1 infection. *AIDS*. 2004;18(8):1109-1116. doi:10.1097/00002030-200405210-00004
- Rodrigues V, Ruffin N, San-Roman M, Benaroch P. Myeloid Cell Interaction with HIV: A Complex Relationship. *Front Immunol*. 2017;8. doi:10.3389/fimmu.2017.01698
- Rodríguez-Da Silva A, Miralles C, Ocampo A, Valverde D. Prevalence study of the genetic markers associated with slow progression of human immunodeficiency virus type 1 in the Galician population (Northwest of Spain). *Enferm Infecc y Microbiol Cl*. 2017; 35(2):104-107. doi:10.1016/j.eimce.2017.01.017
- Rodríguez-Gallego E, Tarancón-Diez L, García F et al. Proteomic Profile Associated With Loss of Spontaneous Human Immunodeficiency Virus Type 1 Elite Control. *J Infect Dis*. 2019; 219(6):867-876. doi: 10.1093/infdis/jiy599.
- Roman F, Deroo S, Franck N, Burgy C, Servais J, Schmit J. A new CX3CR1 genotype with implications for HIV disease progression. *AIDS*. 2001;15(5):659-660. doi:10.1097/00002030-200103300-00022
- Rosás-Umbert M, Llano A, Bellido R et al. Mechanisms of Abrupt Loss of Virus Control in a Cohort of Previous HIV Controllers. *J Virol*. 2018;93(4). doi:10.1128/jvi.01436-18
- Rose K, Marin M, Kozak S, Kabat D. The viral infectivity factor (Vif) of HIV-1 unveiled. *Trends Mol Med*. 2004;10(6):291-297. doi:10.1016/j.molmed.2004.04.008
- Rotger M, Dalmau J, Rauch A et al. Comparative transcriptomics of extreme phenotypes of human HIV-1 infection and SIV infection in sooty mangabey and rhesus macaque. *J Clin Invest*. 2011;121(6):2391-2400. doi:10.1172/jci45235
- Rothberg K, Heuser J, Donzell W, Ying Y, Glenney J, Anderson R. Caveolin, a protein component of caveolae membrane coats. *Cell*. 1992;68(4):673-682. doi:10.1016/0092-8674(92)90143-z
- Saag M, Benson C, Gandhi R et al. Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults. *JAMA*. 2018;320(4):379. doi:10.1001/jama.2018.8431
- Sabin C, Lundgren J. The natural history of HIV infection. *Curr Opin HIV AIDS*. 2013;1. doi:10.1097/coh.0b013e328361fa66
- Sáez-Cirión A, Sinet M, Shin S et al. Heterogeneity in HIV Suppression by CD8 T Cells from HIV Controllers: Association with Gag-Specific CD8 T Cell Responses. *The Journal of Immunology*. 2009;182(12):7828-7837. doi:10.4049/jimmunol.0803928
- Sáez-Cirión A, Hamimi C, Bergamaschi A et al. Restriction of HIV-1 replication in macrophages and CD4+ T cells from HIV controllers. *Blood*. 2011;118(4):955-964. doi:10.1182/blood-2010-12-327106
- Sáez-Cirión A, Bacchus C, Hocqueloux L et al. Post-Treatment HIV-1 Controllers with a Long-Term Virological Remission after the Interruption of Early Initiated Antiretroviral Therapy ANRS VISCONTI Study. *PLoS Pathog*. 2013;9(3):e1003211. doi:10.1371/journal.ppat.1003211
- Sáez-Cirion A, Jacquelin B, Barré-Sinoussi F, Müller-Trutwin M. Immune responses during spontaneous control of HIV and AIDS: what is the hope for a cure?. *Phil. Trans. R. Soc. B*. 2014;369(1645):20130436. doi:10.1098/rstb.2013.0436
- Saha K, Bentsman G, Chess L, Volsky D. Endogenous Production of  $\beta$ -Chemokines by CD4+, but Not CD8+, T-Cell Clones Correlates with the Clinical State of Human Immunodeficiency Virus Type 1 (HIV-1)-Infected Individuals and May Be Responsible for Blocking Infection with Non-Syncytium-Inducing HIV-1 In Vitro. *J Virol*. 1998;72(1):876-881. doi:10.1128/jvi.72.1.876-881.1998
- Sakharkara M, Chow V, Kanguane P. Distributions of Exons and Introns in the Human Genome. *In Silico Biology*. 2004; 4:387-393.
- Salgado M, Brennan T, O'Connell K et al. Evolution of the HIV-1 *nef* gene in HLA-B\*57 Positive Elite Suppressors. *Retrovirology*. 2010;7(1):94. doi:10.1186/1742-4690-7-94
- Salgado M, López-Romero P, Callejas S et al. Characterization of host genetic expression patterns in HIV-infected individuals with divergent disease progression. *Virology*. 2011;411(1):103-112. doi:10.1016/j.virol.2010.12.037
- Sammaibashi S, Yamayoshi S, Kawaoka Y. Strain-Specific Contribution of Eukaryotic Elongation Factor 1 Gamma to the Translation of Influenza A Virus Proteins. *Front Microbiol*. 2018;9. doi:10.3389/fmicb.2018.01446
- Samson M, Libert F, Doranz B et al. Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature*. 1996;382(6593):722-725. doi:10.1038/382722a0

- Sanfilippo C, Cambria D, Longo A et al. SERPING1 mRNA overexpression in monocytes from HIV+ patients. *Inflamm Res*. 2017;66(12):1107-1116. doi:10.1007/s00011-017-1091-x
- Schang L, St. Vincent M, Lacasse J. Five Years of Progress on Cyclin-Dependent Kinases and other Cellular Proteins as Potential Targets for Antiviral Drugs. *Antivir Chem Chemother*. 2006;17(6):293-320. doi:10.1177/095632020601700601
- Schang L. Cyclin-dependent kinases as cellular targets for antiviral drugs. *J Antimicrob Chemother*. 2002;50(6):779-792. doi:10.1093/jac/dkf227
- Schuberth C, Buchberger A. UBXd domain proteins: major regulators of the AAA ATPase Cdc48/p97. *Cellular and Molecular Life Sciences*. 2008;65(15):2360-2371. doi:10.1007/s00018-008-8072-8
- Shungin D, Winkler T, Croteau-Chonka D, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*. 2015; 518 (7538), 187–196. doi: 10.1038/nature14132
- Smith M, Dean M, Carrington M, et al. Contrasting Genetic Influence of CCR2 and CCR5 Variants on HIV-1 Infection and Disease Progression. *Science*. 1997;277(5328):959-965. doi:10.1126/science.277.5328.959
- Sneller M, Justement J, Gittens K et al. A randomized controlled safety/efficacy trial of therapeutic vaccination in HIV-infected individuals who initiated antiretroviral therapy early in infection. *Sci Transl Med*. 2017;9(419):eaan8848. doi:10.1126/scitranslmed.aan8848
- Solloch U, Lang K, Lange V, Böhme I, Schmidt A, Sauter J. Frequencies of gene variant CCR5-Δ32 in 87 countries based on next-generation sequencing of 1.3 million individuals sampled from 3 national DKMS donor centers. *Hum Immunol*. 2017;78(11-12):710-717. doi:10.1016/j.humimm.2017.10.001
- Stephens H. HIV-1 diversity versus HLA class I polymorphism. *Trends Immunol*. 2005;26(1):41-47. doi:10.1016/j.it.2004.11.001
- Stewart G, Ashton L, Biti R et al. Increased frequency of CCR-5 Δ32 heterozygotes among long-term non-progressors with HIV-1 infection. *AIDS*. 1997;11(15):1833-1838. doi:10.1097/00002030-199715000-00007
- Suppiah V, Moldovan M, Ahlenstiel G et al. IL28B is associated with response to chronic hepatitis C interferon-alpha and ribavirin therapy. *Nat Genet*. 2009 Oct;41(10):1100-4. doi: 10.1038/ng.447.
- Swaminathan S, Murray D, Kelleher A. The role of microRNAs in HIV-1 pathogenesis and therapy. *AIDS*. 2012;26(11):1325-1334. doi:10.1097/qad.0b013e328352adca
- Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet*. 2019;20(8):467-484. doi:10.1038/s41576-019-0127-1
- Tanq J, Tanq S, Lobashevsky E et al. Favorable and Unfavorable HLA Class I Alleles and Haplotypes in Zambians Predominantly Infected with Clade C Human Immunodeficiency Virus Type 1. *J Virol*. 2002;76(16):8276-8284. doi:10.1128/jvi.76.16.8276-8284.2002
- Tang Y, Huang S, Dunkley-Thompson J et al. Correlates of spontaneous viral control among long-term survivors of perinatal HIV-1 infection expressing human leukocyte antigen-B57. *AIDS*. 2010;24(10):1425-1435. doi:10.1097/qad.0b013e32833a2b5b
- Tarancon-Diez L, Rodríguez-Gallego E, Rull A et al. Immunometabolism is a key factor for the persistent spontaneous elite control of HIV-1 infection. *EBioMedicine*. 2019;42:86-96. doi:10.1016/j.ebiom.2019.03.004
- Tebit D, Arts E. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis*. 2011;11(1):45-56. doi:10.1016/s1473-3099(10)70186-9
- Teixeira S, de Sá N, Campos D et al. Association of the HLA-B\*52 allele with non-progression to AIDS in Brazilian HIV-1-infected individuals. *Genes Immun*. 2014;15(4):256-262. doi:10.1038/gene.2014.14
- The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015; 526(7571): 68–74.
- Thèze J, Chakrabarti L, Vingert B, Porichis F, Kaufmann D. HIV controllers: A multifactorial phenotype of spontaneous viral suppression. *Clin Immunol*. 2011;141(1):15-30. doi:10.1016/j.clim.2011.07.007
- Thomas D, Thio C, Martin M et al. Genetic variation. *Nature*. 2009 8;461(7265):798-801. doi: 10.1038/nature08463
- Thomas R, Apps R, Qi Y et al. HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. *Nat Genet*. 2009;41(12):1290-1294. doi:10.1038/ng.486
- Tiffany H, Lautens L, Gao J et al. Identification of CCR8: A Human Monocyte and Thymus Receptor for the CC Chemokine I-309. *J Exp Med*. 1997;186(1):165-170. doi:10.1084/jem.186.1.165
- Tolstrup M, Laursen A, Gerstoft J, Pedersen F, Ostergaard L, Duch M. Cysteine 138 mutation in HIV-1 Nef from patients with delayed disease progression. *Sex Health*. 2006;3(4):281. doi:10.1071/sh06002
- Tomescu C, Duh F, Hoh R et al. Impact of protective killer inhibitory receptor/human leukocyte antigen genotypes on natural killer cell and T-cell function in HIV-1-infected controllers. *AIDS*. 2012;26(15):1869-1878. doi:10.1097/qad.0b013e32835861b0
- Troyer J, Nelson G, Lautenberger J et al. Genome-Wide Association Study Implicates PARD3B-Based AIDS Restriction. *J Infect Dis*. 2011;203(10):1491-1502. doi:10.1093/infdis/jir046
- UN Joint Programme on HIV/AIDS (UNAIDS), Global Report: UNAIDS Report on the Global AIDS Epidemic: 2018
- Vaeth M, Feske S. NFAT control of immune function: New Frontiers for an Abiding Trooper. *F1000Res*. 2018;7:260. doi:10.12688/f1000research.13426.1
- Valenzuela-Ponce H, Alva-Hernández S, Garrido-Rodríguez D et al. Novel HLA class I associations with HIV-1 control in a unique genetically admixed population. *Sci Rep*. 2018;8(1). doi:10.1038/s41598-018-23849-7
- Valle-Casuso J, Allouch A, David A et al. p21 Restricts HIV-1 in Monocyte-Derived Dendritic Cells through the Reduction of Deoxynucleoside Triphosphate Biosynthesis and Regulation of SAMHD1 Antiviral Activity. *J Virol*. 2017;91(23). doi:10.1128/jvi.01324-17



- van der Helm J, Geskus R, Lodi S et al. Characterisation of long-term non-progression of HIV-1 infection after seroconversion: a cohort study. *The Lancet HIV*. 2014;1(1):e41-e48. doi:10.1016/s2352-3018(14)70016-5
- van Grevenynghe J, Procopio F, He Z et al. Transcription factor FOXO3a controls the persistence of memory CD4+ T cells during HIV infection. *Nat Med*. 2008;14(3):266-274. doi:10.1038/nm1728
- Van Gulck E, Heyndrickx L, Bracke L et al. Control of viral replication after cessation of HAART. *AIDS Res Ther*. 2011;8(1):6. doi:10.1186/1742-6405-8-6
- van Manen D, Kootstra N, Boeser-Nunnink B, Handulle M, van 't Wout A, Schuitemaker H. Association of HLA-C and HCP5 gene regions with the clinical course of HIV-1 infection. *AIDS*. 2010;24(6):939. doi:10.1097/qad.0b013e328338a16c
- van Manen D, Delaneau O, Kootstra N et al. Genome-Wide Association Scan in HIV-1-Infected Individuals Identifying Variants Influencing Disease Course. *PLoS ONE*. 2011;6(7):e22208. doi:10.1371/journal.pone.0022208
- van 't Wout A, Lehrman G, Mikheeva S et al. Cellular Gene Expression upon Human Immunodeficiency Virus Type 1 Infection of CD4+-T-Cell Lines. *J Virol*. 2003;77(2):1392-1402. doi:10.1128/jvi.77.2.1392-1402.2003
- Vendel A, Lumb K. Molecular Recognition of the Human Coactivator CBP by the HIV-1 Transcriptional Activator Tat. *Biochemistry*. 2003;42(4):910-916. doi:10.1021/bi0270034
- Venter J, Adams M, Myers E et al. The Sequence of the Human Genome. *Science*. 2001;291(5507):1304-1351. doi:10.1126/science.1058040
- Verhelst J, Parthoens E, Schepens B, Fiers W, Saelens X. Interferon-Inducible Protein Mx1 Inhibits Influenza Virus by Interfering with Functional Viral Ribonucleoprotein Complex Assembly. *J Virol*. 2012;86(24):13445-13455. doi:10.1128/jvi.01682-12
- Vesa J, Chaillon A, Wagner G et al. Increased HIV-1 superinfection risk in carriers of specific human leukocyte antigen alleles. *AIDS*. 2017;31(8):1149-1158. doi:10.1097/qad.0000000000001445
- Vidal F, Viladés C, Domingo P et al. Spanish HIV-1-Infected Long-Term Nonprogressors of More Than 15 Years Have an Increased Frequency of the CX3CR1 249I Variant Allele. *JAIDS*. 2005;40(5):527-531. doi:10.1097/01.qai.0000186362.50457.e0
- Vieillard V, Fausther-Bovendo H, Samri A, Debré P. Specific Phenotypic and Functional Features of Natural Killer Cells From HIV-Infected Long-Term Nonprogressors and HIV Controllers. *JAIDS*. 2010;53(5):564-573. doi:10.1097/qai.0b013e3181d0c5b4
- Viq M, Kinet J. Calcium signaling in immune cells. *Nat Immunol*. 2009;10(1):21-27. doi:10.1038/nif.220
- Vigneault F, Woods M, Buzon M et al. Transcriptional Profiling of CD4 T Cells Identifies Distinct Subgroups of HIV-1 Elite Controllers. *J Virol*. 2011;85(6):3015-3019. doi:10.1128/jvi.01846-10
- Vilches C, Parham P. KIR: Diverse, Rapidly Evolving Receptors of Innate and Adaptive Immunity. *Annu Rev Immunol*. 2002;20(1):217-251. doi:10.1146/annurev.immunol.20.092501.134942
- Vingert B, Perez-Patrigeon S, Jeannin P et al. HIV Controller CD4+ T Cells Respond to Minimal Amounts of Gag Antigen Due to High TCR Avidity. *PLoS Pathog*. 2010;6(2):e1000780. doi:10.1371/journal.ppat.1000780
- Visscher P, Wray N, Zhang Q et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*. 2017;101(1):5-22. doi:10.1016/j.ajhg.2017.06.005
- Wang M, Liu Q, Liu C. Correlation of CCR5 and NLRP3 gene polymorphisms with renal damage due to hepatitis C virus-related cryoglobulinemia. *Exp Ther Med*. 2018. doi:10.3892/etm.2018.6558
- Wang X, Kim H, Nakahira K, Ryter S, Choi A. The Heme Oxygenase-1/Carbon Monoxide Pathway Suppresses TLR4 Signaling by Regulating the Interaction of TLR4 with Caveolin-1. *J Immunol*. 2009;182(6):3809-3818. doi:10.4049/jimmunol.0712437
- Wang X, Nadeau P, Lo Y, Mergia A. Caveolin-1 Modulates HIV-1 Envelope-Induced Bystander Apoptosis through gp41. *J Virol*. 2010;84(13):6515-6526. doi:10.1128/jvi.02722-09
- Wang X, Nadeau P, Lin S, Abbott J, Mergia A. Caveolin 1 Inhibits HIV Replication by Transcriptional Repression Mediated through NF- $\kappa$ B. *J Virol*. 2011;85(11):5483-5493. doi:10.1128/jvi.00254-11
- Wang Y, Tan B, Mu R et al. Ubiquitin-associated Domain-containing Ubiquitin Regulatory X (UBX) Protein UBXN1 Is a Negative Regulator of Nuclear Factor  $\kappa$ B (NF- $\kappa$ B) Signaling. *J Biol Chem*. 2015;290(16):10395-10405. doi:10.1074/jbc.m114.631689
- Watts G, Wymer J, Kovach M et al. Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat Genet*. 2004;36(4):377-381. doi:10.1038/ng1332
- Wei Z, Liu Y, Xu H et al. Genome-Wide Association Studies of HIV-1 Host Control in Ethnically Diverse Chinese Populations. *Sci Rep*. 2015;5(1). doi:10.1038/srep10879
- Weihl C, Pestronk A, Kimonis V. Valosin-containing protein disease: Inclusion body myopathy with Paget's disease of the bone and fronto-temporal dementia. *Neuromuscul Disord*. 2009; 19(5):308-315. doi:10.1016/j.nmd.2009.01.009
- Weiss R. Retrovirus classification and cell interactions. *J Antimicrob Chemother*. 1996;37(suppl B):1-11. doi:10.1093/jac/37.suppl\_b.1
- Welter D, MacArthur J, Morales J et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2013;42(D1):D1001-D1006. doi:10.1093/nar/gkt1229
- Wilen C, Tilton J, Doms R. HIV: Cell Binding and Entry. *Cold Spring Harb Perspect Med*. 2012;2(8):a006866-a006866. doi:10.1101/cshperspect.a006866
- Wilson C, Kalams S, Wilkes B et al. Overlapping epitopes in human immunodeficiency virus type 1 gp120 presented by HLA A, B, and C molecules: effects of viral variation on cytotoxic T-lymphocyte recognition. *J Virol*. 1997;71(2):1256-1264. doi:10.1128/jvi.71.2.1256-1264.1997

- Winkler C, Modi W, Smith M et al. Genetic Restriction of AIDS Pathogenesis by an SDF-1 Chemokine Gene Variant. *Science*. 1998;279(5349):389-393. doi:10.1126/science.279.5349.389
- Witwer K, Watson A, Blankson J, Clements J. Relationships of PBMC microRNA expression, plasma viral load, and CD4+ T-cell count in HIV-1-infected elite suppressors and viremic patients. *Retrovirology*.2012;9(1). doi:10.1186/1742-4690-9-5
- Wittke-Thompson J, Pluzhnikov A, Cox N. Rational Inferences about Departures from Hardy-Weinberg Equilibrium. *Am J Hum Genet*. 2005;76(6):967-986. doi:10.1086/430507
- Woelk C, Beliakova-Bethell N, Goicoechea M et al. Gene expression before HAART initiation predicts HIV-infected individuals at risk of poor CD4+ T-cell recovery. *AIDS*. 2010;24(2):217-222. doi:10.1097/qad.0b013e328334f1f0
- Woods M, Kelly J, Hattmann C et al. Human HERC5 restricts an early stage of HIV-1 assembly by a mechanism correlating with the ISGylation of Gag. *Retrovirology*. 2011;8(1):95. doi:10.1186/1742-4690-8-95
- Woods M, Tong J, Tom S et al. Interferon-induced HERC5 is evolving under positive selection and inhibits HIV-1 particle production by a novel mechanism targeting Rev/RRE-dependent RNA nuclear export. *Retrovirology*. 2014;11(1):27. doi:10.1186/1742-4690-11-27
- Wu M, Ao Z, Prasad K et al. IEX-1L, an Apoptosis Inhibitor Involved in NF-B-Mediated Cell Survival. *Science*. 1998;281(5379):998-1001. doi:10.1126/science.281.5379.998
- Wu J, Dwyer D, Dyer W, Yang Y, Wang B, Saksena N. Genome-wide analysis of primary CD4+ and CD8+ T cell transcriptomes shows evidence for a network of enriched pathways associated with HIV disease. *Retrovirology*. 2011;8(1):18. doi:10.1186/1742-4690-8-18
- Wu J, Ruth Sassé T, Wolkenstein G et al. Transcriptome analysis of primary monocytes shows global down-regulation of genetic networks in HIV viremic patients versus long-term non-progressors. *Virology*. 2013;435(2):308-319. doi:10.1016/j.virol.2012.10.026
- Xu W, Han M, Chen D et al. Genome-wide search for the genes accountable for the induced resistance to HIV-1 infection in activated CD4+ T cells: apparent transcriptional signatures, co-expression networks and possible cellular processes. *BMC Med Genomics*. 2013;6(1). doi:10.1186/1755-8794-6-15
- Yang Y, Al-Mozaini M, Buzon M et al. CD4 T-cell regeneration in HIV-1 elite controllers. *AIDS*. 2012;26(6):701-706. doi:10.1097/qad.0b013e3283519b22
- Yao Y, Xu M, Liang L et al. Genome-wide analysis of Epstein-Barr virus identifies variants and genes associated with gastric carcinoma and population structure. *Tumor Biol*. 2017;39(10):101042831771419. doi:10.1177/1010428317714195
- Yoon W, Ma B, Fellay J et al. Apolymorphism in the HCP5 gene associated with HLA-B\*5701 does not restrict HIV-1 in vitro. *AIDS*. 2010;24(1):155-157. doi:10.1097/qad.0b013e32833202f5
- Yuan F, Wang W, Cheng H. Co-expression network analysis of gene expression profiles of HER2+ breast cancer-associated brain metastasis. *Oncol Lett*. 2018. doi:10.3892/ol.2018.9562
- Zahoor M, Xue G, Sato H, Murakami T, Takeshima S, Aida Y. HIV-1 Vpr Induces Interferon-Stimulated Genes in Human Monocyte-Derived Macrophages. *PLoS ONE*. 2014;9(8):e106418. doi:10.1371/journal.pone.0106418
- Zarei Ghobadi M, Mozhgani S, Farzanehpour M, Behzadian F. Identifying novel biomarkers of the pediatric influenza infection by weighted co-expression network analysis. *Viro J*. 2019;16(1). doi:10.1186/s12985-019-1231-8
- Zaunders J, van Bockel D. Innate and Adaptive Immunity in Long-Term Non-Progression in HIV Disease. *Front Immunol*. 2013;4. doi:10.3389/fimmu.2013.00095
- Zhang J, Scadden D, Crumpacker C. Primitive hematopoietic cells resist HIV-1 infection via p21Waf1/Cip1/Sdi1. *J Clin Invest*. 2007;117(2):473-481. doi:10.1172/jci28971
- Zhang L, Zhang Z, Wu X, Jiang Y, Fu Y, Shang H. Transcriptomic meta-analysis identifies gene expression characteristics in various samples of HIV-infected patients with nonprogressive disease. *J Transl Med*. 2017;15(1). doi:10.1186/s12967-017-1294-5
- Zhang W, Wang L, Hong K et al. Frequency of HLA-A\*03 associates with HIV-1 infection in a Chinese cohort. *Sci China Life Sci*. 2013;56(11):1014-1019. doi:10.1007/s11427-013-4555-4
- Zhang X, Justice A, Hu Y et al. Epigenome-wide differential DNA methylation between HIV-infected and uninfected individuals. *Epigenetics*. 2016;11(10):750-760. doi:10.1080/15592294.2016.1221569
- Zhang X. Anti-retroviral drugs: current state and development in the next decade. *Acta Pharm Sin B*. 2018;8(2):131-136. doi:10.1016/j.apsb.2018.01.012
- Zhou L, Pupo G, Gupta P et al. A parallel genome-wide mRNA and microRNA profiling of the frontal cortex of HIV patients with and without HIV-associated dementia shows the role of axon guidance and downstream pathways in HIV-mediated neurodegeneration. *BMC Genomics*. 2012;13(1):677. doi:10.1186/1471-2164-13-677
- Zuñiga J, Buendia-Roldan I, Zhao Y et al. Genetic variants associated with severe pneumonia in A/H1N1 influenza infection. *Eur Resp J*. 2011;39(3):604-610. doi:10.1183/09031936.00020611

# **ANEXOS**

## 7. ANEXOS:

### ANEXO I

**Tabla S1: Genetic variants related to progression and viral control analyzed between LTNPs and control population**

SNP	Gene	Implication	Genotype				p value *	
			LTNP	%	Control	%		
rs1799864	CCR2	Progression	GG	68	78,2	227	85,0	0.1604
			AA	2	2,3	2	0,7	
			AG	17	19,5	38	14,2	
rs361525	TNF	Progression	GG	68	78,2	228	85,4	0.1768
			AG	19	21,8	37	13,9	
			AA	0	0,0	2	0,7	
rs3732379	CXCR1	Progression	CC	50	57,5	138	51,7	0.1939
			CT	34	39,1	104	39,0	
			TT	3	3,4	25	9,4	
rs9368699	C6orf48	Progression	TT	74	85,1	262	98,1	<b>1,61E-05</b>
			TC	13	14,9	5	1,9	
			CC	0	0,0	0	0,0	
rs8192591	NOTCH4	Progression	CC	81	93,1	258	96,6	0.2154
			TC	6	6,9	9	3,4	
			TT	0	0,0	0	0,0	
rs2395029	HCP5	HIV-1 control	TT	70	80,5	257	96,3	<b>1,12E-05</b>
			TG	17	19,5	10	3,7	
			GG	0	0,0	0	0,0	
rs1127888	UBXN6	New	CC	27	31,0	187	70,0	<b>1,04E-10</b>
			CT	46	52,9	71	26,6	
			TT	14	16,1	9	3,4	

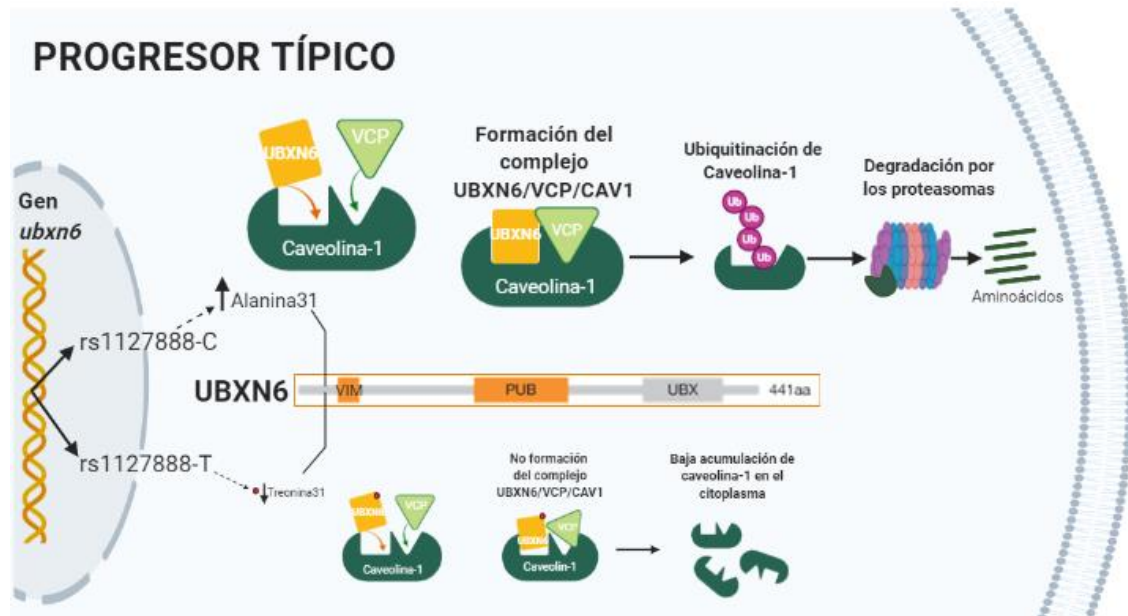
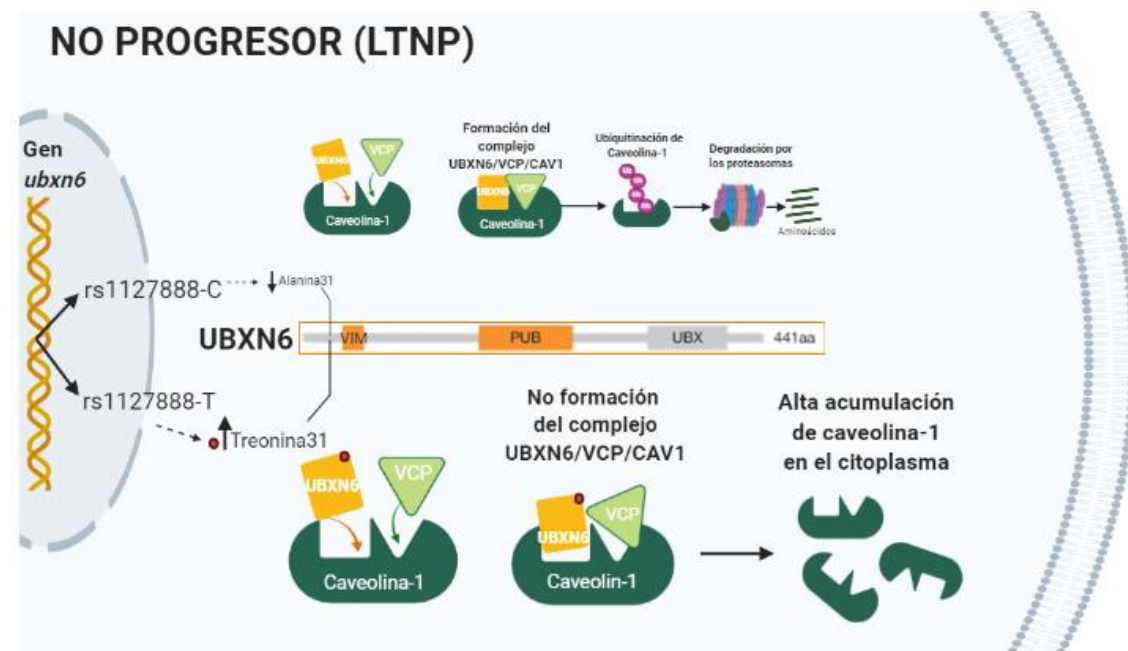
\* p value: 2x3 Fisher exact test

**Tabla S2: Genetic variants related to progression and viral control analyzed between vLTNPs and EC-LTNPs**

SNP*	Gene	Implication	Genotype (%)				
			EC-LTNP	vLTNP	p value**	All LTNP	
rs1799864	CCR2	Progression	GG	80,9	75,8	<b>0,000354</b>	78,7
			AA	0,0	21,2		8,8
			AG	19,2	3,0		12,5
rs361525	TNF	Progression	GG	80,9	75,8	0,592	78,8
			AG	19,2	24,2		21,2
			AA	0,0	0,0		0,0
rs3732379	CXCR1	Progression	GG	53,2	60,6	0,6191	56,2
			AG	42,6	39,4		41,3
			AA	4,3	0,0		2,5
rs3823418	PSORS1C1	Progression	GG	57,5	72,7	0,3835	63,7
			AG	34,0	24,2		30,0
			AA	8,5	3,0		6,3
rs11884476	PAR3B	Progression	CC	78,7	81,8	0,7841	80,0
			CG	21,3	18,2		20,0
			GG	0,0	0,0		0,0
rs333	CCR5	Progression	wt/wt	80,9	66,7	0,1922	75,0
			wt/ $\Delta$ 32	19,2	33,3		25,0
			$\Delta$ 32/ $\Delta$ 32	0,0	0,0		0,0
rs9368699	C6orf48	Progression	AA	87,2	81,8	0,5387	85,0
			GA	12,8	18,2		15,0
			GG	0,0	0,0		0,0
rs8192591	NOTCH4	Progression	CC	91,5	93,9	1	92,5
			TC	8,5	6,1		7,5
			TT	0,0	0,0		0,0
rs2395029	HCP5	HIV-1 control	AA	74,5	84,9	0,4055	78,8
			CA	25,5	15,2		21,2
			CC	0,0	0,0		0,0
rs9264942	HLA-C	HIV-1 control	AG	46,8	63,6	0,1903	53,8
			GG	40,4	21,2		32,4
			AA	12,8	15,2		13,8
rs130065	CCHCR1	HIV-1 control	GG	61,7	66,7	0,8218	63,8
			AG	36,2	30,3		33,8
			AA	2,1	3,0		2,4
rs12198173	TNXB	HIV-1 control	GG	74,5	60,6	0,2157	68,7
			AG	23,4	39,4		30,0
			AA	2,1	0,0		1,3
rs4522556	ITGB6	New factor	GG	72,3	39,4	<b>0,004772</b>	58,8
			AG	23,4	39,4		30,0
			AA	4,3	21,2		11,2

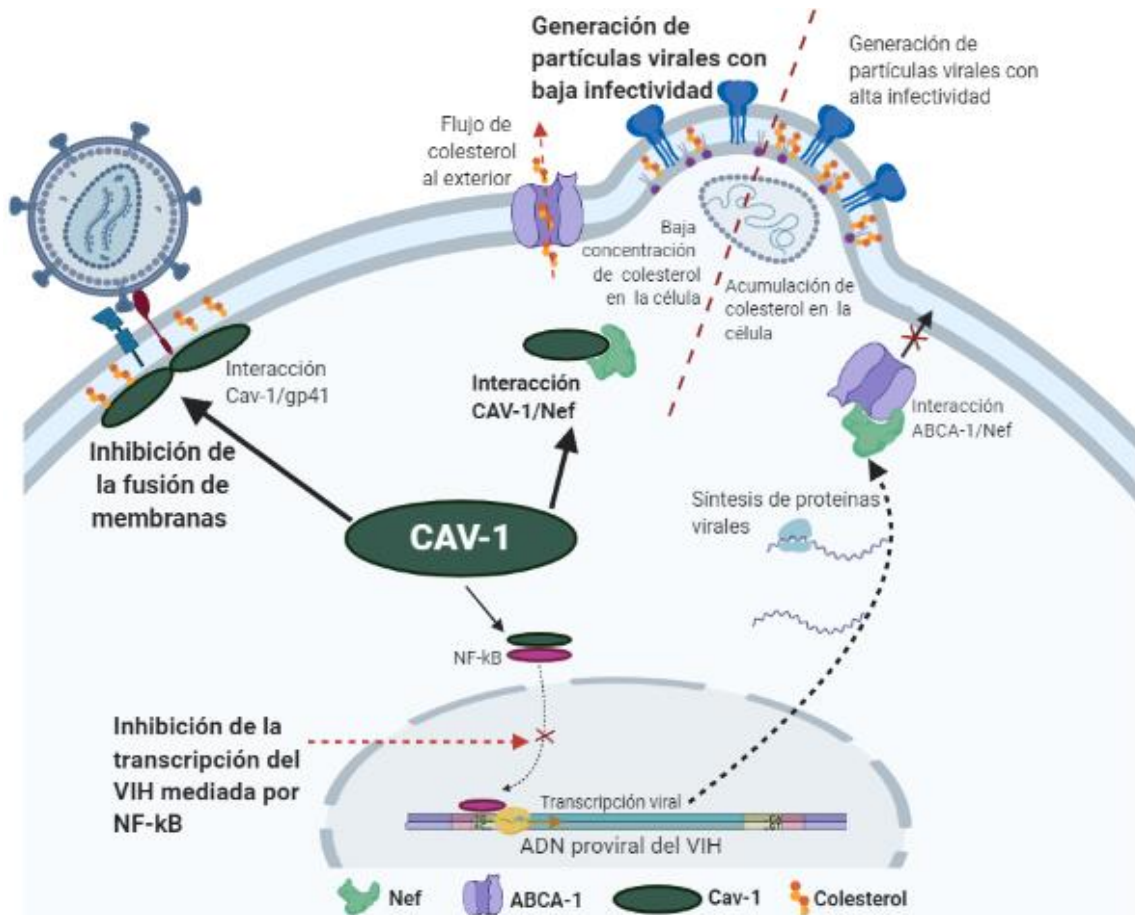
\* These SNPs are included in Infinium BeadChip microarrays (Illumina) used in our analysis.

\*\* p value: 2x3 Fisher exact test

**ANEXO II****A****B**

**Figura S1: Efecto del polimorfismo rs1127888 en la regulación de CAV-1 en la célula.**

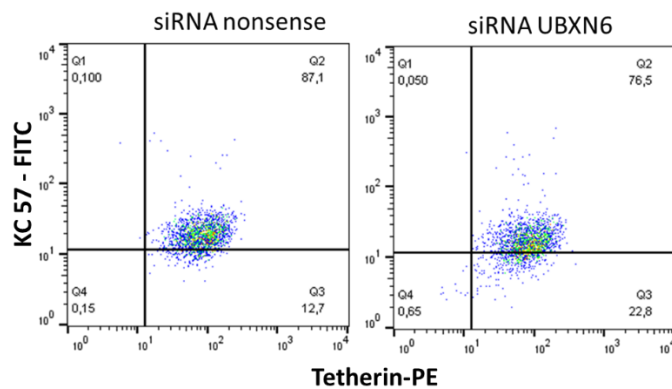
A) La alta frecuencia del alelo C en rs1127888 de *UBXN6* en progresores típicos conllevaría a una mejor formación del complejo UBXN6/CAV-1/VCP y la baja acumulación de CAV-1. B) La alta frecuencia del alelo T en rs1127888 de *UBXN6* en LTNPs conllevaría a una formación deficiente del complejo UBXN6/CAV-1/VCP y la alta acumulación de CAV-1 (Figuras elaboradas con BIORENDER)



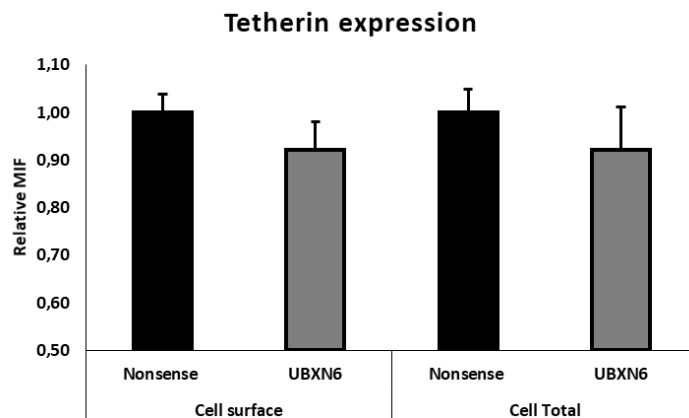
**Figura S2: Mecanismos de inhibición del VIH por CAV-1 en las células infectadas.**

A) CAV-1 puede inhibir la entrada viral al interferir en la fusión de membranas mediante la unión a gp41. B) CAV-1 puede inhibir la proteína Nef del VIH-1, lo cual contribuye al flujo normal de colesterol al exterior de la célula mediante la proteína transportadora ABCA-1. C) CAV-1 puede bloquear la traslocación de p65 NF-κB al núcleo, inhibiendo así la transcripción viral (Figura elaborada con BIORENDER).

A



B



**Figure S3: Effect of *UBXM6* knockdown in HIV infection and BST-2/tetherin expression**

A) HIV-1 infection. HeLa cells were infected with NL4.3-Ren virus (X4 tropic HIV) during 48 hours. HIV-1 infection was measured as a percentage of infected cells. B) Cell surface and total cell expression of BST-2/tetherin. HeLa cells were treated with siRNA nonsense or siRNA UBXM6 as described in Article II. Immunophenotyping was performed with BST-2/CD137-PE antibody and/or KC57 (p24)-FITC, using a FACScalibur flow cytometer (Becton Dickinson). Background staining was assessed with the appropriate isotype- and fluorochrome-matched control mAb and subtracted.



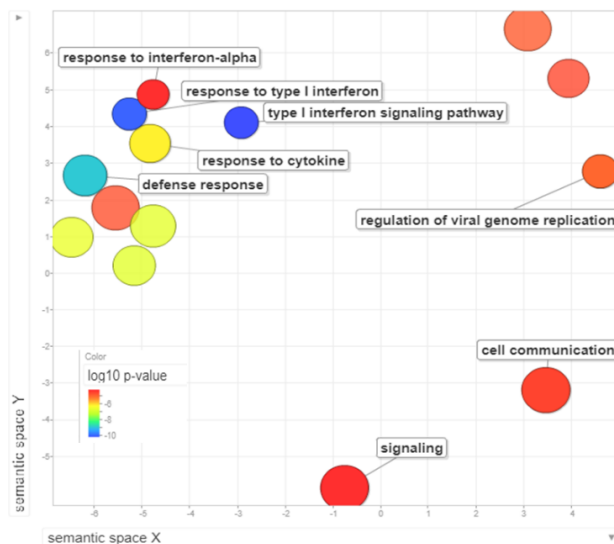
**ANEXO III**

**A**

Analysis Type: PANTHER Overrepresentation Test (Released 20190711)  
 Annotation Version and Release Date: GO Ontology database Released 2019-12-09  
 Analyzed List: upload\_1 (Homo sapiens)  
 Reference List: Homo sapiens (all genes in database)  
 Test Type: FISHER  
 Correction: FDR

GO biological process complete	Homo sapiens REFLIST (20851)	upload_1 (62)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold Enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
response to interferon-alpha (GO:0035455)	24	3	.07	+	42.04	6.90E-05	4.40E-02
type I interferon signaling pathway (GO:0060337)	67	8	.20	+	40.16	5.40E-11	8.60E-07
cellular response to type I interferon (GO:0071357)	67	8	.20	+	40.16	5.40E-11	4.30E-07
response to type I interferon (GO:0034340)	72	8	.21	+	37.37	9.16E-11	4.87E-07
negative regulation of viral genome replication (GO:0045071)	62	4	.18	+	21.70	4.40E-05	3.18E-02
regulation of viral genome replication (GO:0045069)	97	5	.29	+	17.34	1.30E-05	1.38E-02
defense response to virus (GO:0051607)	201	7	.60	+	11.71	2.69E-06	3.06E-03
regulation of viral process (GO:0050792)	210	6	.62	+	9.61	4.30E-05	3.26E-02
regulation of symbiosis, encompassing mutualism through parasitism (GO:0043903)	225	6	.67	+	8.97	6.24E-05	4.32E-02
response to virus (GO:0009615)	285	7	.85	+	8.26	2.46E-05	2.17E-02
innate immune response (GO:0045087)	774	14	2.30	+	6.08	5.62E-08	1.79E-04
cytokine-mediated signaling pathway (GO:0019221)	694	12	2.06	+	5.82	9.17E-07	1.22E-03
defense response (GO:0006952)	1352	20	4.02	+	4.97	1.23E-09	4.89E-06
defense response to other organism (GO:0098542)	962	14	2.86	+	4.89	7.58E-07	1.21E-03
cellular response to cytokine stimulus (GO:0071345)	1031	14	3.07	+	4.57	1.70E-06	2.09E-03
response to cytokine (GO:0034097)	1122	15	3.34	+	4.50	8.15E-07	1.18E-03
response to other organism (GO:0051707)	1330	17	3.95	+	4.30	2.36E-07	6.27E-04
response to external biotic stimulus (GO:0043207)	1332	17	3.96	+	4.29	2.41E-07	5.49E-04
response to biotic stimulus (GO:0009607)	1356	17	4.03	+	4.22	3.10E-07	5.48E-04
positive regulation of cell differentiation (GO:0045597)	990	12	2.94	+	4.08	3.22E-05	2.70E-02
response to external stimulus (GO:0009605)	2449	23	7.28	+	3.16	2.49E-07	4.96E-04
response to stress (GO:0006950)	3612	25	10.74	+	2.33	2.45E-05	2.30E-02
regulation of biological quality (GO:0065008)	4106	27	12.21	+	2.21	1.81E-05	1.80E-02
cell communication (GO:0007154)	5576	32	16.58	+	1.93	3.66E-05	2.91E-02
signaling (GO:0023052)	5473	31	16.27	+	1.90	6.79E-05	4.51E-02

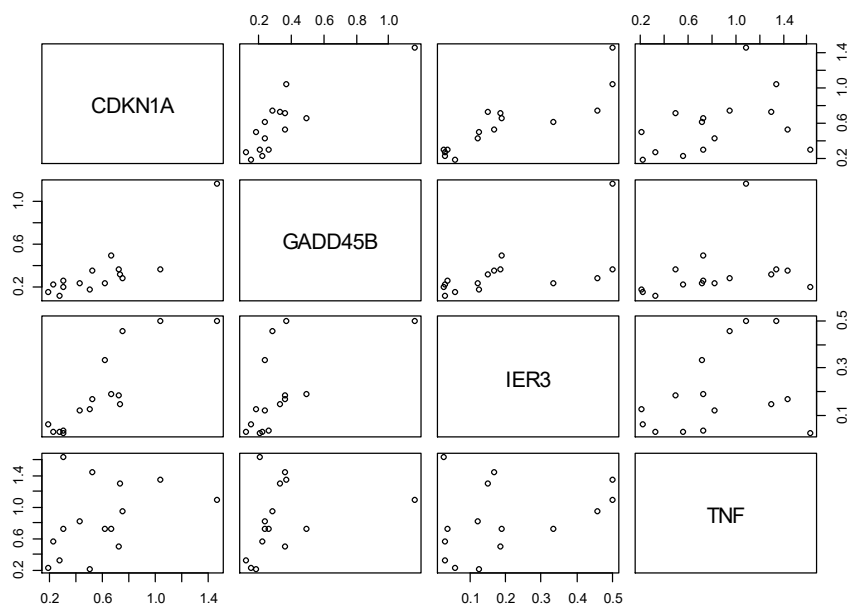
**B**



**Figure S4: Differential biological processes characterize transcriptional profiles between vLTNPs and TPs**

A) Gene ontology (GO) analysis of the 63 differentially expressed transcripts between vLTNPs and TPs (without ART) samples (Figure 7, Article III). Analysis was performed using Gene Ontology Resource, available in <http://geneontology.org/>. B) The REVIGO plot reflects clustering of semantic similarities between the biological GO terms and is color coded according to log<sub>10</sub> (p-value) of the GO term as determined using DAVID analysis of the identified genes within the list of transcripts.

A



B

	CDKN1A	GADD45B	IER3	TNF
CDKN1A		0,829	0,900	0,439
GADD45B	0,00026		0,746	0,507
IER3	0,00001	0,00202		0,239
TNF	0,10321	0,0562	0,38923	

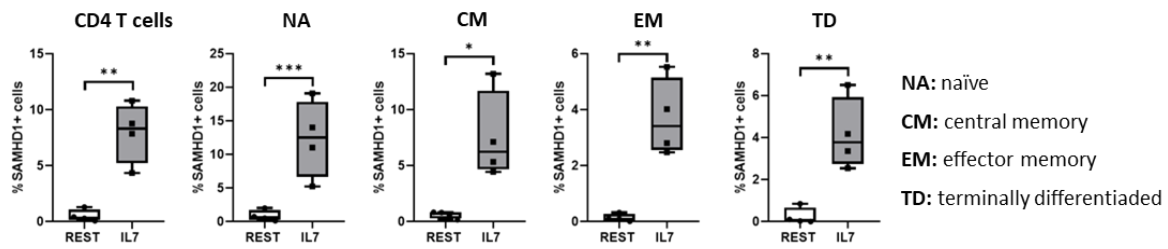
Correlation coefficients

*P value*

**Figura S5: Expression of *CDKN1A*, *GADD45*, *IER3* and *TNF* genes in PBMCs from HIV elite controllers (French cohort CODEX).**

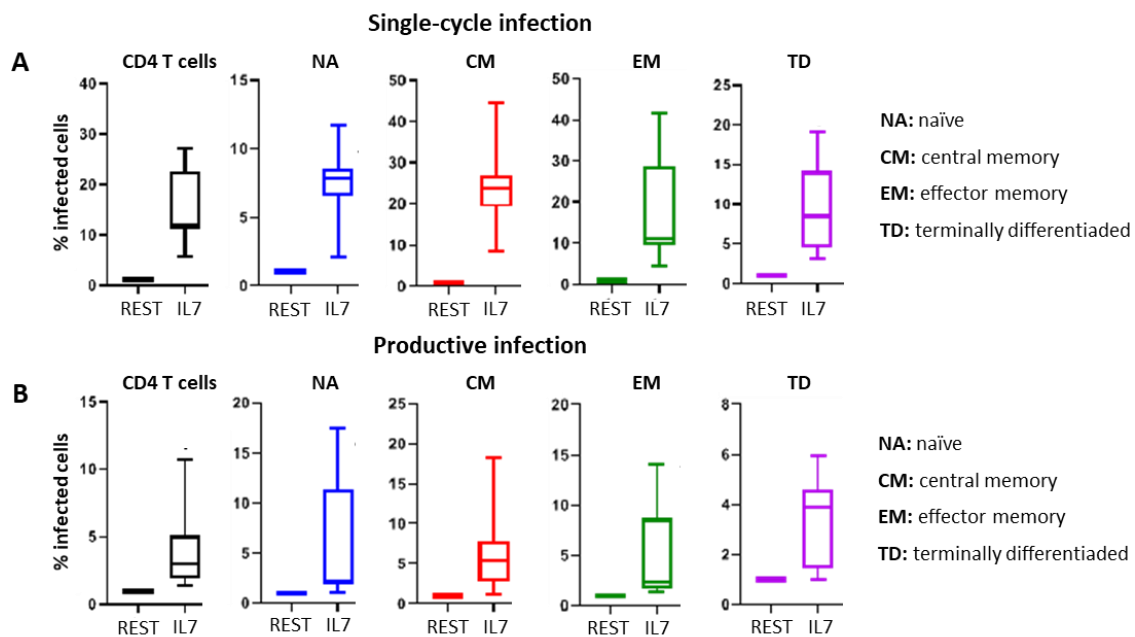
A) Relative expression of *CDKN1A*, *GADD45*, *IER3* and *TNF* in PBMCs from 15 HIV-1 elite controllers. B) Correlation analysis of *CDKN1A*, *GADD45*, *IER3* and *TNF* expression in PBMCs from 15 HIV-1 elite controllers. Analysis was performed with non-parametric Spearman correlation test. ARN extraction was performed with RNeasy Mini Kit (Qiagen). Reverse transcription was performed using GoScript Kit (Promega) starting from 200 ng of total RNA. Expression of each gene was performed with Taqman probes (FAM), using BECN-1 as a housekeeping gene (Applied Biosystems).





**Figura S7: SAMHD1 phosphorylation in CD4 T cell subpopulations.**

SAMHD1 phosphorylation in basal resting (overnight) and IL7 stimulation (4 days) conditions. PBMCs were cultured in RPMI medium. Flow cytometry was performed with SONY SH800 sorter. Immunophenotyping was performed using CD3, CD4, CCR7 and CD45RA antibodies, and propidium iodide as a cell viability marker. SAMHD1 phosphorylation was evaluated using anti phospho-SAMHD1 antibody (Thr592) (Cell Signalling). Background staining was assessed with the appropriate isotype- and fluorochrome-matched control mAb and subtracted. CD4 T cell subpopulations were classified in: naïve (CD45RA+ CCR7+), central memory (CD45RA- CCR7+), effector memory (CD45RA- CCR7-) and terminally differentiated (CD45RA+ CCR7-) cells.



**Figura S8: HIV infection in CD4 T cell subpopulations.**

A) HIV single-round infections. B) HIV productive infections. CD4 T cells were cultured in resting conditions or under IL7 stimulation during 4 days. Single-round and productive infections were performed with HIV-VSVg-GFP pseudotyped virus and Bal virus, respectively, during 72 h. Immunophenotyping was performed using CD3, CD4, CCR7 and CD45RA antibodies, and aqua live as a cell viability marker. Infections were stimulated by GFP (single-round infection) or KC57 (productive infection) expression by flow cytometry, as a percentage of infected cells. Background staining was assessed with the appropriate isotype- and fluorochrome-matched control mAb and subtracted. CD4 T-cell subpopulations were classified in: naïve (CD45RA+ CCR7+), central memory (CD45RA- CCR7+), effector memory (CD45RA- CCR7-) and terminally differentiated (CD45RA+ CCR7-) cells.

**ANEXO IV**

Otras publicaciones realizadas en el periodo de la tesis doctoral:

- 4-deoxyphorbol inhibits HIV-1 infection in synergism with antiretroviral drugs and reactivates viral reservoirs through PKC/MEK activation synergizing with vorinostat.  
**De la Torre-Tarazona HE** , Jiménez R , Bueno P , Camarero S , Román L , Fernández-García JL , Beltrán M , Nothias LF , Cachet X , Paolini J , Litaudon M, Alcamí J , Bedoya LM. Journal of Biochemical Pharmacology. Abril, 2020.
- Environmentally-Friendly Workflow Based on Supercritical Fluid Chromatography and Tandem Mass Spectrometry Molecular Net-working For the Discovery of Potent Anti-Viral Leads From Plants  
Nothias, Louis-Félix; Boutet-Mercey, Stéphanie; Cachet, Xavier; **De La Torre, Erick**; Laboureur, Laurent; Gallard, Jean-François; Retailleau, Pascal; Brunelle, Alain; Costa, Jean; Bedoya, Luis; Roussi, Fanny; Leyssen, Pieter; Alcamí, Jose; Paolini, Julien; Litaudon, Marc; Touboul, David. Journal of Natural Products. Agosto, 2017
- Hydroxytyrosol: A new class of microbicide displaying broad anti HIV-1 activity. Bedoya LM, Beltrán M, Obregón-Calderón P, García-Pérez J, **De La Torre E**, González N, Pérez-Olmeda M, Auñón D, Capa L, Gómez-Acebo E, Alcamí J. AIDS. Septiembre, 2016