2016

A Census of Human RNA-Binding Proteins

Stefanie Gerstberger

A CENSUS OF HUMAN RNA-BINDING PROTEINS AND CHARACTERIZATION OF THE DEDDH RNA EXONUCLEASE NEF-SP IN THE 3' END MATURATION OF 28S RIBOSOMAL RNA

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Stefanie Gerstberger

June 2016

A CENSUS OF HUMAN RNA-BINDING PROTEINS AND CHARACTERIZATION OF THE DEDDH RNA EXONUCLEASE NEF-SP IN THE 3' END MATURATION OF 28S RIBOSOMAL RNA

Stefanie Gerstberger, Ph.D.

The Rockefeller University 2016

Posttranscriptional gene regulation (PTGR) concerns all processes acting directly upon coding and non-coding RNAs, regulating and executing their maturation, ribonucleoprotein assembly, transport, stability and translation. RNA-binding proteins (RBPs) and ribonucleoprotein (RNP) complexes coordinate these processes. RBPs are central to cellular metabolism and their role in human diseases has been widely studied. Recent large-scale quantitative methods such as next-generation sequencing and modern protein mass spectrometry enabled new approaches to dissect PTGR networks and renewed interest in investigations of factors involved in PTGR at a genome-wide level. A census of all coding and noncoding RBPs has previously not been readily available and the number of RBPs was estimated based on few selected protein classes. However, for system-wide analyses of PTGR a comprehensive account of the RBPs is necessary. To address this need, I developed a census of 1,542 manually curated RBPs and categorized their interactions with different classes of RNA, defined the number of factors in different regulatory pathways, and investigated their evolutionary patterns, abundance, and tissue-specific expression. Co-regulated gene expression during developmental processes often gives novel insights into regulatory pathways and components. Furthermore, I showed that by classifying RBPs into their main regulatory RNA pathways we can start to understand the disease phenotypes of proteins involved in the same RNA metabolic pathways. These insights are useful for dissecting

dysregulated PTGR pathways in human diseases and finding new therapeutic targets. Finally, I showed in this chapter that by careful domain analysis, novel RBPs can be predicted and characterized the previously unknown RG/GG-rich RBP FAM98A. Overall, this analysis provides a critical step towards the comprehensive characterization of proteins involved in human RNA metabolism.

In the second part of the thesis I focused on the uncharacterized, highly tissue-specific RNA exonuclease NEF-sp and characterized its function in pre-28S ribosomal RNA processing. Ribosomal RNA biogenesis requires a series of endo- and exonucleolytic processing steps for the production of mature rRNAs. Although the mechanism of 28S 3' end rRNA maturation remains largely unknown in higher eukaryotes, it is thought that the 3' external transcribed spacer (3'ETS) of the large 47S rRNA precursor, containing 18S, 5.8S, and 28S rRNA, is removed in a precise endonucleolytic cleavage reaction, guided by U8 snoRNA. Here I show instead that the 3'ETS is exonucleolytically trimmed by the DEDDh RNA exonuclease NEF-sp in *Drosophila melanogaster*. I characterize for the first time in higher eukaryotes a nuclease that is involved in the removal of the 3' ETS. Interestingly, NEF-sp shows high tissue-specific expression in gonads. Gonad development is arrested in dNEF-sp mutants. Our results demonstrate that exonucleolytic trimming is essential for 28S rRNA maturation in higher eukaryotes and, counterintuitively, the expression of a factor involved in a core RNA metabolic process can be highly regulated. Our findings suggest an additional level of posttranscriptional gene regulation in the maturation of 28S rRNA, mediated by the regulated expression of RNA exonucleases.

Acknowledgements

This work was conducted under the mentorship of Thomas Tuschl and I am profoundly grateful for his guidance and support. He gave me the intellectual freedom to explore a new area in the laboratory and reach out to the Steller lab to complement their expertise with ours. I thank him for his unwavering attention for detail in both experiments and computational analysis, being able to detect flaws in endless rows of data tables, but also having the right intuition about how to solve a problem when I thought I hit a wall.

Thank you past and present Tuschl lab members for your support, advice, discussions, help, and critical eye when it came to writing manuscripts, grants, or my thesis. I would like to thank Markus Hafner for being the most thorough, well-read, and patient scientific editor, his productive scientific discussions, encouragement, and exchange of ideas and experimental design, as well as his other half Astrid Haase for discussions and feedback; Manuel Ascano for experimental guidance early in my PhD; Cindy Meyer and Aitor Garcia for being my lab companions, sharing advice, and experimental expertise; Claudia Bognanni for synthesizing oligos for me, which became one of the most beautiful experiments in my work; Daniel Briskin, Kimberly Bogardus, Bin Zhang from the genomics facility for preparing all Illumina RNA-seq libraries, the monoclonal antibody facility from MSKCC for carrying out the mAb project with me. I would like to particularly thank the 12th floor of the Tuschl lab, our computational unit: Miguel Brown, Pavel Morozov, Manju Kustagi, Hemant Suryawanshi for their advice, discussions, constant support and help, company and friendship.

I would like to thank Hermann Steller and the Steller lab for having been so welcoming for me to join as an unofficial member and have been a warm second home since then. The supportive work environment, the eagerness of every lab member to share reagents and contribute expertise in lab meetings and discussions has been one in a kind. Sigi Benjamin and Joe Rodriguez have been the advice beacons on anything imaginable, from *Drosophila* genetics experiments to

biochemical purifications. Completely selfless you took so much time for me to think with me about project directions and experimental approaches; my thesis would have not been the same without your often ingenious advice, and I would have not been the same without the many encouraging words from you on the way.

I would also like to thank the Christina Leslie lab for letting me join their lab meetings and scientific discussions. Thanks to Steve, Manu, Rafi, and Phaedra for advice and discussions.

I would like to thank my committee Brian Chait, David Allis, Dinshaw Patel, Christina Leslie and Olivier Elemento. You have been a truly outstanding committee. I thank you for your support and guidance, thoughtful advice, direction and solutions during more difficult patches of the project, and continuing support and interest in my success beyond my PhD and now during my application phase. Your support, encouragement and belief in me made an enormous difference for my project, future career and my development as a scientist. I would also like to thank my external examiner Susan Baserga from Yale University for taking part in my defense, incredibly good advice and making this thesis official.

I would like to thank the Dean's office in Rockefeller: Sid Strickland, Emily Harms, Martha Delgado and Cristian Rosario, you were always supportive and generous, caring, protective of us, organized everything in the background and made our life so smooth during our graduate years. I immensely appreciate your work!

Lastly, I would like to thank my friends in New York He Tian, Han, and Mark, Irit (my right brain half), Chaya, Cindy, Aitor, Nyasha, Sasha, Roos, Johannes, Stephanie, Erato, Qinsi, Zeeshan, Miguel, Mehrpoya, Julia, Ming, Manu and Sree. You made New York my home and your friendship elevated my life through good and bad times. My parents and brother, who were always my greatest supporters and gave me strength in moments of doubt; and Amir who has been a sunray in my life in the last few years of my PhD.

Table of Contents

List of Figures

List of Tables

List of General Abbreviations

3'UTRs	3'-untranslated regions. Correspond to the 3' end of mRNAs, specifically the region between the stop codon and the polyA-tail.
5'UTRs	5'-untranslated regions. Correspond to the 5' untranslated end of mRNAs, specifically the region between the transcript start and the start codon.
dsRNA	double-stranded RNA
ncRNAs	Noncoding RNAs, all RNAs that are not encoding proteins. Here also used to specifically group together noncoding RNAs that are not rRNAs, tRNAs, snRNA, or sno/scaRNAs.
nt/nts	nucleotide(s)
ORF	open reading frame
PAR-CLIP	Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation
PCW	Post-conception week. A time measure used to describe stages of human development in prenatal weeks. PCW records the time elapsed since the day of conception. Also commonly used is gestation week, which counts from the day of the last menstrual period. Assuming a normal 28-day menstrual cycle, gestation week and PCW differ by 2 weeks
RBD	RNA-binding domain
RBPs	RNA-binding proteins
RNP	Ribonucleoprotein, protein(s) complexed with RNA as obligate binding partner
rpkm	Reads per kilobase (kb) per million mapped reads. Measure for quantifying single-end read RNA sequencing data per transcript/gene exon model and library size
RRE	RNA recognition element. Short, rarely longer than 4-6 nucleotide (nt) sequence element within RNA targets that are recognized and bound by RBPs.
ssRNA	single-stranded RNA
TFs	Transcription factors

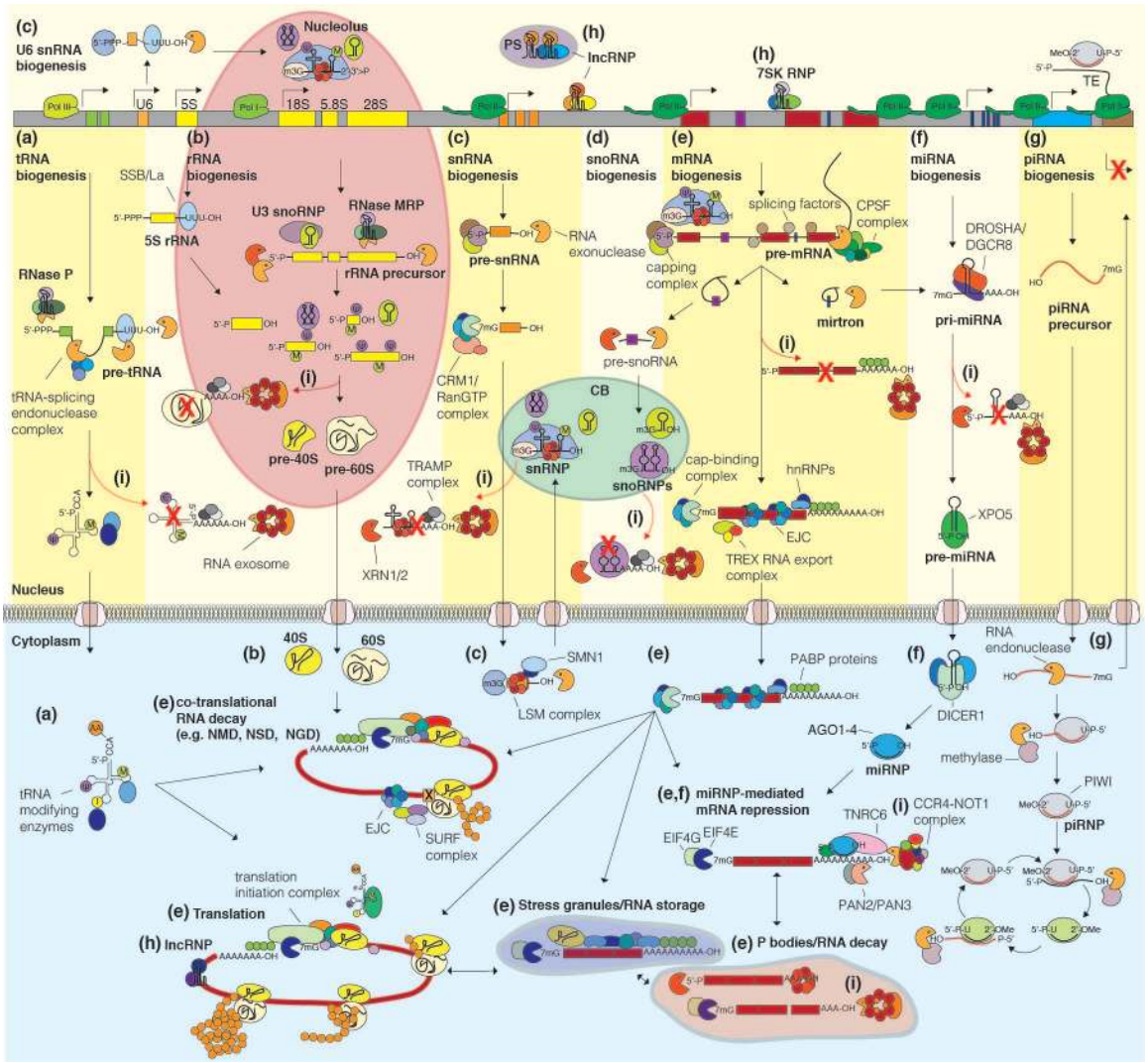
1 Introduction

1.1 Principles of posttranscriptional gene regulation

RNA is an essential constituent of all living organisms and central to decoding the genetic information of every cell. Posttranscriptional gene regulation (PTGR) is a term that refers to the cellular processes that control gene expression at the level of RNA; it encompasses RNA maturation, modification, transport, and degradation. Each of these events is regulated by the formation of different ribonucleoprotein (RNP) complexes with RNA-binding proteins (RBPs) at their core. Consequently, every RNA molecule independent of its ultimate function is at some level subject to PTGR (Figure 1.1).

Initially, it was thought that RNA mainly served either as the template in the form of messenger RNA (mRNA) or as an adaptor or structural component during protein synthesis provided by transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs). With the discovery of catalytic RNAs and a multitude of non-coding RNA (ncRNA) species it was recognized that RNA is a highly versatile molecule carrying out many regulatory functions in the cell, either by acting as a guide to recognize RNA sequence motifs or RNA recognition elements (RREs) present in their target RNAs, or by functioning as a scaffold and assembly platform for recruiting proteins to act synergistically (Cech and Steitz, 2014).

Figure 1.1 Overview of the main PTGR pathways in eukaryotes. Overview of biogenesis, decay and function of the most abundant RNAs: rRNAs, tRNAs, mRNAs, snRNAs, snoRNAs, miRNAs, piRNAs, and lncRNAs. Processes are described from left to right in the diagram. (a) tRNAs are transcribed by Pol III, 5' leader and 3' trailer sequences removed, introns spliced and end-joined. CCA nucleotides are added to 3' ends and nucleotide modifications introduced before tRNA aminoacylation (Maraia and Lamichhane, 2011). (b) 5S rRNA is transcribed by Pol III, while 28S, 18S, and 5.8S rRNAs are transcribed as one transcript by Pol I. The precursor is processed by RNA exo- and endonucleases and RNP RNase MRP, guided by U3 snoRNP. Nucleotide modifications are introduced by snoRNPs. rRNAs are assembled together with ribosomal proteins into ribosomal precursor complexes in the nucleus and transported to the cytoplasm, where they mature to functional ribosomes (Granneman and Baserga, 2004; Lafontaine and Tollervey, 2001; Thomson et al., 2013). (c) Most small nuclear RNAs (snRNAs) are transcribed by Pol II, capped and processed in the nucleus. Exported to the cytoplasm they undergo methylation and assemble with LSM proteins to snRNP particles aided by the SMN complex. Re-imported into the nucleus into the Cajal Body (CB) they undergo final maturation and snRNP assembly (Kiss, 2004). U6 and U6atac snRNAs are transcribed by Pol III and alternatively processed in the nucleus/nucleolus (Mroczek and Dziembowski, 2013). Mature snRNPs form the core of the spliceosome. (d) Small nucleoloar and small cajal body-specific RNAs (snoRNAs and scaRNAs) are processed from mRNA introns, capped, and modified before they assemble into snoRNPs/scaRNPs in the CB. snoRNPs/scaRNPs carry out methylations and pseudouridylations in rRNAs and snRNA or function in rRNA processing (e.g. U3 snoRNA) (Kiss, 2004). (e) mRNAs are transcribed by Pol II, capped, spliced, edited, and polyadenylated in the nucleus. Correctly matured mRNAs are exported into the cytoplasm. Regulatory RBPs control correct translation, monitor stability, decay and localization, and shuttle mRNAs between active translation, stress granules and P bodies (Buchan and Parker, 2009; Dreyfuss et al., 2002; Glisovic et al., 2008; Jackson et al., 2010; Müller-McNicoll and Neugebauer, 2013; Parker and Sheth, 2007). (f) microRNAs (miRNAs) are transcribed from separate genes by Pol II as long pri-miRNA transcripts, or alternatively, are expressed from mRNA introns (mirtrons), and processed into hairpin pre-miRNAs in the nucleus. After transport into the cytoplasm, they are processed into 21-nt double-stranded RNAs. One strand is incorporated into AGO-proteins (miRNPs) and guides them to partially complementary target mRNAs to recruit deadenylases and repress translation (Kim et al., 2009). (g) piwi-interacting RNAs (piRNAs) are ~28 nt long, germline-specific small RNAs. Primary piRNAs are directly processed and assembled from long, Pol II-transcribed precursor transcripts, while secondary piRNAs are generated in the ping-pong cycle by the cleavage of complementary transcripts by PIWI proteins. Mature piRNAs are 2'-O-methylated and incorporated into PIWI-proteins. The piRNA-PIWI complexes (piRNPs) silence transposable elements (TEs) either by endonucleolytic cleavage in the cytoplasm or through transcriptional silencing at their genomic loci in the nucleus (Siomi et al., 2011). (h) Most long noncoding RNAs (lncRNAs) are transcribed and processed similarly to mRNAs. Nuclear lncRNAs play an active role in gene regulation by directing proteins to specific gene loci, where they recruit chromatin modification complexes and induce transcriptional silencing/activation (Ulitsky and Bartel, 2013). Other ncRNAs, e.g. 7SK RNA, regulate transcription elongation rates (Peterlin et al., 2011) or induce formation of paraspeckles (PS) (Fox and Lamond, 2010). Cytoplasmic ncRNAs can modulate mRNA translation (Yoon et al., 2012). (i) Incorrectly processed RNAs are recognized by a number of complexes in the nucleus and cytoplasm that initiate and execute their degradation (Doma and Parker, 2007; Houseley et al., 2006).


(c)
U6 snRNA biogenesis
Nucleolus
(h)
PS
lncRNP
(h)
7SK RNP
TE
(a)
tRNA biogenesis
(b)
rRNA biogenesis
(c)
snRNA biogenesis
(d)
snoRNA biogenesis
(e)
mRNA biogenesis
(f)
miRNA biogenesis
(g)
piRNA biogenesis
SSB/La
U3 snoRNP
RNase MRP
5S rRNA
rRNA precursor
RNase P
pre-tRNA
tRNA-splicing endonuclease complex
pre-40S
pre-60S
pre-snRNA
RNA exonuclease
CRM1/ RanGTP complex
pre-snoRNA
CB
snRNP
snoRNPs
splicing factors
CPSF complex
pre-mRNA
capping complex
mirtron
DROSHA/ DGCR8
pri-miRNA
piRNA precursor
cap-binding complex
hnRNPs
EJC
TREX RNA export complex
XPO5
pre-miRNA
RNA exosome
XRN1/2
TRAMP complex
Nucleus
Cytoplasm
40S
60S
SMN1
LSM complex
PABP proteins
RNA endonuclease
DICER1
AGO1-4
miRNP
methylase
PIWI
piRNP
tRNA modifying enzymes
(e) co-translational RNA decay (e.g. NMD, NSD, NGD)
EJC
SURF complex
(e,f) miRNP-mediated mRNA repression
EIF4G
EIF4E
TNRC6
CCR4-NOT1 complex
PAN2/PAN3
translation initiation complex
(e) Translation
(h) lncRNP
(e) Stress granules/RNA storage
(e) P bodies/RNA decay

Recent advances in RNA sequencing technologies have facilitated the discovery of novel transcripts and we will soon know the precise composition of most cellular transcriptomes. While functional annotation for many RNAs is still in progress, the major classes of RNAs have now been described. The most abundant RNAs, constituting 90% of cellular RNAs by copy number, are shared by all organisms and required for protein synthesis: rRNAs, tRNAs and mRNAs. The remaining 10% are noncoding RNAs (ncRNAs) that mainly serve as guides or molecular scaffolds in a variety of processes including RNA splicing, RNA modification, and RNA silencing. The structure, length and composition of these RNAs and their RNPs is distinct and allows their integration into diverse functions (Table 1.1).

Table 1.1 Classes of RNAs, their sizes and functions. Functional description of the main RNA classes in humans and their length distribution. Additional reviews on biogenesis pathways and RBP components interacting with each class of RNA are referenced.

RNA class	Size (nt)	Biological role (additional reviews on function and biogenesis)
messenger RNA (mRNA)	~200-100,000	Encode the information for protein coding genes, translated by ribosomes (Dreyfuss et al., 2002; Glisovic et al., 2008; Müller-McNicoll and Neugebauer, 2013)
transfer RNA (tRNA)	~70-95	RNA adaptor molecules, transport amino acids to ribosome and recognize specific triplet codons on mRNA (Maraia and Lamichhane, 2011; Simos and Hurt, 1999; Suzuki et al., 2011)
ribosomal RNA (rRNA)	~120-5070	Structural component of ribosomes (Boisvert et al., 2007; Ciganda and Williams, 2011; Granneman and Baserga, 2004; Woolford and Baserga, 2013)
small nuclear RNA (snRNA)	~150	snRNAs U1, U2, U,4, U5 , U6, U12, U4atac, U6atac are core components of the spliceosome; U7 snRNA functions in 3' end maturation of histone RNAs (Kiss, 2004; Matera et al., 2007)
small nucleolar RNA (snoRNA) and small Cajal-body-specific RNA (scaRNA)	50-450	Guide chemical modifications (methylation and pseudouridylation) of rRNAs, snRNAs and snoRNAs (Filipowicz and Pogacić, 2002; Kiss et al., 2006; Matera et al., 2007)
microRNA (miRNA) and small interfering RNA (siRNA)	21-22	Associate with AGO proteins, guide them to target sequences predominantly in the 3'UTRs of mRNAs; induce degradation and translational repression (Bartel, 2009; Kim et al., 2009)
piwi-interacting RNA (piRNA)	~28-32	Associate with PIWI proteins, induce transposon silencing in the germline by guiding PIWI RNP complexes to genomic loci in the nucleus and transposon RNAs in the cytoplasm, leading to epigenetic silencing and ribonucleolytic cleavage of transposon RNA (Kim et al., 2009; Siomi et al., 2011)
long intervening noncoding RNA (lncRNA), 7SK RNA	>200	Recruit chromatin modifiers and remodeling complexes, modulate transcription by recruitment of protein cofactors to transcription starts sites and enhancers, function as molecular scaffolds for nuclear RBPs (Batista and Chang, 2013; Ulitsky and Bartel, 2013); 7SK RNA regulates transcription elongation (Peterlin et al., 2011)
RNase P and RNase MRP	~260-340	Ribonucleolytic RNP complexes that carry out processing of precursor tRNAs, rRNAs, snRNAs, and other noncoding RNAs; RNase P is an RNA-based enzyme, RNase MRP a protein-based ribonuclease (Ellis and Brown, 2009; Esakova and Krasilnikov, 2010; Jarrous, 2002; Xiao et al., 2002)
Y RNA	~ 80-110	Small noncoding RNAs that form an RNP complex with TROVE2 (Ro60) protein and act as RNA-chaperones, have a role in DNA replication and immune response (Hall et al., 2013; Köhn et al., 2013)
signal recognition particle RNA (7SL/SRP RNA)	~300	RNA of the signal recognition particle, which recognizes signal sequences of newly synthesized peptides and targets them to the endoplasmatic reticulum (Akopian et al., 2013)
Vault-associated RNA (vtRNA)	~80-140	Small noncoding RNAs, which are part of the vault RNP complex, thought to be involved in drug resistance and to downregulate mRNA targets through posttranscriptional gene silencing (Berger et al., 2008)
telomerase RNA (telRNA)	~500	RNA component of the telomerase complex TERC, which acts as reverse transcriptase and elongates telomer repeats. TERC is structurally related to box H/ACA snoRNAs (Egan and Collins, 2012)

RBPs execute PTGR, hence the characterization of the proteins transiently or stably interacting with RNAs is a prerequisite for the dissection of RNA regulatory processes. In many cases RBPs form a stable complex with an obligate RNA component, named ribonucleoprotein complex (RNP), which is the basic regulatory unit (e.g. snRNPs, snoRNPs, RNase P, ribosome subunits). However, many other types of RNAs, particularly mRNAs and tRNAs, only transiently associate with RBPs, whose functions are necessary for their proper maturation, localization, and turnover (Dreyfuss et al., 2002; Granneman and Baserga, 2004; Müller-McNicoll and Neugebauer, 2013; Phizicky and Hopper, 2010). Most mRNA-binding proteins (mRBPs) regulate thousands of targets, thus the proper assembly and function of RNA-protein complexes is critical for the maintenance of cellular metabolism in all cells and organisms. For a large fraction of RBPs, we are only starting to understand the complexity of their basic molecular roles, modes of recognition and global targets.

1.2 Aim of this thesis

In the present thesis I focused first on the system-wide analysis of RNA metabolism and second, resulting from this analysis, I characterized the function of the putative 3'-5' RNA exonuclease NEF-sp in *Drosophila melanogaster*. First, I set out to create a foundation for the system-wide study of PTGR factors. I developed a comprehensive census of human RBPs that regulate all coding and noncoding RNAs, and demonstrated its utility to gain insights into patterns in PTGR and to discover novel regulatory roles of RBPs. I created this catalog of RBPs based on structural domain annotations and experimental evidence and defined the number of proteins in different RNA regulatory pathways. I investigated their evolutionary conservation, protein families, and expression across adult tissues using published RNA-seq and microarray gene expression data. I further analyzed the co-expression of RBPs, finding common and unique trends in regulatory RNA pathways. Lastly, I used this categorization to summarize phenotypic commonalities

encountered for diseases caused by mutations in RBPs involved in the same RNA regulatory pathways. From the computational analyses of RBDs and RNA-binding low complexity repeats, I discovered a novel RBP, FAM98A and characterized its binding sites by PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation).

Furthermore, from the curated RBP census I discovered a highly tissue-specific 3'-5' RNA exonuclease NEF-sp (LOC81691). Using a combination of genetic, biochemical and RNA-sequencing experiments, I characterized the function of the *Drosophila melanogaster* homolog dNEF-sp (CG8368) and discovered its role in ribosomal RNA biogenesis in *Drosophila melanogaster*. The findings give unprecedented insights into the 3' end maturation of 28S rRNA in higher eukaryotes.

Together, the presented work developed a foundation for the system-wide analysis of PTGR factors in humans and its conserved homologs in other species, and elucidated mechanistic details of the 3' end maturation of 28S rRNA in higher eukaryotes.

2 A census of human RNA binding proteins

2.1 Introduction

2.1.1 Experimental and bioinformatic approaches towards a census of RBPs

PTGR is essential to cellular metabolism, coordinating maturation, transport, stability and degradation of all classes of RNAs. Mechanistically, each of these events is regulated by the formation of different RNP complexes with RBPs at their core. Among the first ribonucleoprotein complexes discovered was the ribosome in the 1950s (Darnell 2011; Steitz, 2008). Early biochemical studies on heteronuclear RNPs (hnRNPs) found that specific structural domains within proteins conferred direct and specific RNA-binding, which led to the definition of the first conserved, canonical RNA-binding domain (RBD) in RBPs (Burd and Dreyfuss, 1994). Different strategies were initially employed towards the identification of RBPs. Early approaches for the isolation of RBPs used gel electrophoresis of UV-crosslinked nuclear extracts, RNA pulldown or conventional chromatography to recover associated RBPs from cell lysates, followed by their mass spectrometric identification or immunodetection (Ascano et al., 2013; Dreyfuss et al., 1984; Pinol-Roma et al., 1988). To distinguish direct RNA-protein interactions from proteins associating with assembled RNPs UV-crosslinking became a refined method to isolate RNA-protein complexes during harsher purification steps, allowing the reduction of protein-protein-mediated background binders (Ascano et al., 2013). Alternatively, *in vitro* assays of recombinantly expressed candidate proteins were performed to interrogate RNA-binding properties. These allowed the quantitative assessment of binding preferences to specific RNA targets *in vitro*.

The definition of RBDs in proteins was facilitated by the growing amount of experimental and structural data, as well as the completion of sequenced genomes from all kingdoms of life, which allowed sequence alignments of multiple, related proteins and computational RBD predictions

across organisms based on sequence information (Finn et al., 2010; Haft et al., 2001; Letunic et al., 2009; Marchler-Bauer et al., 2013; Murzin et al., 1995; Tatusov et al., 2000; Wilson et al., 2009). These algorithms used Hidden Markov probabilistic models (HMM) that determine the likelihood of a specific amino acid sequence based on multiple sequence alignments. HMMs enable the detection of structural domains in uncharacterized protein sequences across organisms based on probability calculations of observed states of amino acid sequences, assuming an unobserved (hidden) state of a defined archetype structural domain sequence. At least 600 structural domains have been defined in RNA-related processes on Pfam by now (Finn et al., 2010). Early counts of RBPs used predictions selecting a small number of single-stranded-RNA-binding domains (ssRBDs, e.g. KH, RRM, PUF, S1, Table 5.1), often named canonical RBDs. These arrived at ~500 RBPs in human and mouse (Cook et al., 2011; Galante et al., 2009; McKee et al., 2005), ~300 in *D. melanogaster* (Gamberi et al., 2006; Lasko, 2000), and ~250-500 in *C. elegans* (Lasko, 2000; Lee and Schedl, 2006; Tamburino et al., 2013). An estimated ~700 RBPs in humans was reached when including additional RBDs involved in other aspects of RNA metabolism (Anantharaman et al., 2002). These approaches suggested a complex regulatory network controlled by RBPs and recognized that a census of RBPs was a prerequisite for the interpretation of synergistic and competitive action of RBPs on their targets.

Other approaches to obtain an estimate of the proteins involved in PTGR are automated functional annotations, including the Gene Ontology (GO) project (Ashburner et al., 2000) and the Kyoto Encyclopedia of Genes and Genomes database (KEGG) (Kanehisa and Goto, 2000), which integrate literature reports, database entries, and structural features. Using GO annotation, we arrived at ~1,900 human RBPs. While these gene groups are useful for gene set pathway analyses, they are not designed to establish a census of RBPs, as they include falsely assigned proteins due to inferred participation in biological processes or exclude valid proteins due to absence of annotation.

High-throughout experimental methods to determine the number of RBPs in different organisms, such as RBP-immunoprecipitation (RIP) coupled with cDNA array hybridization of recovered RNA (Tenenbaum et al., 2000), or SELEX-based (Systematic Evolution of Ligands by Exponential Enrichment) RNA ligand selections (Stoltenburg et al., 2007), enabled researchers to gain more global insights into specific RNA targets and RBP/RNPs preferential binding, but they were still limited to a few hundred targets. Protein microarrays allowed increased throughput for probing RNA-binding capabilities of a fraction of the proteome *in vitro*, using RNA probes of defined sequence (Scherrer et al., 2010; Siprashvili et al., 2012; Tsvetanova et al., 2010). The recent development of large-scale quantitative methods, especially next-generation sequencing and modern protein mass spectrometry (Ascano et al., 2012a; Gerstberger et al., 2013; Konig et al., 2011; Mann, 2006; Wang et al., 2009), now facilitates genome-wide identification of RBPs, their protein cofactors, and RNA targets at an unprecedented scale. In addition, deep sequencing approaches utilizing immunoprecipitation of RBPs, with or without *in vivo* RNA-protein crosslinking (CLIP- and RIP-seq, respectively) (Ascano et al., 2012a; Konig et al., 2011), as well as *in vitro* evolution methods (Ray et al., 2013; Stoltenburg et al., 2007), revealed binding spectra of RBPs and showed that a large proportion of RBPs binds to thousands of transcripts in cells at defined binding sites.

In an effort to experimentally validate RBPs across proteomes, four recent studies employed a combination of *in vivo* UV RNA-protein crosslinking followed by poly(A) RNA pulldown and protein mass spectrometry. These characterized the mRNA-binding proteome in HEK293 and HeLa human cell lines, mouse embryonic stem cells, and yeast (Ascano et al., 2013; Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013; Mitchell et al., 2013). The approach identified ~800 mRNA-binding proteins (mRBPs) in human HEK293 and HeLa cell lines, respectively (Baltz et al., 2012; Castello et al., 2012), 555 in mouse embryonic stem cells (mESCs) (Kwon et al., 2013), and 200 mRBPs in yeast (Mitchell et al., 2013). A total of ~1100

putative mRBPs were identified in the human and mouse datasets, ~600 of them were found in all three of them (Figure 2.1). A significant portion of these (64%) overlapped with known GO-classified RBPs (Figure 2.1).

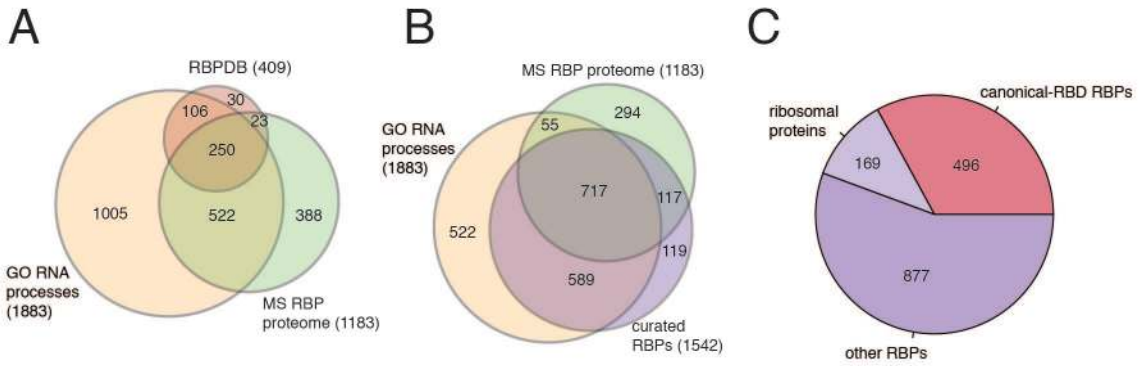


Figure 2.1 Results from different cataloging efforts for defining RBPs. (A) Venn diagram showing the overlap of proteins with RNA-related Gene Ontology (GO) categories (Ashburner et al., 2000) (orange), the human RNA-binding proteome identified by RNA-crosslinking and mass spectrometry studies (MS RBP proteome, green) (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013), and the RBPDB database of human RBPs with canonical RBDs (Cook et al., 2011) (red). (B) Venn diagram showing the overlaps of GO RBPs (orange), MS RBP proteome (green), and the curated RBP list presented here based on domain analysis of RBDs and experimental evidence of RNA-binding found in the literature (violet). (C) Composition of RBPs in the curated census: Canonical-RBD RBPs (containing canonical RBDs (Cook et al., 2011; Lunde et al., 2007), red) (Table 5.1), ribosomal proteins (bright violet), other RBPs (dark violet).

Many of the residual mRBP candidates did not contain previously described RBDs and require further experimental validation, while other known and expressed RBPs were missed due to the sensitivity of the experiments. However, in comparison to earlier predictive counts of the number of mRBPs using only canonical RBDs (Cook et al., 2011) (Figure 2.1, Table 5.1), this approach expanded the mRBP proteome from an estimated 400 to ~1,100 putative, experimentally derived candidates. With increasing sensitivity, approaches like these may represent the most suitable method to experimentally identify novel RBPs in proteome-wide experiments in different cell types. Comparative large-scale studies isolating the many ncRNA-binding proteins have not been undertaken yet, reflecting the predominant focus in PTGR centering around mRBPs. However, PTGR is not limited to mRNA regulation and a predominant

part comprises processes acting on ncRNAs. In this respect, it may not be surprising that among the ~150 RBPs listed in the Online Mendelian Inheritance (OMIM) (Hamosh et al., 2005) linked to human diseases, only a third are described as binding mRNAs, while others target diverse ncRNAs.

In summary, bioinformatic and experimental methods advanced enough to allow the investigation of RBPs at a systems-wide level, but perhaps due to the complexity of RNA metabolism, PTGR research is often focused on selected pathways and ignored that the fundamental RNA regulatory processes are tightly interconnected. As a consequence, at the onset of this project we did not have a clear understanding of the identity and number of all genes involved in PTGR. To address this need, I synthesized the available knowledge of PTGR factors through computational analyses to arrive at a curated census of the 1,542 proteins involved in RNA metabolism (Figure 2.1 B-C). The work presents an important step towards the comprehensive characterization of PTGR and the results of the analysis are presented here.

2.2 Computational methods

2.2.1 Selection of RNA-binding proteins and transcription factors

To define the total number of RBPs in humans, the complete set of protein sequences of the human proteome (105,237 isoforms) in Ensembl database (release v75) was searched for the presence of protein domains defined by Pfam A (Pfam HMM profiles, release v27) (Finn et al., 2010) using the domain-search algorithm hmmer3.0 (Eddy, 1998). We decided to set a confidence cut-off at e-value <0.01 for protein domain annotations. Next, I selected 799 Pfam domains (see Table 5.2) that were described to be involved in RNA-related processes (either directly RNA-binding/ processing or typically found in RNP complexes) and filtered out all protein isoforms that contained the selected RNA-binding domains (RBDs) (Figure 2.2).

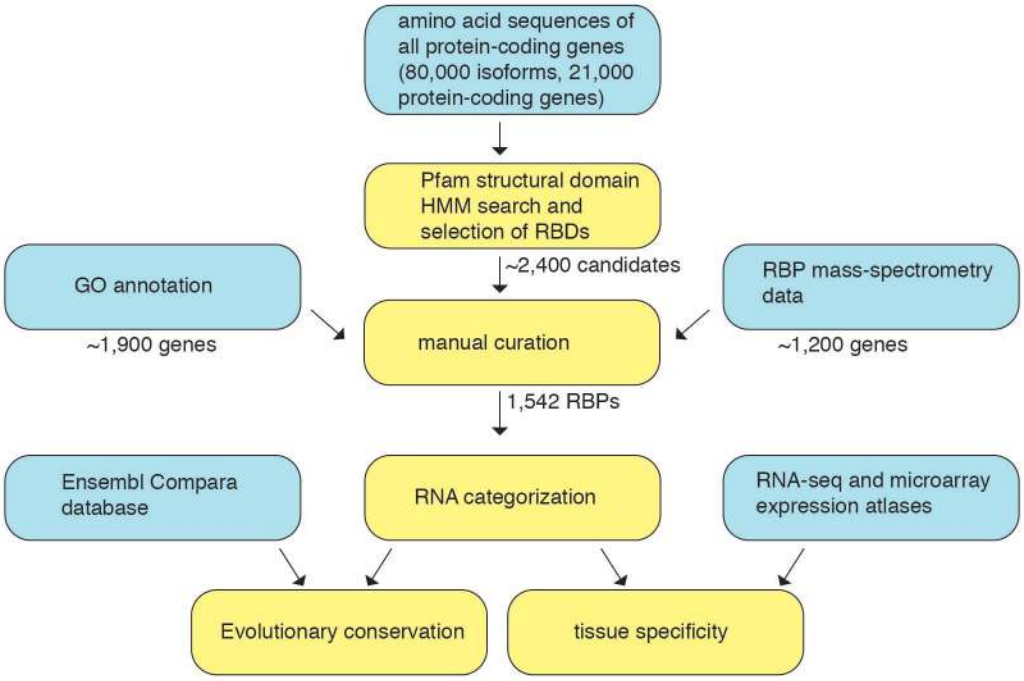


Figure 2.2 Diagram for generating a curated list of human RBPs. 80,000 protein-coding mRNA isoforms as curated by Ensembl were searched with defined HMM models for 799 RNA-binding domains (Table 5.2). The candidates were merged and compared with mass-spectrometry data from RNA-protein crosslinking experiments and Gene Ontology annotations. The resulting candidate genes were manually curated to arrive at a final census of 1,542 RBPs. The RBPs were further analyzed for their evolutionary conservation, families, tissue specificity, and expression trends across developmental stages.

For the final gene list I took one representative protein isoform per gene, selecting either the isoform with the highest number of RBDs or, in cases of an equal number of RBDs per gene in each isoforms, the longest isoform was selected. This procedure resulted in 2,103 putative RBPs. In a second filtering step I removed zinc finger proteins of the KRAB-, SCAN, and BTB domain classes, which are exclusively found in DNA-binding proteins. I manually inspected the residual 1,962 proteins using information available on NCBI's Entrez Gene and publication records to further remove proteins, which were unlikely to be involved in RNA metabolism based on their literature reports, homolog and paralog annotation, or presented pseudogenes and gene duplications. Finally, I manually added RBPs based on literature reports of the proteins or their orthologs, if they were missed by domain annotation, refined the selection of RBDs, and added known RBPs, which lacked a Pfam-defined structural protein domain. This resulted in a final census of 1,542 proteins (Figure 2.2, Table 5.3).

I compared structural conservation of RBPs with the other main gene regulatory group, transcription factors (TFs). For this comparison, I downloaded the curated TF dataset from Vaquerizas *et al.* (Vaquerizas et al., 2009) and matched the gene names and IDs to the current Ensembl v75 release, removing non-protein-coding genes and archived gene IDs. Of the 1,987 listed genes we could trace 1,874 genes. Duplicated genes were only listed once and genes, which were also present in our curated RBP list (such as e.g. ZFP36 protein family), were removed and classified as TF misannotations and predominant RBPs. The final list of TFs contains 1,704 proteins (see Table 5.5).

2.2.2 Expression analysis across 16 tissues

I used RNA-seq data from the human body map generated by Illumina, which comprised single and paired-end sequencing reads from poly(A)-selected RNA, isolated from 16 adult human tissues (http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-513/). I aligned the raw reads

with tophat2 and cufflinks2 against the human genome (hg19) using default parameters (Kim et al., 2013a; Trapnell et al., 2010). With a minimum expression threshold of rpkm=1 log2(rpkm) expression values were calculated for RBPs, TFs, and the remaining protein-coding genes across the 16 tissues. For the cumulative abundance analysis of RBPs and TFs, I added rpkm values for both groups in each tissue. For the subdivision of mRBPs and ribosomal proteins I used the manually curated annotation categories (see next section). Downstream analysis was performed in R using ggplot2.

2.2.3 Categorization of RNA targets for RBPs

I manually inspected the list of 1,542 RBPs and defined categories for RNA targets based on literature reports as ribosomal proteins, diverse targets, mRNA-, tRNA-, rRNA-, snoRNA-, snRNA-, and ncRNA-binding. Proteins with reported RNA-binding properties but unknown natural targets, were grouped into an unknown targets category. All other noncoding RNAs were grouped together in the ncRNA-binding category, including miRNAs, piRNAs, 7SK, lncRNAs, MRP RNA and RNase P, 7SL, vtRNAs, Y RNAs, viral RNAs and telomerase RNA. RBPs binding DNA/RNA hybrids were categorized into the diverse category, together with RNA nucleases and RNA exosome components. Proteins that are part of RNP complexes, but may not bind RNA directly, were assigned to their main RNA pathways, e.g. candidates involved in splicing or translation processes were classified as mRBPs.

2.2.4 RBP family definition, targets, and conservation analysis

I used Ensembl Compara (Vilella et al., 2009) to retrieve paralog information for the current Ensembl v75 proteome and grouped protein families based on an average of 20% sequence identity of the query and target percentage identities. I hand-curated cases where family relationships were not well defined and overlapped. I furthermore removed pseudogenes, read-

through transcripts, cDNA clones, and open reading frames (ORFs), which had no protein-coding entry on NCBI's Entrez Gene. Using this approach, I defined 1,111 RBP families and 554 TF families. Target categories for RBP families were defined by manually inspecting the annotation categories of the individual members in each family and then assigning the common target group per family based on the most frequently found target group in the family.

For the conservation analysis of paralogous protein families, the number of homologs and average homologies to human were queried through the Ensembl Compara database, which defined sequence identities and evolutionary relationships of proteins across organisms. The identity scores were retrieved for RBPs and TFs from 10 selected species (macaque, mouse, rat, *Xenopus*, zebrafish, zebrafinch, chicken, fruitfly, *C. elegans*, and *S. cerevisiae*). The average homology for protein families was determined based on sequence identities in three subsequent steps. First, for each homolog the average identity between the target and query score was calculated. Next, if one gene had more than one homolog in the queried species, the geometric average of the identity scores was taken to arrive at a final average homolog conservation score per gene. Third, per family I took the geometric averages of the average identity scores calculated for each individual member to give an average homology score per protein family. I grouped families into 4 different conservation categories with average identity scores of 0-20%, 20-30%, 40-60%, 60-75%, and 75-100%. The average conservation of each RNA target group was determined for RBP families conserved in *S. cerevisiae*.

2.2.5 Tissue specificity analysis of RBPs and RBP families

To assess tissue specificity of RBPs and TFs across human tissues, I used a microarray tissue atlas by Dezso *et al.* (Dezso et al., 2008) that profiled expression of 16,867 genes across 31 human tissues using the ABI Human Genome Survey Microarray platform version 2. I directly used the processed data deposited on GEO for further analysis, selecting per gene, if more than

one probe was present, the probe with the highest reported minimum intensity value. I defined a tissue specificity score *S* that assesses the deviation from a uniform distribution as information content. *S* measures, analogous to the definition of sequence logos in information theory, the difference between the logarithm of the total number of tissues and the Shannon entropy of the expression values for a gene.

$$S = H_{max} - H_{obs} = log_2(N) - (-\sum_{i=1}^{N}[p_i \times log_2(p_i)])$$

where the relative frequency $p_i = \frac{x_i}{\sum_{i=1}^{N} x_i}$;

x_i is the microarray expression level for gene in tissue *i*, H_{max} = maximum possible entropy, H_{min} = observed entropy, N is the number of tissues, in this case 31. The scale of the tissue specificity score *S* depends on the number of tissues, in the current dataset the maximum *S* (H_{max}) is reached when $S = log_2(31) = 4.95$. The plots show the maximum expression value per gene against its tissue specificity score *S*. Assessing tissue specificities of RBP families, I examined scores of the profiled genes per family and defined four categories of expression profiles for RBP families. In the first category I defined tissue-specific protein families, in which all members had tissue specificity scores $S \geq 1$. This score was set based on the tissue specificity scores of known tissue-specific RBPs, such as the PIWI proteins exclusively expressed in the germline. In the second category, I defined families with tissue-specific members, in which not all, but at least one member had a tissue specificity score $S \geq 1$. For the third category I binned gradient RBP families, if all their members had tissue specificity scores between $0.3 > S > 1$. In the final category of ubiquitous families, I grouped all protein families with tissue specificity scores for all their members $S < 0.3$.

2.2.6 Expression analysis of RBPs across developmental stages

For analyzing the variation of expression of RBPs during human fetal ovarian development, I used a microarray study by Houmard *et al.* (Houmard et al., 2009), which profiled expression of 17,508 genes during different gestational weeks using the Affymetrix Human Genome U133 Plus 2.0 Array platform. The publically available processed data was used for further analysis, selecting, if more than one probe per gene was present, the probe with the highest minimum expression value across the profiled stages. The geometric average for biological replicates of developmental stages was computed. Tissue specificity scores were calculated as described above across the developmental stages. 1,461 of the 1,542 RBPs profiled in this dataset could be extracted and selected the top 200 RBPs with highest scores *S*, i.e. which showed the highest deviation from a uniform distribution across the developmental stages, were selected for the representation of differentially expressed RBPs. In order to display dynamics of expression on a comparable, scale I further normalized the microarray intensity values for each gene to fold changes over the mean intensities across the profiled stages (Fold change = $\frac{x_i}{\sum_{i=1}^{N} x_i} \times {}^{1}/_{N}$, where x_i = microarray intensity for stage i, N = number of stages). Heat maps show unsupervised clustering of the $\log_2$(fold change over mean) expression values.

For the analysis of RBP expression in brain, I downloaded the publically available, processed RNA-sequencing data for human brain development from the brainspan atlas (http://www.brainspan.org) and selected 12 stages of hippocampus development as model for assessing RBP expression changes during neural development. Geometric average for biological replicates of the developmental stages were computed. To calculate tissue-specificity scores, a pseudo counts of 0.1 was added to each rpkm value and tissue specificity scores calculated as described above. A minimum expression cut-off of the sum of rpkm values across the developmental stages was set to $\sum_{i=1}^{N}(x_i) \geq 12$, where N = total number of profiled stages, and x_i = rpkm value in stage i. 1,402 RBPs were above this threshold. The top 200 RBPs with highest

tissue specificity scores *S* were selected for the heat map analysis. Rpkm values were further transformed to fold changes over the mean by dividing the rpkm values over the mean rpkm across the developmental stages. This allowed comparison of expression changes of different genes at a similar scale ($Fold\ change = \frac{x_i}{\sum_{i=1}^{N} x_i} * 1/N$, where x_i= rpkm for stages i, N = number of stages). Unsupervised clustering of the $\log_2$(fold change over mean) was used for the heat map representation. Pearson correlation and heat maps were generated using the gplots package in R.

2.2.7 Analysis of PAR-CLIP Libraries

FAM98A PAR-CLIP reads were adapter extracted, clipped with length of at least 20 nts and mapped to the hg19 human genome with Bowtie 0.12.9 (Langmead et al., 2009) (Bowtie parameters "-v 1 -m 10 --all --best –strata"), allowing for one mismatch in read alignments and up to 10 multimatches in the genome. Processing and annotation of clusters to the ENCODE GRCh38 genome annotation was performed using the PARalyzer software as described in Corcoran *et al.* (Corcoran et al., 2011)(http://www.genome.duke.edu/labs/ohler/research/PARalyzer/). Downstream analysis was performed in R. Shuffling of protein-coding ORFs was performed using the shuffling algorithm in the HMMER-3.0 suite using dinucleotide shuffling to preserve the dinucleotide composition (Eddy, 1998).

2.2.8 Analysis of FAM98A RNA-sequencing data

Illumina HiSeq 2000 100-base-pair (bp) single-end sequence reads were aligned to the reference genome (GRCh38/ hg19) using TopHat version 2.0.5 (Kim et al., 2013a) in default settings. Gene expression was estimated using the Ensembl GRCh38 gene model and the Cufflinks software version 2.0.2 (Trapnell et al., 2012b). Downstream cumulative distribution and statistical analyses were performed in R.

2.3 Experimental methods

2.3.1 Cell lines and plasmids

pENTR4-FAM98A was generated by restriction enzyme digest with SalI and NotI according to standard cloning procedures. The resulting pENTR4 vector was subsequently recombined into the pFRT/TO/FLAG/HA-DEST destination vectors using Gateway LR recombinase according to manufacturer's protocol (Invitrogen). Primers for FAM98A cDNA amplification were designed containing the BP recognition overhangs: FAM98A-forward: 5'-ACGCGTCGACATGGAGTGTGACCTCATGGAG; FAM98A-reverse: 5'- ATAGTTTAGCG-GCCGCTCAACTAGTGTAATGTCTTCCCTG. Cell lines were established according to the manufacturer's protocol for generating stable cell lines using the FlpIn system (Invitrogen). 50,000 cells/ml HEK293T FlpIn cells were seeded in a 12-well plate in antibiotic-free DMEM (Dulbecco's Modified Eagle Medium (DMEM) (Invitrogen, Life Technologies) supplemented with 10% heat inactivated fetal bovine serum (Thermo Scientific), 2 mM L-Glutamine, 100 U/ml penicillin, 100 U/ml streptomycin). 0.1 µg pFRT/TO/FLAG/HA-FAM98A-DEST destination vector was co-transfected with 0.9 µg pOG44 plasmid containing the Flp recombinase enzyme and 2 µl Lipofectamine 2000 in 50 µl Opti-MEM according to the manufacturer's instructions (Invitrogen). After 24 hrs Flp-In T-REx HEK293 cells expressing doxycycline-inducible FLAG/HA-tagged FAM98A were supplemented with a hygromycin and blasticidin selection medium (Dulbecco's Modified Eagle Medium (DMEM) (Invitrogen, Life Technologies) supplemented with 10% heat inactivated fetal bovine serum (Thermo Scientific), 2 mM L-Glutamine, 100 U/ml penicillin, 100 U/ml streptomycin, 100 µg/ml hygromycin, 15 µg/ml blasticidin). Colonies selected were after a few days, tested for inducible FLAG/HA-FAM98A expression by Western blot analysis, and cell line stocks were flash frozen in liquid nitrogen with 50% DMSO and stored at -196°C in liquid nitrogen. Stable cell lines were generally maintained in cell culture in DMEM supplemented with 10% heat inactivated fetal bovine serum (Thermo

Scientific), 2 mM L-Glutamine, 100 U/ml penicillin, 100 U/ml streptomycin, 100 µg/ml hygromycin, 15 µg/ml blasticidin. Parental HEK293FlpIn-TRex cells were maintained in DMEM supplemented with 10% FBS, 2 mM L-Glutamine, 100 U/ml penicillin and 100 mg/ml streptomycin, 15 mg/ml blasticidin, and 100 mg/ml zeocin.

2.3.2 RNA isolation and cDNA preparation

Total RNA was isolated with TRIzol following the manufacturer's protocol. First strand cDNA was synthesized from 5 µg total RNA using oligo(dT) priming and the Superscript III First Strand synthesis kit following the manufacturer's instructions (Invitrogen).

2.3.3 RNA phenol/chloroform isolation and ethanol precipitation

One volume of phenol/chloroform/isoamyl alcohol (25:24:1, phenol buffered at pH 4.3) was added to one volume of an RNA solution and vortexed vigorously for 15 sec. The phases were separated by centrifugation at 12,000 g at 4°C for 2 min. The upper aqueous layer was removed to a new tube and re-extracted with an equal volume of ice-cold chloroform. The mixture was vortexed for 15 sec and the phases were separated by centrifugation, 12,000 g at 4°C for 2 min. The upper aqueous layer was removed and added to a new tube, and the salt concentration adjusted to a final of 0.3 M NaCl. 3 volumes of 100% EtOH were added to the solution, mixed thoroughly, and stored at -20°C for 10 min (long RNAs) or 1 hr/overnight (short RNAs). RNA was precipitated by centrifugation at 12,000 g at 4°C for 30 min. The supernatant was removed and the pellet dissolved in 10-50 µl water.

2.3.4 Radiolabeling of oligo size markers

For 5' end radiolabeling, 100 pmoles of RNA size markers (19 nt: 5' pCGUACGCGGGUUUAAACGA; 35 nt: 5' pCUCAUCUUGGUCGUACGCGGAAUAGUUU

AAACUGU) was incubated with 5 pmoles of γ-^{32}P ATP and 10 units of T4 PNK (NEB) in 1 x T4 PNK buffer (70 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT) in a 20 µl reaction for 15 min at 37°C; followed by addition of 1,000 pmoles of non-radiolabelled ATP for another 5 min. The reaction was stopped by 1 volume of stop buffer (8 M urea, 10 mM EDTA, bromophenol). Labeled oligoribonucleotides were separated on an 18% polyacrylamide/ 8M urea gel at 28 W for 1 hr; the gel was exposed on a phosphorimaging screen, and gel pieces with the labeled oligomers excised. The RNA was eluted from the excised gel pieces with 3x the gel volume of 0.3 M NaCl solution overnight at 4°C, ethanol precipitated and the RNA pellet collected by centrifugation at >12,000g for 30 min at 4°C as described above.

2.3.5 PAR-CLIP

PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) was performed as previously described in Hafner *et al.* (Hafner et al., 2010a; 2010b). Expression of FLAG/HA-tagged FAM98A in stable FAM98A HEK293 cell lines was induced by addition of doxycycline (1 µg/ml) 16 hrs prior to harvesting.

4-SU incorporation and in vivo crosslinking:

FAM98A expressing cells were incubated with 100 µM 4-thiouridine (4SU) nucleoside analog for 16 hrs, the medium decanted and the cells irradiated on ice with a dose of 0.15 J/cm^2 of at 365 nm UV light in a Spectrolinker XL-1500 (Spectronics Corporation) equipped with 365 nm light bulbs or similar device. Cells were collected, frozen in liquid nitrogen, and stored at -80°C. FAM98A cell pellets were lysed in 3x the volume of the cell pellet in NP-40 lysis buffer [50 mM HEPES-KOH pH 7.2, 150 mM KCl, 0.5% (v/v) NP-40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)]. The lysate was incubated 30 min on ice and cleared by centrifugation at 12,000 g for 30 min at 4°C. The cleared supernatant was filtered through a 5 µm membrane syringe filter (Pall).

RNase T1 digest: In the first RNase T1 digest cell lysates were incubated with RNase T1 (Fermentas) at a final concentration of 1 U/µl at 22°C for 15 min and cooled on ice before proceeding. 50 µl magnetic Dynabeads Protein G coupled (Invitrogen) were used per 5 ml cell pellet. Magnetic beads were washed 3x in 1 ml PBS-Tween 20 0.1% (PBS-T) and incubated with 25 µl anti FLAG mouse monoclonal antibody (M2, 1 mg/ml, Sigma) in 200 µl PBS-T on a rotating wheel for 1 hr at room temperature.

Immunoprecipitation: Antibody-conjugated beads were washed 2x with PBS-T, resuspended in 200 µl lysis buffer and added to the lysate. Beads were incubated in the cell lysates on a rotating wheel for 1 hr at 4°C. After immunoprecipitation, magnetic beads were collected in a magnetic particle collector and washed 3x in 1 ml IP wash buffer [50 mM HEPES-KOH pH 7.5, 300 mM KCl, 0.05% (v/v) NP-40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)].

2nd RNase T1 digest: In a second RNase T1 digest beads were resuspended in one bead volume IP wash buffer and incubated with RNase T1 to a final concentration of 100 U/µl for 15 min at 22°C.

Dephosphorylation: Prior to dephosphorylation of protein-bound RNA fragments, the magnetic beads were washed 3x in 1 ml high salt wash buffer [50 mM HEPES-KOH pH 7.5, 500 mM KCl, 0.05% (v/v) NP-40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)], and resuspended in 1 bead volume dephosphorylation buffer (NEB buffer #3, 50 mM Tris-HCl pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT). Calf intestinal alkaline phosphatase (CIP) (10 U/µl, NEB) was added to a final concentration of 0.5 U/µl, and the suspension incubated for 10 min at 37ºC. Beads were washed 2x in 1 ml phosphatase wash buffer (50 mM Tris-HCl, pH 7.5, 20 mM EGTA, 0.5% (v/v) NP-40), and 2x in 1 ml T4 polynucleotide kinase (PNK) buffer without DTT (50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$).

Phosphorylation with radioactive γ -^{32}P-ATP: For radiolabeling of protein-bound RNA segments beads were resuspended in one bead volume of PNK buffer with DTT (50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 5 mM DTT). Radioactive γ-^{32}P-ATP (0.01 mCi/µl) was added to a final concentration of 0.5 µCi/µl (1.6 µM ATP) and the reaction incubated with T4 PNK (10 U/µl) added to a final concentration of 1 U/µl. The suspension was incubated at 37°C for 30 min. Non-radioactive ATP was added to a final concentration of 100 µM and the reaction incubated for another 5 min at 37ºC.

SDS-PAGE separation and electro-elution: Beads were washed 5x with 1 ml PNK buffer without DTT, resuspended in 70 µl of 4x SDS-PAGE loading dye (10% glycerol (v/v), 50 mM Tris-HCl pH 6.8, 2 mM EDTA, 2% SDS (w/v), 100 mM DTT, 0.1% bromophenol blue) and incubated at 90°C for 5 min. The supernatant was separated by SDS-PAGE, loaded onto a Novex Bis-Tris 4-12% polyacrylamide gel (Nupage, Invitrogen), and the gel run at 200 V for 45 min in 1x Nupage MOPS running buffer according to the manufacturer's instructions (Invitrogen). The gel was exposed for 1 hr on a phosphorimaging screen and a band excised at the correct protein size. The excised gel pieces were transferred into a dialyzer tube (Novagen) and the residual volume filled with 800 µl 1x SDS running buffer (25 mM Tris base pH 8.3, 0.192 M glycine, 0.01% SDS). The RNA-protein complex was electro-eluted from the gel in SDS running buffer at 100 V for 1 hr.

Proteinase K digest: The electro-eluted supernatant was transferred to a new tube and mixed with 2x Proteinase K buffer (100 mM Tris-HCl, pH 7.5, 150 mM NaCl, 12.5 mM EDTA, 2% (w/v) SDS), followed by addition of Proteinase K (Roche) to a final concentration of 1.2 mg/ml and the reaction incubated for 30 min at 55°C. The reaction mix was phenol-chloroform extracted, ethanol precipitated (to aid precipitation 1 µl glycogen was added), and the RNA redissolved in 10 µl water.

***3'adapter ligation*:** RNA was incubated with 2 µl of 10x RNA ligase buffer without ATP [0.5 M Tris-HCl, pH 7.6, 0.1 M $MgCl_2$, 0.1 M 2-mercaptoethanol, 1 mg/ml acetylated BSA (Sigma)], 6 µl 50% DMSO and 1 µl of 100 µM adenylated RNA 3'adapter 21.930 (5'-AppTCGTATGCCGTCTTCTGCTTGT), heat denatured at 90°C for 1 min and placed on ice. As ligation control 40 fmol of 1:100 dilution of 5'-^{32}P-labeled RNA size markers were prepared in a separate reaction. 1 µl of truncated T4 RNA ligase Rnl2 (1-249)K227Q ligase (1 µg/µl, plasmid for recombinant expression available from addgene.org) was added per reaction and incubated on ice at 4°C overnight. The reaction was stopped by addition of 20 µl 2x formamide stop mix (98.8% deionized formamide, 1% (v/v) 0.5 M EDTA pH 8.0, 0.2% Bromophenol blue), denatured at 90°C for 1 min, and loaded onto a 18% polyacrylamide/8M urea gel. The RNA was separated by electrophoresis in 1x TBE buffer (0.045 M Tris base, 0.045 M boric acid, 0.001 M Na_2EDTA) at 28 W for 45 min. The gel was exposed on a phosphorimaging screen for 1 hr and gel bands of the 3' ligation product excised in the size range of the ligated size marker control. The RNA was eluted from the gel with 400 µl 0.3 M NaCl at 4°C overnight, ethanol precipitated with 3 volumes, the pellet collected by centrifugation at 12,000 g and resuspended in 9 µl water.

5'adapter ligation: 1 µl of 100 µM RNA 5' adapter 26.68 (5'-GUUCAGAGUUCUACAGUCCGACGAUC), 2 µl of 10x RNA ligase buffer with ATP (0.5 M Tris-HCl pH 7.6, 0.1 M $MgCl_2$, 0.1 M 2-mercaptoethanol, 1 mg/ml acetylated BSA, 2 mM ATP) and 6 µl 50% aqueous DMSO were added to the RNA and the mixture denatured at 90°C for 1 min. The reaction was immediately placed on ice for 2 min, followed by addition of 2 µl T4 RNA ligase Rnl1 (Fermentas) and incubated at 37°C for 1 hr. The ligation reaction was stopped by addition of 20 µl formamide stop mix, heat inactivated at 90°C for 1 min, and separated on a 15% polyacrylamide/ 8M urea gel in 1x TBE running buffer at 28 W for 45 min. The gel was exposed on a phosphorimage screen for 1 hr and gel bands excised at the correct size range of the 5' ligated product (assessed by the control ligation of size marker). RNA was eluted from the gel

with 400 µl 0.3 M NaCl solution overnight at 4°C. To facilitate the recovery of the ligation product 1 µl of 100 µM 3' primer 21.929 (5'-CAAGCAGAAGACGGCATACGA) was added during the elution as carrier. The mixture was ethanol precipitated and re-dissolved in 5.6 µl water.

Reverse transcription: The RNA ligation product was denatured for 30 sec at 90°C and transferred to 50°C. cDNA reaction mix was prepared as following: 1.5 µl 0.1 M DTT, 3 µl 5x first-strand synthesis buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM $MgCl_2$), 4.2 µl 10x dNTPs (2 mM dATP, 2 mM dCTP, 2 mM dGTP, 2 mM dTTP) and the mix added to the RNA. The reaction was incubated for 3 min at 65°C before addition of 0.75 µl Superscript III reverse transcriptase (Invitrogen) and the reaction was incubated for 2 hrs at 50°C. The cDNA product was diluted with 85 µl water and stored at -20°C.

PCR amplification of cDNA library: The PCR reaction mix was prepared per sample as following: 40 µl of the 10x PCR buffer (100 mM Tris-HCl pH 8.0, 500 mM KCl, 1% Triton-X-100, 20 mM $MgCl_2$, 10 mM 2-mercaptoethanol), 40 µl 10x dNTPs, 2 µl of 100 µM 5' primer 44.32 (5'- AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA), 2 µl of 100 µM 3' primer 21.929, 272 µl H_2O. 89 µl of PCR reaction mix was used for a pilot PCR while the remaining reaction mix was kept for 3 large scale PCRs. 10 µl cDNA product, 1 µl Taq polymerase (5 U/µl) and 89 µl reaction mix were added to a final volume of 100 µl. PCR amplification was carried out under following cycle conditions: 45 sec at 94°C denaturation, 85 sec at 50°C annealing, 60 sec at 72°C extension. Aliquots were taken every second cycle between 12-28 cycles, mixed with 5x DNA loading dye (0.2% bromophenol blue, 0.2% xylene cyanol FF, 50mM EDTA pH 8, 20% Ficoll type 400) and analyzed on a 2.5% agarose gel for 1.5-2 hrs at 180 V. An expected length of about 95-110 nts and a lower band at 65 nts corresponding to the direct 3'adapter-5'adapter ligation product were separated. The optimal cycle number was obtained by choosing the cycle number 5 cycles away from the saturation level of the PCR amplification.

Large scale PCRs at the optimal cycle were performed, the PCR product was directly ethanol precipitated, the DNA pellet dissolved in 60 µl of 5x DNA loading dye, and separated on a 2.5% agarose gel for 1.5-2 hrs at 180 V. The upper band corresponding to the cDNA library product was excised from the gel and eluted using the QiaQuick gel extraction kit according to the manufacturer's instructions (Qiagen). The DNA was eluted in 30 µl water and 10 µl submitted for HiSeq sequencing.

2.3.6 siRNA-mediated knockdown

siRNAs with oligo U overhangs were chemically synthesized in the laboratory by C. Bognanni. The siRNAs were designed using the algorithms from Dharmacon siDESIGN (http://dharmacon.gelifesciences.com/design-center/) and the Whitehead siRNA design center (http://sirna.wi.mit.edu/home.php) and were antisense to the ORF of FAM98A. Sequences of the sense strands of the three different siRNAs were the following: FAM98A siRNA#1: 5'-GCUAAGAGCCAGACAGAAAUU; FAM98A siRNA#2: 5'-GGAGAAAGCU-GCUAAUAAAUU; FAM98A siRNA#3: 5'-GGGAAAAGAUAGAAGCAAUUU. dsRNA duplexes were annealed in 10 mM Tris pH 7.5, 20 mM NaCl with a final siRNA concentration of 20 µM by heating them for 1 min at 95°C and letting them cool down overnight at room temperature. For siRNA-mediated knockdown 2.5x10^5 Flp-In T-REx HEK293 cells/well were seeded in 6-well plates in DMEM antibiotic-free medium (DMEM supplemented with 10% FBS, 2 mM L-Glutamine, 100 U/ml penicillin, 100 U/ml streptomycin) and transfected with a 3.3 nM final concentration for each dsRNA duplex using Lipofectamine RNAiMAX following the manufacturer's instructions (Invitrogen). All siRNAs were pooled to a final dsRNA concentration of 10 nM (3.3 nM per siRNA duplex) and transfected together to obtain efficient knockdown with reduced off-target effects. Cells were harvested 72 hrs post-transfection and the knockdown efficiency was evaluated on the protein level by Western blot analysis.

2.3.7 RNA-sequencing

Total RNA from knockdown and overexpression experiments performed in biological duplicates was used as input for poly(A) purification and cDNA library construction using the TruSeq version 1.5 kit (Illumina). cDNA was barcoded using the Illumina Multiplexing Sample Preparation Oligonucleotide kit and sequenced on an Illumina HiSeq 2000 sequencer in a 100-base-pair (bp) single-end sequencing run.

2.3.8 Western blot analysis

Cells were lysed in NP-40 lysis buffer [50 mM HEPES-KOH pH 7.2, 150 mM KCl, 0.5% (v/v) NP-40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)] and cleared by centrifugation at 12,000 g for 15 min at 4°C. Total protein concentration of supernatants was assessed by BCA assay (Pierce). 40 µg lysate was mixed in 4x SDS loading dye (10% glycerol (v/v), 50 mM Tris-HCl, pH 6.8, 2 mM EDTA, 2% SDS (w/v), 100 mM DTT, 0.1% bromophenol blue), incubated at 90°C for 2 min, and samples were separated by SDS-PAGE at 30 mA per gel using standard Tris base glycine running buffer (25 mM Tris base pH 8.3, 0.192 M glycine, 0.01% SDS). After electrophoresis, proteins were blotted onto nitrocellulose membranes (Hybond-ECL, GE Life Science), pre-wetted in transfer buffer (25 mM Tris base, 190 mM Glycine, 20% MeOH, 0.05% SDS), and semi-dry transferred (Bio-Rad) at 250 mA for 1 hr. Nitrocellulose membranes were taken through a standard immunoblot protocol, followed by enhanced chemiluminescent detection (Crescendo ECL, Millipore) using a Lumimager (Fuji, LAS-3000). Following primary antibodies were used: rabbit polyclonal anti-FAM98A (1:500, abcam), mouse monoclonal anti-HA (HA-7, 1:1000, Sigma-Aldrich), mouse monoclonal anti-HA (HA.11, 1:1,000, Covance). Horseradish-peroxidase-conjugated polyclonal goat secondary antibodies raised against rabbit or mouse immunoglobulin (Dako, P0448 and P0447) were used at

a titer of 1:5,000, in conjunction with the appropriate species primary antibody, for immunoblot analyses.

2.3.9 Immunofluorescent stainings and microscopy

HEK293 cells expressing FLAG/HA-tagged FAM98A cells were grown on Lab-Tek II Chamber slides and induced with 1 μg/mL doxycycline 24 hrs before fixation. Chamber slides were rinsed with PBS and cells were fixed in 4% paraformaldehyde/PBS for 15 min at room temperature. Cells were permeabilized in PBS supplemented with 0.1% Triton-X100 (PBS-T) for 5 min, blocked with 5% normal goat serum in PBS-T for 30 min at room temperature and subsequently incubated for 1 hr with anti-HA antibody solution (Sigma-Aldrich, HA-7, 1:1000 in 5% normal goat serum in PBS-T). Chamber slides were washed 3x for 10 min in PBS-T, and incubated for 1 hr in PBS-T with Hoechst stain (1:1000), Alexa Fluor 488 anti-rabbit IgG (H + L) (1:500), and Alexa-647 Phalloidin (1:300, in 5% normal goat serum in PBS-T). Chamber slides were washed in PBS 3x for 10 at room temperature and disassembled according to the manufacturer's instructions. VectaShield mounting medium (Vector laboratories Inc.) was used for mounting. Single-layer images were recorded on a Zeiss LSM-710 confocal microscope.

2.4 Results

2.4.1 Defining the RBP repertoire: Generation of a curated list of RBPs

To annotate RBPs in the human genome, we defined proteins as RBPs if they contained domains known to directly interact with RNA or which resided within well-characterized RNPs, even if they were not directly contacting RNA, in some structurally characterized conformations. Considering that most RNPs are dynamically assembled and disassembled, transient sequence-unspecific RNA contacts are still plausible, examples include components of the exosome complex, the polyadenylation and cleavage complex, the spliceosome, and the ribosome. To select RBPs with known defined structural RBDs, I extracted 799 Pfam-defined protein domains (Table 5.2) known to be RNA-binding or exclusively found in RNA-related proteins and used their protein domain Hidden Markov models to search the human genome (~20,500 protein-coding genes) for proteins containing these RBDs (Eddy, 1998; Finn et al., 2010). From these candidates, I filtered out proteins with established RNA-unrelated functions, mostly DNA-binding zinc finger proteins, and manually added RBPs missed by domain searches but clearly defined in the literature. This resulted in a final census of 1,542 RBPs (Table 5.3), or 7.5% of all protein-coding genes in human, which constitutes the basis of subsequent analysis described in this chapter. This catalog provides a fresh starting point for future curation efforts, yet is likely to change as experimental studies continue to uncover novel RBPs or recognize that candidate RBPs comprising established RBDs evolved to adopt new functionalities unrelated to RNA-binding.

2.4.2 Analysis of structural features of RBPs

RBPs are commonly classified based on their specific RBDs as their structure and function provides some insight into their binding preferences and targets. Many excellent reviews have covered the different RBD families and their modes of RNA-binding (Anantharaman et al., 2002; Arcus, 2002; Auweter et al., 2006; Burd and Dreyfuss, 1994; Chang and Ramos, 2005; Curry et

al., 2009; Draper and Reynaldo, 1999; Gerstberger et al., 2014b; Glisovic et al., 2008; Jankowsky, 2011; Kim and Bowie, 2003; Kuchta et al., 2009; Linder and Jankowsky, 2011; Lunde et al., 2007; Masliah et al., 2013; Meister, 2013; Mihailovich et al., 2010; Rajkowitsch et al., 2007; Rocak and Linder, 2004; Singh and Valcarcel, 2005; Sommerville, 1999; Tanner and Linder, 2001; Tharun, 2009; Valverde et al., 2008; Wang et al., 2002; Wilusz and Wilusz, 2005), and the depth of their discussion is limited to the insights gained from this analysis. RBDs are deeply conserved across bacteria, archaea, and eukaryotes. The 1,542 human RBPs contain a repertoire of ~600 structurally distinct RBDs. Among the structural RBDs, only 20 have more than 10 human gene members, while most of them have on average one or two members (Figure 2.3, Table 5.4). mRBPs predominantly comprise the large RBD classes, mirroring the rapid expansion of mRNA metabolic processes in the evolution of higher eukaryotes (e.g. alternative splicing and polyadenylation) (Chen and Manley, 2009; Keren et al., 2010). 405 of the estimated 692 mRBPs contain an RRM, KH, DEAD, dsrm, or zinc finger domain (Figure 2.3). In contrast, the 169 ribosomal proteins have 119 distinct domains exclusively found in ribosomal proteins. This diversity of RBDs complicates both the task of defining a census, as well as *de novo* identification of RBPs, and explains why earlier approximations based on the few large structural groups underestimated the number of proteins involved in PTGR.

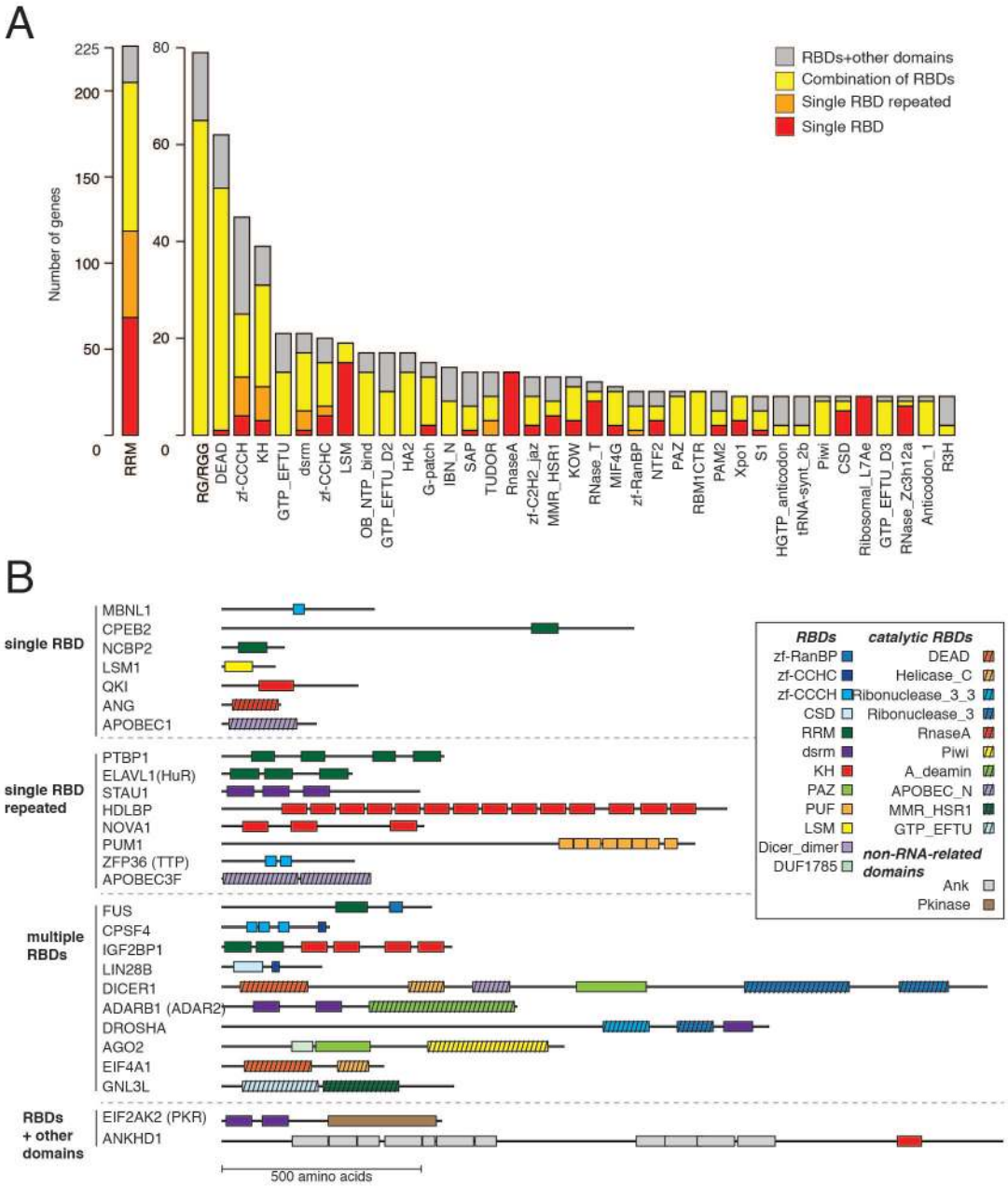
Prototypical ssRNA-binding domains interact with their targets in a nucleobase-sequence-specific manner, typically binding between 4-8 nucleotides (Glisovic et al., 2008; Lunde et al., 2007; Singh and Valcarcel, 2005). Specificity is introduced mainly by hydrogen bonding and Van-der-Waals interactions of the nucleobases with the protein side chains or the carbonyl and amide groups of the main chain (Auweter et al., 2006), often leaving the RNA phosphate backbone exposed to the solvent. Additional base stacking interactions with aromatic amino acids or positively charged residues in cationic π-interactions serve to increase affinity. dsRNA-binding proteins achieve specificity through recognition of shape of RNA secondary

structure, such as stem loops (Masliah et al., 2013). Non-sequence-specific RBDs generally interact with the negatively charged phosphate backbone, leaving the bases exposed to the solvent. To achieve specificity, these RBPs can interact with cofactors recruiting them to specific targets, as has been observed for many RNA helicases (Auweter et al., 2006; Rocak and Linder, 2004).

While many RBDs and DNA-binding domains derive from a few common superfamily folds, such as the oligonucleotidyl transferase fold (Kuchta et al., 2009) and the oligosaccharide-binding fold (OB-fold) (Arcus, 2002), RBDs largely diversified throughout evolution. Oligonucleotidyl transferase fold proteins include enzymatic RBPs such as TUTases, poly(A) polymerases, RNA ligases, tRNA CCA-adding enzymes and immune-stimulatory 2',5'-oligoadenylate synthases (Kuchta et al., 2009). RBDs of OB-fold superfamily are the S1, PAZ, and CSD domains (Arcus, 2002; Lunde et al., 2007; Murzin, 1993).

Analyzing the number and abundance of structural domain classes, it became evident that most RBDs had only one member. Only 4% of all RBD classes found in human have more than 8 members. Further analysis of this newly curated census showed that members of the 26 most abundant RBD classes constituted a third of the 1,542 RBPs - most of them were mRBPs. Most of the prominently studied "canonical" RBDs (Lunde et al., 2007), such as the PUF (2 proteins), S1, CSD, and PIWI domains (8 members each), did not represent highly abundant RBD classes in humans and their overrepresentation in PTGR studies does not mirror their abundance in the genome. One of the largest functional groups, ribosomal proteins and proteins involved in ribosome maturation, possessed almost all unique domains for each protein and could not be classified into large families of related structural organization (Korobeinikova et al., 2012). In conclusion, mainly mRNA-related processes showed an expansion of protein families and are frequently found in small families of paralogs.

Figure 2.3 Overview of most abundant RNA-binding domains (RBDs) with selected examples in humans. (A) Counts of proteins with Pfam-defined RBDs with 8 or more members in humans. Domain names are listed according to Pfam nomenclature; additional information can be found in Table 5.4. In addition, low complexity RG/RGG repeat regions in RBPs are shown, defined as by at least three RG/RGG repeats spaced 10 amino acids or less apart. Counts are further subdivided to indicate the number of genes containing (1) one RBD as the only structural domain in the encoded protein (red), or (2) repeats of the same class RBD (orange), or (3) one or more RBDs in combination with different class RBDs (yellow), or (4) combinations of the RBD with one or more domains unrelated to RNA-metabolic function, e.g. protein kinase domains (grey). (B) Domain structure organization of representative RBPs, scaled by amino acid length and categorized into the domain combination classes listed in (A).


A
Number of genes
RBDs+other domains
Combination of RBDs
Single RBD repeated
Single RBD
RRM
RG/RGG
DEAD
zf-CCCH
KH
GTP_EFTU
dsrm
zf-CCHC
LSM
OB_NTP_bind
GTP_EFTU_D2
HA2
G-patch
IBN_N
SAP
TUDOR
RnaseA
zf-C2H2_jaz
MMR_HSR1
KOW
RNase_T
MIF4G
zf-RanBP
NTF2
PAZ
RBM1CTR
PAM2
Xpo1
S1
HGTP_anticodon
tRNA-synt_2b
Piwi
CSD
Ribosomal_L7Ae
GTP_EFTU_D3
RNase_Zc3h12a
Anticodon_1
R3H
B
single RBD
MBNL1
CPEB2
NCBP2
LSM1
QKI
ANG
APOBEC1
single RBD repeated
PTBP1
ELAVL1(HuR)
STAU1
HDLBP
NOVA1
PUM1
ZFP36 (TTP)
APOBEC3F
multiple RBDs
FUS
CPSF4
IGF2BP1
LIN28B
DICER1
ADARB1 (ADAR2)
DROSHA
AGO2
EIF4A1
GNL3L
RBDs + other domains
EIF2AK2 (PKR)
ANKHD1
500 amino acids
RBDs
zf-RanBP
zf-CCHC
zf-CCCH
CSD
RRM
dsrm
KH
PAZ
PUF
LSM
Dicer_dimer
DUF1785
catalytic RBDs
DEAD
Helicase_C
Ribonuclease_3_3
Ribonuclease_3
RnaseA
Piwi
A_deamin
APOBEC_N
MMR_HSR1
GTP_EFTU
non-RNA-related domains
Ank
Pkinase

A common feature of the large mRNA-binding domain classes is their frequent occurrence in multiple repeats or in combination with other RBDs. The modular design also allows for rapid evolutionary adaptation of proteins to new RNA targets (Lunde et al., 2007), which in some cases poses interesting questions about the regulatory functions of RBPs and evolution of targets. For example, while most KH-domain-containing RBPs have one or two KH domains, the HDLBP gene expanded from 7 KH repeats in *S. cerevisiae* to 14 KH repeats in humans (Figure 2.3).

Overall, this analysis highlighted the total number of structural RBD classes present in humans and their members, and highlighted that many RBD classes are currently far from a comprehensive characterization. Even among the established large RBD classes, the function of many individual members, including RRM proteins, helicases (DEAD, HA2), zinc finger proteins (zf-CCCH, zf-CCHC, jaz-like zf-CH2H2), RNA nucleases (RNase A, RNase T, RNase Zc3h12a), and putative translation factors (GTP-EFTU) have not been characterized in their RNA-binding capacity and physiological roles. For at least a third of all 1,542 RBPs their biological functions are unknown or merely inferred from homologs.

2.4.3 Abundance of RBPs across tissues in comparison to other proteins

The importance of PTGR is revealed by analyzing RNA-seq expression levels of RBPs relative to the residual proteome across 16 human tissues of the Illumina body map (http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-513/). The majority of RBPs were ubiquitously expressed and typically at higher levels than average cellular proteins (Figure 2.4 A), consistent with previous analyses (Kechavarzi and Janga, 2014; Vaquerizas et al., 2009). While RBPs and TFs encoded a similar number of genes (1,542 and 1,704 genes, respectively), the cumulative abundance of rpkm expression levels of RBPs contributed up to 20% of the expressed, protein-coding transcriptome, whereas TFs constituted only up to 3% by transcript

abundance (Figure 2.4 B). The equivalent of 10-12% of the expressed transcriptome originates from the 169 ribosomal proteins but represents only about 0.8% of protein-coding genes in the genome. In contrast, the transcripts encoded by the 692 mRBP genes, representing 3% of protein-coding genes, accounted for 4-5% of the transcriptome, while all remaining RBPs (tRNA-, rRNA-, etc.) contributed the remaining 4-5%. The discrepancy between RBPs, TFs as the main gene regulatory protein groups, and other cellular protein groups, such as cytoskeletal proteins (CSK proteins), was also illustrated when one compares the number of total genes present in the genome (Figure 2.4 C) to their relative expression in the transcriptome (Figure 2.4 D). Scaling each by total mRNA abundance, the group of transcription factors displays the largest reduction in percentage going from gene presence to expression (from 6.2% to 2.4%), while ribosomal proteins show the largest percentage increase (1.1% to 10.3%). All other RBPs are relatively unchanged in their contribution and CSK proteins increase by 50% in smooth muscle cells.

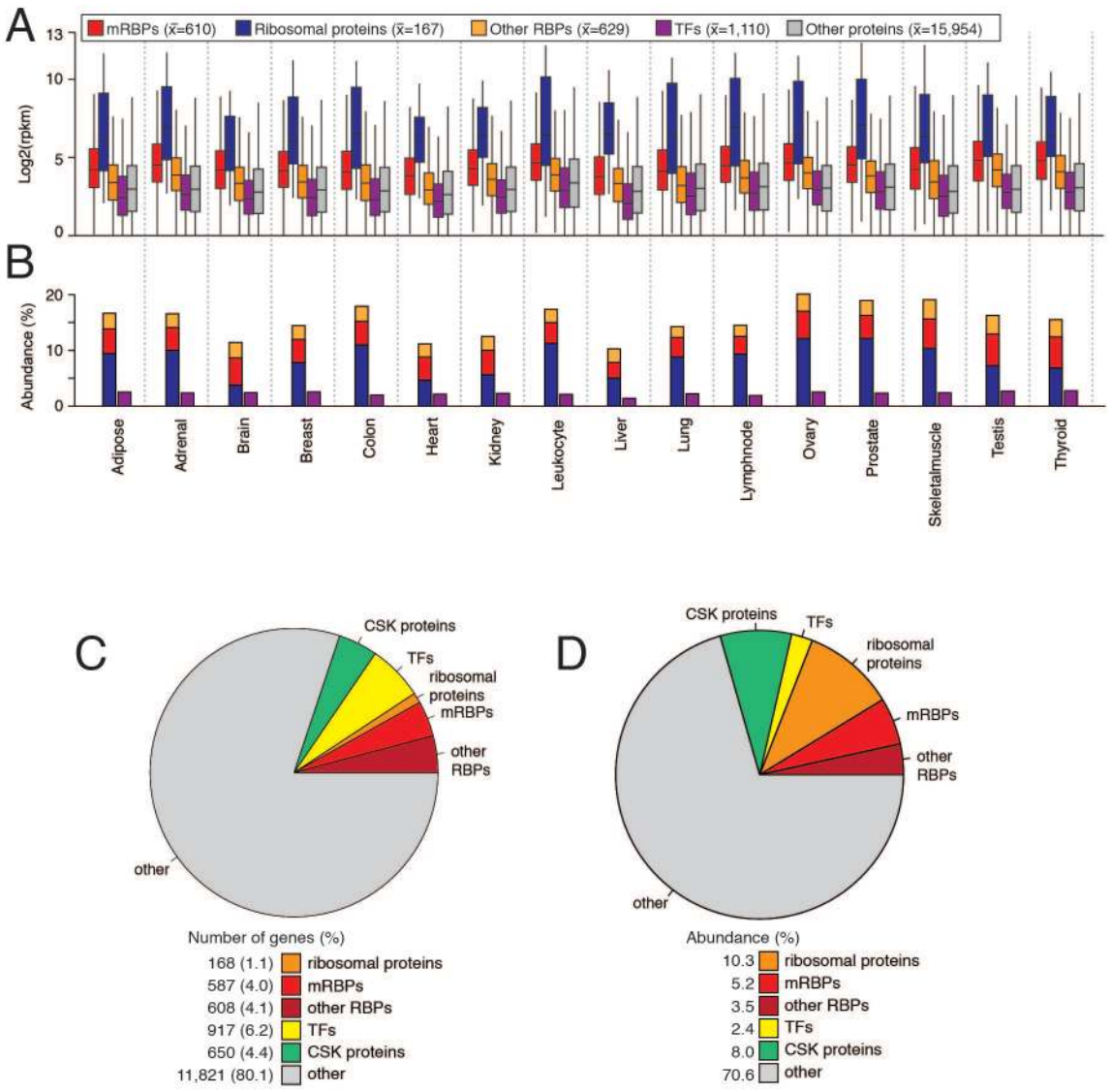


Figure 2.4. Transcript abundance of RBPs and TFs across 16 different human tissues.
(A) Distribution of gene expression levels of protein-coding genes measured by RNA-seq with rpkm expression values ≥ 1. Shown as subgroups are mRNA-binding proteins (mRBPs) (red), ribosomal proteins (blue), the remaining RBPs (orange), TFs (magenta), and the residual protein-coding transcriptome (grey). For each group, the mean number of expressed proteins across the tissues is shown in brackets. (B) Cumulative abundance of RBPs (blue, red, orange) and TFs (magenta) as fraction of all RNA-seq reads. (C) Pie chart of the number of protein-coding genes in the genome encoding for mRBPs (red), ribosomal proteins (orange), other RBPs (dark red), transcription factors (yellow), cytoskeletal proteins (green), other proteins (grey). (D) Pie chart of the rpkm cumulative expression in muscle cells of all genes in (C) expressed with rpkm ≥ 1.

These relative mRNA abundances illustrated the central importance of translation-related processes in the cell. Tumors and immortalized cell lines express mRBPs and ribosomal proteins at even higher levels than normal tissues. The increased demand for continuous protein production, changes in nucleolar size (the site of ribosome and rRNA biogenesis), and translational activities have long been considered as hallmarks of cancer (Boisvert et al., 2012; Montanaro et al., 2008; Ruggero and Pandolfi, 2003; Zhang et al., 2015a). More generally, altered translational activity has been observed in a wide range of human pathologies and has also recently been connected to neurodegenerative diseases, including Parkinson's and Alzheimer's disease (Klein and Westenberger, 2012; Ma et al., 2013; Martin et al., 2014; Scheper et al., 2007b). Given the central importance of protein translation and ribosome biogenesis for energy metabolism and cellular growth, understanding disease-related changes in PGTR pathways is not only of diagnostic and possibly prognostic value, but also enables therapeutic approaches. Consequently, targeting of PTGR pathways has been explored in drug development of inhibitors to block translation initiation or ribosome biogenesis, such as the anti-cancer drug silvestrol inhibiting the translation initiation factors EIF4A1 and 2 (Grzmil and Hemmings, 2012; Hein et al., 2013; Silvera et al., 2010; Skrtić et al., 2011).

2.4.4 Categorization of RBPs into target subclasses

A classification of RBPs by interacting RNA targets is useful as it isolates individual PTGR pathways and can also explain similar phenotypes for RBP genes in human diseases. I grouped the set of 1,542 human RBPs based on literature reports (Figure 2.5) into mRNA-, rRNA-, tRNA, snRNA-, snoRNA-binding, and a residual ncRNA-binding category (Table 5.3). Further categories were introduced to define protein components of the ribosome, diverse RBPs interacting indiscriminately with many types of RNAs (such as the RNA exosome in general RNA turnover) and unknown-target RBPs (proteins with known RBDs or some experimental data

on RNA-binding, which lacked information on specific targets). RBPs with more than one target class were also found and emerging transcriptome-wide binding studies reveal that it may be common for certain RBPs to interact with and regulate multiple classes of RNAs, e.g. LIN28 proteins bind let-7 pre-miRNA, mRNAs, and snoRNAs, while DGCR8 binds dsRNA regions within pre-miRNAs and mRNAs (Cho et al., 2012; Hafner et al., 2013; Macias et al., 2012; Wilbert et al., 2012). For simplification we grouped RBPs, when known, by the predominantly interacting RNA class based on literature reports, homolog conservation and structural domain information.

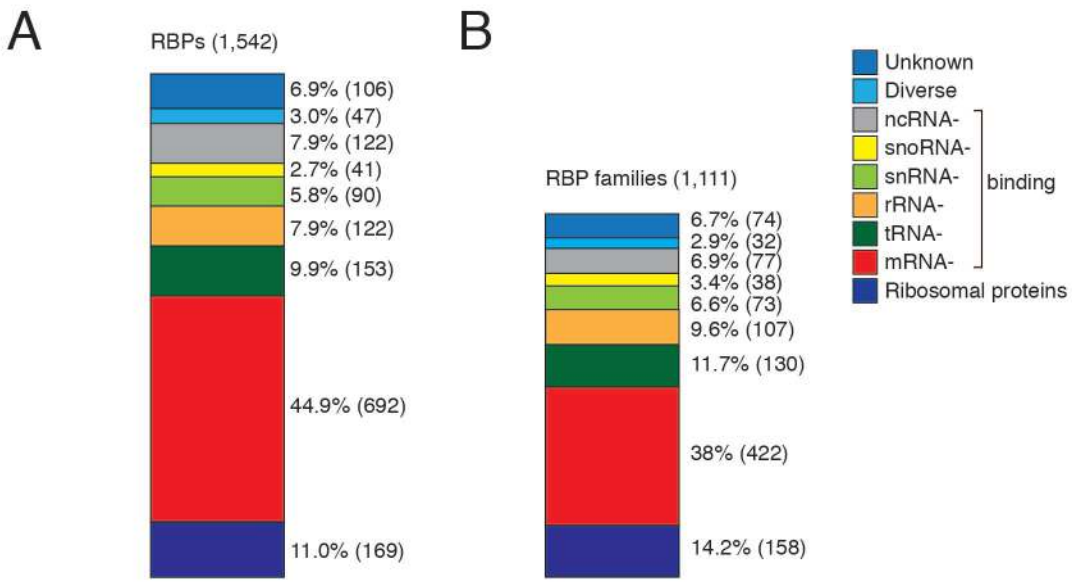


Figure 2.5 Target RNA classification of RBPs and of RBP paralogous families. (A) RBPs were grouped by their respective target RNAs: ribosomal proteins (dark blue), mRNA- (red), tRNA- (dark green), rRNA- (orange), snRNA- (light green), snoRNA- (yellow), ncRNA-binding (grey), diverse targets (light blue), unknown targets (blue). Percentage and counts (in brackets) of RBPs in the category are shown. (B) Same as in (A) but paralogs are grouped into families defined by 20% sequence identity according to Ensembl Compara.

Almost all categories of RBPs are directly or indirectly invested in the process of protein synthesis: 692 proteins were mRNA-binding, 169 ribosomal proteins, and 130 proteins were involved in biogenesis and delivery of charged tRNAs to the ribosome. Another 90 proteins were

involved in biogenesis of snRNAs or the formation of snRNPs (snRNA-protein complexes) and 122 and 41 RBPs took part in rRNA- and snoRNA biogenesis, respectively, which may be an underestimation given that rRNA biogenesis and some orphan snoRNAs or snoRNA-like lncRNAs have yet to be fully characterized (Bratkovič and Rogelj, 2014; Henras et al., 2008; Tafforeau et al., 2013; Yin et al., 2012). 122 RBPs were grouped together that interacted with the remaining ncRNAs (including all remaining ncRNA categories, e.g. miRNAs, piRNAs, lncRNAs, the RNA components of MRP and RNase P etc.. For the different categories see Table 1.1 and 5.3). These proteins associate with a range of ncRNAs, some of which are involved in gene regulation, e.g. miRNAs, piRNAs, and lncRNAs, and control gene expression through posttranscriptional RNA degradation, transcriptional silencing/activation of gene loci, and translational repression/activation. Others act as structural and catalytic components of RNP complexes (RNase MRP and P nucleases, telomerase RNP), or form RNP complexes of unknown functions, e.g. vault RNAs and Y RNAs. 47 RBPs, mostly RNA nucleases involved in general RNA turnover, were categorized into the diverse target RNA group.

The categorization of proteins into different pathways became highly useful for breaking down the complexity of PTGR into basic units, and allows now detailed analyses of changes taking place in different regulatory pathways in system-wide PTGR studies. Furthermore, it facilitates exploration of areas in PTGR, which have been relatively overlooked. For example, our curation highlights that the functions of many rRNA and tRNA biogenesis factors in humans are largely inferred from distant homologs and domain relationships (Hopper et al., 2010; Kiss, 2004; Phizicky and Hopper, 2010). As these processes are essential for cellular life and highly conserved, core functions often remain the same. However, with increasing complexity of organisms, protein factors and their family members, and/or the spectrum of target RNAs of pathway components evolved, grew in size, and diverged. The recently discovered new roles of tRNA methylases in mRNA and ncRNA metabolism may represent such an example (Hussain et

al., 2013; Sibbritt et al., 2013). Future characterization of these proteins will be important to delineate the specificities and targets of these factors in higher eukaryotes.

2.4.5 Conservation of RBP and TF families

Phylogenetic relationships of RBPs reveal the creation of gene families during evolution, a process that in principle allows for diversification of RNA metabolic pathways. In most instances, however, human paralogs are functionally overlapping, with similar or even identical binding sites (Ascano et al., 2012a; Ray et al., 2013; Spencer et al., 2006; Todd et al., 2001), and the evolution of paralogs represent an alternative to facilitating regulation across cell types. Understanding family relationships and examining protein families together, instead of singular members, is therefore a valuable approach.

To define redundant properties of RBPs, I compared the evolutionary characteristics of RBPs to TFs, the regulatory group of proteins controlling gene expression at the transcriptional level. I used the phylogenetic homology classification as already defined by the Ensembl Compara database (Vilella et al., 2009) and further grouped together paralogs with even closer homology. Most RBP paralogs shared 20-70% sequence identity and by this criteria known functionally related RBPs grouped together, such as members of the CPEB1-4 cytoplasmic polyadenylation family, which shared ~25% sequence identity. We refer to these grouped paralogs as ‘paralogous RBP families’ throughout the text. We found a minimum of 20% to be the best threshold to group functionally related or redundant proteins into families. Sequence identity cut-offs below 20% included predominantly functionally unrelated, distant paralogs, which were not in the same RNA regulatory pathways, while higher cut-offs often missed family members of known functional similarity.

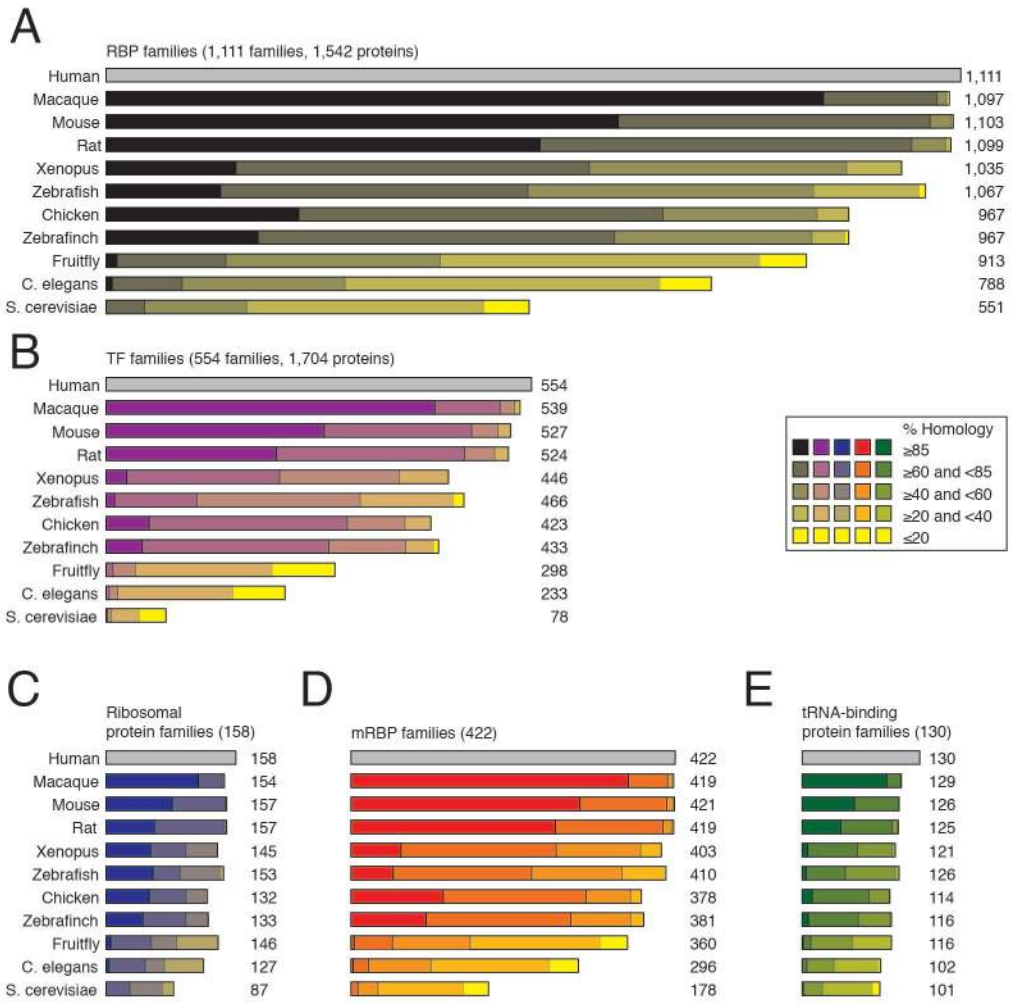


Figure 2.6 Evolutionary conservation of RBP and TF families. (A, B) Number of human RBP and TF families conserved across 10 species and their percentage identity score, (A) RBP (black) and (B) TF families (magenta). The number of families with different degrees of conservation are binned into 5 categories, color-coded in black-yellow and magenta-yellow shades: (i) ≥85% homology, (ii) ≥60% and <85% homology, (iii) ≥40% and <60% homology, (iv) ≤20% homology. (C,D,E) Number of paralogous families and their degree of sequence identity in humans. (C) ribosomal proteins (blue), (B) mRBPs (red), and (D) tRNA-binding proteins (green).

Consistent with their high structural diversity, the 1,542 RBPs formed 1,111 families, in which individual members within RBP families generally have the same RNA target class (Figure 2.6). In contrast, the 1,704 human TFs, which diverged more recently than RBPs [census generated by Vaquerizas *et al.* (Vaquerizas et al., 2009)], formed only 554 protein families by the above stated homology criteria (Table 5.4). RNA- and DNA-binding domains often originated from common superfamily folds, such as the OB-fold (oligonucleotide/oligosaccharide-binding fold), nucleotidyltransferase, zinc finger, or DNA/RNA helicase domains (Arcus, 2002; Jankowsky and Fairman-Williams, 2010; Krishna et al., 2003; Kuchta et al., 2009). However, despite a shared evolutionary history, TFs expanded late in evolution into large families by multiple gene duplications (Vaquerizas et al., 2009) and the current homology grouping resulted in TF families with up to 50 members per family and 2.5 members on average, while RBPs diversified early and paralogous families comprised 1.3 members on average, with the largest RBP families including up to 10 members. Paralogous RBP families were well conserved across eukaryotes and 50% of the human RBP families were also present in *S. cerevisiae* (Figure 2.6). This finding is consistent with previous observations that at least 200 distinct RBPs were present in the lowest common ancestor of animals (LCA), and 80 orthologous groups of RBPs were traceable even to the lowest universal common ancestor (LUCA) (Anantharaman et al., 2002; Kerner et al., 2011). In striking contrast, few TFs were conserved across evolution, and only 14% of the human TF families were found in *S. cerevisiae.* Even the most rapidly expanding group, mRBPs, had 178 of 422 (42%) mRBP families conserved in yeast. While the expansion of TFs traced organismal complexity, possibly enabling the development of new functionalities (Vaquerizas et al., 2009), evolutionary stability of RBPs went along with the early establishment of core RNA metabolic processes in all cellular systems (Anantharaman et al., 2002).

2.4.6 Conservation of RBP families interacting with different RNA classes

With the growing number of RBP families in higher eukaryotes the relative size of the various RBP families committed to different RNA targets remained constant across phylogenies (38% mRBPs, 12% tRNA-binding, 14% ribosomal proteins, Figure 2.6). However, RBP families in different target groups displayed varying levels of evolutionary conservation. Most RBP classes, including families in rRNA, tRNA, snRNA and snoRNA pathways, displayed average homologies of 31% between human and yeast (e.g. tRNA-binding proteins in Figure 2.6). Ribosomal proteins were among the most conserved, with an average homology of 51%, in contrast to the factors involved in maturation and processing of rRNAs, which were conserved only to 31%, reflecting the increasing divergence of rRNA biogenesis factors in higher eukaryotes (Granneman and Baserga, 2004; Phipps et al., 2011). ncRNA- and mRNA-binding protein families displayed the lowest conservation, with 20% and 27%, respectively, and only a fifth of all ncRNA-binding families had homologous families in yeast.

2.4.7 Phylogenetic comparisons of small ribosomal and KH-domain proteins

Visualization of the evolutionary relationships of RBP families facilitates systems biology approaches to dissect their regulatory roles by giving an intuitive graphic representation of the conservation of proteins. They highlight closely related homologs and provide a glimpse into function of uncharacterized RBPs when the homolog has been characterized. Phylogenetic comparison of the predominantly mRNA-binding KH-domain-containing proteins and the proteins of the small subunit of the cytosolic ribosome illustrate the differences in their evolutionary trajectory and I used these extensively to judge family relationships and conservation among RBP families (Figure 2.7).

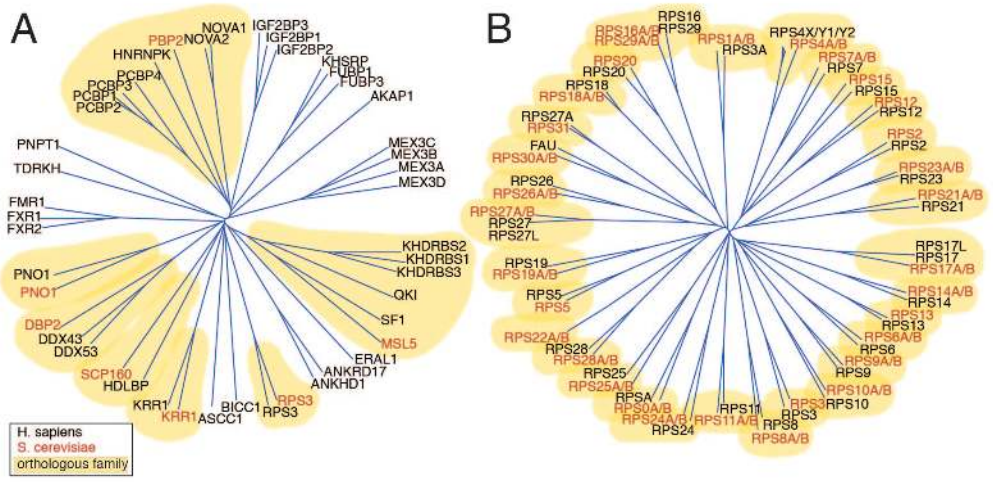


Figure 2.7 Phylogenetic trees of RBP families highlight their evolutionary history. Phylogenetic trees of (A) KH-containing proteins and (B) ribosomal proteins of the small subunit. Branch lengths are scaled to the sequence identity of the proteins. *S. cerevisae* proteins are marked in red, human proteins in black, homologous families with conserved members in *S. cerevisae* are highlighted in yellow.

KH proteins experienced multiple gene expansions, as noted earlier for mRBPs, and evolved new RBP families at the later metazoan stages. Thereby they expanded and diversified, evolving new factors involved in various regulatory pathways, such as mRNA splicing, translational regulation, and transport. KH protein families contain between one to four members in human, and possess generally one distantly related homolog in yeast (Figure 2.7). Multiple family members often have redundant biological functions and RNA target spectra. For example, members of the FMR1 family (FMR1, FXR1, FXR2) or the IGF2BP1 family (IGF2BP1, 2, and 3) show >90% identical RNA-binding specificities (Ascano et al., 2012b; Hafner et al., 2010a). In contrast, cytosolic ribosomal proteins display an unusually high conservation, not too surprising, given that the process of protein translation is conserved to such a high degree between prokaryotes and all clades of eukaryotes that functional details of translation determined in bacteria are almost identical to higher systems (Dever and Green, 2012; Melnikov et al., 2012; Wool et al., 1995).

The ~90 ribosomal cytosolic proteins are highly similar in structure between yeast and human and show late divergence in evolution, illustrated for the phylogenetic tree of small ribosomal subunit proteins (Figure 2.7). With on average 55% protein identity, 96% of all human ribosomal proteins have direct one-to-one, or due to a whole genome duplication in yeast, one-to-two or two-two matching homologs (Anger et al., 2013; Wool, 1979; Wool et al., 1995). From the level sequence identity and number of homologs we often also gain functional insights: high conservation of RBPs with 1:1 conserved homologs (or in the case of *S. cerevisiae* gene duplication 1:2) often points to either the RBP being part of a highly conserved RNP complex with a structural RNA component at its core (e.g. ribosome, snRNPs, snoRNPs), or towards a conserved enzymatic process with little mechanistic divergence (e.g. tRNA splicing). In contrast, RBPs involved in mRNA-gene regulatory processes display much higher redundancy and less sequence conservation. Thus, intuitively one can make a prediction about the likely process encountered from the phylogenetic tree. The ~80 human mitochondrial ribosomal proteins form the exception to the rule (Matthews et al., 1982). The majority (80%) have no homologs in yeast and the few that do display low conservation of 22% sequence identity. The low conservation reflects their evolutionary history. Mitochondrial ribosomes were acquired through eubacterial endosymbiosis and rapidly evolved independently across species with major remodeling events happening later during evolution (Cavdar Koc et al., 2001; O'Brien, 2003). In contrast, the nuclear encoded RBPs show characteristic conservation patterns and protein family expansions for the different RNA regulatory processes.

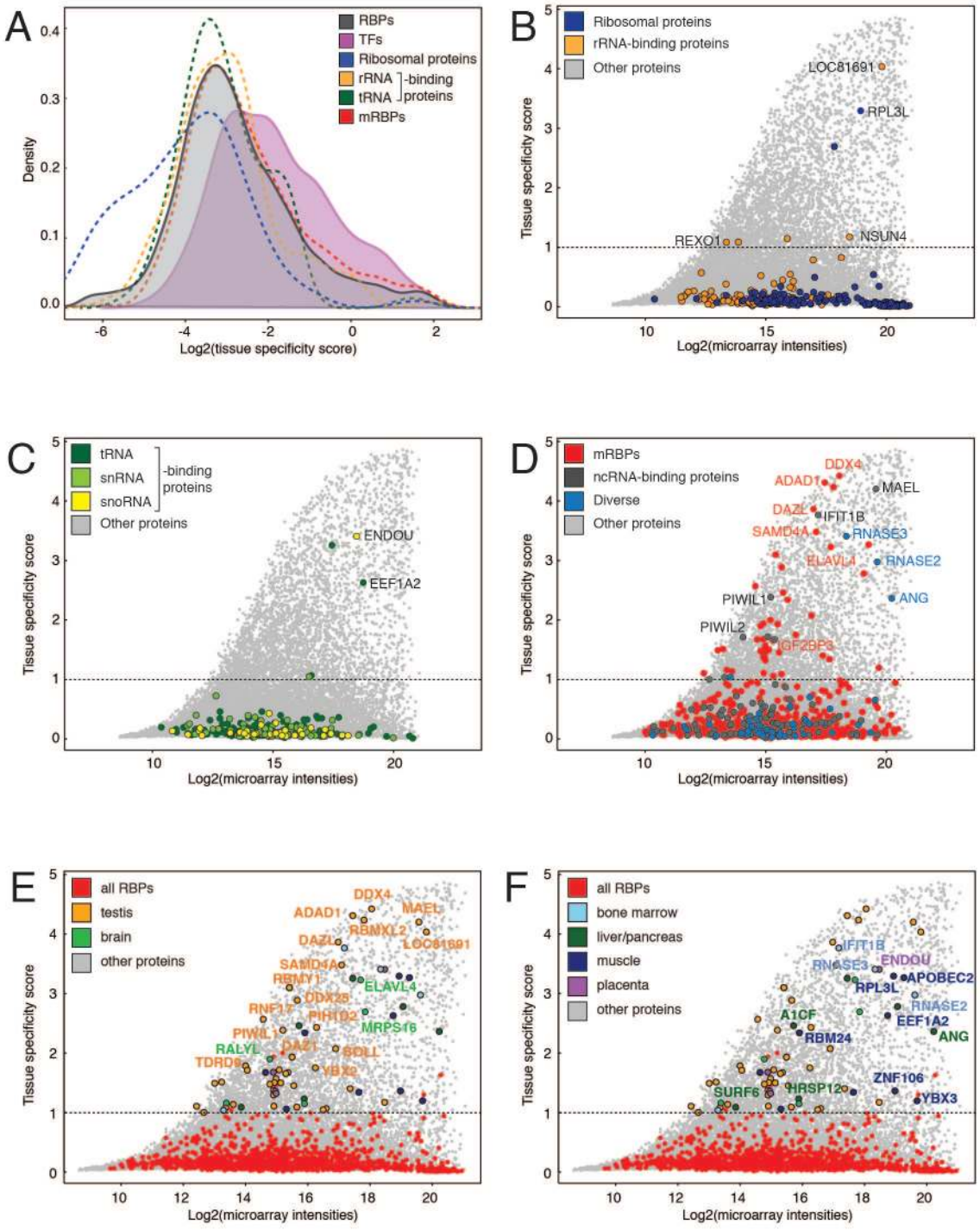
2.4.8 Tissue specificity of RBPs

Tissue-specific expression, phyletic age, and cellular functions of proteins often correlate. While ancient and highly conserved genes are widely expressed and support basic cellular functions, more recently evolved genes, such as TFs and secreted proteins, are expressed in a species- and

tissue-specific manner (Freilich et al., 2005; Ramsköld et al., 2009; Winter et al., 2004). I investigated mRNA expression of 1,441 RBPs and 1,463 TFs profiled in a microarray study measuring transcript levels of 16,867 protein-coding genes across 31 human tissues (Dezso et al., 2008). To analyze protein abundance, we assumed that transcript abundance approximated protein abundance in the cell as previously shown by large-scale transcriptomic and proteomic studies (Guo et al., 2010; Schwanhausser et al., 2011). Tissue-specific variation of RBP isoforms due to alternative splicing and alternative cleavage and polyadenylation were not considered for this analysis as they are not well understood. We favored microarray over RNA-seq studies because the former profiled larger collections of different human tissues and organs. I calculated a tissue specificity score *S* for every gene on the array across the profiled samples, which ranged from 0 for ubiquitously expressed to 5 for highly tissue-specific proteins. Based on the score of established tissue-specific RBPs, such as the germline-specific PIWI-family (*S*=1.7-2.3) or neuronal members of the ELAVL family (*S*=3.2) (Li et al., 2007; Thomson and Lin, 2009), I set an empirical cut-off score of 1 for referring to tissue-specific genes, at which 6% of RBPs and 13% of TFs showed some level of tissue-specific expression (Figure 2.8 A). As expected, ribosomal proteins (Figure 2.8 A,B), as well as general components of the spliceosome, RNA transport, and turnover machineries were ubiquitously expressed across tissues. Furthermore, while tissue-specific variation has been reported for some tRNAs and snoRNAs (Castle et al., 2010; Dittmar et al., 2006; Plotkin and Kudla, 2011), the biogenesis factors of tRNAs and snoRNAs, as well as snRNA and rRNA maturation pathways, were generally uniformly expressed across tissues (Figure 2.8 C). The majority of tissue-specific RBPs consisted of mRNA- and ncRNA-binding proteins, as well as a range of RNA nucleases with diverse target specificity (Figure 2.8D). Perhaps unexpectedly, some tissue-specific outliers were observed among rRNA biogenesis factors and ribosomal proteins, including the uncharacterized, muscle-enriched ribosomal multicopy gene RPL3L, a homolog of RPL3 (Figure 2.8B). These selectively

expressed RBPs may reflect extra-ribosomal roles (Warner and McIntosh, 2009) or tissue-specific adaptations in composition of ribosomes that regulate translation of subsets of proteins (Xue and Barna, 2012).

Figure 2.8 Tissue specificity of RBPs across 31 human tissues and organs. A tissue specificity score *S* was calculated from mRNA expression levels of 1,441 RBPs and 1,463 TFs profiled in a human microarray tissue atlas assessing expression across 31 tissues (Dezso et al., 2008). (A) Densities of the log2 transformed tissue specificity scores are shown for RBPs (black), TFs (magenta), ribosomal proteins (dark blue), mRNA- (red), tRNA- (dark green), rRNA-binding (orange) proteins. The densities of RBPs and TFs are filled in shades of their original color to highlight their shifts in distribution. (B) Log2 maximum expression intensity value of a gene versus *S* for ribosomal proteins (dark blue), and rRNA-binding proteins (orange) compared to the residual proteome (grey). Tissue-specific genes were defined as genes with scores $S \geq 1$ (dashed line). Selected genes are highlighted. (C) Same as (B) for tRNA- (dark green), snRNA- (light green), and snoRNA-binding proteins (yellow). (D) Same as (B) for mRBPs (red), ncRNA-binding (dark grey), and diverse target RBPs (blue). (E) Log2 maximum expression intensity value of a gene is plotted against its tissue specificity score *S*. RBPs are highlighted in red, all other proteins in grey. Tissue-specific RBPs with maximum expression in testis (orange) and brain (green) are highlighted. (F) Same as in (E) highlighting tissue-specific RBPs expressed in bone marrow (sky blue), liver/pancreas (dark green), muscle (blue), placenta (violet).


A
RBPs
TFs
Ribosomal proteins
rRNA -binding proteins
tRNA -binding proteins
mRBPs
Density
Log2(tissue specificity score)
B
Ribosomal proteins
rRNA-binding proteins
Other proteins
LOC81691
RPL3L
REXO1
NSUN4
Tissue specificity score
Log2(microarray intensities)
C
tRNA
snRNA
snoRNA
-binding proteins
Other proteins
ENDOU
EEF1A2
Tissue specificity score
Log2(microarray intensities)
D
mRBPs
ncRNA-binding proteins
Diverse
Other proteins
DDX4
ADAD1
MAEL
DAZL
IFIT1B
SAMD4A
RNASE3
ELAVL4
RNASE2
PIWIL1
ANG
PIWIL2
IGF2BP3
Tissue specificity score
Log2(microarray intensities)
E
all RBPs
testis
brain
other proteins
DDX4
ADAD1
MAEL
RBMXL2
DAZL
LOC81691
SAMD4A
RBMY1
ELAVL4
DDX25
RNF17
PIWIL2
MRPS16
PIWIL1
DAZ1
RALYL
SOLL
TDRD6
YBX2
Tissue specificity score
Log2(microarray intensities)
F
all RBPs
bone marrow
liver/pancreas
muscle
placenta
other proteins
IFIT1B
RNASE3
ENDOU
APOBEC2
RPL3L
RNASE2
A1CF
EEF1A2
ANG
RBM24
ZNF106
SURF6
HRSP12
YBX3
Tissue specificity score
Log2(microarray intensities)

The majority of profiled RBPs, including mRBPs, displayed no tissue specificity; 80% of all RBPs (1,144 of 1,441 RBPs) had scores lower than 0.3. Among the paralogous RBP families, 808 (77%) of the 1,049 families, or 277 (68%) of 409 mRBP families, were ubiquitously expressed with S<0.3, while only 20 families were tissue-specific with $S \geq 1$, and 53 had at least one tissue-specific member (Figure 2.9).

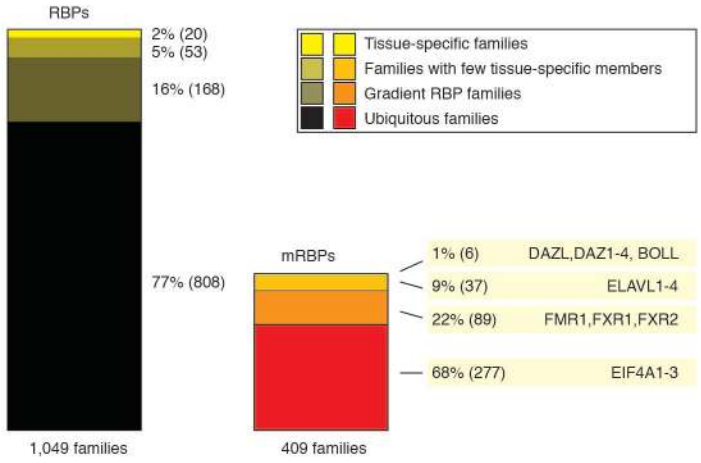


Figure 2.9 RBP families with tissue-specific or ubiquitous expression. Expression of paralogous families profiled in the tissue atlas: 1,049 RBP families, of which 409 are mRBP families, are scaled to relative size. Families are binned into different categories of expression, color-coded in shades of black-yellow and red-yellow. Representative paralogous families are highlighted for mRBPs. 2% of RBP and 1% of mRBP families displayed tissue-specific expression for all their members (yellow), 5% and 9% respectively had one or more members with tissue specificity score $S \geq 1$, 16% and 22% of families had members with tissue-specificity scores ranging between $0.3 > S > 1$, classified here as gradient RBP families, and 77% of RBPs (black) and 68% of mRBPs (red) display little variation in expression, with $S < 0.3$, named here ubiquitous RBP families.

Few tissues contained specialized RBPs and 90% of the 82 tissue-specific RBPs were identified in germline, brain, muscle, bone marrow, placenta or liver (Figure 2.8 E-F). The largest fraction, 47 proteins, were enriched in adult testis, where they contribute to gametogenesis and fertility through regulation of transposon silencing, mitosis, meiosis, stem cell maintenance, and differentiation (Kang and Han, 2011; Kotaja and Sassone-Corsi, 2007; Luteijn and Ketting, 2013; Seydoux and Braun, 2006; Siomi et al., 2011; Voronina et al., 2011) (Figure 2.8 E). These

proteins were also expressed during fetal ovary development, at the stage of mitotic and meiotic cell divisions and germ cell expansion (Houmard et al., 2009). 2% (20 families) of all RBP families were exclusively tissue-specific and most of these families were expressed in the germline, such as the DAZ1-4 and PIWI family, and other RBPs involved in the piRNA pathway (Brook et al., 2009; Reynolds and Cooke, 2005; Siomi et al., 2011; Thomson and Lin, 2009) (Figure 2.8 D-F).

Instead of RBP families where all paralogs are tissue-specifically expressed, the larger proportion of tissue-specific RBPs belongs to sequence families with at least one ubiquitously expressed member. 5% (52 families) of all RBP families were broadly expressed with one or more highly tissue-specific member. Examples included a number of helicase families with germline-specific paralogs, such as the tissue-specific DDX4 protein, comprising ubiquitously expressed family members DDX3X and DDX3Y, the MOV10L1 helicase and its ubiquitously present paralog MOV10, or the tissue-specific helicase DDX25 with the ubiquitous paralogs DDX19A and DDX19B (Dufau and Tsai-Morris, 2007; Frost et al., 2010; Lasko, 2013; Zheng et al., 2010). Also, most members of the secreted, vertebrate-specific RNase A family displayed high tissue-specificity and were expressed in bone marrow cells and liver, where they have a role in immune response and angiogenesis (ANG, RNASE2, and RNASE3) (Rosenberg, 2011) (Figure 2.8D). Other families in this group were the mRNA splicing and regulatory families ELAVL1-4 and IGF2BP1-3, which had ubiquitously expressed paralogs (ELAVL1, IGF2BP2) and highly tissue-specific members (ELAVL3-4, IGF2BP1, IGF2BP3) in the brain, germline, and liver (Li et al., 2007; Simone and Keene, 2013; Yisraeli, 2005) (Figure 2.8D). For 16% of RBP families, here named gradient RBP families, individual members were ubiquitously expressed with tissue-specificity scores below 1, but displayed, while not classified as tissue-specific, some degree of differential expression across tissues (Figure 2.9). Loss-of-function of proteins in these families often affects the tissue of highest expression most strongly. A prominent example for

RBPs with this expression pattern is the FMR1 family comprising three ubiquitously expressed members, FMR1, FXR1, and FXR2, with redundant target specificities (Ascano et al., 2012b). Of these proteins, FMR1 has the highest expression levels in brain, thyroid, and gonads, and FXR1 and FXR2 are most abundant in skeletal muscle and testis. Thus, even though activity of this protein family is necessary in every tissue, loss-of-function of FMR1 mainly affects the brain and gonads and causes mental retardation and premature ovarian insufficiency in Fragile X syndrome (FXS) or Fragile X-associated ataxia syndrome (FXTAS) (Wang et al., 2012), while mouse knockout models for FXR1 are embryonic lethal due to skeletal muscle defects (Mientjes et al., 2004). For families with some tissue-specific variation, the closely related paralogs often bind the same sites on target RNAs with similar affinities, such as the members of the FMR1 and ELAVL families, which have identical binding sites in cell culture models (Ascano et al., 2012a; 2012b; Simone and Keene, 2013). The redundant functions of the ubiquitous paralogs can therefore complicate the dissection of the role of the tissue-specific proteins and may require technically challenging experimental designs, such as generation of animals with tissue-specific knock-in/out of family members.

In conclusion, 98% of paralogous RBP families were ubiquitously expressed and their deep evolutionary conservations supports their predominant basic cellular function. Of these a fifth are families with tissue-specific and ubiquitous paralogs or ubiquitous members that are enriched in some tissues. Only 2% of families are tissue-specific for all paralogs, suggesting for these, similar to the evolution of TFs, a strictly cell-type specific contribution to PTGR pathways. Cell-type specific expression levels of an RBP and its paralogs must be considered when choosing a system to study regulatory networks and targets.

2.4.9 Co-regulated expression of RBPs in common pathways

2.4.9.1 Dynamic complexes of RBPs

RBPs assemble into dynamic RNP complexes that mature, process, regulate, or transport RNAs. Remodeling RNA structure to keep RNAs accessible to other RBPs and enzymatic RNA processing complexes, RBPs and RNPs also act as RNA chaperones, prevent aggregation, misfolding, and incomplete processing, and facilitate movement of RNA targets to required locations in the cell across cellular compartments. As a consequence, abundance of RBPs and their constituents differentially affects RNA regulation (Dreyfuss et al., 2002; Glisovic et al., 2008; Keene, 2007; Mitchell and Parker, 2014; Müller-McNicoll and Neugebauer, 2013). For example, the abundance of various splicing factors can influence mRNA splicing patterns (Chen and Manley, 2009; Kalsotra and Cooper, 2011; Kornblihtt et al., 2013; Smith and Valcarcel, 2000; Wahl et al., 2009), while U1 snRNP levels control alternative polyadenylation sites (Berg et al., 2012; Kaida et al., 2010). The competition among RBPs with similar or overlapping target specificity can also define regulatory modes. For example, ELAVL1 protein antagonizes miRNA regulation on a number of mRNA targets (Mukherjee et al., 2011), LIN28 protein competes with the miRNA-processing machinery to suppress pre-let-7 miRNA processing (Cho et al., 2012; Hafner et al., 2013; Wilbert et al., 2012), and PUM proteins synergize with miR-221/222 to destabilize the CDKN1B mRNA (Kedde et al., 2010). Similarly, multiple RBPs are involved in localization and transport of RNA to distinct RNP granules, which contain highly concentrated subsets of RNAs and RBPs and act in the storage and/or degradation of mRNAs (Anderson and Kedersha, 2009; Kedersha and Anderson, 2007). The central position of PTGR regulatory networks in cellular processes shows in genetic knockouts of RBPs that are often lethal or affect all tissues, consistent with their high conservation, number of targets, and low tissue-specificity. Selective expression of a single RBP typically does not result in differentiation or dedifferentiation into distinct cell types, in contrast to TFs (Hanna et al., 2010). Instead, the

interactions of many RBPs in regulatory complexes determine specificity of PTGR processes. Hence, groups of RBPs in common PTGR pathways are often co-expressed and can drive coordinate expression of targets in cells, tissues, and across developmental processes (Cirillo et al., 2014; McKee et al., 2005; Mittal et al., 2009). Specific expression of RBPs can be used to deduce their putative roles and to identify novel components of regulatory pathways. In the next two sections I investigated co-regulated RBP expression and used available gene expression data across several developmental stages in human ovary, testis, and brain to identify correlated expression of groups of RBPs that potentially act in the same developmental pathways.

2.4.9.2 Co-expression of RBPs required for ovarian development

The germline presents a unique system for functional studies of process-specific RBPs, as it has a highly specialized RNA metabolism. At least 50 tissue-specific RBPs contribute to differentiation and maintenance of germ cells (Seydoux and Braun, 2006) and many are involved in germline-specific piRNA-induced transposon silencing, alternative polyadenylation and translational regulation affecting hundreds of mRNA targets (Di Giammartino et al., 2011; Lianoglou et al., 2013; MacDonald and McMahon, 2010; Norbury, 2013; Seydoux and Braun, 2006; Siomi et al., 2011). Between 8 to 20 weeks of gestation, oogonia proliferate and their numbers increase from 0.6 to 6 million cells (Oktem and Urman, 2010). By 20 weeks of gestation, primordial oocytes enter meiosis and arrest in the diplotene stage of meiosis I prophase I until oogenesis resumes in puberty. I examined RBP expression in a microarray study of 9-18 week human fetal ovary (Houmard et al., 2009). Expression of germline-specific RBPs peaked at 14, 16.4, 16.9, and 18 weeks (Figure 2.10), and displayed highly correlated expression dynamics, reaching Pearson coefficients close to 1 (Figure 2.10).

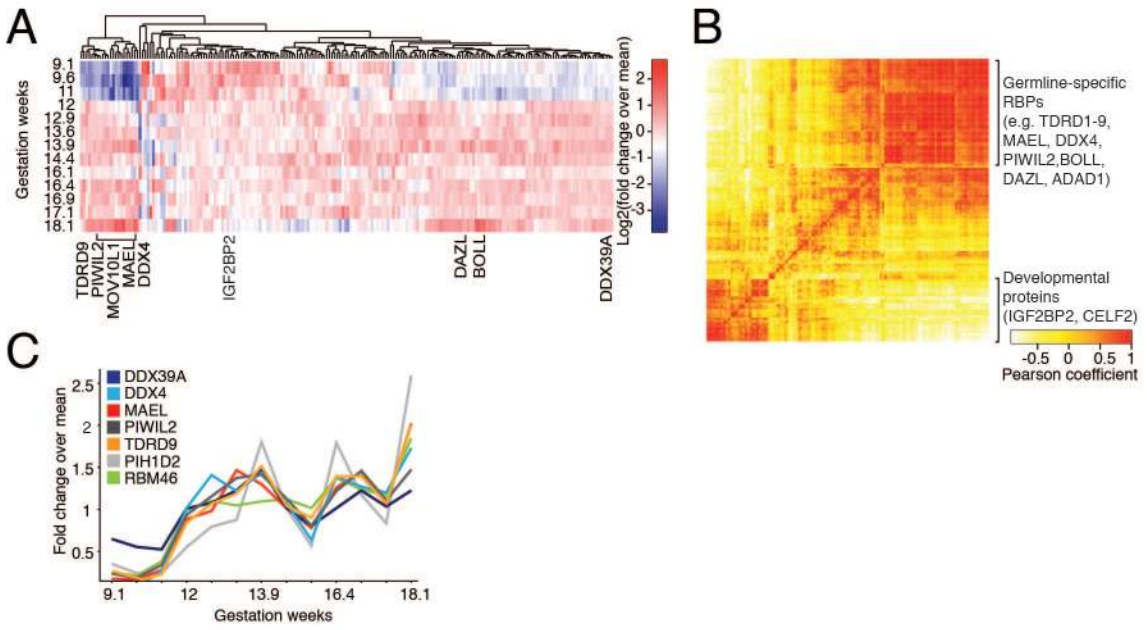


Figure 2.10 Expression of RBPs across 9 gestational stages of fetal human ovarian development. The top 200 most differentially expressed RBPs from a microarray study profiling human fetal gonad development are shown (Houmard et al., 2009). For each gene microarray, intensity values were normalized to relative fold changes by dividing the expression value by the mean expression value across developmental stages. (A) Heat map of the log2 transformed relative fold changes of the RBPs shown sorted by unsupervised clustering. Some gonad-specific RBPs are indicated. (B) Pearson correlation map indicates correlated expression changes of the 200 selected RBPs. Functionally related RBPs in gonad development cluster into a distinct expression group. (C) Plot showing the normalized expression changes of selected genes relevant in gonad development.

While some of the expression changes may be attributed to changing percentages of tissue composition of germline and somatic cells, the increase in expression for known germline-specific RBPs was evident and also correlated with their high tissue-specific expression in adult testis, confirming a role in germline development for both sexes. The expression dynamics were clearly distinct from differentially expressed, somatic RBPs with functions unrelated to germline development. For instance, the IGF2BP1-3 proteins, required during embryogenesis and organ development, were highly expressed at week 9 before expression levels rapidly decreased (Bell et al., 2013; Yisraeli, 2005). All constituents of the piRNA pathway (Siomi et al., 2011) were upregulated in the course of germline development, including piRNA biogenesis factors such as

the RNA endonucleases MAEL and PLD6, the RNA helicases DDX4, DDX39A (UAP56 homolog), and MOV10L1, and most members of the Tudor protein family (TDRD1-9, RNF17) (Siomi et al., 2011) (Figure 2.10). In addition to the piRNA pathway, we observed coordinated expression of the established germline-specific translational regulators DAZ1-4, DAZL, and BOLL, which are also essential regulators of gametogenesis (Brook et al., 2009; Kee et al., 2009). The expression dynamics of groups of RBPs during ovarian development and expression patterns mirrors their role in germline development derived from genetic experiments. In conclusion, the clustering of expression profiles of RBPs across developmental stages allowed us to investigate novel regulatory roles of RBPs not previously studied in germline development, including RBM46, PIH1D2, ADAD1, and PNLDC1 (Figure 2.10). Resulting from this clustering analysis I discovered one novel factor, LOC81691 (NEF-sp), which displayed differential expression across development identical to gonad-specific RBPs and characterized this factor in detail in Chapter 3.

2.4.9.3 Co-expression of RBPs in brain development

Neurons demonstrate unique alternative splicing and polyadenylation of mRNAs (Chen and Manley, 2009; Di Giammartino et al., 2011; Li et al., 2007; Lianoglou et al., 2013; Norbury, 2013). Furthermore, the considerable length of neuronal projections makes mRNA transport and local translation at neuronal dendrites indispensable for development, synaptic plasticity, and long-term memory (Bramham and Wells, 2007). Not surprisingly, many RBPs regulating splicing, RNA transport, storage, and translation are critical for neuroplasticity (Jung et al., 2014; Kandel et al., 2014; Sutton and Schuman, 2006). In order to capture brain-specific PTGR networks, I examined the expression dynamics of RBPs at different fetal and postnatal stages in human hippocampus development using RNA-seq data from the BrainSpan database (http://www.brainspan.org). While more than 75% of all protein-coding genes were reported to be expressed in brain (Hawrylycz et al., 2012), the expression of thousands was found to be either

restricted to a particular cell-type or to be temporally regulated (Hawrylycz et al., 2012; Miller et al., 2014). We found in our analysis that RBPs were generally expressed ubiquitously, but we detected distinct groups of RBPs that were upregulated at different developmental stages, consistent with previous studies (McKee et al., 2005; Mody et al., 2001).

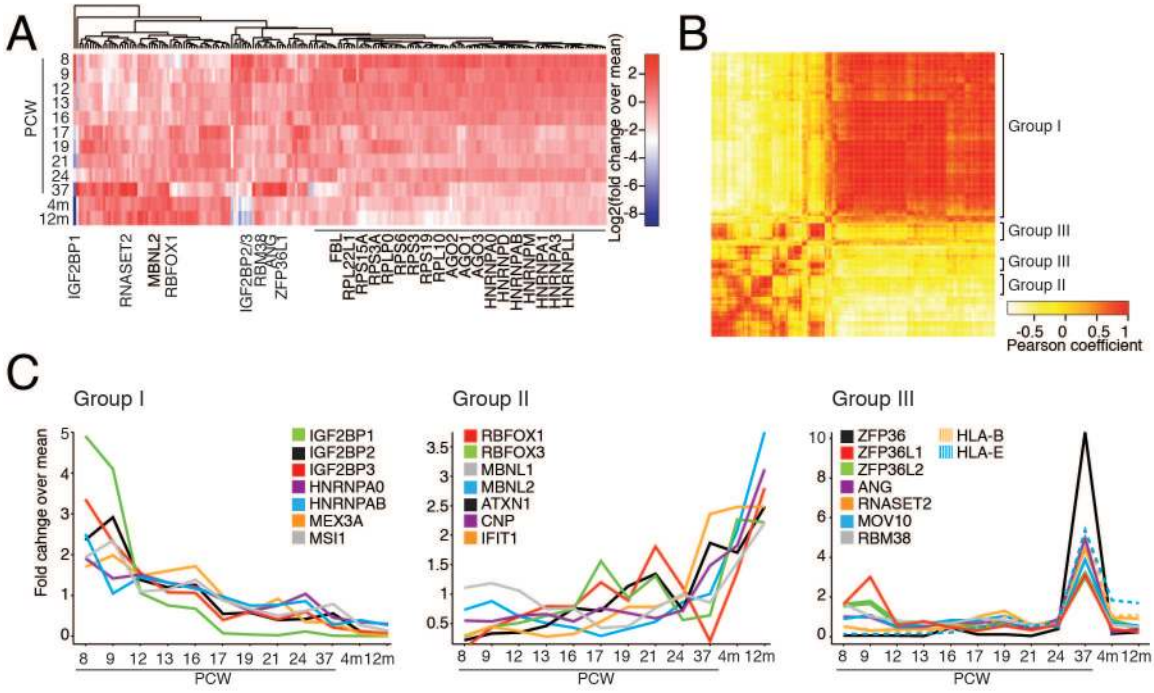


Figure 2.11 Expression of RBPs across fetal human hippocampus development. The top 200 most differentially expressed RBPs across 12 stages of human hippocampus development ranging from post-conception week (PCW) 9 up to 12 months after birth profiled by RNA-seq (Houmard et al., 2009). For each gene, rpkm values were normalized to relative fold changes by dividing the expression value by the mean expression value across developmental stages. (A) Heat map of the log2 transformed relative fold changes of the RBPs sorted by unsupervised clustering. (B) Pearson correlation map indicates correlated expression changes of the 200 selected RBPs. (C) Characteristic expression fold changes across developmental stages for genes in the three different groups. Group I: Genes with high expression levels at early PCW, which rapidly decrease at later stages. Group II: Genes with low expression at early PCW and rapidly increasing levels at late PCW and postnatal stages. Group III: Genes with a single, high expression peak at 37 PCW.

The largest cluster of differentially expressed RBPs in hippocampus contained ~100 RBPs, which were highly expressed during early development, but then were rapidly down-regulated at later stages (Figure 2.11, group I). This cluster included proteins required for the regulation of developmentally relevant pathways, such as the IGF2BP1-3 family (Bell et al., 2013; Yisraeli, 2005) and LIN28 (Thornton and Gregory, 2012), as well as general factors involved in translation, mRNA splicing, transport, and rRNA biogenesis.

Opposite in trend, a group of ~20 proteins displayed low expression in the first fetal post-conception weeks (PCW), which rapidly increased during the period coinciding with hippocampal development. This group comprised multiple splicing regulators required for neuronal function, such as the RBFOX family, which contributes to the characteristic splicing pattern of many neuronal transcripts (Gehman et al., 2011; Lovci et al., 2013) (Figure 2.11, group II). Another distinct group of ~20 RBPs, enriched in RNA nucleases and mRNA-regulatory proteins involved in inflammatory and innate immune responses, was highly expressed at 37 PCW (Figure 2.11, group III), coinciding with a maturation wave of pyramidal neurons and synaptogenesis at 34-36 PCW (Arnold and Trojanowski, 1996). Consistent with a neuronal function, mutations in the RNA nucleases ANG and RNASET2 are found in patients with the neurological diseases ALS and cystic leukoencephalopathy, respectively (Greenway et al., 2006; Henneke et al., 2009; Skorupa et al., 2012; Thiyagarajan et al., 2012). Strikingly, the mRNA-regulatory protein ZFP36 (tristetraprolin, TTP) was the most specifically upregulated RBP in the hippocampus by more than 200-fold. ZFP36 is known to destabilize mRNAs coding for cytokines and other inflammatory immune genes by recruiting the CCR4-NOT1 complex to AU-rich elements in the 3'UTR of targets, leading to deadenylation and subsequent degradation of mRNAs (Brooks and Blackshear, 2013; Fabian et al., 2011; 2009; Mukherjee et al., 2014; Reyes-Turcu and Grewal, 2012). Recently, it was found that cytokines and other immune regulatory proteins are expressed in the developing and adult nervous system, where they are required for

normal brain development and synaptic plasticity (Boulanger, 2009; Deverman and Patterson, 2009). Indeed, coinciding with ZFP36, immune regulatory genes such as the members of the major histocompatibility complex class I (e.g. HLA-A/B/C/E/F) were also selectively expressed at 37 PCW (Figure 2.11, group III) (Zhang et al., 2013). Whether or not the molecular function of ZFP36 remains the same in neurons is unknown, but the coordinated expression of these regulatory RBPs may imply a biological role during neural development and synaptic plasticity.

2.4.10 RBPs in human diseases

Disease phenotypes of RBPs may correlate with tissue-specific expression, e.g. loss-of-function of germline-specific proteins causes infertility (Reynolds and Cooke, 2005) and loss of the mRBP Fragile X mental retardation (FMR1) protein causes severe phenotypes in the tissues where it is most enriched (Wang et al., 2012). However, highly tissue-specific pathologies are often observed for loss-of-function of RBPs with no specificity in expression at all. These tissue-specific phenotypes may be explained by either (1) tissue-specific expression of critical RNA targets and cofactors of the RBP, or (2) a greater sensitivity to expression changes of PTGR components in general for the affected tissue.

Within the previously categorized RBP regulatory groups we found as general trend that RBPs, which interacted with the same RNA classes (e.g. mRNA-, tRNA-binding, etc.), displayed similar pathologies. Understanding diseases involving RBPs in the context of their RNA pathway was therefore the most important factor for predicting and interpreting their disease phenotypes. For example, ribosomopathies, such as Diamond-Blackfan anemia and Shwachman-Diamond syndrome, are caused by defects in ribosomal proteins and rRNA biogenesis factors, severely affecting the bone marrow and skin (Narla and Ebert, 2010), while mutations in mRBPs are found in multiple neurodegenerative and neuromuscular diseases, affecting mRNA metabolism in neurons and, in particular, motor neurons (Cooper et al., 2009; Lukong et al., 2008; Scheper et al.,

2007b). Here, mutations in mRBPs or their RNA targets cause impaired RNA transport and translation, often leading to protein/RNA aggregation and inefficient clearance of neuronal RNA/protein granules, which cause a range of neuropathological diseases (Buchan et al., 2013; Cooper et al., 2009; Kim et al., 2013b; Lagier-Tourenne et al., 2010; Liu-Yesucevitz et al., 2011; Lukong et al., 2008; Ramaswami et al., 2013; Scheper et al., 2007b). Examples include amyotrophic lateral sclerosis (ALS), caused by mutations in the mRBPs TDP43, FUS, and HNRNPA2B1/A1 (Kim et al., 2013b; Lagier-Tourenne et al., 2010), leading to prion-like accumulation of proteins in RNA granules in motor neurons, and spinocerebellar ataxia, caused by polyglutamine expansions in ATXN proteins leading to protein aggregations (Banfi et al., 1994; Orr et al., 1993). Defect in RNA transport and protein translation, such as loss-of-function of FMR1, and defects in mRNA splicing, caused by loss-of-function of e.g. the splicing RBFOX1, both manifest in neurological phenotypes as mental retardation and autism (Voineagu et al., 2011; Wang et al., 2012). mRNA repeat expansions typically lead to sequestration of RBPs, often splicing factors and have been linked to muscular diseases caused by dysregulated splicing, such as myotonic dystrophies, mental retardation, and ataxia (Echeverria and Cooper, 2012). Loss-of-function of the snRNP assembly factor SMN1 directly affects the spliceosome and causes spinal muscular atrophy (SMA), a motor neuron disease (Cooper et al., 2009). Loss-of-function mutations in tRNA splicing components and aminoacyl tRNA synthetases typically cause encephalopathies and the neurological Charcot-Marie tooth disease (Budde et al., 2008; Scheper et al., 2007b; Yao and Fox, 2013). A number of RNA/DNA nucleases critical for nucleic acid clearance have been implicated in the autoimmune disease Aicardi-Goutieres syndrome (Crow et al., 2006b; Rice et al., 2009).

In summary, our target categorization of RBPs allowed to detect and interpret patterns in RNA metabolic diseases. From this analysis it became clear that, instead of RBP expression, the interacting RNA targets are often a better predictor for the disease pathologies observed and it is

more useful to separate RBPs by their targets in order to interpret disease phenotypes and affected organs. Furthermore, for novel RBPs with unknown targets involved in human diseases, the disease phenotypes can point towards the likely dysregulated RNA pathways. In Table 2.1 I generated a human disease table of RBPs listed by OMIM and separated the RBPs by their dominant RNA targets. This analysis gives an overview of recurring patterns and phenotypes encountered in the different target groups. In the sections below, I briefly summarized the main characteristic phenotypes of diseases typically encountered in each RNA category.

2.4.10.1 Diseases of mRBPs

Most of the ~150 RBPs currently listed in the OMIM database (Hamosh et al., 2005) are mRBPs. Mutations in mRBPs typically display neurological and neuromuscular dysfunctions due to dysregulation of splicing, translation, localization or protein aggregation (Cooper et al., 2009; Hanson et al., 2011; Kapeli and Yeo, 2012; Lukong et al., 2008; Ule, 2008). Family members tend to have overlapping phenotypes, reflecting their functional redundancies. For instance, the paralogs RBM20 and MATR3 are both involved in myopathies due to dysregulated splicing of their targets (Guo et al., 2012; Senderek et al., 2009).

RNA gain-of-function diseases do not necessarily occur within an mRNA coding for an RBP, but they commonly lead to altered mRBP binding patterns, thereby they directly affect PTGR (Cooper et al., 2009; Echeverria and Cooper, 2012; Nelson et al., 2013). In these disorders, repeat expansions in introns or UTRs of mRNAs lead to sequestration of mRBPs in the nucleus, thereby causing dysregulation of their respective targets. The myotonic dystrophies DM1 and DM2 are caused by repeat expansions in the 3'UTR of DMPK and the intron of CNBP (ZNF9), which sequester the mRBPs CELF1 and MBNL1 and their paralogs (Echeverria and Cooper, 2012). FXTAS (Fragile X-associated tremor/ataxia syndrome), caused by trinucleotide repeat

expansions in the 5'UTR of the FMR1 mRNA, directly leads to loss-of-function of FMR1 (Hagerman, 2013).

Another common disease mechanism encountered for RBPs are mutations that lead to prion-like aggregation of cytoplasmic or shuttling mRBPs into RNA granules. Defective clearance and dysregulation in assembly and disassembly of cytoplasmic RNP granules have been found to be the underlying cause in a range of neurodegenerative disorders. For example, cytoplasmic inclusion of TARDBP/TDP43 and FUS have been found in amyotrophic lateral sclerosis, a motor neuron disease leading to muscle atrophy (Anthony and Gallo, 2010; Lagier-Tourenne et al., 2010; Lee et al., 2012; Ling et al., 2013). Accumulation and inefficient removal of these RNA-protein granules leads to cellular stress predominantly affecting neuronal cells (Buchan et al., 2013; Li et al., 2013; Liu-Yesucevitz et al., 2011; Ramaswami et al., 2013).

2.4.10.2 Mitochondrial RBPs in disease

Generally, mitochondrial RBPs, such as translation elongation factors GFM1 and TSFM, cause deficiencies in oxidative phosphorylation, that manifest themselves on a physiological level as neurological and muscular myopathies (Smeitink et al., 2006; Smits et al., 2010; Yao and Fox, 2013).

2.4.10.3 Diseases involving snRNA-binding proteins

Mutation or loss-of-function of snRNA-binding and assembly factors are known to lead to defects in assembly of spliceosomal U snRNPs and thus ultimately cause mRNA splicing defects. Overall, only a few snRNA-binding proteins have been linked to human disease so far. Loss-of-function of the snRNP assembly factor SMN causes spinal muscular atrophy (SMA). SMN proteins form a multimeric complex with Gemin proteins, which carries out assembly of snRNPs and other RNP complexes in the cytoplasm (Battle et al., 2006; Paushkin et al., 2002). Autosomal

recessive loss-of-function of the SMN1 gene is the molecular cause for SMA, affecting one in 6000 births (Gubitz et al., 2004). In the autosomal recessive disorder, the SMN1 locus is deleted and SMN2 becomes the main transcription locus. Due to a point mutation in SMN2 very low amounts of full-length SMN protein are produced, which leads to highly skewed ratios of snRNPs and results in global aberrant splicing patterns (Cooper et al., 2009; Zhang et al., 2008). While snRNP assembly defects are detected in all tissues upon SMN1 deletion, the physiological phenotype manifests itself mainly in motor neurons (Cooper et al., 2009; Glisovic et al., 2008; Liu-Yesucevitz et al., 2011), similar in phenotypes to defects in mRBP splicing factors.

Interestingly, loss-of-function of RBPs in the U2, U12 and U4/U6-U5 snRNP complexes specifically cause myelodysplastic syndromes (Lindsley and Ebert, 2013) and retinitis pigmentosa, a retinal degeneration leading to blindness caused by incorrect splicing of mRNAs encoding for photoreceptors (Daiger et al., 2013). Why mutations in components of the general splicing machinery display highly tissue-specific phenotypes in the eye remains unclear (Singh and Cooper, 2012; Wang and Cooper, 2007).

2.4.10.4 Diseases of tRNA-binding proteins

Disease-causing mutations in tRNA-binding proteins are found in the tRNA maturation and aminoacylation pathways and show predominantly neurological phenotypes (Scheper et al., 2007b) (Table 2.1). Mutations in a number of cytoplasmic tRNA synthetases cause Charcot-Marie-Tooth disease, affect the peripheral nervous system and lead to muscular atrophy (Antonellis and Green, 2008; Yao and Fox, 2013). Loss-of-function of components of the TSEN tRNA-splicing endonuclease lead to pontocerebellar hypoplasia, a sometimes fatal underdevelopment of the cerebellum that causes intellectual disability and impairs muscle control and motor skills (Budde et al., 2008). Mutations in cytoplasmic tRNA aminoacyl synthetases lead

to inefficient translation (Scheper et al., 2007b). The disease phenotypes for cytoplasmic tRNA aminoacyl synthetases overlap with mRBP diseases.

2.4.10.5 Diseases of rRNA-biogenesis and ribosomal proteins

Loss-of-function of rRNA biogenesis factors and ribosomal proteins are generally embryonically lethal and only few diseases, classified as ribosomopathies, are known for these RBPs (Narla and Ebert, 2010; Ruggero and Pandolfi, 2003). Ribosomopathies commonly show growth retardation, organ malformation and frequently bone marrow failure (Liu, 2006). Examples of ribosomopathies include mutations in the SBDS involved in rRNA biogenesis that causes Shwachman-Bodian syndrome, patients also show exocrine pancreatic dysfunction; mutation in SBDS leads to a deficit in neutrophils (Boocock et al., 2003). A number of mutations in ribosomal proteins cause Diamond-Blackfan anemia, a disease which impairs red blood cell formation (Narla and Ebert, 2010). A number of other ribosomopathies are caused by mutations in RNA Pol I components or rRNA-specific transcription factors, such as in Treacher-Collins syndrome (Dauwerse et al., 2011; Edwards et al., 1997) [reviewed in (McCann and Baserga, 2013)].

2.4.10.6 Diseases of snoRNA-binding proteins

snoRNA-binding proteins are required for the maturation of rRNAs, snRNAs, and the H/ACA-snRNA-like telomerase RNA. snoRNPs introduce nucleotide modifications in their targets, which are essential for viability (Watkins and Bohnsack, 2012). Thus loss-of-function of snoRNPs leads to deficient nucleotide modifications in rRNAs, snRNAs and telomerase RNA (Bachellerie et al., 2002; Filipowicz and Pogacić, 2002). As a consequence, snoRNP disease phenotypes overlap with ribosomopathies, as well as genetic diseases of components involved in telomerase assembly, such as TERT and WRAP53 of the telomerase complex. Defects in snoRNA

biogenesis manifest themselves in the severe developmental disorder dyskeratosis congenita, lead to bone marrow failure, growth retardation, neurological defects and premature aging (Filipowicz and Pogacić, 2002; Ruggero et al., 2003; Smogorzewska and de Lange, 2004).

2.4.10.7 Diseases of microRNA pathway components

Mutations in miRNA-binding proteins are found in different cancers and developmental disorders (Kaneko et al., 2011; Merritt et al., 2008; Perron and Provost, 2009). Mutations, loss-of-function, or reduced levels of DICER1, TARBP2, and XPO5 have been found in pleuropulmonary blastomas, ovarian and other cancers (Hill et al., 2009; Melo et al., 2010; 2009; Zhang et al., 2006).

2.4.10.8 Autoimmune diseases caused by RBPs

In recent years it has become evident that nucleic acids play a central role in autoimmune and cellular-stress-related diseases. A number of nucleases and RBPs specific for DNA/RNA hybrids, or which have overlapping DNA/RNA specificity, display autoimmune disease phenotypes. Mutations or loss-of-function in three RNA/DNA nucleases, SAMHD1, the RNase H2 complex, and TREX1, lead to development of the autoimmune disease Aicardi-Goutieres syndrome (AGS), a neurodevelopmental disorder causing white matter abnormalities and cerebral atrophy. Symptomatically, AGS overlaps with the autoimmune disorder systemic lupus erythematosus (SLE) (Crow et al., 2006b). In both cases, a failure to remove accumulating nucleic acids is central to disease development and activates the innate immune system by type I interferon signaling (Atianand and Fitzgerald, 2013; Rabe, 2013). The triphosphatase SAMHD1 possesses 3'-to-5' exonuclease activity for ssRNA, ssDNA, and DNA/RNA hybrids. Its antiviral and autoimmune-suppressive function has been attributed to its role in the removal of nucleotides and nucleic acids in the cell (Beloglazova et al., 2013). The heterotrimeric RNase H2 complex

endonucleolytically cleaves DNA/RNA hybrids and is thought to be required for the removal of Okazaki fragments during DNA replication (Cerritelli and Crouch, 2009; Rabe, 2013). Mutations in all three subunits of the RNase H2 complex (RNASEH2A, RNASEH2B, RNASEH2C) have been found to cause AGS (Crow et al., 2006b; Rabe, 2013). The 3'-to-5' ssDNA and ssRNA exonuclease TREX1 is involved in the degradation of ssDNA fragments during replication and antiviral defense (Rabe, 2013; Yuan et al., 2015). Mutations in TREX1 are found in AGS patients (Crow et al., 2006a) and TREX1 knockout mice accumulate endogenous retroelements. The accumulating nucleic acids are thought to trigger a subsequent interferon response (Stetson et al., 2008). Loss-of-function of the dsRNA-editing ADAR enzyme has also been shown to cause AGS by an as yet unknown mechanism (Rice et al., 2012).

Notably, while not directly related to loss-of-function mutations in RBPs, many autoantibodies against other RBPs and even RNA have been detected in autoimmune diseases, pointing to a central importance of a dysregulation of nucleic acid/RNA metabolism in the mechanism of autoimmune diseases (DeHoratius et al., 1975; Gelpi et al., 1992; Gold et al., 1988; Hendrick et al., 1981; Pettersson et al., 1984). It is thought that dysregulation of RNA clearance mechanisms triggers innate immune responses and leads to apoptosis and release of RBP-RNA complexes into circulation. There these granules mobilize the immune system to develop autoantibodies against self-RNA-protein complexes (Gaipl et al., 2005; Muñoz et al., 2010). The Ro60 complex, consisting of the TROVE2 (Ro60) protein and Y RNAs, was among the first identified targets of autoimmune antibodies in SLE patients was the Ro-RNP particle (Hendrick et al., 1981; Lerner et al., 1981). The Ro60 complex plays a regulatory role in DNA replication and stress response, removing misfolded RNAs, and mice lacking Ro60 develop lupus-like syndromes (Chen and Wolin, 2004; Hall et al., 2013; Sim and Wolin, 2011). Cleavage of tRNAs and Y RNAs accompanies cellular stress response and apoptosis (Hall et al., 2013; Köhn et al., 2013; Nawrot et al., 2011; Phizicky and Hopper, 2010) and these stress-induced small RNA

fragments may also act as immune-stimulatory RNAs. Autoantibodies against RBPs associating with these RNAs have been found in serum of SLE patients, stressing the importance of efficient clearance of circulating RNP granules in autoimmune diseases.

Dysfunctional nucleic acid clearance has not only been associated with autoimmune diseases, but also with neurological defects of the peripheral nervous system, as seen in loss-of-function of the RNA exosome component EXOSC3 and RNASET2 (Henneke et al., 2009; Monti et al., 2008; Wan et al., 2012). Loss of these general RNA turnover factors more closely resembles loss of mRBPs or tRNA-binding proteins.

Table 2.1 Overview of RBPs involved in human genetic diseases. Overview of RBPs with identified genetic disease-causing mutations collected in the OMIM database (Hamosh et al., 2005), categorized into their main RNA target groups. Mitochondrially localized proteins are indicated with (mt). Proteins within the same RBP family are written in one line, family members also involved in the same disease are highlighted in bold, family members involved in other diseases are highlighted in brown and listed elsewhere again in a separate category.

RBP class	Disease category	RBP family	Disease	Genetic mutation	Reference
mRNA-binding	cancer	**EWSR1**, **FUS**, **TAF15**	Ewing sarcoma, soft tissue tumors	gene fusion	(Gill et al., 1995; Ichikawa et al., 1994; May et al., 1993; Panagopoulos et al., 1994; 1999)
		TPR	gastric, thyroid carcinoma, sarcoma	gene fusion	(Dean et al., 1987; Gonzatti-Haces et al., 1988)
	muscular/ cardiac disease	**CNBP**, ZCCHC13	myotonic dystrophy	RNA repeat expansion sequesters RBPs	(Liquori et al., 2001)
		MBNL1, **MBNL2**, **MBNL3**	myotonic dystrophy	sequestered RBP in repeat expansion	(Fardaei, 2002; Mankodi et al., 2001; Miller, 2000)
		CELF1, CELF2-6	myotonic dystrophy	sequestered RBP in repeat expansion	(Roberts et al., 1997; Timchenko et al., 1996)
		MATR3, RBM20	cardio-/distal myopathy	missense mutation	(Brauch et al., 2009; Senderek et al., 2009)
		PABPN1, PABN1L	muscular dystrophy	polyalanine expansion leading to protein aggregation	(Brais et al., 1998)
	neurological disease	AFF1, **AFF2**, AFF3, AFF4	mental retardation	deletion, loss-of-function through repeat expansion in mRNA	(Knight et al., 1993; Stettner et al., 2011)
		ATXN1, ATXN1L	spinocerebellar ataxia	polyglutamine expansion leading to protein aggregation	(Banfi et al., 1994; Orr et al., 1993; Servadio et al., 1995)
		ATXN2, ATXN2L	spinocerebellar ataxia, susceptibility to late-onset Parkinson disease, susceptibility to amyotrophic lateral sclerosis (ALS)	polyglutamine expansion leading to protein aggregation	(Cancel et al., 1997; Elden et al., 2010; Gwinn-Hardy et al., 2000; Pulst et al., 1996)
		DYNC1H1,DNAH1-11, DNAH17, DYNC2H1	Charcot-Marie-Tooth disease, mental retardation, spinal muscular atrophy (SMA)	missense mutation	(Harms et al., 2012; Vissers et al., 2010; Weedon et al., 2011)
		EIF2B1	leukoencephalopathy with vanishing white matter	missense mutation	(van der Knaap et al., 2002)
		EIF2B2	leukoencephalopathy with vanishing white matter	missense, nonsense mutation	(Leegwater et al., 2001)
		EIF2B3	leukoencephalopathy with vanishing white matter	missense mutation	(van der Knaap et al., 2002)
		EIF2B4	leukoencephalopathy with vanishing white matter	missense mutation	(van der Knaap et al., 2002)
		EIF2B5	leukoencephalopathy with vanishing white matter	missense mutation	(Fogli et al., 2002; Leegwater et al., 2001; van der Knaap et al., 2002)
		EIF4G1, EIF4G2, EIF4G3	Parkinson disease	missense mutation	(Chartier-Harlin et al., 2011)
		FMR1, FXR1, FXR2	fragile X mental retardation syndrome (FXS), fragile X tremor/ataxia syndrome (FXTAS), premature ovarian failure	deletion, repeat expansion leading to protein loss-of-function (FXS) or RNA-gain-of-function (FXTAS)	(Devys et al., 1992; Gedeon et al., 1992; Hagerman et al., 2001; Kremer et al., 1991; Murray et al., 1998; Wöhrle et al., 1992)

				sequesters RBPs	
		EWSR1, **FUS**, **TAF15**	amyotrophic lateral sclerosis (ALS)	missense mutation leading to prion-like protein aggregation	(Kwiatkowski et al., 2009; Vance et al., 2009)
		DAZAP1, **HNRNPA2B1**, HNRNPA0, HNRNPAB, HNRNPA1L2, **HNRNPA1**, HNRNPA3, HNRNPD, HNRNPDL	amyotrophic lateral sclerosis (ALS)	missense mutation leading to prion-like protein aggregation	(Kim et al., 2013b)
		TARDBP	amyotrophic lateral sclerosis (ALS)	missense mutation leading to protein aggregation	(Sreedharan et al., 2008)
		IGHMBP2	distal spinal muscular atrophy (DSMA1)	missense mutation	(Grohmann et al., 2001)
		LRPPRC	Leigh syndrome	missense mutation	(Mootha et al., 2003)
		MECP2	Rett syndrome, X-linked mental retardation	missense, nonsense mutation, frameshift, deletion	(Amir et al., 1999; Cheadle et al., 2000; Huppke et al., 2000; Wan et al., 1999)
		MTPAP, PAPD4, TUT1, ZCCHC6, ZCCHC11	spastic ataxia	missense mutation	(Crosby et al., 2010)
		PARK7	Parkinson disease	missense mutation	(Bonifati et al., 2003)
		PQBP1	Renpenning syndrome 1	frameshift	(Kalscheuer et al., 2003)
		PRKRA, **TARBP2**	dystonia	frameshift, missense mutation	(Camargos et al., 2008; Seibler et al., 2008)
		RANBP2, RGPD1-6, RGPD8	acute, infection-induced susceptibility to encephalopathy	missense mutation	(Neilson et al., 2009)
		NOVA1, NOVA2	paraneoplastic opsoclonus-myoclonus ataxia (POMA)	autoantibodies	(Buckanovich et al., 1996)
		ELAVL1, ELAVL2, **ELAVL3, ELAVL4**	paraneoplastic neurological disorders, encephalomyelitis, neuropathy	autoantibodies	(Sakai et al., 1994)
		UPF3A, **UPF3B**	mental retardation	frameshift, missense, nonsense mutation	(Tarpey et al., 2007)
		TIA1, TIAL1	Welander distal myopathy	missense mutation	(Hackman et al., 2012)
		RBFOX1, RBFOX2, RBFOX3	mental retardation, epilepsy	deletion, breakpoint	(Bhalla et al., 2004; Martin et al., 2007)
	neurological/ developmental disease	**GLE1**	lethal congenital contracture syndrome	splice site mutation, missense mutation	(Nousiainen et al., 2008)
	developmental disease	**BICC1**	susceptibility to renal dysplasia	missense, nonsense mutation	(Kraus et al., 2012)
		EEF2, **EFTUD2**	mandibulofacial dysostosis with microcephaly	splice site mutation, nonsense, missense mutation, frameshift	(Bernier et al., 2012; Gordon et al., 2012; Lines et al., 2012)
		EIF2AK1, EIF2AK2, **EIF2AK3**	Wolcott-Rallison syndrome, multiple epiphyseal dysplasia	missense, nonsense mutation, splice site mutation	(Brickwood et al., 2003; Delépine et al., 2000; Durocher et al., 2006)
		FTO	growth retardation, developmental delay	missense mutation	(Boissel et al., 2009)
		NR0B1, NR0B2	congenital adrenal hypoplasia	deletion, missense mutation	(Muscatelli et al., 1994; Yanase et al., 1996)
		RBM5, RBM6, **RBM10**	TARP syndrome	frameshift, missense mutation	(Johnston et al., 2010)
		RBM11, **SF3B4**	acrofacial dysostosis	missense, nonsense mutation, frameshift	(Bernier et al., 2012; Czeschik et al., 2013)

		SKIV2L	trichohepatonenteric syndrome 2	missense, nonsense mutation	(Fabre et al., 2012)
	infertility	**BOLL**, **DAZ1-4**, **DAZL**	azoospermia	deletion	(Reijo et al., 1995)
	metabolic disease	**AUH**, ECH1, ECHS1, ECHDC2, ECHDC3	3-methylglutaconic aciduria	nonsense mutation, frameshift, splice site mutation	(IJlst et al., 2002; Ly et al., 2003)
		C12ORF65 (mt)	combined oxidative phosphorylation deficiency, spastic paraplegia-55 (SPG55)	frameshift, missense, nonsense mutation	(Antonicka et al., 2010; Shimazaki et al., 2012)
		GFM1 (mt)	combined oxidative phosphorylation deficiency	missense, nonsense mutation	(Coenen et al., 2004; Valente et al., 2007)
		TSFM (mt)	combined oxidative phosphorylation deficiency	missense mutation	(Smeitink et al., 2006)
		EEFSEC, **TUFM** (mt)	combined oxidative phosphorylation deficiency	missense mutation	(Valente et al., 2007)
		SECISBP2, SECISBP2L	abnormal thyroid metabolism	missense mutation	(Dumitrescu et al., 2005)
	hematologic disease	**FIP1L1**	spontaneous hypereosinophilic syndrome	deletion leading to gene fusion	(Cools et al., 2003; Griffin et al., 2003)
		U2AF1, U2AF1L4 ZRSR1, ZRSR2	myelodysplastic syndrome	missense mutation	(Graubert et al., 2012)
	Immunological/ skin disease	ADAD1, ADAD2, ADAT, **ADAR**, ADARB1, ADARB2	Aicardi-Goutieres syndrome (AGS), dyschromatosis symmetrica hereditaria 1 (DSH1)	missense, nonsense mutation	(Miyamura et al., 2003; Rice et al., 2012)
tRNA-binding	cancer/ metabolic disease	ELAC1, **ELAC2**	prostate cancer, combined oxidative phosphorylation deficiency	missense, nonsense mutation, frameshift	(Haack et al., 2013; Tavtigian et al., 2001)
	muscular/ metabolic/ hematologic disease	**PUS1**	myopathy, lactic acidosis and sideroblastic anemia 1	missense, nonsense mutation	(Bykhovskaya et al., 2004; Fernandez-Vizarra et al., 2007)
		YARS2 (mt)	myopathy, lactic acidosis and sideroblastic anemia 2	missense mutation	(Riley et al., 2010)
	neurological disease	**AARS, AARS2** (mt)	Charcot-Marie Tooth disease	missense mutation	(Latour et al., 2010; Lin et al., 2011)
		AIMP1, YARS	hypomyelinating leukodystrophy, Charcot-Marie Tooth disease	frameshift, missense mutation, deletion	(Feinstein et al., 2010; Jordanova et al., 2006)
		CLP1	Pontocerebellar hypoplasia	missense mutation	(Karaca et al., 2014)
		KARS	Charcot-Marie Tooth disease, deafness	missense mutation, frameshift	(McLaughlin et al., 2010; Santos-Cortez et al., 2013)
		GARS	Charcot-Marie Tooth disease	missense mutation	(Antonellis et al., 2003)
		ANG, RNASE1-4, RNASE6-8	amyotrophic lateral sclerosis (ALS)	missense mutation	(Greenway et al., 2006)
		DARS2 (mt)	leukoencephalopathy	frameshift, missense, nonsense mutation, splice site mutation	(Scheper et al., 2007a)
		NSUN2	mental retardation	nonsense, missense mutation, splice site mutation	(Abbasi-Moheb et al., 2012; Khan et al., 2012)
		FTSJ1	mental retardation	frameshift, splice site	(Freude et al., 2004;

				mutation	Ramser et al., 2004)
		RARS, **RARS2** (mt)	pontocerebellar hypoplasia	splice site mutation, missense mutation	(Edvardson et al., 2007; Rankin et al., 2010)
		SEPSECS	pontocerebellar hypoplasia	missense mutation	(Agamy et al., 2010)
		TSEN2	pontocerebellar hypoplasia	missense mutation	(Budde et al., 2008)
		TSEN34	pontocerebellar hypoplasia	missense mutation	(Budde et al., 2008)
		TSEN54	pontocerebellar hypoplasia	missense, nonsense mutation, deletion	(Budde et al., 2008; Cassandrini et al., 2010)
	metabolic disease	**EEF2**, **EFTUD2**	spinocerebellar ataxia	missense mutation	(Hekman et al., 2012)
		AARS, **AARS2** (mt)	combined oxidative phosphorylation deficiency	missense mutation	(Götz et al., 2011)
		MTO1 (mt)	combined oxidative phosphorylation deficiency	frameshift, missense mutation	(Ghezzi et al., 2012)
		EARS2 (mt)	combined oxidative phosphorylation deficiency	missense mutation, insertion	(Steenweg et al., 2012; Talim et al., 2013)
		FARS2 (mt)	combined oxidative phosphorylation deficiency	missense mutation	(Elo et al., 2012; Shamseldin et al., 2012)
		SARS2 (mt)	hyperuricemia, pulmonary hypertension, renal failure, and alkalosis	missense mutation	(Belostotsky et al., 2011)
		TRMU (mt)	liver failure, deafness	missense mutation	(Guan et al., 2006; Zeharia et al., 2009)
	opthalmologic disease/ hearing loss	**HARS**, **HARS2** (mt)	Usher syndrome, Perrault syndrome	missense mutation	(Pierce et al., 2011; Puffenberger et al., 2012)
rRNA-binding	developmental disease	**EMG1**	Bowen-Conradi syndrome	missense mutation	(Armistead et al., 2009)
		MURC, **PTRF**, PRKCDBP, SDPR	lipodystrophy, muscular dystrophy	frameshift	(Hayashi et al., 2009; Shastry et al., 2010)
	developmental / hematologic disease	**SBDS**	Shwachman-Diamond syndrome	frameshift, missense, nonsense mutation	(Boocock et al., 2003; Nakashima et al., 2004)
	opthalmologic disease	**WDR36**	open angle glaucoma	missense mutation	(Monemi et al., 2005)
snRNA-binding	neurological disease	**SMN1, SMN2**	spinal muscular atrophy (SMA)	missense, nonsense mutation, frameshift, deletion	(Cobben et al., 1995; Gambardella et al., 1998; Hahnen et al., 1997; Lefebvre et al., 1995; Parsons et al., 1996; Sossi et al., 2001)
		RBM28	alopecia, neurologic defects, endocrinopathy syndrome	missense mutation	(Nousbeck et al., 2008)
	skin disease	**SART3**	porokeratosis	missense mutation	(Zhang et al., 2005)
	opthalmologic disease	**SNRNP200**, ASCC3	retinitis pigmentosa	missense mutation	(Zhao et al., 2009)
		PRPF3	retinitis pigmentosa	missense mutation	(Chakarova et al., 2002)
		PRPF31	retinitis pigmentosa	splice site mutation/deletion, missense mutation	(Vithana et al., 2001)
		PRPF6	retinitis pigmentosa	missense mutation	(Tanackovic et al., 2011)
		PRPF8	retinitis pigmentosa	missense mutation	(McKie et al., 2001)
		RP9	retinitis pigmentosa	missense mutation	(Keen et al., 2002)

snoRNA-binding	neurological disease	**NOP56**	spinocerebellar ataxia	RNA repeat expansion sequesters RBPs	(Kobayashi et al., 2011)
	hematologic/ neurodevelopmental/ developmental disorder	**DKC1**	dyskeratosis congenita	missense mutation, deletion, intron insertion, splice site mutation	(Heiss et al., 1998; Kanegane et al., 2005; Knight et al., 1999; 2001; Pearson et al., 2008; Vulliamy et al., 1999)
		NHP2	dyskeratosis congenita	missense mutation	(Vulliamy et al., 2008)
		NOP10	dyskeratosis congenita	missense mutation	(Walne et al., 2007)
	skin disease	**USB1**	poikiloderma with neutropenia	deletion, splice site mutation, frameshift	(Tanaka et al., 2010; Volpi et al., 2010)
Cyto-solic ribo-somal proteins	neurological disease	**RPL10**, RPL10L	autism	missense mutation	(Klauck et al., 2006)
	hematologic disease	**RPL11**	Diamond-Blackfan anemia	frameshift, deletion, splice site mutation, nonsense mutation	(Gazda et al., 2008)
		RPL35A	Diamond-Blackfan anemia	missense, nonsense mutation, deletion	(Farrar et al., 2008)
		RPL5	Diamond-Blackfan anemia	missense, nonsense mutation, frameshift, splice site mutation	(Gazda et al., 2008)
		RPS10	Diamond-Blackfan anemia	missense, nonsense mutation, frameshift	(Doherty et al., 2010)
		RPS17, **RPS17L**	Diamond-Blackfan anemia	missense, frameshift	(Cmejla et al., 2007; Gazda et al., 2008)
		RPS19	Diamond-Blackfan anemia	missense, nonsense mutation, frameshift	(Draptchinskaia et al., 1999; Matsson et al., 1999)
		RPS24	Diamond-Blackfan anemia	nonsense mutation, frameshift	(Gazda et al., 2006)
		RPS26	Diamond-Blackfan anemia	missense mutation, splice site mutation, frameshift	(Doherty et al., 2010)
		RPS7	Diamond-Blackfan anemia	splice site mutation	(Gazda et al., 2008)
Mito-chondr. (mt) ribo-somal proteins	metabolic disease	**MRPL3**	combined oxidative phosphorylation deficiency	missense mutation	(Galmiche et al., 2011)
		MRPS16	combined oxidative phosphorylation deficiency	nonsense mutation	(Miller et al., 2004)
		MRPS22	combined oxidative phosphorylation deficiency	missense mutation	(Saada et al., 2007)
lncRNA-binding	cancer	**BRCA1**	breast, ovarian, pancreatic cancer	missense, nonsense mutation, deletion, frameshift	(Al-Sukhni et al., 2008; Castilla et al., 1994; Simard et al., 1994)
	developmental disorder	EZH1, **EZH2**	Weaver syndrome 2	missense mutation, frameshift	(Gibson et al., 2012)
miRNA-binding	cancer	**DICER1**	pleuropulmonary blastoma, goiter with testicular tumors, embryonal habdomyosarcoma	missense, nonsense mutation, frameshift	(Foulkes et al., 2011; Hill et al., 2009; Rio Frio et al., 2011)
		XPO5	colorectal cancer	frameshift, insertion	(Melo et al., 2010)
	cancer/ developmental disorder	**SMAD1**, SMAD2, **SMAD3**, **SMAD4**, SMAD5, **SMAD6**, **SMAD7**, SMAD9	colorectal cancer, Loeys-Dietz syndrome, juvenile polyposis, pancreatic cancer	missense, nonsense mutation, frameshift	(Broderick et al., 2007; Howe et al., 1998; Regalado et al., 2011; Schutte et al., 1996; van de Laar et al., 2011)
		PRKRA,	colorectal cancer	frameshift	(Melo et al., 2010)

		TARPBP2			
	developmental disorder	DIS3, DIS3L, **DIS3L2**	Perlman syndrome	deletion, splice site mutation, missense mutation	(Astuti et al., 2012)
	neurological disease	**SNIP1**	psychomotor retardation, epilepsy, craniofacial dymorphism	missense mutation	(Puffenberger et al., 2012)
	pulmonary disease	**SMAD1,** SMAD2, **SMAD3**, **SMAD4**, SMAD5, **SMAD6**, **SMAD7**, **SMAD9**	pulmonary hypertension, aortic valve disease (AOVD2)	missense, nonsense mutation	(Drake et al., 2011; Nasim et al., 2011; Tan et al., 2012)
telRNA-binding	developmental / cardio-vascular/ pulmonary disease	**TERT**	coronary artery disease, dyskeratosis congenita, aplastic anemia	missense mutation, frameshift	(Armanios et al., 2007; Marrone et al., 2007; Tsakiri et al., 2007; Yamaguchi et al., 2005)
	developmental disorder	**WRAP53**	dyskeratosis congenita	missense mutation	(Zhong et al., 2011)
7SL-RNA-binding	hematologic disease	**SRP72**	bone marrow failure	missense mutation	(Kirwan et al., 2012)
immune-stimul-atory RNA-binding	metabolic disease	**OAS1**, OAS2, OAS3, OASL	susceptibility to diabetes mellitus	missense mutation	(Tessier et al., 2006)
	cancer	**RNASEL**	prostate cancer	nonsense, missense mutation	(Carpten et al., 2002; Casey et al., 2002)
RNA/ DNA-hybrid-binding	autoimmune/ neurological disease	**RNASEH2A**	Aicardi-Goutieres syndrome (AGS)	missense mutation	(Crow et al., 2006b; Rice et al., 2013)
		RNASEH2B	Aicardi-Goutieres syndrome (AGS)	missense mutation	(Crow et al., 2006b)
		RNASEH2C	Aicardi-Goutieres syndrome (AGS)	missense mutation	(Crow et al., 2006b)
		SAMHD1	Aicardi-Goutieres syndrome (AGS), Chilbain lupus 2	missense mutation	(Ravenscroft et al., 2011; Rice et al., 2009)
diverse targets	neurological disease	**EXOSC3**	pontocerebellar hypoplasia, spinal motor neuron degeneration	missense mutation, deletion	(Wan et al., 2012)
		RNASET2	leukoencephalopathy, cancer	missense mutation, deletion, splice site mutation	(Henneke et al., 2009)
unknown targets	cardiac disease	**CALR3**, CALR, CANX, CLGN	cardiomyopathy	missense mutation	(Chiu et al., 2007)
	connective tissue/ skin/ muscular disease	**PLEC**, DSP, EPPK1	epidermolysis bullosa simplex with muscular dystrophy	insertion, deletion, missense, nonsense mutation	(Koss-Harnes et al., 2002; McLean et al., 1996; Pulkkinen et al., 1996; Selcen et al., 2011; Smith et al., 1996)
	developmental disorder	**ASCC1**	barrett esophagus	missense mutation	(Orloff et al., 2011)
	neurological disease	**APTX**, PNKP	ataxia	deletion, splice site mutation, missense mutation	(Amouri et al., 2004; Criscuolo et al., 2005)
	pulmonary disease	**DNAAF2**	ciliary dyskinesia	insertion, nonsense mutation	(Omran et al., 2008)
RBP-interact-ing proteins	ophthalmol. disease	**TDRD7**	cataract	frameshift	(Lachke et al., 2011)

2.4.11 FAM98A is a novel RG/RGG-rich RBP

Several decades ago, arginine/glycine-rich structural repeats were identified to be often present in heteronuclear RBPs (hnRNPs) [reviewed in (Thandapani et al., 2013)]. Consecutive repeats of RG/RGG amino acid motif, separated by few amino acids in between, created a functional unit defined as an RG/RGG box and were enriched in RBPs and some extracellular proteins such as collagens. These RG/RGG boxes were subsequently shown to influence RBP transport through posttranslational modifications of the RG residues and also to directly bind to RNA (Kiledjian and Dreyfuss, 1992; Shen et al., 1998). The prion-like, undefined structural properties of RG/RGG rich RBPs causes them to easily aggregate in RNA-protein granules, and a number of them have been implicated in diseases with prion-like aggregation into stress granules (Kim et al., 2013b; Li et al., 2013; Ramaswami et al., 2013). The location of RG/RGG boxes is often conserved within these RBPs, suggesting functional conservation (Thandapani et al., 2013). However, because of their loose structural definition, these low-complexity regions are not defined as structural domains by public protein domain databases such as Pfam, SMART, or InterPro, and currently missed by protein domain annotations (Apweiler et al., 2001; Finn et al., 2010; Letunic et al., 2009).

We noticed the density of RG/RGG boxes in a number of RNA transport proteins such as FUS, EWSR1, HNRNPU, and hypothesized that if RG-rich repeats were RNA-binding or characteristic of RBPs, their presence may be a sufficient predictor for RBPs. To identify *de novo* RBPs based on RG/RGG-rich regions, I conducted a genome-wide search of RG/RGG repeats in human proteins. I scanned all human protein isoforms for the presence of at least three RG/RGG motifs and examined the distribution of RG/RGG motifs across known RBPs, comparing them to the residual proteome (Figure 2.12 A). On average distances of <20 amino acids were found between RG/RGG motifs in RBPs (Figure 2.12 A). Based on the distribution of RG repeats in RBPs, I defined a conservative approach to identify and classify RG/RGG boxes in proteins, if a

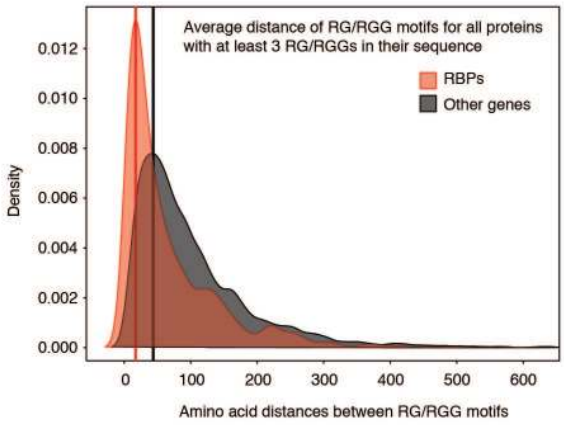
minimum of at least three RG/RGG amino acid motifs were found within a distance of ≤10 amino acids between each RG/RGG amino acid sequence. This gave a local density measure of RG/RGG boxes independent the total number of RG/RGG motifs and the length of the protein. From this definition I ranked the number of genes with one or more RG/RGG box. Further excluding collagen genes, 369 proteins had at least one or more RG/RGG repeat. The highest enriched Gene Ontology pathway was mRNA-binding and indeed 80 of those proteins were known characterized RBPs in our RBP census. In fact, RG repeats are a commonly encountered structural feature in RBPs and the second most abundant structural motif in our curated census, after the RRM domain (Figure 2.3 A). The top five RG/RGG rich proteins in humans were all RBPs: ZC3H4, TAF15, EWSR1, CHTOP, SYNCRIP with 21, 20, 20, 19 and 16 RG-boxes respectively. Of the 98 proteins with at least three RG/RGG boxes, 50 were known RBPs; the uncharacterized FAM98A and B protein were among the 48 putative RBP candidates.

The FAM98 protein family is absent in *S. cerevisae*, but highly conserved across higher eukaryotes, from *C. elegans*, fruitfly, to vertebrates. FAM98 proteins contain one conserved protein domain of unknown function (DUF2465) (Finn et al., 2010) (Figure 2.12 B). FAM98A has two paralogs in humans, FAM98B and C. FAM98C remains uncharacterized, but FAM98B was recently identified in a complex with the HSPC117 (RtcB) tRNA splicing ligase (Popow et al., 2011). FAM98A contains six RG/RGG boxes, FAM98B contains three, and FAM98C has none. While the number of RG/RGG boxes varies across organisms, the location of the RG/RGG repeats is generally conserved in FAM98A: one RG/RGG box lies within the conserved DUF2465 domain and two other boxes lie in the C-terminal region (Figure 2.12 C). To test whether the FAM98 family may be an RBP family, I chose FAM98A as candidate, since it had the highest count of RG/RGG boxes and their density most likely suggested a role in RNA-binding. To assess the closest structurally homologous relationship of the unknown domain of FAM98A, I ran a Phyre2 domain folding prediction of the conserved DUF2465 domain of

FAM98A, which most strongly resembled the kinetochore Hec1/human NDC80 protein (Kelley et al., 2015) (Figure 2.12 B). Hec1 is required for spindle checkpoint signaling, which assures correct chromosome segregation during cell division (Martin-Lluesma et al., 2002). A previous computational study also noted the structural similarities of FAM98 proteins with Hec1 kinetochore and other microtubule associated proteins (Schou et al., 2014).

Figure 2.12 FAM98A is a conserved RG-rich mRBP. (A) Density plot of the geometric mean distances of RG/RGG repeats of proteins with at least three RG/RGG repeats. RBPs are shown in red, other proteins in grey. The highest density of RG/RGG amino acid repeats is 17 for RBPs and 43 for non-RBPs. (B) Phyre2 predicted domain folding of the conserved DUF2465 domain of FAM98A shows similarities to the Hec1 kinetochore protein. (C) FAM98A is a conserved protein across vertebrates and invertebrates. Domain distribution of the DUF2465 (green) and RG repeats (dark blue) and conservation of FAM98A proteins are shown across human, mouse, *Xenopus* and *Drosophila*. (D) Schematic overview of PAR-CLIP: First photoactivatable thioribonucleosides are incorporated into nascent transcripts, RNA-protein complexes are crosslinked *in vivo* at UV 365 nm. After cell lysis and limited RNase T1 treatment, RNA–RBP complexes are immunoprecipitated. The crosslinked RNA segments are recovered, converted into cDNA libraries and deep sequenced. (E) PAR-CLIP 4-SU crosslink and Western blot of immunoprecipitated FLAG/HA-tagged FAM98A protein in HEK293 cells. (F) Length distribution of FAM98A PAR-CLIP reads mapping to mRNA genes. Reads mapping to the human mRNA reference annotation are split into T>C crosslinked reads (red), reads mapping with distance 0 to the reference annotation (white), reads mapping with distance 1, but which contain transitions other than T>C (grey), and reads mapping with distance 2 (black). The x-axis shows read length, the y-axis shows read number. (G) Pie chart of the distribution of PARalyzer clusters with ≥20 reads for all annotation categories. (H) Pie chart of the distribution of clusters with ≥20 reads in mRNA genes alone, given are also the number of total mRNA targets. (I) Immunostaining of FLAG/HA-FAM98A shows cytoplasmic localization in HEK293 cells.

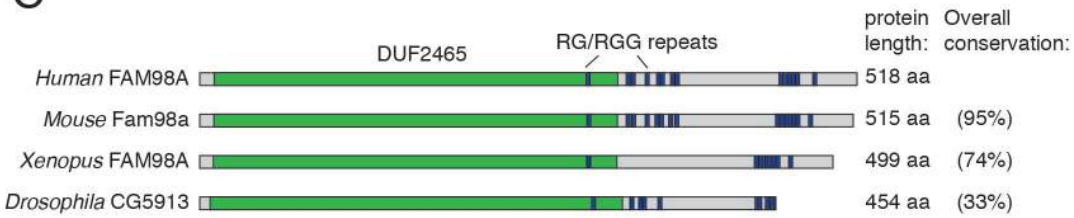
A



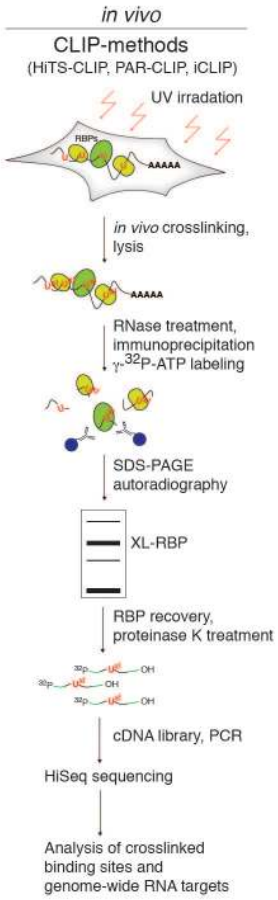
B

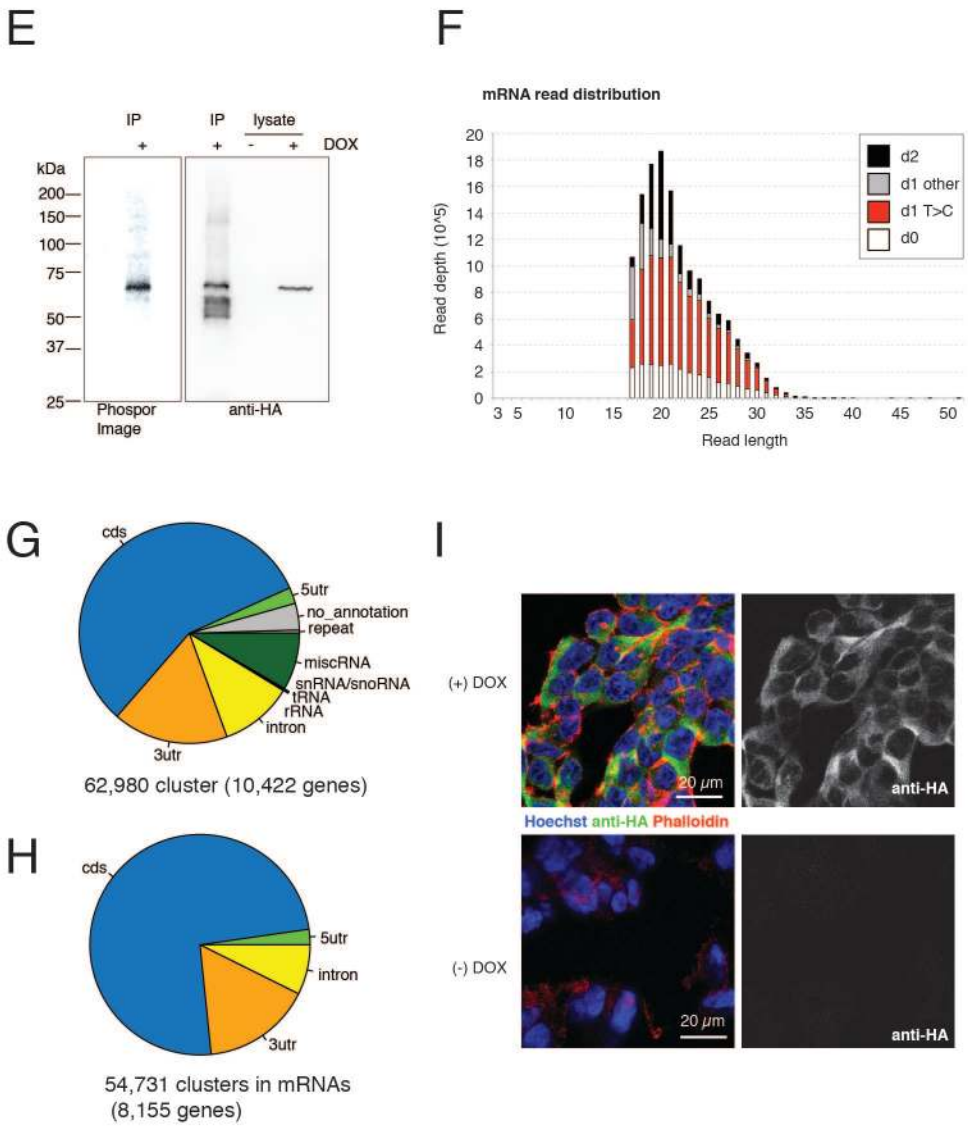
Phyre2 model

C



D




E
IP
IP
lysate
+
+
-
+
DOX
kDa
200
150
100
75
50
37
25
Phospor Image
anti-HA
F
mRNA read distribution
Read depth (10^5)
20
18
16
14
12
10
8
6
4
2
0
3 5 10 15 20 25 30 35 40 45 50
Read length
d2
d1 other
d1 T>C
d0
G
cds
5utr
no_annotation
repeat
miscRNA
snRNA/snoRNA
tRNA
rRNA
intron
3utr
62,980 cluster (10,422 genes)
H
cds
5utr
intron
3utr
54,731 clusters in mRNAs
(8,155 genes)
I
(+) DOX
20 μm
anti-HA
Hoechst anti-HA Phalloidin
(-) DOX
20 μm
anti-HA

To test whether FAM98A had direct RNA-binding activity, I generated stable HEK293 cell lines with doxycycline inducibly expressing FLAG/HA-FAM98A protein and performed PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation). In the PAR-CLIP protocol 4-thiouridine (4-SU) metabolically-labeled RNA is crosslinked at 365 nm to interacting protein components *in vivo* (Hafner et al., 2010a). The photochemical reactivity of the thio group at the 4-position of the uridine analogue shows increased efficiency of photo-induced crosslinking compared to the natural nucleoside (Sontheimer, 1994). Immunoprecipitation of the crosslinked RBP-RNA complex with an RBP-specific antibody, followed by cDNA library preparation of the RNA for HiSeq sequencing, allows us to identify transcriptome-wide targets of RBPs (Figure 2.12 D). During crosslinking of the photoreactive 4-thiouridine with aromatic amino acids the chemical structure of the uridine analogue changes such that the reverse transcription step of the library preparation results in a significant enrichment of T to C (T>C) transitions on the DNA level. Hence, PAR-CLIP sequences intrinsically contain the information of specific crosslinking events. This is important as during the RT-PCR step the reverse transcriptase is in many cases blocked as soon as protein-RNA crosslinks are encountered. As a consequence, about 70-90% of background RNA sequences co-purified during immunoprecipitation will be amplified more frequently and make peak detection of real crosslinking events of RNA to protein very difficult. Consequently, T>C transitions created by crosslinking of 4-SU to protein targets become an essential parameter for the identification of specific crosslink events and distinguish specific RNA-protein interactions from background RNA. For a successful experiment we expect the largest fraction of distance 1 mismatches against the genome to be T>C transitions.

Using PAR-CLIP I established that the previously unknown FAM98A protein was an RBP and showed that FAM98A crosslinked with high efficiency to the mRNA-transcriptome. The results are described in the next few sections.

2.4.11.1 FAM98A is a cytoplasmic mRBP

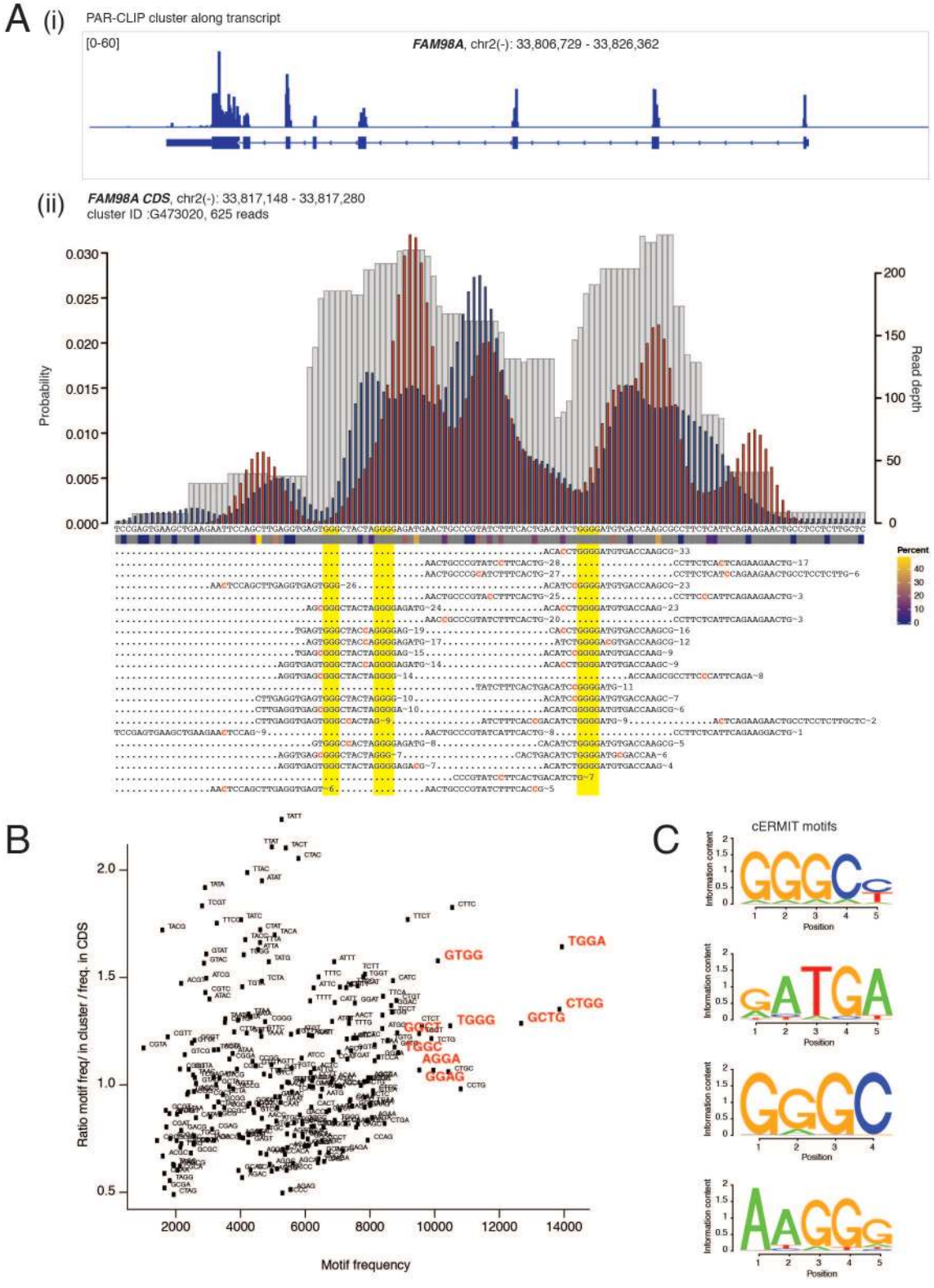
The FAM98A showed good crosslinking capacity as assessed by phosphorimaging and the resulting PAR-CLIP library had an unusually high T>C crosslinking efficiency, with thousands of targets, detecting significant enrichment of T>C transitions over background nucleotide transitions in sequence reads (Figure 2.12 E, F). To extract the binding regions, sequence reads were grouped by PARalyzer, which identifies regions with locally enriched T>C conversions in PAR-CLIP reads over background mismatch errors (Corcoran et al., 2011). Defining a cut-off of ≥20 reads per cluster, 62,980 clusters (or binding regions) were identified by PARalyzer, 54,731 (87%) of those were located in the coding regions, introns, 5'UTR or 3'UTRs of mRNAs (Figure 2.12 G,H). These corresponded to 10,422 total genes; 8,155 of these were protein-coding (Figure 2.12 H). By immunofluorescent analysis FAM98A was predominantly localized in the cytoplasm, which was in agreement with the PAR-CLIP results of FAM98A predominantly binding to mature mRNAs (Figure 2.12 I).

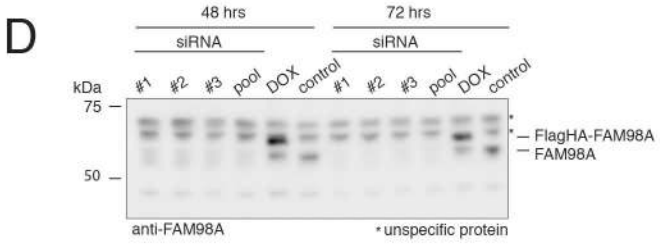
2.4.12 FAM98A binds to G-rich regions in mRNA targets

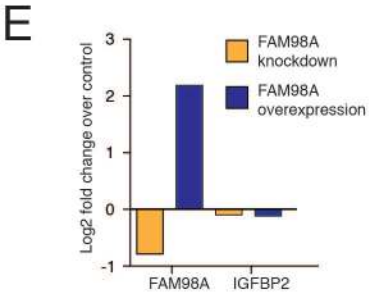
Further inspection of the extracted PAR-CLIP binding regions revealed that FAM98A bound broadly along the entire transcript, in coverage similar to RNA-seq data, and preferentially bound to G-rich regions (Figure 2.13 A). To quantify enrichment of sequence motifs in the binding sites, I used kmer cluster analysis to calculate the enrichment of 4mers along a 20-nt sliding window within clusters of coding regions, normalizing them over 4mer counts within a 20-nt sliding window of shuffled coding regions of the complete human transcriptome. By this analysis the most frequently enriched motifs contained variations of GG, GGG, GNGG, GNNG (Figure 2.13 B). I confirmed this G enrichment using the motif finding algorithm cERMIT, which ranks clusters based on T>C conversion specificity to give an evidence based estimate of preferred

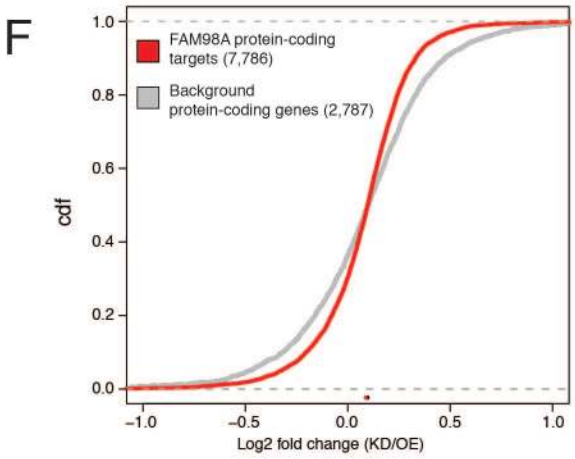
binding sites (Georgiev et al., 2010). The top cERMIT motifs of FAM98A clusters were G-rich motifs (Figure 2.13 C).

Figure 2.13 PAR-CLIP target clusters, motif enrichment, and RNAi knockdown of FAM98A. (A) (i) PAR-CLIP cluster along the FAM98A gene. (ii) Representative PAR-CLIP cluster of FAM98A within the FAM98A gene. Shown is the signal T base enrichment (red), background intrinsic sequence T enrichment (blue), and read coverage (grey). G-rich regions are highlighted in yellow in sequence alignments of reads. Number of reads is given behind each read, T>C transitions are highlighted in red. T>C frequency transitions are shown in a heat map from blue to yellow below the cluster. (B) Enrichment of 4mers in FAM98A PAR-Clip clusters normalized over shuffled coding regions. (C) cERMIT motif enrichments resulting from FAM98A PAR-CLIP clusters. (D) Western blot of knockdowns (three different siRNA duplexes and all three pooled together) and overexpression of FAM98A, compared to control parental HEK293 FlpIn T-Rex cells. (E) Log2 fold down- (orange) and up-regulation (blue) of FAM98A mRNA and control RBP gene (IGF2BP1) assessed by RNA-seq. (F) Cumulative distribution plot for mRNA abundance of FAM98A targets (red) and background mRNAs (black) in knockdown versus overexpression experiments.


A (i)
PAR-CLIP cluster along transcript
[0-60]
FAM98A, chr2(-): 33,806,729 - 33,826,362
(ii)
FAM98A CDS, chr2(-): 33,817,148 - 33,817,280
cluster ID :G473020, 625 reads
Probability
Read depth
Percent
B
Ratio motif freq/ in cluster / freq. in CDS
Motif frequency
GTGG
TGGA
CTGG
TGGG
GCTG
TGGC
AGGA
GGAG
C
cERMIT motifs
Information content
Position


D
48 hrs
72 hrs
siRNA
siRNA
kDa
#1
#2
#3
pool
DOX
control
#1
#2
#3
pool
DOX
control
75
50
FlagHA-FAM98A
FAM98A
anti-FAM98A
* unspecific protein


E
Log2 fold change over control
3
2
1
0
-1
FAM98A knockdown
FAM98A overexpression
FAM98A
IGFBP2


F
cdf
1.0
0.8
0.6
0.4
0.2
0.0
FAM98A protein-coding targets (7,786)
Background protein-coding genes (2,787)
−1.0
−0.5
0.0
0.5
1.0
Log2 fold change (KD/OE)

2.4.13 FAM98A knockdown does not affect target mRNA stability

To study putative regulatory effects of FAM98A on mRNA stability, I recorded mRNA expression profiles of siRNA-mediated knockdown of endogenous FAM98A or overexpression of FLAG/HA-FAM98A in HEK293 cells. Successful knockdown of all three siRNAs was assessed by Western blot analysis and the pooled knockdown submitted for RNA-sequencing. siRNA-mediated knockdowns reduced protein levels of FAM98A >10-fold, while doxycycline induced overexpression of FAM98A slightly reduced expression of endogenous FAM98A, but increased overall FAM98A levels approximately 2-fold (Figure 2.13 D). FAM98A mRNA levels were reduced 1.7-fold, while FAM98A overexpression resulted in >4-fold increase in mRNA levels. To assess the affect of FAM98A knockdown on mRNA stability of its targets, I analyzed the fold change differences in targets and nontargets mRNA levels in knockdown and overexpression experiments. The cumulative distribution of FAM98A PARalyzer-defined PAR-CLIP targets and non-targets showed no significant changes on mRNA stability of its targets (Figure 2.13 F). At an expression cut-off of rpkm $\geq$ 0.1 most expressed mRNAs (74%) were bound by FAM98A. Adding PARalyzer-defined PAR-CLIP binding sites with less than 20 reads resulted in an even larger fraction of the expressed transcriptome being bound to FAM98A. We concluded that FAM98A broadly bound most expressed mRNAs in HEK293 cells. Furthermore, at the protein level, targets displayed no measurable changes in abundance upon FAM98A knockdown or overexpression by Western blot analysis. Both assays, measuring mRNA and protein abundance, showed no indication that FAM98A had a role in regulating mRNA or protein stability, or translation. We reasoned that either FAM98A had an independent role in other RNA metabolic pathways (such as RNA transport) or that the family members of FAM98A could exert some redundancy upon RNAi knockdown of FAM98A and that knockdown of one member, but not the entire family was not sufficient to see regulatory effects. However, in our experience most studied RBPs with a role in mRNA or protein stability show measurable regulatory effects upon

knockdown or overexpression even if the expression of their paralogs are unaltered in the cell. Given the results we decided that the experimental conditions and design had to be significantly changed to elucidate the function of FAM98A *in vivo*. With a lacking phenotype upon loss-of-function and an unidentified process to pursue, I decided to not further investigate the effects of FAM98A expression levels on protein levels using methods such as quantitative mass spectrometry methods with stable isotope labeling (SILAC) (Mann, 2006). Our analysis of the PAR-CLIP experiment together with the RNA-seq data showed that generally abundant protein-coding mRNAs were bound by FAM98A and and that all unbound mRNAs were expressed at on average lower levels than bound targets. Thus, a regulatory role of FAM98A for specific subsets of mRNAs seemed unlikely and FAM98A may have a general role in RNA translation or transport.

2.4.14 Summary and discussion on the physiological role of FAM98A

I identified, using low-complexity domain definitions, the novel RBP FAM98A and confirmed its RNA-binding ability by PAR-CLIP. A number of proteins of unknown function with high numbers of RG/RGG boxes still remain uncharacterized and it will be interesting to characterize their role in RNA-binding in the future. Interestingly, also a number of chromatin regulatory proteins such as MBD2 (methyl-CpG binding protein 2), BRWD3 (bromodomain and WD repeat domain containing protein 3), KMT2B (Histone-lysine N-methyltransferase 2B) are also highly RG/RGG rich (12,11,11). For MBD2 (Tan and Nakielny, 2006) RNA-binding capability has been demonstrated, but physiological RNA targets have not been characterized and a putative role in RNA-binding of other chromatin regulatory proteins with RG-rich regions would be interesting to investigate further.

In conclusion, we found that FAM98A bound to most expressed mRNAs with some enrichment in G-rich regions. FAM98A is a cytoplasmic protein that binds broadly along the

entire mature mRNA transcript, predominantly in the coding region. Overexpression or knockdown of FAM98A did not alter mRNA or protein abundance of its targets. The predominant localization of PAR-CLIP clusters in coding regions could support a putative role in protein translation, where it perhaps is required to assist unwinding of extended secondary structures and thereby to facilitate protein translation. To measure translation changes, a more global assay such as ribosome profiling may present a better approach to investigate a putative role of FAM98A during translation (Ingolia, 2014).

Independent of my own studies, two proteome-wide pulldown studies of RBPs also isolated FAM98A (Baltz et al., 2012; Castello et al., 2012) and one study verified its binding activity *in vitro* (Strein et al., 2014). However, their findings disagree with our results in localization, which was reported to be predominantly nuclear, and RNA-binding capability, detecting low RNA affinity of FAM98A in GFP-multiTrap essays. In contrast, a proteomic study of proteins involved in 5'terminal oligopyrimidine (TOP) motif RNAs found FAM98A associated with the translation regulatory factor LARP and the mTOR complex and supports our finding that FAM98A is localized to the cytoplasm (Tcherkezian et al., 2014).

FAM98 is highly conserved in metazoan with 33% sequence conservation of the human FAM98A protein to the *Drosophila melanogaster* FAM98 homolog CG5913. I tested the viability of a loss-of-function genetic mutant of CG5913. CG5913 was not an essential gene and homozygous mutants did not show any physiological abnormalities (data not shown). Together, given our data, as well as the domain similarities to kinetochore proteins such as Hec1, and previous studies on RG-rich proteins, it seems likely that FAM98A is a nonessential, cytoplasmic (or nuclear/cytoplasmic) shuttling/transport RBP or could have a general function in protein translation. Recent findings for the paralog FAM98B support the idea of a shuttling protein and find that FAM98B is a component of a shuttling RNA-transporting complex with HSPC117/DDX1/C14orf166 (Pérez-González et al., 2014).

2.5 Chapter 2 Discussion

A census of human RBPs is critical to organize our current molecular and genetic understanding of PTGR. This catalog provides researchers with a newly curated resource to guide their investigations of PTGR processes and to systematically study RBPs. An analogous catalog that assesses the abundance and classifies all expressed RNAs, i.e. the RBP targets, across tissues and cell types is still missing and would represent a useful addition to this census. Among the 20,500 protein-coding genes in humans, our curated census estimates that 7.5% (1,542 genes) are directly involved in RNA metabolism by binding and/or processing RNA or comprising essential components of RNPs. RBPs are structurally diverse and include many distinct classes of RBDs. In contrast, the three most abundant DNA-binding domains account for 80% of the 1,704 TFs in humans (Vaquerizas et al., 2009). The three most abundant RBDs only accounted for 20% of all RBPs in our census.

Based on our target RNA categorization, nearly 50% of RBPs acted in mRNA metabolic pathways and 11% constituted ribosomal proteins, while the rest were involved in the diverse number of ncRNA metabolic processes. From the categorization of RBPs into RNA pathways, we can deduce the percentage of RBPs committed to different pathways and interpret expression changes of RBPs by regulatory process. The target-based categorization of RBPs also assists the interpretation of disease phenotypes and mutations emerging from rapidly increasing patient genome sequencing and may guide future functional studies.

Analyzing expression data across multiple tissues in humans and also other organisms I found that the majority of RBPs were ubiquitously expressed at higher levels than the residual protein-coding transcriptome, and represented up to a fifth of the total expressed protein-coding transcripts encoded for RBPs. We conclude that RNA metabolism not only constitutes one of the most conserved processes in the cell, but is also one of the cellular processes with the highest protein copy number demand.

Lastly, I demonstrated how structural data mining could be used to discover novel RBPs. Through a simple count of RG/RGG repeats across the human proteome I discovered the novel RBP FAM98A, a metazoan-conserved protein, and characterized its RNA-binding affinity by PAR-CLIP. FAM98A predominantly bound mature mRNAs in G-rich regions. Based on domain homologies, the domain of unknown function (DUF2465) in FAM98A may point towards a regulatory role in kinetochore/RNA-transport related processes, while its predominant localization to coding regions may also point to a putative role in protein translation. Further studies are needed to determine its physiological need in the cell.

Many details of PTGR remain to be revealed, including the dissection of newly discovered RNA regulatory processes, such as nuclear noncoding RNAs transcriptionally regulating gene expression and chromatin conformation or the dynamic regulation of posttranscriptional modifications of mRNAs (Cech and Steitz, 2014; Meyer and Jaffrey, 2014; Ulitsky and Bartel, 2013). However, even at the basic biochemical level we still have an incomplete understanding of how binding specificity is achieved, and how the regulatory function of an individual RBP is influenced by synergy and competition with other RBPs. How is PTGR executed in the cell with such high precision? A balanced approach of detailed biochemical and functional studies paired with complex, systems biology methods will ultimately lead to an understanding of the principles underlying PTGR networks. The more recent development of next generation sequencing-based methods aids the investigation of PTGR networks, such as RIP and CLIP-based methods (Ascano et al., 2012a; Konig et al., 2011; McHugh et al., 2014), ribosome profiling (Ingolia, 2014), *in vivo* RNA-secondary structure profiling (Ding et al., 2014; Rouskin et al., 2014; Wan et al., 2014), small and long RNA-sequencing (Jan et al., 2011; Ozsolak and Milos, 2011; Wang et al., 2009), or 3'-end sequencing methods profiling of alternative polyadenylation sites and poly(A) tail lengths (Chang et al., 2014; Jan et al., 2011; Lianoglou et al., 2013; Subtelny et al., 2014). These studies reveal an unanticipated complexity in RBP binding

and targeting and highlight the need to experimentally dissect PTGR networks in various cellular systems. Much of our current efforts are still focused on developing the tools and collecting large-scale datasets to understand the breadth of regulatory mechanisms of RBPs and their targets. With this census I hope to have created a foundation for the system-wide study of PTGR factors, which facilitates our understanding and interpretation of their gene regulatory patterns.

3 The DEDDh RNA exonuclease NEF-sp is involved in the 3'ETS removal of 28S rRNA

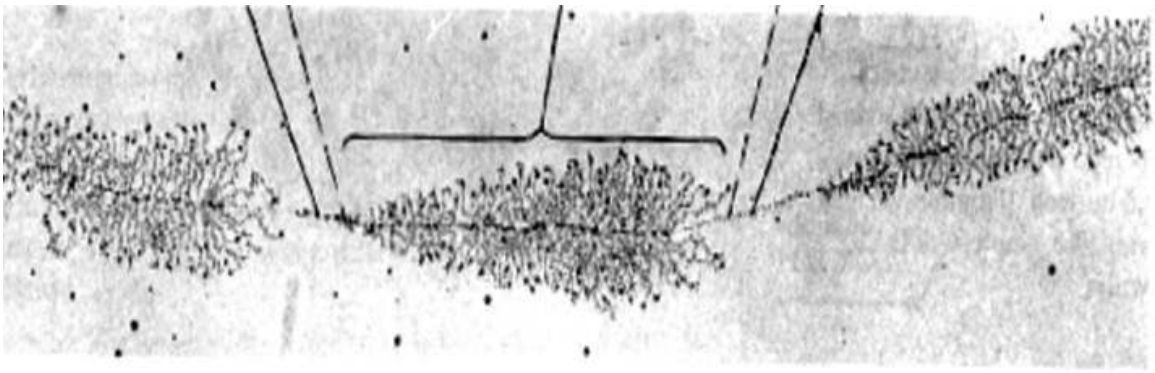
3.1 Introduction

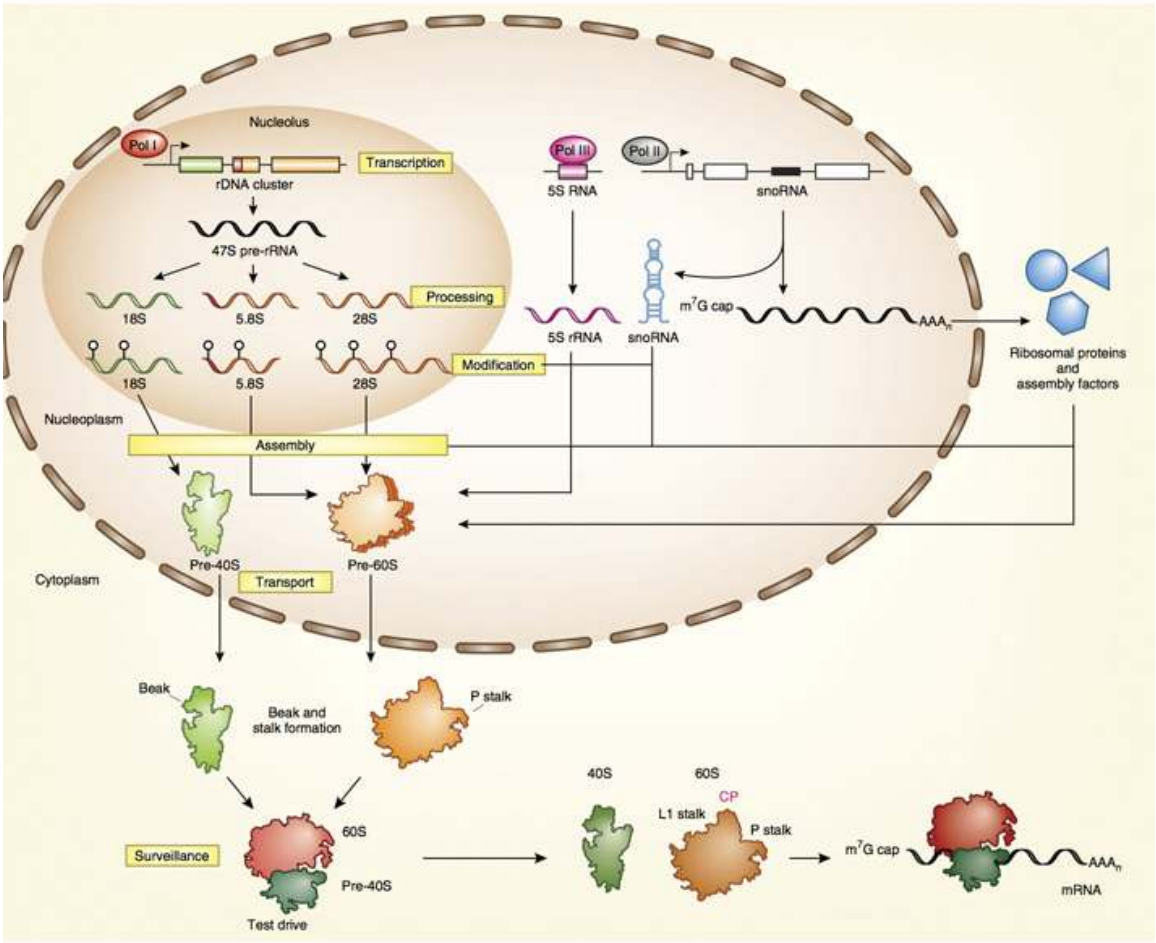
3.1.1 Ribosomes

Ribosomes are highly conserved RNA-protein complexes central to protein synthesis. They decode information of mRNAs and translate it into proteins, catalyzing peptide bond formation. Ribosomes are composed of two subunits, the small (40S in eukaryotes) and large (60S in eukaryotes) ribosomal subunit. The large subunit catalyzes the peptide bond formation, also referred to as peptidyl transferase activity, while the small subunit has the decoding function of the ribosome [reviewed in (Lafontaine and Tollervey, 2001)]. The large ribosomal subunit contains three ribosomal RNAs, 5S, 5.8S, and 28S rRNA, while 18S rRNA forms the RNA component of the small ribosomal subunit. Eukaryotic ribosomes have 80 core ribosomal proteins (Anger et al., 2013), the large ribosomal subunit contains 47 proteins, while the small subunit possesses 33 proteins; together the human ribosome forms a 4.3 MDa complex (Anger et al., 2013; Khatter et al., 2015). The majority of cellular metabolism is committed to ribosome synthesis and about 90% of all cellular transcribed RNA is rRNA (Warner, 1999). The number and production of ribosomes determine protein synthesis and growth rates of cells. In highly growth-stimulated cells, such as cancerous cells, ribosomal proteins and rRNA biogenesis factors are often expressed at higher levels. Nucleolar size (the site of rRNA production) and rRNA transcription are increased and lead to a higher number of ribosomes and protein synthesis in the cell (Boisvert et al., 2012; Montanaro et al., 2008; Ruggero and Pandolfi, 2003; Zhang et al., 2015a). Several oncogenes, such as p53 and Myc, directly regulate ribosome biogenesis (Ruggero and Pandolfi, 2003). Mutations that lead to dysfunctional ribosome assembly are generally lethal

early on during development. However, some defects in ribosomal proteins or rRNA biogenesis have been found to occur in several genetic diseases, commonly referred to as ribosomopathies [reviewed in (Hannan et al., 2013; Sondalle and Baserga, 2014)]. These show predominantly pathologies in bone marrow malfunction and growth defects. Several mutations in ribosomal proteins of the small (RPS7, 10, 17, 17L, 19, 24, 26) and large (RPL5, 10, 11, 35A) subunit cause Diamond-Blackfan anemia, a disorder that affects the production of red blood cells, while mutations in the ribosome biogenesis factors SBDS and EMG1 cause Shwachman-Diamond syndrome and Bowen-Conradi syndrome, which affect the production of white blood cells (neutrophils) and cause growth defects (Gerstberger et al., 2014b; Narla and Ebert, 2010; Sondalle and Baserga, 2014) (also see Table 2.1). Mutations in RNA polymerase I components, as well as rDNA transcription factors, such as TCOF1, have also been found in a number of genetic diseases causing growth and developmental defects, mental retardation, craniofacial and limb abnormalities (Hannan et al., 2013).

Figure 3.1 Overview of rRNA biogenesis in eukaryotes. Electron microscopy image of transcribed rRNA transcription units of amphibian oocytes and schematic overview of rRNA biogenesis steps in eukaryotes [adapted from (Lafontaine, 2015), image by (Miller and Beatty, 1969)]. Three out of four rRNAs are transcribed in the nucleolus by Pol I as a long 47S precursor (47S pre-rRNA). The precursor is further processed, cleaved and modified to give the mature 18S, 5.8S and 28S rRNAs, which are assembled into the pre-40S (green) and pre-60S (orange) ribosomal subunits. 5S rRNA (pink) is transcribed by Pol III in the nucleus and incorporated into maturing 60S subunits. 80 ribosomal proteins, >200 auxiliary factors and 200 snoRNAs are required for assembly. Pre-60S subunits require more nuclear maturation steps than pre-40S subunits and are exported after the 40S export to the cytoplasm. Final structural modifications take place in the cytoplasm (Panse and Johnson, 2010), such as formation of the beak on the small subunit and the stalk on the large subunit (Nerurkar et al., 2015).




Nucleolus
Pol I
Transcription
rDNA cluster
47S pre-rRNA
18S
5.8S
28S
Processing
Modification
Nucleoplasm
Assembly
Pre-40S
Pre-60S
Cytoplasm
Transport
Pol III
Pol II
5S RNA
snoRNA
5S rRNA
m7G cap
AAAn
Ribosomal proteins and assembly factors
Beak
Beak and stalk formation
P stalk
Surveillance
60S
Pre-40S
Test drive
40S
60S
CP
L1 stalk
P stalk
m7G cap
AAAn
mRNA

3.1.2 rRNA transcription, processing, assembly and export

Eukaryotic ribosome biogenesis is a highly orchestrated process and involves >200 assembly factors in *S. cerevisiae* and >400 in humans (Lafontaine, 2015; Tafforeau et al., 2013). By complexity and number of proteins involved, ribosome biogenesis is perhaps the most complex RNA metabolic pathway in the cell. rRNA processing is highly conserved among species, but additional protein factors, and processing pathways have evolved in higher eukaryotes concomitant with increasing organismal complexity. Due to its relative organismal simplicity, most of our knowledge on rRNA biogenesis results from the functional studies in *S. cerevisae*.

The main site of rRNA biogenesis is the nucleolus. Here, three of the four ribosomal RNAs, 18S, 5.8S and 28S, are transcribed by RNA polymerase I as a long 47S (35S in *S. cerevisiae*) precursor transcript, posttranscriptionally cleaved, nucleotide modified and assembled into early ribosomal complexes. ~200 proteins have been found in *S. cerevisiae* nucleoli (Woolford and Baserga, 2013) and up to ~4500 proteins have been detected in human nucleoli (Ahmad et al., 2009). Most of these are likely to be involved directly or indirectly in rRNA biogenesis and assembly (Mullineux and Lafontaine, 2012). 5S rRNA is transcribed separately by RNA polymerase III in the nucleus and assembled into pre-60S ribosomal complexes (Ciganda and Williams, 2011; Lafontaine, 2015). rDNA gene loci are organized in tandemly repeated arrays, with copy numbers varying between 400-100,000 copies in eukaryotic cells (Prokopowich et al., 2003). 400-600 copies are found in *S. cerevisiae* and humans, which are distributed between the nucleolar organization regions (NORs) (Carmo-Fonseca et al., 2000; Stults et al., 2009). About 50% of all rDNA loci are actively transcribed, while the others are nonfunctional or transcriptionally inactive repeats (Boulon et al., 2010). When visualized by electron microscopy, nascent precursor transcripts branch off from actively transcribed rRNA genes and form characteristic Christmas tree structures, at which ends RNP protein complexes, also termed terminal knobs, carry out rRNA processing (Kass and Sollner-Webb, 1990; Miller

and Beatty, 1969; Mougey et al., 1993) (Figure 3.1). In yeast, these have been identified as the small subunit (SSU) processome complexes [(Dragon et al., 2002), reviewed in (Granneman and Baserga, 2004; Henras et al., 2014; Woolford and Baserga, 2013)]. The long rRNA precursor is co- and post-transcriptionally processed, with endo- and exonucleolytic cleavage reactions taking place that remove the internal and external transcribed spacer regions (5'ETS, 3'ETS, ITS1, ITS2). Site-specific RNA modifications, such as methylation or pseudouridylation, are introduced by enzymes or snoRNPs guided by sequence specific basepairing of the snoRNAs to their target site. Instead of nucleotide modifications, some snoRNAs, such as U3 snoRNA, guide cleavage of rRNA precursors at specific sites [reviewed in (Henras et al., 2008)]. Concomitantly with posttranscriptional processing, rRNAs and ribosomal proteins are assembled into the ribosomal precursor complexes of the small and large subunit, assisted by a plethora of transport, folding and assembly complexes in the nucleolus, nucleus and cytoplasm. 18S rRNA assembles into the small subunit, while 5S rRNA, 28S, and 5.8S rRNA are incorporated into the large ribosomal subunit.

3.1.3 RNA exo- and endonucleases involved in rRNA processing

rRNA cleavage sites within the long precursor have been mapped in detail in yeast and the main pre-rRNA processing intermediates have also been characterized in higher eukaryotes [reviewed in (Mullineux and Lafontaine, 2012)] (Figure 3.2). The characterization, however, of the enzymes catalyzing the cleavage reactions has been challenging in all organisms. For some sites we do not know the identity of the cleavage factors, and for others understanding the overlapping substrate specificity has resulted in some uncertainty about how substrates are recognized and how processing is regulated (Mullineux and Lafontaine, 2012). Table 3.1 summarizes the current knowledge of RNA endo- and exonucleases involved in pre-rRNA processing.

Table 3.1 Overview of RNA exo- and endonucleases in rRNA processing. RNA exo- and endonucleases characterized in rRNA biogenesis in *S. cerevisae* and higher eukaryotes, their cleavage sites and nuclease domains are listed [based on (Henras et al., 2014; Zuo and Deutscher, 2001)].

S. cerevisiae **RNA nuclease**	***S. cerevisiae*** **cleavage site**	**Nuclease domain (endo/exo)**	**Mammalian homolog (human)**	**Mammalian cleavage site**	***Drosophila melanogaster*** **homolog**
NOB1	D	PIN (endo)	NOB1	3	CG2972
RNase MRP (MRP RNA in complex with POP1, POP5, RPP20, RPP25, RPP30, RPP38 and RPP40)	A_3	Catalytic RNA (endo)	RNase MRP (POP1, POP4, POP5, POP7, RPP14, RPP21, RPP25, RPP25L, RPP30, RPP38, RPP40)	-	RNase MRP (l(1)G0045, CG8038, CG14057, Rpp20, CG9422, Rpp30)
RCL1 (SSU component)	A_0, A_1, A_2	RNA 3'terminal phosphate cyclase-like domain (endo)	RCL1	E	Rtc1
Utp24 (Fcf1p)/cofactor Utp23 (Faf1p) (SSU component)	A_0, A_1, A_2	PINc (endo)	FCF1	SSU exosome component, putative A', A0, 1	Bka
RNT1	B_0	Ribonuclease III (endo)	DROSHA	-	Drosha
Rrp6 (RNA exosome)	5.8S+30>6S	DNA_pol_A_exo1/DEDD/ RNase D (3'-5' exo)	EXOSC10		Rrp6
Rrp44 (Dis3) (RNA exosome)	7S>5.8S+30	PIN/RNR (endo, 3'-5' exo)	DIS3	7S>5.8S+40	Dis3
Rat1p/cofactor Rai1p	C_2>C_1, A_3>B_{1S}	Xrn1/ 5PX (5'-3' exo)	XRN2/ cofactor DXO	2>5.8S rRNA, 4>28S rRNA,	Rat1/ putative cofactors CG912/ cuff
Rrp17p	A_3>B_{1S}	Nop25 domain (5'-3' exo)	NOL12	Putative 2>5.8S rRNA	viriato
Ngl2p	6S>5.8S rRNA	Exo_endo_phos/ DEDD/ Ccr4p (3'-5' exo)	-	-	
Rexp1	28S rRNA, 5S rRNA, 5.8S rRNA(?) (Piper et al., 1983; 1987; van Hoof et al., 2000)	DEDDh RNase T (3'-5' exo)	REXO1, NEF-sp	unknown	CG42666 (dREXO1), CG12877 (dREXO1), CG8368 (dNEF-sp)
Rex2p	5.8S rRNA(?) (van Hoof et al., 2000)	DEDDh RNase T (3'-5' exo)	REXO2	-	CG10214
Rex3p	5.8S rRNA(?) (van Hoof et al., 2000)	DEDDh RNase T (3'-5' exo)	-	-	-
Rex4p	ITS1 processing (Faber et al., 2004)	DEDDh RNase T (3'-5' exo)	REXO4	(ITS1 regulation)	CG6833
-	-	DEDDh RNase T (3'-5' exo)	ISG20L2	5.8S+40>6S rRNA	-
-	-	DEDDh RNase T (3'-5' exo)	Eri1	6S>5.8S rRNA	-

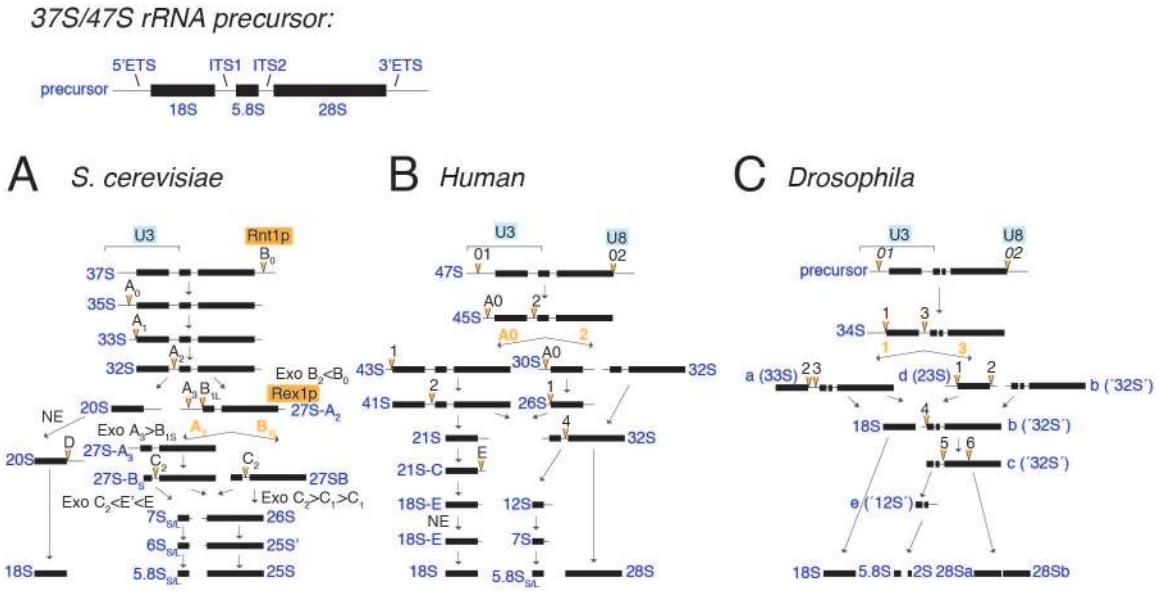


Figure 3.2 pre-rRNA processing in *S. cerevisiae*, humans, and *Drosophila*. rRNA size nomenclature is shown in dark blue, U3 and U8 snoRNAs are highlighted in light blue, RNA nucleases characterized in 28S rRNA 3' end processing are highlighted in orange, endonucleolytic cleavage sites are shown with orange triangles, exonucleolytic cleavages are shown in direction of the cleavage sites. (A) rRNA processing in *S. cerevisiae*: The majority of nascent transcripts is already co-transcriptionally cleaved at sites A_0, A_1, A_2 dependent on U3 snoRNP, yielding 20S and 27S-A_2 precursors. Alternatively, the full-length 35S pre-rRNA precursor is released by cleavage by the dsRNA endonuclease Rnt1p and processed posttranscriptionally. Cleavage A_2 involved Rcl1p, while RNase MRP is involved in cleavage at A_3. Nuclear export (NE) of the 20S precursor precedes final maturation of 20S to 18S rRNA by the RNA endonuclease Nob1 in the cytoplasm. The 5'-3' exonucleases Rrp17p and Rat1p trim the 5' ends of 27S-A_3, 26S, 25S', while the 3'-5' exonucleases of the nuclear exosome trim $7_{S/L}$ and $6_{S/L}$ rRNA precursors in the nucleus, and the yeast-specific cytoplasmic RNA exonuclease Ngl2p matures the 3' ends of 5.8S rRNA. The 3'-5' exonuclease Rex1p (RNH70) trims off the 3'ETS overhangs from 28S rRNA. (B) rRNA processing in humans/mammalian cells [processing steps are adapted from (Henras et al., 2014)]: The rRNA precursor is fully transcribed before it is posttranscriptionally processed. The precursor is initially cleaved at sites 01 and 02. The 45S precursor can be processed in two different pathways, either first in the 5'ETS (at site A0 and 1) or first in the ITS1 (site 2). Cleavage at site 2 depends on Rcl1p, endonucleolytic cleavage of Nob1 of the 18S-E precursor generates the mature 18S rRNA. The 5'-3' exonuclease XRN2 trims 5' overhangs of 32S and 28S precursors after cleavage 2 and 4. Isg20l2 and the nuclear exosome trim different 5.8S rRNA precursors, namely the 12S, 7S, and 5.8S+40, while the final 5.8S exonucleolytic maturation of 5.8S takes place in the cytoplasm processed by Eri1. (C) rRNA processing in *D. melanogaster*: Few studies have investigated rRNA processing in *D. melanogaster*, but the main rRNA processing intermediates have been characterized (Long and Dawid, 1980). Similar to mammalian and *Xenopus laevis* pre-rRNA processing, *Drosophila* also contains U8 snoRNA (Peculis, 1997). Drosha is not involved in rRNA processing (Smibert et al., 2011). MRP RNase is involved in cleavage at site 3 (Schneider et al., 2010). Conceptionally, the two alternative processing pathways are conserved among mammalian and *Drosophila* rRNA processing.

3.1.4 pre-rRNA processing in *S. cerevisiae*

In *S. cerevisiae* transcription terminates 210 nts downstream of the mature 28S rRNA site. The precursor transcript is co-transcriptionally cleaved by the dsRNA-binding endonuclease Rnt1p, which results in the 35S precursor with a 3'ETS of 10-20 nts in length (Kressler et al., 1999) (Figure 3.2 A). The resulting primary 35S rRNA precursor is cleaved in the 5'ETS at site A_0, A_1, and A_2, a process that requires the small subunit processome (SSU) in complex with U3 snoRNA (Woolford and Baserga, 2013). At present the enzyme carrying out endonucleolytic cleavage remains unknown, but both the endonucleases Rcl1 and Utp24 have been implicated as putative catalytic factors at A_1 and A_2, most recent studies favor Rcl1 as the catalytic factor based on its mammalian homolog that is also involved in ITS1 cleavage (Henras et al., 2014; Mullineux and Lafontaine, 2012; Woolford and Baserga, 2013). After A_2, 20S intermediates are exported to the cytoplasm, where Nob1 endonuclease cleaves the remaining 3' end of ITS1 from the precursor to generate the mature 18S rRNA (Henras et al., 2014). Following 5'ETS cleavage, precursors are cleaved either at A_3 by RNase MRP or cleaved at B_{1L} by an unknown RNA endonuclease. Some uncertainty about the required enzymatic factors also exists at A_3, since cleavage at A_3 is not strictly dependent on RNase MRP and cleavage by MRP seems to be not conserved in higher eukaryotes (Henras et al., 2014; Kressler et al., 1999). After cleavage at A_3 or B_{1L}, the resulting 27S intermediate is 3' end processed by RNH70 (also known as Rex1p), which exonucleolytically trims the 3'ETS overhangs of the 27S precursor (Kempers-Veenstra et al., 1986; Mullineux and Lafontaine, 2012; Piper et al., 1984; Skowronek et al., 2014; van Hoof et al., 2000) (highlighted in Figure 3.2 A). 27S intermediates cleaved at A_3 possess 5' extensions of ITS1, which are 5'-3' exonucleolytically removed by Rat1p/Rai1 or Rrp17p to generate the mature end of 5.8S rRNA (Hage et al., 2008; Henras et al., 2014; Oeffinger et al., 2009). After A_3 or B_{1L} cleavage, the resulting 27S-$B_{S/L}$ intermediates undergo endonucleolytic cleavage at C_2, releasing the 7S precursor and the mature 25S rRNA. The mature 25S rRNA is transported to the

cytoplasm, while the 7S precursor is trimmed by the RNA exosome, assisted by unwinding of secondary structures by the RNA helicase Mtr4p. The nuclear exosome component Rrp6 further trims 7S to generate the 6S precursor, which is then exported to the cytoplasm. The yeast specific RNA exonuclease Ngl2p exonucleolytically 3'-5' trims the final extensions, generating the mature 5.8S rRNA.

3.1.5 pre-rRNA processing in higher eukaryotes

In *Xenopus laevis* RNA Pol I transcribes the entire ribosomal gene repeat unit far into the 3' external spacer and terminates ~200 nts upstream of the rDNA promoter of the next rDNA unit (Labhart and Reeder, 1986). A processing event at a site ~250 nt downstream of the mature 28S rRNA (T2) forms the 7,800 kb precursor (Labhart and Reeder, 1987). Similarly, near-promoter termination sites -200 nts upstream of the initiation site have also been found in mouse rDNA repeats, but rRNA transcription is thought to terminate ~560 nts downstream of the 28S 3'end (Grummt et al., 1986; Henderson and Sollner-Webb, 1986). In humans, precursor transcripts extend to ~350 nts downstream of 28S rRNA (Bartsch et al., 1987; Gurney, 1985; Kuhn and Grummt, 1989). In *Drosophila melanogaster* transcription termination can extend throughout the entire 3'ETS and spacer region without a fixed termination point, however the long 47S precursor transcript has a defined length of ~8 kb (Long and Dawid, 1980; Mandal and Dawid, 1981; Tautz and Dover, 1986).

In mammals, the 47S precursor is first cleaved at site 01, ~600 nts downstream of the transcription start site, followed by cleavage at 02 (also known as site 6) (Figure 3.2 B). This generates the 45S precursor, which already contains the mature 3' end of 28S rRNA. Two alternative pathways for maturation can take place then. In pathway 1 the precursor undergoes cleavage at site A0 and 1, generating the 41S precursor. The 41S precursor has already the mature 5' and 3' ends of 18S and 28S rRNA defined (Figure 3.2 B). Internal cleavage 2 in the ITS1

generates 21S, the 3' extended 18S precursor, and 32S, the 5.8-28S precursor. Cleavage at site 2 depends on the PES1/BOP1/NOL12 complex (Henras et al., 2014). Alternatively, in pathway 2, cleavage at 2 precedes cleavage at A0 and 1, generating first the 32S precursor of 28S and a 30S precursor of 18S, which is then rapidly cleaved by A0 and 1 to generate the 21S precursor. Cleavage at A', A0 and 1 depend on the U3 SSU processome (Henras et al., 2014). Ultimately, both pathways arrive at the 21S and 32S precursors. The 21S rRNA precursor undergoes cleavage at site E, which is dependent on RCL1, followed by a series of exonucleolytic trimming reactions before it is exported to the cytoplasm, where Nob1 endonucleolytically cleaves off the 3' extension to generate mature 18S rRNA. The 5.8S-28S containing 32S precursor undergoes endonucleolytic cleavage at site 4, generating the 12S precursor of 5.8S rRNA and a 5' extended 28S rRNA. 28S rRNA is 3'-5' exonucleolytically trimmed by Xrn2. The 12S rRNA precursor undergoes a series of 3'-5' exonucleolytic trimming reactions, first to by the RNA exosome and Isg20L2, which generate the 7S precursor in the nucleus. The nuclear RNA exosome component Rrp6 further trims the 7S to generate the 6S precursor. The 6S precursor is transported to the cytoplasm, where Eri1 3'-5' exonucleolytically trims the final ends to form mature 5.8S rRNA (Ansel et al., 2008).

Pre-rRNA processing in *Drosophila* has been characterized in less detail, and only the main rRNA precursor intermediates have been described (Long and Dawid, 1980), however the basic outline, as well as conserved protein factors closely mirror mammalian pre-rRNA processing steps (Figure 3.2 C). Additionally, *Drosophila* 5.8S rRNA is split into two pieces, a shortened 5.8S rRNA and an additional 2S rRNA downstream, presenting essentially the 3' end of 5.8S rRNA (Pavlakis et al., 1979). 2S rRNA basepairs with 5.8S and 28S rRNA and is also incorporated into the large ribosomal subunit. Furthermore, insect 28S rRNAs contains a natural hydrolytic internal cleavage site, which cleaves 28S rRNA into two fragments 28Sa and 28Sb (Long and Dawid, 1980).

3.1.6 28S rRNA maturation in *S. cerevisiae*

The 3' ends of the primary precursor transcript are processed co-transcriptionally in yeast by the dsRNA-specific RNA endonuclease Rnt1p. The resulting precursor transcripts terminate 10-20 nt downstream of the mature 28S 3' end (Henras et al., 2014; Kressler et al., 1999) (Figure 3.2 A). The 3'-5' DEDDh RNA exonuclease RNH70 (Rex1p) exonucleolytically trims the 3' overhangs of 28S during late ribosome biogenesis (Mullineux and Lafontaine, 2012). RNH70 has also been shown to trim the 3' trailer of 5S rRNAs in yeast (Kempers-Veenstra et al., 1986; Piper et al., 1983; 1984; Skowronek et al., 2014; van Hoof et al., 2000). While Rnt1p is essential for rRNA biogenesis in yeast, RNH70 is a nonessential enzyme and deletion strains are viable without any growth defects (Henras et al., 2008; Kufel et al., 1999; van Hoof et al., 2000). The 28S rRNA is released from the precursor after cleavage at site C2 by an unknown RNA endonuclease. Rat1p does a final trimming by exonucleolytically removing 5'-3' overhangs on 28S rRNA to generate the mature 5' end (Wang and Pestov, 2011).

3.1.7 28S rRNA maturation factors in higher eukaryotes

In metazoans, under steady state conditions the majority of the precursor is already processed at the mature 3' end of 28S rRNA (site 02), by an unknown nuclease (Labhart and Reeder, 1986). Removal of the 3'ETS is thought to take place as precise endonucleolytic cleavage reaction at the mature 28S rRNA end and depends on the presence of U8 snoRNA. Drosha, the homolog of Rnt1p, is functionally not conserved in 3'ETS cleavage and processes microRNA precursors in the nucleus (Johanson et al., 2013; Lee et al., 2003). Knockdown of Drosha with antisense oligonucleotides in HeLa cells had no effect on 47S precursor levels, but a small increase observed for the 12S and 32S intermediates suggested a minor role of Drosha in ITS2 cleavage (Wu et al., 2000). However, an *in vivo* knockout of Drosha in mice or in *Drosophila* had no

effects on precursors or mature rRNA species, leading to the conclusion that Drosha has no role in pre-rRNA processing (Chong et al., 2008; Smibert et al., 2011).

Absent in yeast, U8 snoRNA is conserved across verte- and invertebrates and displays base complementarity to the 28S 3'end and the 5.8-28S ITS2 junction. U8 is thought to assist long range pre-rRNA folding and correct positioning of the RNAs for 3'ETS and ITS2 cleavage (Peculis, 1997; Peculis and Steitz, 1993). U8 snoRNA depleted *Xenopus laevis* oocytes show depleted cytoplasmic 28S rRNA pools and cleavages at the 28S-3'ETS border, ITS1 and ITS2 sites are inhibited. Cleavage at T1 (site 02 in mammals) is completely blocked and rRNA precursors are extended by ~800 nts to the termination site T2 in U8 depleted oocytes. The inhibition leads to accumulation of aberrant 32S and 36S intermediates with extended nucleotides at both 3' and 5' ends (Peculis and Steitz, 1993). Currently, apart from the U8 snoRNP, only two other protein factors are implicated in the 3' end maturation of 28S rRNA in mammalians: the DEAD box helicase Ddx51 with its interaction partner Nog1, while an additional helicase, Ddx27, is required for correct 3'ETS end definition (Kellner et al., 2015; Srivastava et al., 2010). The nuclease that removes the 3'ETS remains unknown in metazoans. To generate the 5' end of 28S rRNA cleavage takes place at site 4 in the ITS2 of the 32S precursor by an unknown RNA endonuclease, thereby releasing the 28S precursor. The resulting 28.5S rRNA precursor is 5'-3' exonucleolytically trimmed by Xrn2 to generate the mature 5' end (Wang and Pestov, 2011).

3.1.8 Role of snoRNPs in rRNA modification and processing

rRNAs undergo extensive co- and posttranscriptional RNA modifications. These modifications are essential in eukaryotes, required for cell growth, and strengthen RNA stability, specific secondary structural folds, and influence ribosome activity (Granneman and Baserga, 2004; Henras et al., 2008; Sharma and Lafontaine, 2015; Watkins and Bohnsack, 2012). snoRNPs catalyze site-specific 2'-O-methylations and pseudouridylations in rRNAs, other snoRNAs,

telomerase RNA and snRNAs. Alternatively, some snoRNPs are required for rRNA processing by guiding cleavage and assisting folding of rRNA precursors. snoRNPs are ribonucleoprotein complexes composed of small nucleolar RNAs (snoRNAs) of ~150 nts (range from 50-450 nts) bound to their obligate protein partners. Based on their conserved sequence motifs and secondary structure, snoRNAs can be grouped into two classes, box C/D and box H/ACA snoRNAs (Filipowicz and Pogacić, 2002) (Figure 3.3).

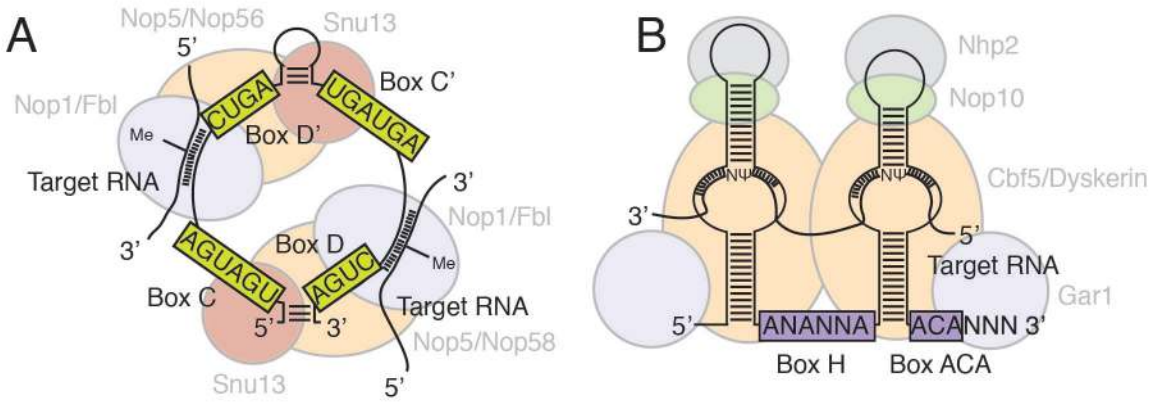


Figure 3.3 Structure of box C/D and H/ACA snoRNPs. Box C/D snoRNPs 2'-O-methylate their targets, while box H/ACA snoRNPs isomerize uridines to pseudouridines. (A) Schematic structure of the box C/D family snoRNAs bound to their protein binding partners. In eukaryotes box C/D snoRNPs form heterodimers with the complex partners Nop58 or Nop56, Fbl, Snu13. Guiding 2′-O-methylation involves base-pairing of the 10-21 nt-long sequence positioned upstream of box D (or D′) to the target RNA, the nucleotide positioned 5 bp upstream of the D/D′ box is selected for methylation. (B) Box H/ACA snoRNAs form a double loop. Within the pseudouridylation pocket 3-10 nucleotides are complementary to the nucleotides in the target RNA flanking the isomerization site. Figure adapted from [(Fatica and Tollervey, 2003; Filipowicz and Pogacić, 2002; Watkins and Bohnsack, 2012)].

Each class associates with a distinct set of four evolutionary conserved core proteins (human orthologs official gene name given in parentheses): box C/D snoRNAs bind to Snu13 (SNU13), Nop56 (NOP56), Nop58 (NOP58), and the 2'-O-methylase Nop1 (Fibrillarin, FBL); box H/ACA snoRNAs protein complex partners are Nhp2 (NHP2), Nop10 (NOP10), Gar1 (GAR1), and the pseudouridine synthase catalytic subunit Cbf5 (DKC1). snoRNAs are transcribed by RNA Pol II, and are predominantly located in introns of mRNA genes in higher

eukaryotes (Dieci et al., 2009). Through complementary base pairing, the RNA component of snoRNPs guides the complex site-specifically to the target rRNAs to introduce nucleotide modifications. Most snoRNAs are involved in RNA modification, but a subset assists RNA folding or cleavage during rRNA processing (Watkins and Bohnsack, 2012). U3 snoRNP is a component of the SSU processome and is essential for cleavage in the 5'ETS at A0, A1 and A2 (in yeast) (Phipps et al., 2011). In higher eukaryotes (not present in *S. cerevisiae*) U8 snoRNA is required for 28S rRNA maturation and cleavage of the 3'ETS. Few other snoRNAs have been characterized in the context of 90S pre-ribosomal complex processing, where they assist folding and interact with the 35S rRNA precursor (C/D: U14, U22; H/ACA: snR30/U17, snR10) (Henras et al., 2008).

3.1.9 Tissue-specificity of rRNAs, ribosomal proteins and rRNA processing factors

Ribosomal proteins and rRNA biogenesis factors are among the highest expressed proteins in the cell. Based on the analysis in Chapter 2, we find that they are generally expressed ubiquitously across human tissues with very little tissue-specific variation in mRNA levels (Gerstberger et al., 2014a) (Figure 2.8). However, tissue-specific differences in pre-rRNA intermediates, the processing pathways, rRNA species (e.g. ratio of $5.8S_{S/L}$ rRNA), as well as heterogeneous compositions of ribosomes have been described (Lafontaine, 2015; Mullineux and Lafontaine, 2012). Heterogeneous ribosomes arise from tissue-specific ribosomal protein composition through incorporation of different ribosomal protein isoforms or variations in the rRNA sequences, resulting from stage-specific transcription of different rDNA gene loci (Lafontaine, 2015; Xue and Barna, 2012). Furthermore, rRNA modifications and posttranslational modifications of ribosomal proteins introduce additional levels of heterogeneity (Mauro and Edelman, 2002; Ramagopal, 1992). Incorporation of specific ribosomal protein paralogs affects the production and function of yeast ribosomes (Parenteau et al., 2011). Tissue-specific and

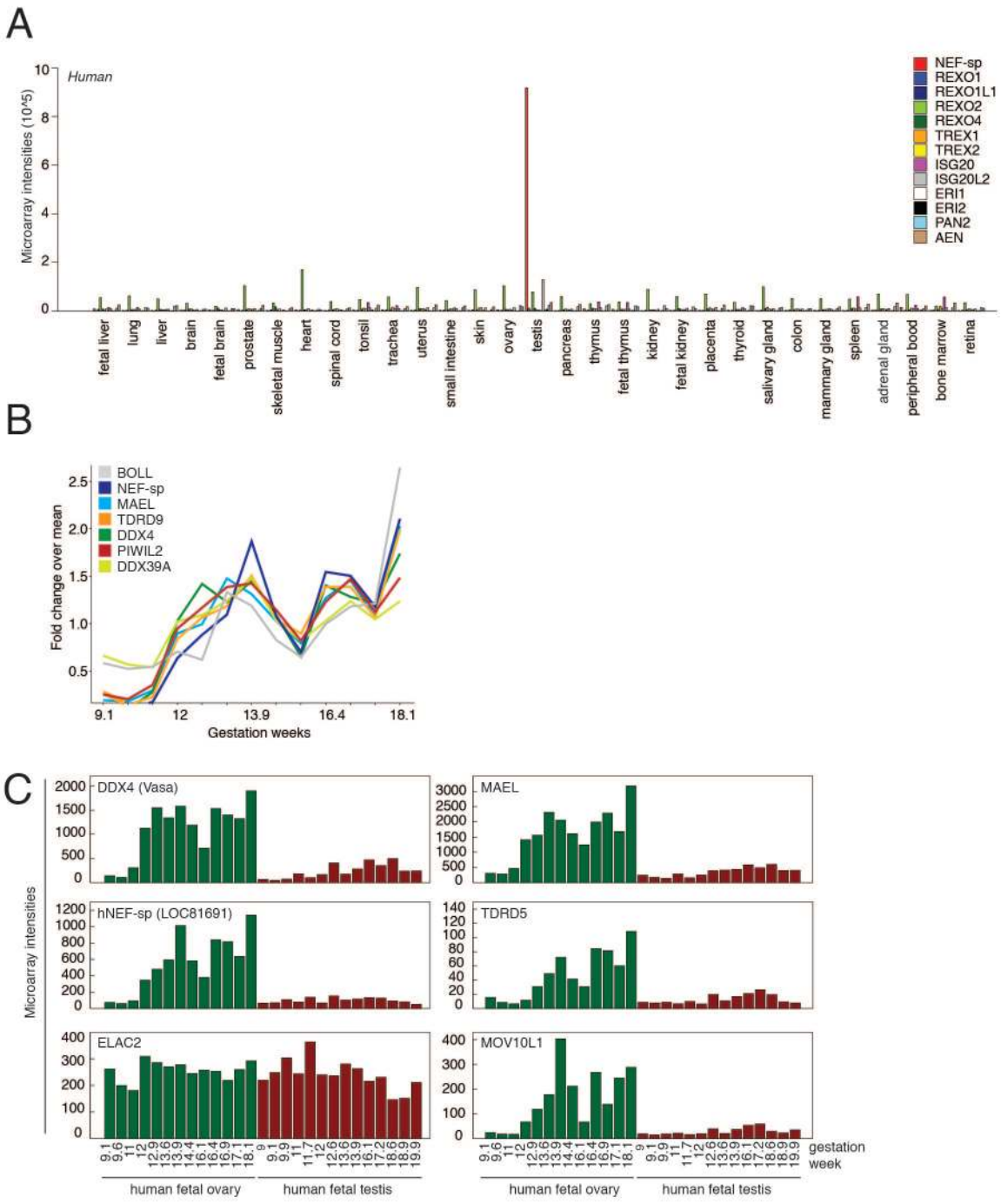
stress-response tailoring of ribosomes by tissue-specific incorporation of ribosomal proteins has been shown to enhance translation of specific mRNA targets (Kondrashov et al., 2011; Vesper et al., 2011; Xue and Barna, 2012; Zhang et al., 2015b). Other extraribosomal roles for ribosomal proteins have also been reported (Warner and McIntosh, 2009). Most of these cases are incidental and a systematic survey of ribosome composition in different cell types and tissues, their expressed rRNA and ribosomal protein isoforms, as well as the level of protein and rRNA modifications has been challenging (Lafontaine, 2015; Xue and Barna, 2012). Profiling protein and rRNA isoforms, as well as their posttranslational modifications, across cell types, developmental stages and tissues remains an important goal for elucidating the underlying principles of ribosome composition in adaptation to different developmental and tissue-specific needs.

Furthermore, while tissue-specific regulation of ribosomal proteins has been reported so far, much less is understood about the processes of posttranscriptional regulation during rRNA biogenesis. We know very little about the regulation and redundancy of rRNA processing and the involved enzymes, their stoichiometry, and activity. Given the functional redundancies of some exonucleases, as well as the regulation of rRNA biogenesis factors by growth stimulatory factors (e.g. mTOR, p53, myc) (Buszczak et al., 2014), the question arises how rRNA biogenesis is regulated by signaling pathways and maintains adaptation and homeostasis in the cell. A detailed understanding of ribosome composition and the different rRNA processing pathways will be important in order to explain the observed heterogeneity of rRNAs and the resulting functional consequences for protein translation.

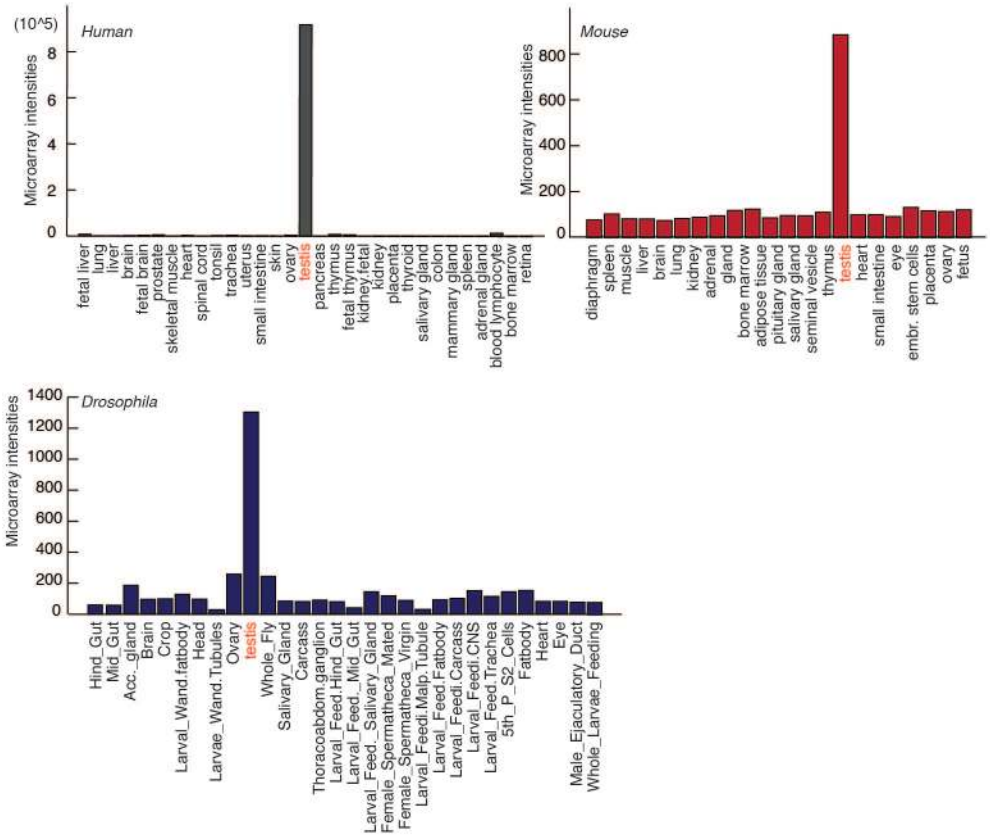
3.1.10 NEF-sp is a conserved, tissue-specific RNA exonuclease uncharacterized in higher eukaryotes

In the previous genome-wide gene expression analysis of RBPs across human 31 tissues (Chapter 2), I noticed the uncharacterized protein LOC81691, human NEF-sp RNA exonuclease (hNEF-sp), among the most tissue-specific RBPs (top 3 in our ranking of tissue-specific RBPs in Chapter 2) with a tissue specificity score ~4 (Gerstberger et al., 2014a) (Figure 2.8 B). hNEF-sp, a member of the DEDDh RNase T exonuclease family, displayed unique expression in comparison to all other members of the RNase T class and had exceptional tissue specificity in adult human testis and human fetal ovaries, where it was found among the top 100 expressed genes (Figure 3.4 A-C).

Figure 3.4 NEF-sp is highly expressed in gonads in humans, mouse and *Drosophila*. (A) Microarray expression profiles for human DEDDh RNA exonucleases across 31 human tissues. NEF-sp is the only DEDDh RNA exonuclease with such high and selective tissue-specific expression. (B) Plot showing the normalized expression changes of selected RBPs relevant in female ovarian gonad development and NEF-sp in comparison [analysis is the same as in Figure 2.10 C, (Houmard et al., 2009)]. (C) Microarray intensity levels for germline-specific RBPs DDX4, MAEL, TDRD5, MOV10L1, in comparison to NEF-sp and the ubiquitous RBP ELAC2 as control across gestational stages in human ovary and testis development. (D) Microarray intensities of NEF-sp mRNA levels across 31 human adult tissues, 32 mouse adult tissues and 30 *Drosophila* embryonic, larval and adult tissues.


A
Human
Microarray intensities (10^5)
NEF-sp
REXO1
REXO1L1
REXO2
REXO4
TREX1
TREX2
ISG20
ISG20L2
ERI1
ERI2
PAN2
AEN
fetal liver
lung
liver
brain
fetal brain
prostate
skeletal muscle
heart
spinal cord
tonsil
trachea
uterus
small intestine
skin
ovary
testis
pancreas
thymus
fetal thymus
kidney
fetal kidney
placenta
thyroid
salivary gland
colon
mammary gland
spleen
adrenal gland
peripheral bood
bone marrow
retina
B
BOLL
NEF-sp
MAEL
TDRD9
DDX4
PIWIL2
DDX39A
Fold change over mean
Gestation weeks
C
DDX4 (Vasa)
MAEL
hNEF-sp (LOC81691)
TDRD5
ELAC2
MOV10L1
Microarray intensities
gestation week
human fetal ovary
human fetal testis

D


(10^5)
Human
Microarray intensities
0
2
4
6
8
fetal liver
lung
liver
brain
fetal brain
prostate
skeletal muscle
heart
spinal cord
tonsil
trachea
uterus
small intestine
skin
ovary
testis
pancreas
thymus
fetal thymus
kidney fetal
kidney
placenta
thyroid
salivary gland
colon
mammary gland
spleen
adrenal gland
blood lymphocyte
bone marrow
retina
Mouse
Microarray intensities
0
200
400
600
800
diaphragm
spleen
muscle
liver
brain
lung
kidney
adrenal gland
bone marrow
adipose tissue
pituitary gland
salivary gland
seminal vesicle
thymus
testis
heart
small intestine
eye
embr. stem cells
placenta
ovary
fetus
Drosophila
Microarray intensities
0
200
400
600
800
1000
1200
1400
Hind_Gut
Mid_Gut
Acc._gland
Brain
Crop
Larval_Wand.fatbody
Head
Larvae_Wand.Tubules
Ovary
testis
Whole_Fly
Salivary_Gland
Carcass
Thoracoabdom.ganglion
Larval_Feed.Hind_Gut
Larval_Feed._Mid_Gut
Larval_Feed._Salivary_Gland
Female_Spermatheca_Mated
Female_Spermatheca_Virgin
Larval_Feedi.Malp.Tubule
Larval_Feed.Fatbody
Larval_Feedi.Carcass
Larval_Feedi.CNS
Larval_Feed.Trachea
5th_P_S2_Cells
Fatbody
Heart
Eye
Male_Ejaculatory_Duct
Whole_Larvae_Feeding

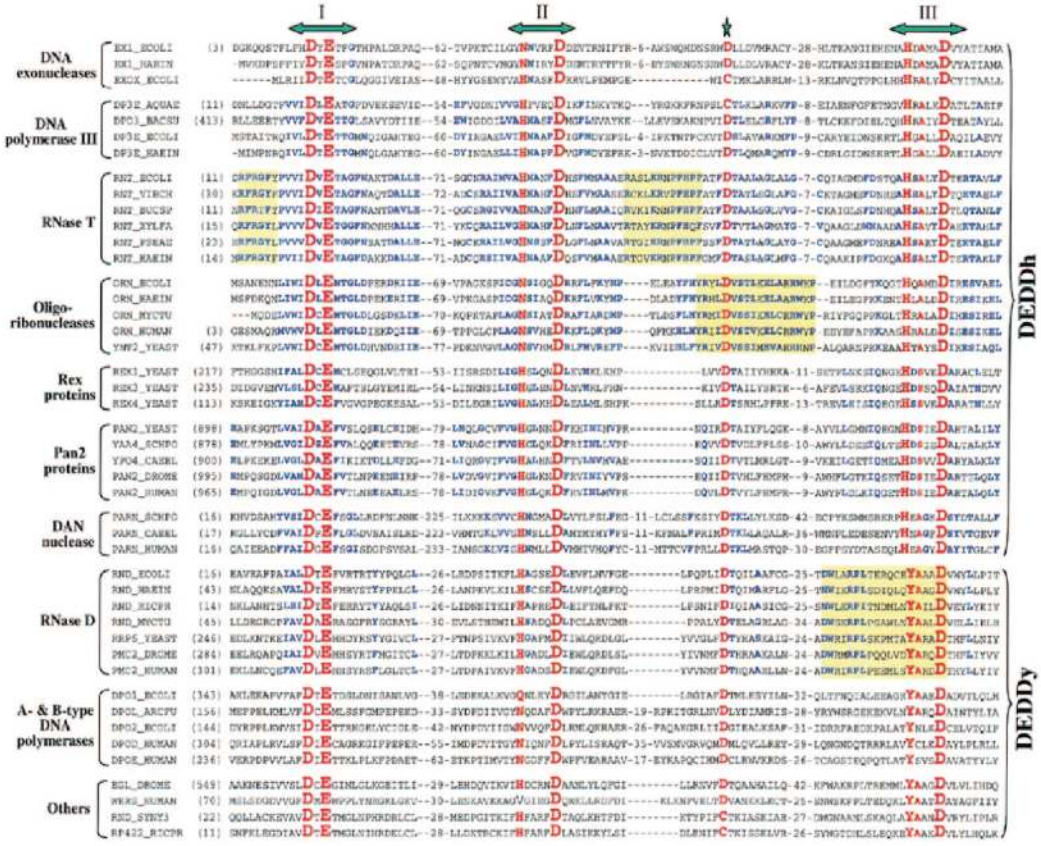
Its expression clustered with gonad-specific RBPs, such as the PIWI proteins, DDX4, or BOLL proteins (Gerstberger et al., 2014a) and suggested to us a putative role of NEF-sp during gonad development (Figure 3.4 B). Supporting a conserved role during gonad development, the observed tissue-specific of NEF-sp in testis was found across organisms, in mouse (2610020H08Rik, mNEF-sp) and *Drosophila melanogaster* (CG8368, dNEF-sp).

3.1.11 DEDDh RNase T class exonucleases

The RNase-T class of single-stranded 3'-5' RNA exonucleases is highly conserved across prokaryotes and eukaryotes and belongs to the DEDD 3'-5' exonuclease superfamily, also known as DnaQ-like or RNase D superfamily (Aravind and Koonin, 2001a; Hsiao et al., 2011; Zuo and Deutscher, 2001) (Figure 3.5 A).

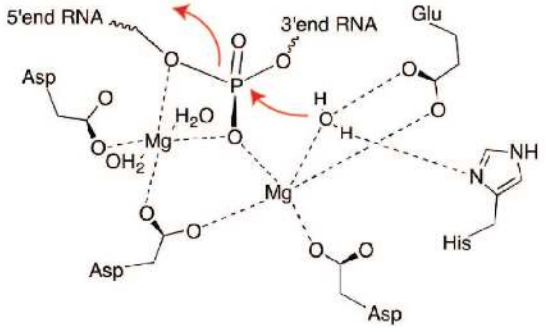
Figure 3.5 Active site of DEDD RNA exonucleases. (A) ClustalX multiple sequence alignments for bacterial, viral and eukaryotic DEDD RNA exonucleases. Two main classes of DEDD RNA exonucleases exist, DEDDy and DEDDh, where four acidic catalytic residues coordinate a divalent cation (e.g. Mg^{2+}, Mn^{2+}), and one general base residue, tyrosine (y) or histidine (h) deprotonates the coordinated water molecule in the active site. Figure adapted from (Zuo and Deutscher, 2001). The three conserved DEDD regions are indicated on the top (green arrow), the conserved residues are highlighted in red. Residues that are only highly conserved in subfamilies are highlighted in yellow. (B) Schematic representation of the active site residues coordinating two Mg^{2+} residues and one water molecule for nucleophilic attack at the 5' phosphate of the target RNA. Figure adapted from bacterial RNase T active site (Hsiao et al., 2011).

A



B

RNase T active site:



Members of the DEDD family contain four acidic catalytic residues that chelate two Mg^{2+} ions and coordinate a water molecule for nucleophilic attack on the phosphate bond of nucleic acids (Beese and Steitz, 1991a; 1991b; Zuo et al., 2007). RNase T exonucleases contain a DEDDh catalytic domain with a conserved histidine residue that acts as a general base and deprotonates the chelated water molecule for nucleophilic attack (Figure 3.5 B). The bacterial RNase T protein forms a homodimer and is involved in the maturation of tRNAs, 5S and 23S rRNA, and other noncoding RNAs (Li and Deutscher, 1995; Li et al., 1998; 1999). In *S. cerevisiae* the RNase T class has five members, Rex1p (RNH70), Rex2p, Rex3p, Rex4p and Pan2p (Figure 3.6). The Rex1-Rex4 proteins have been described in rRNA, tRNA and small nuclear noncoding RNA metabolism. Rex1p was demonstrated to trim the 3' ends of tRNAs, 5S, 23S (Copela et al., 2008; Ozanick et al., 2009; Piper and Stråby, 1989; Piper et al., 1983; 1987). In addition, Van Hoof *et al*. reported that deletion of Rex1p affected 5.8S rRNA, $U5_L$ snRNA and RNase P maturation (van Hoof et al., 2000). In the same study Rex2 was shown to also affect processing of 5.8S rRNA, RNase P, U4 and U5S/L snRNA; and Rex3 trimmed RNase MRP, P, U5L snRNA and 5.8S rRNA (van Hoof et al., 2000). Rex4p has been reported in ITS1 processing (Faber et al., 2004). In higher eukaryotes additional RNase T families evolved, namely the Eri1 family (ERI1-3), the ISG20 family (ISG20, ISG20L2, AEN), and the ssDNA/ssRNA-specific TREX1 family involved in ssDNA 3' end repair (TREX1-2) (Stetson et al., 2008; Yuan et al., 2015).

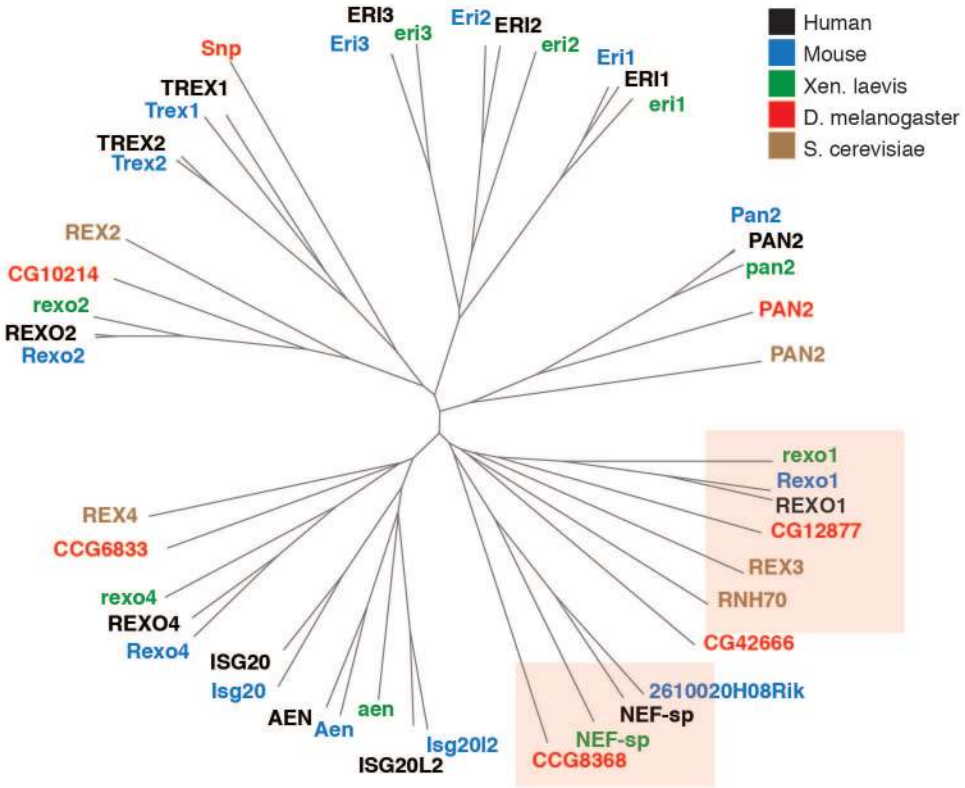


Figure 3.6 Phylogenetic tree of DEDDh RNA exonucleases. Phylogenetic tree generated from multiple sequence alignments of DEDDh RNA exonucleases across human (black), mouse (blue), *Xen. laevis* (green), *D. melanogaster* (red), *S. cerevisae* (brown). The NEF-sp and its paralog REXO1 family are highlighted in red. Branch lengths are scaled according to sequence identity.

Very few RNase T class members have been characterized in higher eukaryotes. PAN2 is cytoplasmic, nonessential for viability, and regulated mRNA stability through deadenylating the long (~200 nts) poly(A) tail of mRNAs; however, it does not efficiently remove the last 20-25 nts, a process which is catalyzed by the Ccr4-Not complex (Wolf and Passmore, 2014). Eri1 was originally discovered to have a negative regulatory effect on siRNAs in *C. elegans* (Kennedy et al., 2004). In *S. pombe* Eri1 was shown to negatively regulate the RNA-induced silencing (RITS) complex (Buhler et al., 2006; Iida et al., 2006). In mouse and *C. elegans* Eri1 trims 5.8S rRNA as the final maturation step in the cytoplasm (Ansel et al., 2008; Gabel and Ruvkun, 2008). Metazoan Eri1 also trims the polyuridine tails of histone mRNAs (Dominski et al., 2003; Hoefig

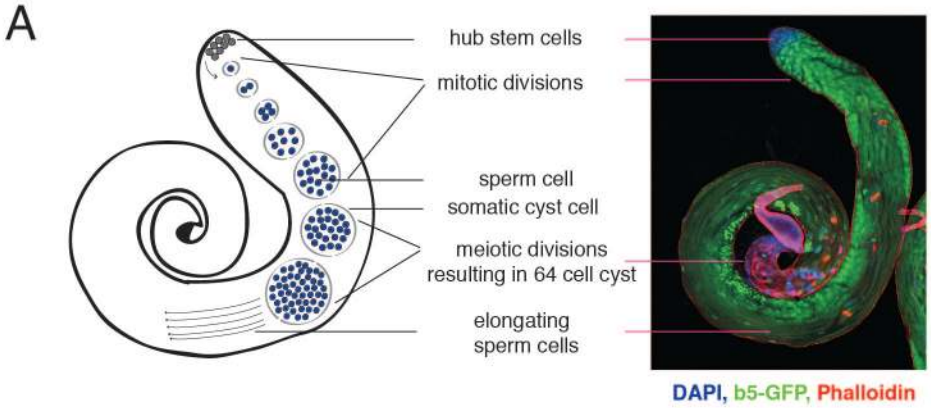
et al., 2013). Furthermore, Eri1 affects miRNA homeostasis and Eri1-/- knockout mice show a twofold increase in total miRNA abundance (Thomas et al., 2012; 2014). The other two Eri family members, Eri2 and Eri3, remain uncharacterized. Interferon stimulated exonuclease gene 20 kDa (ISG20) was initially identified as interferon- or double-stranded RNA induced exonuclease (Espert et al., 2004; Gongora et al., 1997). ISG20 was found to specifically degrade single-stranded viral RNA (Espert et al., 2003) and was shown to associate with snRNAs and snoRNAs in the nucleolus and Cajal body, suggesting a role of ISG20 in snRNA/snoRNA and rRNA maturation (Espert et al., 2006). The nucleolar exonuclease ISG20L2 has been shown to trim 7S rRNA precursor transcripts (Coute et al., 2007; Henras et al., 2014). The third member of the ISG20 family, ISG20L1 was renamed to nuclear apoptosis enhanced nuclease (AEN) as it was initially characterized as p53 induced RNA exonuclease, which degrades both double and single-stranded RNA and DNA in the nucleus (Kawase et al., 2008).

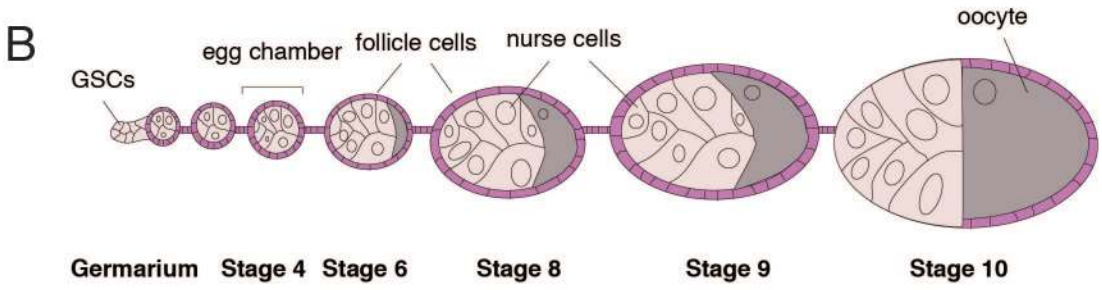
The REX1-4 family remains largely uncharacterized in higher eukaryotes. Only REXO2 has been described as oligoribonuclease in mitochondrial RNA and DNA turnover, where it shows exonuclease activity for small (≤ 5 nts) oligomers and appears to have similar in function to its bacterial homolog, the ORN nuclease (Bruni et al., 2013; Nguyen et al., 2000).

3.1.12 *Drosophila* testis and ovarian development

To characterize the role of NEF-sp in gonad development we used *Drosophila melanogaster* as a model system. Figure 3.7 illustrates the developmental stages of germ cell development in the context of the anatomical structures of female and male gonads of *Drosophila*.

Figure 3.7 *Drosophila melanogaster* testis and ovary development. (A) *Drosophila* testis development: Adult testes form a coiled tube with the seminal vesicle at the basal end and the apical tip forming the hub with the germline stem cell (GSC) niche. Spermatogenesis occurs throughout larval and adult life, and all stages of differentiation are represented in a single adult testis. GSCs divide asymmetrically to self-renew themselves and produce a spermatogonium. The gonialblast undergoes four rounds of mitosis with incomplete cytokinesis and is encapsulated by two somatic cyst cells. The resulting 16-cell cyst enters two meiotic divisions with incomplete cytokinesis yielding 64 interconnected haploid round spermatids. During terminal spermatid differentiation, major morphological changes take place: spermatids elongate, nuclei condense and migrate to the posterior end; and actin-based cones form an individualization complex that drives expulsion of the cytoplasmic contents into cystic bulges, which grow in size as they increasingly collect more material during their travel away from the nuclei towards the spermatid tail. The thereby formed waste bag is subsequently degraded (Steller, 2008; White-Cooper, 2010). Confocal image shows actin staining (Phalloidin, red), DNA (Hoechst, blue), proteoasome (b5-GFP, green). Figure courtesy Sigi Benjamin. (B) *Drosophila* ovary development: Egg-chamber development begins in the germarium. At the anterior tip, specialized somatic follicle cells create a niche that supports 2-3 GSCs. The GSCs divide asymmetrically to produce two daughter cells, a cystoblast and a daughter stem cell that remains in the niche. The cystoblast goes through four rounds of mitosis that give rise to 16 germline cells; these remain connected as a result of incomplete cytokinesis. Of the 16 cystoblast cells, one differentiates into the oocyte and enters meiosis, while the other 15 become polyploid nurse cells. Follicle cells then surround the 16-cell cyst and the oocyte moves into the most posterior position. At this point, the nascent egg chamber buds off the germarium to form stage 1. In stages 2–7 nurse cells undergo polyploidization, follicle cells go through mitotic division and the size of the egg chamber increases. As the egg chamber continues to grow, it is pushed further posterior within the ovary as successive egg chambers bud off from the germarium. Egg chambers remain connected to each other by chains of stalk cells. Polar cells, specialized follicle cells, start to differentiate at each end of the egg chamber. They stop dividing soon after they become specified, while all other follicle cells continue to divide until stage 6. The oocyte begins to accumulate yolk at stage 8. By stage 10, the oocyte occupies half the total volume of the egg chamber. The nurse cells nourish the oocyte by transferring cytoplasm. In the final few hours of oogenesis, the nurse cells undergo apoptosis and the follicle cells produce the eggshell. Figure adapted from (Montell, 2003).


A
hub stem cells
mitotic divisions
sperm cell
somatic cyst cell
meiotic divisions
resulting in 64 cell cyst
elongating
sperm cells
DAPI, b5-GFP, Phalloidin


B
GSCs
egg chamber
follicle cells
nurse cells
oocyte
Germarium
Stage 4
Stage 6
Stage 8
Stage 9
Stage 10

Drosophila melanogaster testis is a long tube that coils at the basal end around a seminal vesicle. The stem cell niche is located at the apical tip of the testis, also called the hub (Figure 3.7 A). At the hub, germ stem cells (GSCs) divide asymmetrically to generate one cell that remains the hub and a daughter gonialblast that is displaced from the niche and differentiates. Each gonialblast is surrounded by two somatic cyst cells, which arise through asymmetric division from cyst stem cells. A gonialblast undergoes four rounds of mitotic divisions to produce a cluster of 16 spermatogonial cells. The mitotic divisions are synchronized and cytokinesis incomplete, resulting in 16 cells interconnected through cytoplasmic bridges (ring canals). A cytoplasmic, germline-specific membranous structure, the fusome, branches throughout the connected mitotic cells and regulates mitotic divisions (stained by 1B1, probing for the adducing-like protein *hts*) (Lin et al., 1994). These 16 spermatogonial cells progress through premeiotic S phase and then undergo a prolonged G2 phase in which the cells grow substantially. Genes that are needed for the development of spermatocytes and spermatids are transcribed at this time (de Cuevas and Matunis, 2011). Two meiotic divisions finally lead to 64 round spermatids in incomplete cytokinesis, which enter terminal differentiation. During spermatid elongation, histones become replaced with protamines, the nuclei condense and transform into needle-shaped structures and migrate towards the seminal vesicle, where they form the spermatid head. Spermatid terminal differentiation begins as actin cones form around the nuclei. The actin cones move along the cyst from the nuclei along the length of the cyst removing excess cytoplasm and organelles into a cystic bulge. At the end of the individualization process, excess cellular material is deposited in a waste bag and each mature spermatid is enclosed by its own membrane (Fuller, 1998) (Figure 3.7 A). Male gonad development starts during embryogenesis. Mitotic divisions are completed at the end of 2nd instar larval development and wild type testes have undergone all four mitotic divisions forming the 16-cell cyst (Sheng et al., 2009). Importantly, transcription stops at the end of meiosis and posttranscriptional regulation, including RNA processing, nuclear RNA export,

translation and RNA decay become critical for controlling protein levels during subsequent stages of sperm development both in mice and *Drosophila* (Bettegowda and Wilkinson, 2010).

Each *Drosophila* ovary is composed of 12 to 16 ovarioles. In each ovariole, stem cells are located in the germarium, which forms the most apical structure, followed by the developing egg chambers, which are arranged in a linear fashion with the most mature egg chamber at the most distal end (Eliazer and Buszczak, 2011; Kirilly and Xie, 2007) (Figure 3.7 B). The germarium contains three types of stem cells, germline stem cells (GSCs), somatic stem cells (SSCs) and escort stem cells (ESCs). GSCs undergo asymmetric self-renewing divisions; the daughter is displaced away from the niche and differentiates. The newly formed cystoblast undergoes four mitotic divisions with incomplete cytokinesis to form an interconnected 16-cell cyst. Differentiated ESCs wrap around the cystoblasts after they move away from the niche and fully encase them until the 16-cell stage. At the 16-cell stage, ESCs are replaced by somatic epithelial follicle cells, and the 16-cell cystoblast buds off the germarium as individual egg chamber (stage 1) (Montell, 2003). The follicle cells are produced by 2-3 SSCs located in the midway of the germarium. Stage 1 is reached relatively late during 2nd instar development. As a consequence 2nd instar female gonads are ~10-fold smaller than male gonads (Kerkis, 1931; Toomey et al., 2013). Of the 16 germline cells, one differentiates into the oocyte and enters meiosis, while the other 15 become polyploid nurse cells, which provide nutrients to the oocyte. The nurse cells nourish the oocyte by transferring cytoplasm. Stages 2–7 (between 5-7 days of larval development) are characterized by polyploidization of the nurse cells (through endocyclic divisions), mitotic division of the follicle cells and increases in the size of the egg chamber (Montell, 2003; Toomey et al., 2013; Williamson and Lehmann, 1996). The oocyte begins to accumulate yolk at stage 8, mostly provided by the cytoplasmic contents of the nurse cells. By stage 10, the oocyte occupies half the total volume of the egg chamber. In the final few hours of oogenesis, the nurse cells undergo apoptosis and the follicle cells produce the eggshell (Montell, 2003).

3.2 Methods

3.2.1 Fly stocks

Flies were kept at a 12-h light/dark cycle. All crosses were performed at 22-25°C, unless stated otherwise. The following stocks were used for this study (Bloomington Drosophila Stock Center (BDSC) and Vienna Drosophila Resource Center (VDRC) number (#) given in parentheses): PBac{PB}CG8368 (Exelixis at Harvard Medical School), CG8368^{M100} (BDSC #1306), CG8368 P{KK101144}VIE-260B (VDRC #108563), Dbp73D (DDX51, VDRC #108310), CG5033 (BOP1, BDSC #35020), CG42666 (REXO1, BDSC #33666), Nop5 (BDSC #55262), Rps3 (BDSC #31625), CG8414 (NOL9, BDSC #44435), white (BDSC #33623), PAN2 (BDSC #53249), Dis3 (VDRC #108013), Rrp6 (VDRC #108548), CG6833 (VDRC #104314), CG6937 (NIFK, BDSC 57467), Drosha (VDRC #108026, knockdown viable), Rs1, (DDX27, BDSC #32363), pUf68 (BDSC #25951) CG12877 (VDRC #20265, knockdown viable). The sequencing strain (Dmel WGS sequence, BDSC #2057) was used as wild type control. Two genetic mutants PBac{RB}CG12877^{e00300} and CG12877^{f00666} were used for knockout analysis of CG12877 (REXO1). RpS2-GFP *CB02294*, RpL13A *CC01920*, RpL10Ab *CB02653* protein trap lines were a gift from A. Spradling (Buszczak et al., 2007) and were crossed into dNEF-sp^{C04255} and dNEF-sp^{M100} mutant backgrounds. UAS-GFP-RpL11 and UAS-RFP-RpL26 lines were a gift from P. DiMario (Rosby et al., 2009) and were crossed with a Tubulin-Gal4 (TubGal4) driver line. For additional genetic mapping of the *C04255* and *M100* mutants we used complementation mapping with the deficiency lines Df(3L)BSC411 (BDSC #24915, cross was lethal) and Df(3L)BSC410 (BDSC #24914, cross was viable) to confirm that the region causing lethality was encompassing the dNEF-sp genetic locus.

3.2.2 Genetic mapping of dNEF-sp mutants

Total DNA was isolated from homozygous mutant larvae and sequencing strain control larvae using the Roche genomic DNA extraction kit (Roche). To confirm the PiggyBac transposon insertion in *C04255,* PCR fragments were amplified with specific primers within the PiggyBac transposon and within the coding region of dNEF-sp. To identify point mutations in *M100,* the entire coding region was PCR amplified and PCR products were sent for DNA sequencing (Genewiz) and analyzed in SeqMan Pro (LaserGene).

3.2.3 Generation of transgenic dNEF-sp rescue lines

For the generation of the dNEF-sp pUAST-attB (FBmc0003002) constructs the following primers were used for amplification of the dNEF-sp ORF: forward 5'-GGAATTGGGAATTCATGAAGGAACATATGTCCACCAA; reverse 5'-GAGCCGCGGCCGCCTAACTTTCCATAGTCTGATTCGATC. For introduction of point mutations in the catalytic site following primers were used (and their reverse complement):

5'-CAGGAGAACATCGATGGTCGAGATTCCATTGAGGATTCGC;

5'- CGCAGTCCTATGTTCGGCGTTGCATGTGAAATGTGTCACACGGA;

5'-CGCAGTCCTATGTTCGGCGTTGATTGTGCAATGTGTCACACGGA;

5'-CAGGAGAACATCGATGGTGCAGATTCCATT GAGGATTCGC.

Two genomic constructs were cloned into the pattB plasmid: forward 6.5 kb: TGGGAATTCGTAACACATACCATCCATGTTG; forward 4.7 kb: 5'-TGGGAATTCGGCGTCATCGCTGAGATC; reverse: 5'-GAGCCGCGGCCGCTAGTGATCCTGACCAGGGCTT. Primers were designed so that they included the promoter region of both or one isoform, respectively. Both constructs were sufficient for rescue of lethality. The human NEF-sp ORF was amplified from human cDNA and cloned into pUAS-attB using following primers: hNEF-sp-pUAS-forward: 5'-

ATCCGAATTCATGGAGCCAGAGAGGGAAGGGA; hNEF-sp-pUAS-reverse: 5'-AGTGCGGCCGCTCACGAACACAGGCCTGGGCCA; GFP cDNA and FLAG/HA cDNA was PCR amplified and cloned with NdeI into a natural N-terminal NdeI restriction site in dNEF-sp. GFP-forward: 5'-AAACATATGGTGAGCAAGGGCGAGG; GFP-reverse: 5'-TGGACATATGAAGCTTCTTGTACA; FLAG/HA-forward: 5'-AAACATATGGACTACAAGGACGACGA; FLAG/HA-reverse: 5'-GCTTGGTGGACATATGTGATATCTGGTTCA. The cDNA product was inserted at the N-terminus of the cloned dNEF-sp genomic construct with NdeI restriction digest according to standard protocols. PCR amplified inserts were cloned into the attB vector using EcoRI and Not1 restriction digest according to standard protocols. pUAS-attB and pattB DNA constructs were sent for injection to Bestgene Inc., and using the phiC31 integrase transgenesis system (Bischof et al., 2007), and site-specifically integrated to generate transgenic flies at the 25C7 landing site in the attP40 strain.

3.2.4 Clonal analysis of RNAi clones

For the generation of RNAi clones UAS-shRNA-dNEF-sp/Cyo;Sb/TM6B males were crossed to yw hsFlp;Sco/Cyo;UAS#Red47a#1tub<+GFP<Gal4/TM6B females, and embryos were heat shocked after 24 hrs egg laying at 37°C for 1 hr. Non-TM6B 3rd instar larvae were dissected for clonal analysis.

3.2.5 *In vivo* shRNA knockdown

UAS-shRNAs were crossed to Sp/Cyo;TubGal4/TM6B. Second instar larvae were collected after 72 hrs and dissected for immunostaining and RNA in situ hybridization. UAS-GFP-RpL11/Cyo;TubGal4/TM6B was crossed to the shRNA lines above and 2nd instar larvae were further processed for RNA-sequencing or dissected for immunofluorescent staining as described.

3.2.6 Immunofluorescence and microscopy

Adult flies and 2nd or 3rd instar larvae were dissected in PBS at room temperature. Tissues were fixed in 4% paraformaldehyde/PBS solution for 30 min at room temperature and permeabilized with PBS/Triton-X100 0.2% for 30 min, followed by blocking with 5% normal goat serum in PBS/Triton-X100 0.2% for 30 min at room temperature. Primary antibodies were incubated overnight at 4°C in 5% normal goat serum in PBS/Triton-X100 0.2%, tissues were washed at least 3x 10 min in PBS/Triton-X100 0.2% before they were incubated for 1-2 hrs with AlexaFluor secondary antibodies (1:500), Hoechst dye (1:1000), and Alexa-Phalloidin dye (1:300) in 5% normal goat serum in PBS/Triton-X100 0.2%. Tissues were washed in PBS/Triton-X100 0.2% for 45 min and mounted with VectaShield mounting media. Images were taken with a Zeiss LSM 710 confocal microscope.

3.2.7 RNA *in situ* hybridization

Larvae were dissected in PBS and tissues fixed in 4% paraformaldehyde/PBS solution for 1 hr at 4°C. Tissues were permeabilized in PBS-Triton-X100 0.2% for 30 min before incubation with hybridization buffer (75 mM Tris-HCl pH 8.5, 50% (v/v) formamide, 1 M NaCl, 1 x Denhardt's reagent, 250 µg/ml Baker's yeast, 500 µg/ml salmon sperm DNA, 2.5 mM Chaps, 0.5% (v/v) Tween-20) for 30 min at room temperature. Alexa-647-fluor- or Alexa-488-conjugated LNA probes were added to the prehybridization solution at a final concentration of 20 nM and tissues were incubated overnight at 45 °C. After hybridization, samples were washed in wash buffer I (75 mM Tris-HCl pH 8.5, 50% (v/v) formamide, 250 mM NaCl, 0.1% (v/v) Tween-20) and wash buffer II (75 mM Tris-HCl pH 8.5, 50 mM NaCl, 0.1% (v/v) Tween-20) for 30 min each. Tissues were incubated with PBS-Triton-X100 0.2% and Hoechst dye (1:1000) for 30 min, washed 3x 10 min with PBS-Triton-X100, and mounted in VectaShield mounting media. Images were taken with a Zeiss LSM 710 confocal microscope. LNA probe sequences were based on the validated

Northern blot probes, synthesized and labeled in the laboratory by C. Bognanni (5= LNA-A, 7=LNA-G, 8=LNA-T, 6=amino-modifier-C6-PDA): 3'ETS (repetitive sequence in trailer): 5'-6T7TTTGGCTACTCTT7ATAAAA6; 3'ETS-2 (specific region in trailer): 5'-6AAATT7ATGACGAGCT7TTTG6; 28S: 5'- 6TCGAATCATCAA7CAAAGGATAAGC6; 18S: 5'- 6CAA7CATATAACTACT7GCAGG6; ITS1: 5'- 6CACCATTTTACTG7CATA-TATCAATTCCTTCAATAAATG6.

3.2.8 Cell culture methods

S2 cells: S2 cells were maintained at 25°C in Schneider's insect medium, supplemented with 10% fetal bovine serum, 100 U/ml penicillin, 100 µg/ml streptomycin in 15-cm cell culture dishes.

Sf9 cells: Sf9 cells were maintained at 25°C in supplemented Grace's Insect Medium (supplemented with 10% FBS, 0.1% Pluronic-F68, 100 U/ml penicillin, 100 µg/ml streptomycin) in spinner flasks. *HEK293 cells:* FlpIn T-REx HEK293 were maintained as described in Chapter 2 and (Spitzer et al., 2013). Expression of NEF-sp proteins was induced by supplementing medium with 1 mg/ml doxycycline for 16 hrs prior to any analysis.

3.2.9 Generation of stable cell lines

HEK293 cell lines: dNEF-sp, mNEF-sp and hNEF-sp were amplified from *Drosophila* total fly cDNA, mouse and human testis cDNA and cloned into pENTR4 using the BP clonase recombinase system. *Drosophila* dNEF-sp (CG8368) has a singular ORF, the ORFs of mNEF-sp (2610020H08Rik) isoform NM_028129.2 and hNEF-sp (LOC81691) isoform NM_030941.2 were amplified. The resulting pENTR4 vector was subsequently recombined into the pFRT/TO/FLAG/HA-DEST destination vectors using Gateway LR recombinase according to manufacturer's protocol (Invitrogen). Following primers were used: dNEF-sp-forward:

5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAAGGAACATATGTCCACCAA; dNEF-sp-reverse: 5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAACTTTCCA-TAGTCTGATTCGATC; mNEF-sp-forward: 5’-GGGGACAAGTTTGTACAAAAAAG-CAGGCTTCATGAAAACATTTCACTTCCCC; mNEF-sp-reverse: 5’-GGGGACCACT-TTGTACAAGAAAGCTGGGTCCTAGGCCAGGGTGCGACC; hNEF-sp-forward: 5’-GTACAAAAAAGCAGGCTTCATGGAGCCAGAGAGGGAAGGGACCGA; hNEF-sp reverse: 5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACGAACACAGGCCTGG-GCCAG. Cell lines were established according to the manufacturer’s protocol for the generation of stable cell lines using the FlpIn system (Invitrogen) (see Chapter 2).

3.2.10 T7 *in vitro* transcription and shRNA-mediated knockdown in S2 cells

Four primers were designed to amplify a 500 nts template, 2 forward and 2 reverse primers. To amplify ssRNA strands into each direction the forward primer contained the T7 promoter and the reverse lacked the T7 promoter sequence; this ensured transcription from the PCR template into one direction only. A 500 nt segment was selected within the ORF and amplified from 100 ng *Drosophila* cDNA in a 100 µl PCR reaction. For amplification of the sense strand the primer pairs [CG8368-sense-T7, CG8368-antisense] were used, for amplification of the antisense strand the primer pairs [CG8368-sense, CG8368-antisense-T7] were used. The PCR product was diluted 1:50 in water. 10 µl diluted PCR product were incubated in a 200 µl reaction with T7 RNA polymerase at 37°C for 24 hrs [T7 reaction mix: 128 µl H_2O, 20 µl T7 reaction buffer (400 mM Tris-HCl pH 7.9, 10 mM spermidine, 260 mM $MgCl_2$, Triton X-100 0.1%), 40 µl 5x NTP (A/C/G/UTP=5/5/8/2 mM), 1 µl DTT (1 M), 1 µl T7 RNA polymerase (1U/µl)]. The T7 amplification product was purified by RNeasy Mini Kit (Qiagen). RNA strands were annealed by incubating sense and antisense strands at 500 nM final concentration each in 10 mM Tris-HCl pH 7.5, 20 mM NaCl solution, heating them for 1 min at 95°C and letting them cool down overnight

at room temperature. Primers used: CG8368-sense-T7: 5'-GCGTAATACGACTCACTATAGCAGCACGAGCGCAACGAGAAGA; CG8368-antisense-T7: 5'-GCGTAATACGACTCACTATAGCAGGTTCTGGATGTCGGTTAGAA; CG8368-sense: 5'-CAGCACGAGCGCAACGAGAAGA; CG8368-antisense: 5'-CAGGTTCTGGATGTCGGTTAGAA.

3.2.11 Western blot analysis

For Western blots on whole tissues approximately ~50-100 µg dissected tissues or total larvae/flies were lysed in lysis buffer [50 mM HEPES-KOH pH 7.4, 150 mM NaCl, 0.05% Triton-X100, complete EDTA-free protease inhibitor cocktail (Roche)] using a 1 ml tissue grinder. Lysates were cleared by centrifugation at >12,000 g for 10 min at 4°C. Protein concentrations of supernatants were determined by BCA assay (Pierce). 40 µg lysate was mixed with 4x sample buffer (200 mM Tris-HCl, pH 6.8, 200 mM dithiothreitol (DTT), 8% SDS, 24% glycerol, 0.04% bromphenol blue) and incubated for 2 min at 95°C. Samples were separated by SDS-PAGE for 1 hr at 30 mA per gel, using standard SDS Tris base-glycine running buffer. After electrophoresis, proteins were blotted onto nitrocellulose membranes (Hybond-ECL, GE Life Science), pre-wetted in transfer buffer (25 mM Tris base, 190 mM Glycine, 20% MeOH, 0.05% SDS), and semi-dry transferred (Bio-Rad) at 250 mA for 1 hr. Protein membranes were taken through a standard immunoblot protocol followed by enhanced chemiluminescent detection (Crescendo ECL, Millipore) using a Lumimager (Fuji, LAS-3000). For Western blots on HEK293, S2 or Sf9 cells lysates were similarly processed either in Triton-X100 buffer [50 mM HEPES-NaOH pH 7.4, 150 mM NaCl, 0.05% Triton-X100, complete EDTA-free protease inhibitor cocktail (Roche)] or NP-40 buffer [50 mM HEPES-KOH pH 7.2, 150 mM KCl, 0.5% (v/v) NP-40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)].

3.2.12 Immunoprecipitations

Immunoprecipitations were carried out for Western blot analysis and mass spectrometric identification of protein interaction partners. For a typical FLAG immunoprecipitation 1-5 ml cell pellet of HEK293 or S2 cells was lysed in 3 volumes of lysis buffer [S2 lysis buffer: 50 mM HEPES-NaOH pH 7.4, 150 mM NaCl, 0.05% Triton-X, complete EDTA-free protease inhibitor cocktail (Roche), HEK293 lysis buffer: 50 mM HEPES-KOH pH 7.2, 150 mM KCl, 0.5% (v/v) NP-40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)] and incubated for 30 min on ice. Lysates were cleared by centrifugation at 12,000 g at 4°C for 30 min. Per 5 ml cell lysate 50 µl magnetic Dynabeads Protein G coupled (Invitrogen) were affinity-conjugated to 25 µl of FLAG antibody (1 mg/ml, M2, Sigma-Aldrich) in 200 µl PBS-Tween-20 0.1% at room temperature for 30 min. The beads were washed 2x in PBS-Tween-20 0.1% and incubated with the lysate on a rotating wheel for 1 hr at 4°C. If the immunoprecipitation experiment was intended for mass spectrometry, prior to immunoprecipitation lysates were RNase digested with RNase T1 at a final concentration of 1 U/µl RNase T1 (Fermentas) at 22°C for 15 min. After immunoprecipitation, samples intended for Western blot analysis were washed 3x in wash buffer (lysis buffer minus detergent). Samples submitted for mass spectrometry were washed 3x in IP wash buffer [50 mM HEPES-KOH pH 7.5, 300 mM KCl, 0.05% (v/v) NP-40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)] and taken through a second RNase T1 digest. In the second RNase T1 digest beads were resuspended in one bead volume (50 µl) IP wash buffer and incubated with RNase T1 to a final concentration of 100 U/µl at 22°C for 15 min. After the second digest beads were washed 2x in IP wash buffer and affinity-bound FLAG/HA-NEF-sp protein was eluted with FLAG-peptide in a 200x FLAG peptide elution reaction (60 µg FLAG peptide per 50 µl beads, Sigma-Aldrich). FLAG peptide was added to IP wash buffer in a final elution volume of 60 µl, and FLAG/HA protein was eluted on a rotating wheel at 4°C for 1 hr. The supernatant was transferred, mixed with 20 µl 4x SDS loading dye

(200 mM Tris-HCl, pH 6.8, 200 mM dithiothreitol (DTT), 8% SDS, 24% glycerol, 0.04% bromophenol blue), and incubated for 2 min at 90°C. 5 µl of the elution mixture were used for Western blot analysis. For mass spectrometric analysis 40 µl of the elution was separated on a Nupage 4-12% Bis-Tris SDS-PAGE (Invitrogen), and the gel stained with Colloidal Blue (Invitrogen) according to the manufacturer's instructions. For Western blot analysis beads were resuspended with 50 µl SDS loading dye and incubated at 90°C for 2 min. 5 µl of supernatant was loaded onto an SDS polyacrylamide gel and separated by gel electrophoresis.

3.2.13 Antibodies

Immunofluorescent staining: The following antibodies were used (dilutions noted in parentheses): mouse monoclonal anti-Hts (1:20, 1B1) and rat monoclonal anti-VASA (1:20) (Developmental Studies Hybridoma Bank, Iowa), rabbit monoclonal anti-Fibrillarin (1:500, EPR10823(B), abcam), rabbit polyclonal anti-phospho-Histone H3 (Ser 10) (1:50, Upstate), rabbit polyclonal anti-phospho-H2Av pS137 (1:500, Rockland), Alexa-488, Alexa-546, and Alexa-647 (Molecular Probes) fluorescence-conjugated secondary antibodies were used at 1:500 dilution.

Western blot analysis: Following primary antibodies were used (dilutions noted in parentheses): rabbit polyclonal anti-dNEF-sp (1:1000 affinity-purified serum), rabbit monoclonal anti-Fibrillarin (1:1000, EPR10823(B), abcam), mouse monoclonal anti-Tubulin (DM1A) (1:1000, Sigma-Aldrich), mouse monoclonal anti-HA (1:1000, HA-7, Sigma-Aldrich, 1:1000, Covance, MMS-101P). Following secondary antibodies were used: HRP conjugated polyclonal goat anti-Armenian hamster heavy and light chain conjugated to horseradish peroxidase (HRP) (1:5000, Jackson ImmunoResearch), polyclonal goat anti-mouse light chain conjugated to HRP (1:2000, 1 mg/ml, Southern Biotechnology Associates, Fisher), polyclonal goat anti-mouse and anti-rabbit conjugated to HRP (1:5000, Dako).

3.2.14 Development of monoclonal antibodies

3.2.14.1 Standard ELISA

Sf9 baculoviral purified mNEF-sp protein (denatured) was used in standard ELISAs; as negative control we used FMRP protein, which was also purified from Sf9 cells [gift of M. Ascano (Ascano et al., 2012b)]. The antigen was diluted in 5.2 ml binding buffer (final concentration 1 µg/ml antigen, but concentration for general assays ranges from 0.1 µg/ml-100 µg/ml depending on antigen, sensitivity, purity, etc.). Binding buffer: 0.1 M $NaHCO_3$ pH 8 (8 ml 0.2M Na_2CO_3, 17 ml 0.2M $NaHCO_3$, 25 ml H_2O). Per well 50 µl antigen solution was plated with a multi-channel pipette into protein-binding plates designed for ELISA (Nunc Maxisorb Immuno plates). Plates were sealed and incubated at 4°C overnight. The residual binding solution was flicked out and dried on paper towels, rinsed 3x with wash buffer (PBS-Tween 0.1%) and filled with 100-150 µl blocking buffer (PBS-Tween 0.1%, 1% BSA, 0.2% azide). The plates were incubated at room temperature for 1 hr or overnight at 4°C. Blocking solution was removed and 50-100 µl antibody solution (1:10-1:10,000 of bleeds, for final screening bleeds were diluted 1:100 as determined by previous test ELISAs, and 1:10 dilution of hybridoma supernatants in blocking buffer) for 1-2 hr at room temperature. As controls following was included: (a) sera from terminal bleed (positive control), (b) known reactive monoclonal antibody (mouse anti-FLAG, M2, Sigma) (positive control), (c) negative pre-immune bleed (negative control), (d) fusion blank (negative control), (e) HT fusion media negative control. Both the antigen of interest and the control antigen were screened under these conditions. Plates were rinsed off with wash buffer 3x and incubated with 50 µl secondary antibody solution (antibody in 1:5000 dilution in blocking buffer) per well for 1 hr at room temperature (goat anti-Armenian hamster heavy and light chain conjugated to horse radish peroxidase (HRP), Jackson ImmunoResearch, goat anti-mouse κ-chain conjugated to HRP (1 mg/ml), Southern Biotechnology Associates, Fisher). Secondary antibody was flicked off and the plates rinsed 4x with wash buffer. The liquid was flicked off and the plates dried on paper

towels. For ELISA development 50 µl TMB substrate solution was added (Turbo TMB, BioFx, Sci-Quest) and the plates were incubated for 2-10 min. 50 µl 2N Sulfuric acid was added to stop the reaction. Plates were read on a SpectraMax 190 microplate reader (Molecular Devices) at 450 nM and analyzed using the Softmax Pro software (Molecular Devices).

3.2.14.2 Sandwich ELISAs

This assay was done to screen supernatants of hybridoma clones, identifying clones that recognized native, folded mNEF-sp protein. Since mNEF-sp was insoluble and could only be purified in denaturing conditions from bacterial and Sf9 baculoviral expression systems, but was soluble in HEK293 cells, we needed this assay to screen hybridomas at a high throughput scale to detect clones that recognized the native, folded protein, so that these clones could also be used in immunoprecipitation assays. Blocking buffer and wash solution were the same as in standard ELISA conditions. The capture antibody used for coating was diluted at 1 µg/ml (can range from 1-5 µg/ml per assay) in 5.2 ml binding buffer. The capture antibody should come from a species other than the immunized animal; in this case we used rabbit anti-FLAG monoclonal antibody (F7425, Sigma). Coating and blocking was performed as in standard ELISA. For antigen capturing, final concentration of antigen should range between 100 ng/ml to 20 µg/ml total protein depending on the purity of sample. In our case we used total HEK293 lysate expressing doxycycline induced mNEF-sp protein and as negative control we coated plates with uninduced HEK293 lysate. 50 µl FLAG/HA-mNEF-sp HEK293 cell lysate (5 µg/µl lysate concentration) was added per well and the supernatant incubated for 1 hr at room temperature. Supernatants were flicked out and the wells washed 3x with wash buffer. In the primary incubation step plates were incubated with 50 µl supernatant solution (1:100 bleeds, 1:10 hybridoma supernatants, same controls were used as in standard ELISA) per well for 1-2 hr at room temperature. Supernatants were flicked off the plates, plates were washed 3x with wash buffer, and dried on paper towels.

Secondary antibodies are usually used in 0.3-1µg/ml or a 1:3000 dilution of polyclonal anti-sera, diluted in PBS-Tween 0.1% (not in blocking solution as presence of azide may interfere with the enzymatic activity of the peroxidase). To detect binding of the hybridomas to mNEF-sp, we selected a secondary antibody detecting Armenian hamster heavy and light chain (goat anti-Armenian hamster heavy and light chain conjugated to horseradish peroxidase (HRP), Jackson ImmunoResearch, goat anti-mouse κ-chain conjugated to HRP (1 mg/ml), Southern Biotechnology Associates, Fisher). In our case all secondary antibodies were used in 1:3000 dilution in PBS-Tween 0.1%. 50 µl of antibody solution was added to each well and incubated for 1 hr at room temperature. The secondary antibody solution was removed by flicking off the supernatant onto paper towels, plates were washed 3x in PBS-Tween 0.1%, dried and incubated with 50 µl Turbo TMB solution (Turbo TMB, BioFx, Sci-Quest) per well for 10 min at room temperature. After 10 min the reaction was quenched with 50 µl 2N sulfuric acid. Plates were read in a SpectraMax 190 microplate reader (Molecular Devices) at 450 nm and the data analyzed with the Softmax Pro software (Molecular Devices).

3.2.14.3 Immunoprecipitations with hybridoma supernatants

In these immunoprecipitations the goal was to determine which hybridoma clone supernatant was able to immunoprecipitate mNEF-sp protein from doxycycline induced HEK293 cell lines. A mixture of protein A/G agarose beads was used for the conjugation to Armenian hamster hybridoma fusions as hamster IgGs show higher binding affinity to protein A versus G (F. Weiss-Garcia personal communication and online information by ThermoScientific on Protein A/G binding characteristics). For each hybridoma supernatant 50 µl A/G beads (Pierce, ThermoScientific) were used per immunoprecipitation. Beads were washed 2x with 1 ml PBS-Tween 0.1% and collected by centrifugation at 1,000 g for 1 min. Beads were resuspended in 450 µl doxycycline induced FLAG/HA-mNEF-sp HEK293 lysate (10 µg/µl, measured by BCA

assay) and 150 µl hybridoma supernatant or 5 µl pre- or terminal bleeds were added to the bead lysate. The lysate was incubated on a rotating wheel for 2 hrs at 4°C. As negative control beads with lysate minus hybridoma supernatants and beads incubated with blank fusion were used. As positive control mouse monoclonal anti-FLAG antibody was used in 1:100 dilution (1 µg/ml, M2, Sigma-Aldrich), and 1:100 dilution of terminal bleed (5 µl). After immunoprecipitation the beads were washed 5x with 1 ml PBS-Tween, and 40 µl 2x SDS buffer was added to yield a final volume of ~60 µl. The suspension was incubated for 2 min at 95°C. The supernatant was resolved by SDS-PAGE and Western blot analysis was carried out as previously described, probing for the presence of FLAG/HA-mNEF-sp protein with mouse monoclonal anti-HA antibody (1:1000, 16B12, Covance). For probing against the mouse monoclonal FLAG antibody polyclonal goat anti-light chain mouse Peroxidase-AffiniPure antibody was used (1:2000, Jackson Immunoresearch lab), for probing against the hamster hybridoma fusions goat anti-Armenian hamster heavy and light chain conjugated to horseradish peroxidase (HRP) was used (Jackson ImmunoResearch, goat anti-mouse κ-chain conjugated to HRP (1 mg/ml), Southern Biotechnology Associates, Fisher). Signal was detected by enhanced chemiluminescent detection (Crescendo ECL, Millipore) using a Lumimager (Fuji, LAS-3000).

3.2.14.4 Validation of clones by Western blot analysis

In this final screen we tested the supernatant of hybridoma clones, which were positive in all previous tests, to see whether they also could recognize mNEF-sp in Western blot analysis. 50 µg doxycycline induced mNEF-sp HEK293 lysate (5 µg/µl lysate concentration as determined by BCA assay), as well as uninduced HEK293 control lysate, was separated on SDS-PAGE and transferred by Western blot onto a nitrocellulose membrane as previously described. Membranes were stained by Ponceau S stain (Sigma-Aldrich) to assess equal transfer, and cut into pieces such that one induced and one induced HEK293 lysate lane was side-by-side. Membranes were taken

through a standard Western blot analysis protocol as previously described. As primary antibody solution hybridoma supernatants were diluted 1:2 in PBS-Tween 0.1% and incubated for 1 hr at room temperature. As positive control one strip was incubated with monoclonal mouse anti-HA (1:1000, HA-7, Sigma-Aldrich), and with terminal bleed (1:100). Secondary polyclonal goat anti-Armenian hamster conjugated to HRP (1:2000, Jackson Immuno Research) was used for detecting signal in the hybridoma bleeds, and secondary polyclonal goat anti-mouse HRP-conjugated antibody (1:5000, Dako) was used to probe against the HA monoclonal antibody positive control. Secondary antibodies were incubated for 1 hr at room temperature, membranes washed in PBS-Tween 0.1% for 30 min and signal detected by enhanced chemiluminescent detection (Crescendo ECL, Millipore) using a Lumimager (Fuji, LAS-3000).

3.2.15 Development of polyclonal antibodies

Full-length 6xHis-tagged-dNEF-sp protein was purified from baculoviral Sf9 expression as described. Polyclonal antisera were generated in two rabbits TX2078 and TX2079 (Covance). For Western blot analysis affinity-purified TX2079 serum was used (1:1000).

3.2.16 Protein quantification

Protein concentration of cell lysates was quantified using bicinchoninic acid assay (BCA assay, Pierce), and using a bovine serum albumin (BSA) standard curve (1, 2, 5, 10, 15, 20 µg BSA). Absorbances were measured at A_{562}. Protein concentrations of recombinantly purified proteins were assessed by in-gel quantification in reference to a BSA standard curve.

3.2.17 Recombinant protein purification from *E. coli*

mNEF-sp was insoluble and degraded in BL21-CodonPlus(DE3)-RIL. Full-length protein could only be co-purified under denaturing conditions with smaller degradation products. dNEF-sp and

hNEF-sp proteins were unstable in bacteria and gave no full-length products. mNEF-sp protein was purified from bacteria solely for antibody generation and injected into Armenian hamsters. For injection the protein could be denatured and had to be stored glycerol-free, and after dialysis it was snap frozen and stored at -80°C prior to injection. mNEF-sp was cloned into the HisSUMO-pET28a bacterial expression vector (Novagen). The plasmids containing the DNA inserts of interest were transformed into *E. coli* strain BL21- CodonPlus(DE3)-RIL (Stratagene) grown in Luria-Bertrani (LB) medium supplemented with 30 mg/mL kanamycin. BL21(DE3)-RIL expressing pET28a-mNFE-sp protein were grown to OD_{280} 0.5 at 37°C before IPTG induction at 1 mM final concentration. Induced BL21(DE3) were grown overnight at 37°C before cells were harvested and the cell pellet flash frozen and stored at -80°C. Cells were lysed in lysis buffer (100 mM Tris HCl pH 8, 500 mM NaCl, 5 mM $MgCl_2$, 10 mM imidazole, 2 mM 2-mercaptoethanol, 50 µg/ml lysozyme) for 30 min on ice and sonicated. The lysate was cleared by centrifugation at 12,000 g for 30 min at 4°C. The supernatant was discarded, the pellet resuspended in loading buffer (100 mM Tris HCl pH 8, 500 mM NaCl, 5 mM $MgCl_2$, 10 mM imidazole, 8M urea) and batch-purified, incubating the solution with 2 mL Cobalt resin (HisPur Cobalt IMAC resin, ThermoScientific) for 1 hr at 4°C. The resin was washed with 12x column volumes lysis buffer, 6x column volumes wash buffer I (100 mM Tris HCl pH 8, 500 mM NaCl, 5 mM $MgCl_2$, 30 mM imidazole, 8M urea), 6x column volumes wash buffer II (100 mM Tris HCl pH 8, 500 mM NaCl, 5 mM $MgCl_2$, 70 mM imidazole, 8M urea) and eluted in 6 column volumes elution buffer (100 mM Tris HCl pH 8, 500 mM NaCl, 5 mM $MgCl_2$, 150 mM imidazole, 8M urea). The protein was dialyzed in 1 L dialysis buffer (100 mM Tris HCl pH 8, 500 mM NaCl, 5 mM $MgCl_2$, 10 mM imidazole), snap frozen and stored at -80°C.

3.2.18 Recombinant protein purification from Sf9 cells

PCR-amplified ORFs of the *Drosophila* wild type and the DADAH dNEF-sp transgenic mutant constructs (see section 3.2.3) were recombined by BP recombinase into pENTR4 and then recombined into the pDEST10 baculoviral expression vector using the Gateway LR recombinase. Baculoviral production and amplification was performed as described in the Bac-to-Bac Manual (Invitrogen). Sf9 cells were used for recombinant virus production, amplification and expression of recombinant dNEF-sp proteins. Recombinant viruses were amplified to ~2-10^8 plaque-forming units per ml titer. 1 × 10^9 Sf9 cells were infected at a multiplicity of infection (MOI) of 5 and placed back into fresh Supplemented Grace's Insect Medium (supplemented with 10% FBS, 0.1% Pluronic-F68, 100 U/ml penicillin, 100 µg/ml streptomycin) into a spinner flask at a final concentration of 1 × 10^6 cells per ml medium for 3–4 days prior to collection. 5 ml cell pellets were lysed in lysis buffer [50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 10 mM imidazole, 10% glycerol, 0.1% Triton-X100, 1x EDTA-free Protease Inhibitor Cocktail (Roche)] for 30 min at 4°C. The lysate was cleared by centrifugation at 19,000 g for 30 min at 4°C and filtered through a 5-µm membrane (Pall). The supernatant was incubated with 2 ml Cobalt resin (HisPur Cobalt IMAC resin, ThermoScientific) on a rotating wheel for 2 hrs at 4°C. The resin was washed with 12 column volumes wash buffer-10 (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 10 mM imidazole, 10% glycerol) and 6 column volumes wash buffer-30 (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 30 mM imidazole, 10% glycerol). The protein was eluted in 6 column volumes with elution buffer-150 (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 150 mM imidazole, 10% glycerol). Fractions were flash frozen in 50% glycerol and stored at -80°C.

3.2.19 Transient protein expression in S2 cells

ORFs of HA/FLAG-dNEF-sp wild type and HA/FLAG-dNEF-sp^{DEDAH} catalytic dead mutant were cloned into pAc5.1/V5-His A (addgene). The dNEF-sp ORF was amplified from the pUAS-attB injection constructs in a 2-step PCR introducing an N-terminal HA/FLAG tag into both mutant and wild type proteins. (FLAG-dNEF-sp-forward: 5'-GACTACAAGGACGACGATGACAAGATGAAGGAACATATGTCCAC; HA/FLAG-forward: 5'-TCAGAATTCATGTACCCTTATGACGTGCCCGATTACGCTGACTACAAGGAC; dNEF-sp-reverse: 5'-TGCGGCCGCCTAACTTTCCATAGTCTGATTCGATCCG.) The ORF was double digested with EcoRI and NotI following standard cloning procedures. Cells were grown in 15 cm dishes. At ~80% S2 cell confluency 15 µg HA/FLAG-dNEF-sp-wt-pAc5.1 or Ha/FLAG-dNEF-sp^{DEDAH}-pAc5.1 was transfected with 30 µl TransIT-Insect transfection reagent (Mirus) in 1.5 ml serum-free medium according to the manufacturer's instructions. Cells were harvested after 48 hrs and expression was verified by Western blot analysis probing against the HA antigen.

3.2.20 RNA extraction and cDNA preparation

50 µg cells or 2nd instar larvae or dissected ovaries and testes where directly homogenized in 1 ml TRIzol (Life Technologies) and total RNA was isolated according to the manufacturer's protocol. Poly(dT) amplified cDNA was prepared from 5 µg total RNA, using oligo(dT) amplification and the Superscript III First Strand synthesis kit following the manufacturer's instructions (Invitrogen).

3.2.21 RNA and DNA phenol/chloroform extraction and ethanol precipitation

One volume of phenol/chloroform/isoamyl alcohol (25:24:1, phenol buffered at pH 4.3 for RNA, phenol buffered at pH 7.4 for DNA) was added to one volume of an RNA/DNA solution and vortexed vigorously for 15 sec. The phases were separated by centrifugation at 12,000 g at 4°C

for 2 min. The upper aqueous layer was removed to a new tube and re-extracted with an equal volume of ice-cold chloroform. The mixture was vortexed for 15 sec and the phases were separated by centrifugation, 12,000 g at 4°C for 2 min. The upper aqueous layer was removed and added to a new tube, and the salt concentration adjusted to a final of 0.3 M NaCl. 3 volumes of 100% EtOH were added to the solution, mixed thoroughly, and stored at -20°C for 10 min (long RNAs/DNAs) or 1 hr/overnight (short RNAs/DNAs). RNA/DNA was precipitated by centrifugation at 12,000 g at 4°C for 30 min. The supernatant was removed and the pellet dissolved in 10-50 µl water.

3.2.22 NEF-sp PAR-CLIP

For small-scale PAR-CLIP tests in HEK293 cells, five 15-cm plates were grown and processed as previously described in Chapter 2. After protein separation by SDS-PAGE, instead of gel electroporation, crosslinked RNA-protein complexes were transferred onto a nitrocellulose membrane (HighBond ECL, GE Healthcare Life Sciences) in a semi dry transfer system with 1x transfer buffer (25 mM Tris base, 190 mM Glycine, 20% MeOH, 0.05% SDS) at 250 mA for 1 hr. The membrane was exposed for 1 hr on a phosphorimaging screen. For PAR-CLIP in S2 cells, pellets expressing dNEF-sp wt and mutant and untransfected control cells were lysed in 3 volumes of Triton-X100 lysis buffer [50 mM HEPES-NaOH pH 7.4, 150 mM NaCl, 0.05% Triton-X, complete EDTA-free protease inhibitor cocktail (Roche)] for 30 min on ice. Lysates were cleared by centrifugation at >10,000 g for 30 min at 4°C. The remaining part of the protocol was followed as previously described.

3.2.23 *In vitro* RNA nuclease assays

For 5' end radiolabeling of oligonucleotides, 100 pmoles of RNA or DNA oligonucleotides were incubated with 5 pmoles of γ-^{32}P ATP and 10 units of T4 PNK (NEB) in 1 x T4 PNK buffer (70

mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT) in a 20 µl reaction for 15 min at 37°C; followed by addition of 1,000 pmoles of non-radiolabelled ATP and incubation for another 5 min. To generate circular RNAs, ssRNAs were labeled as described, phenol/chloroform extracted and ethanol precipitated, and self-ligated by incubating them with T4 Rnl1 (0.2 µg/µl) in 50 µl of buffer containing 50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 0.2 mM ATP, 0.1 mg/ml acetylated BSA, 15% DMSO for 1 hr at 37°C. All reactions were stopped by addition of 1 volume of stop buffer (8 M urea, 10 mM EDTA, bromophenol blue). Labeled oligoribonucleotides were separated on a 18% polyacrylamide/ 8M urea gel at 28 W for 1 hr, eluted with 400 µl 0.3 M NaCl solution overnight, and ethanol precipitated. Exonuclease assays were performed in 50 mM KOAc, 20 mM Tris-OAc pH 7.9, 10 mM MgOAc, 1 mM DTT (NEB buffer 4) in the presence of 1 pmol 5'-^{32}P-labeled oligomers and 20 or 25 nM recombinant protein in 50 µl total reaction volume. Samples were incubated at 25°C and 10 µl aliquots were taken at 1, 5, 10 and 30 min and mixed directly with stop mix. We used commercially available bacterial RNase T (ExoT, NEB) (0.2U) and laboratory-purified (20 nM) C3PO [gift from M. Ascano (Tian et al., 2011)] as exonucleolytic and endonucleolytic enzyme control. Samples were loaded on a 18% polyacrylamide/ 8M urea gel, run for 1 hr at 28 W, and exposed on a phosphorimaging screen.

3.2.24 Northern Blot analysis

5 µg of total RNA was heated for 5 min at 65°C, snap cooled and separated the RNA on a 1.2% formaldehyde agarose gel in MOPS running buffer (200 mM MOPS, 50 mM NaOAc, 10 mM EDTA) for 3 hrs at 75 V. As single-stranded RNA size marker we used Ambion Millenium RNA marker. The gel was incubated for 30 min in an alkaline solution (50 mM NaOH, 10 mM NaCl) and neutralized for 15 min in 0.1 M Tris-HCl pH 7.0. RNA was transferred to a Zeta-Probe GT membrane in 20x SSC (Na_3Citrate $2H_2O$, 3M NaCl) in a downward-blotting procedure for 1.5

hrs. Transfer efficiency was assessed with a methylene blue solution (0.04%, methylene blue hydrate, 0.5 M NaOAc). Membranes were UV-crosslinked and incubated with 20 ml hybridization buffer (5x SSC, 1% SDS, Dextran sulfate 10 %, formamide 25 %, 250 µg/ml Baker's yeast RNA) for 1 hr at 65°C. 30 pmol oligonucleotide probes were 5'end labeled as previously described and purified using an Illustra MicroSpin G-25 column. Radioactive probes were added to the prehybridized membrane solution for 1 hr at 65°C, and then incubated overnight at 37°C. The membrane was washed with 2x SSC/1% SDS and 0.2x SSC/1% SDS solutions for 30 min each at 37°C before it was exposed on a phophorimaging screen for 1 hr. Membranes were stripped in 0.1x SSC/0.5% SDS at 85°C for 1 hr and reprobed for all probes. Following oligonucleotide probes were used: 5'ETS: 5'-TTCGAACAATGCGAGGTCGGCAA; ITS1 (Zhang et al., 2014): 5'-CACCATTTTACTGGCATATATCAATTCCTTCAATAAATG; 3'ETS-repeat: 5'- TGTTTGGCTACTCTTGATAAAA; 3'ETS-1: 5'-AAATTGATGACGAGC TGTTTG; 28Sa: 5'-ACTTAGGACCGACTAACTCGTGA; 28Sb: 5'-TCGAATCATCAAGC AAAGGATAAGC; 18S: 5'-CAAGCATATAACTACTGGCAGG; 7SL: 5'-TGGAAGGTT GGCAGCTTCTGTAATCA.

3.2.25 RNA extraction and Illumina total RNA and mRNA-seq

2nd instar larvae or dissected ovaries and testes were directly homogenized in 1 ml TRIzol with a tissue grinder and total RNA was isolated according to the manufacturer's protocol. poly(A) purification and total RNA cDNA library construction was performed using the TruSeq version 1.5 kit (Illumina). cDNA was barcoded using the Illumina Multiplexing Sample Preparation Oligonucleotide kit and analyzed an Illumina HiSeq 2000 in a 100-base-pair (bp) single-end sequencing run.

3.2.26 Hydro-seq

19-nt 19.39 (5'-CGUACGCGGGUUUAAACGA) and 24-nt 24.60 (5'-CGUACGCGGAAUAGUUUAAACUGU) oligo size markers were ^{32}P-5'end-labeled with radioactive ATP by PNK as described. The RNA size markers contain a PmeI restriction endonuclease recognition site (underlined). After PCR-amplification the cDNA libraries were digested with PmeI to avoid sequencing of the size markers. Briefly, size markers were individually radiolabeled in a 10 µl reaction containing 1 µM RNA, 10 U T4 polynucleotide kinase and 50 µCi γ -^{32}P-ATP (6,000 Ci/mmol) at 37 ºC for 15 min. The labeled size markers were separated on a 15% polyacrylamide/8M urea gel, gel extracted in 0.3 M NaCl solution at 4°C overnight, and ethanol precipitated. RNA was redissolved in 10 µl water and 19-nt and 24-nt marker were combined 1:100 in water. A known amount of calibrator RNA was added to the experiments to follow the success of the library preparation. The calibrator oligoribonucleotides have no match to the human or mouse genome. 0.5 fmol each of the ten following calibrator oligoribonucleotides were added to 2 µg of total RNA. The preparation of a calibrator cocktail requires the use of carrier oligonucleotide to prevent surface adsorption during preparation of the dilution series in the nanomolar concentration range (11-nt oligodeoxynucleotide 11.6, 5'-TCGAAGTATTC). Following calibrator RNA sequences were used (p, 5'-monophosphate): *01:* 5'-pGUCCCACUCCGUAGAUCUGUUC; *02:* 5'-pGAUGUAACGAGUUGGAAUGCAA; *03:* 5'-pUAGCAUAUCGAGCCUGAGAACA; *04:* 5'-pCAUCGGUCGAACUUAUGUGAAA; *05:* 5'-pGAAGCACAUUCGCACAUCAUAU; *06:* 5'-pUCUUAACCCGGACCAGAAACUA; *07:* 5'-pAGGUUCCGGAUAAGUAAGAGCC; *08:* 5'-pUAACUCCUUAAGCGAAUCUCGC; *09:* 5'-pAAAGUAGCAUCCGAAAUACGGA; *10:* pUGAUACGGAUGUUAUACGCAGC.

7 ml of carrier solution containing 500 nM 11-nt carrier of DNA oligo 11.6 was prepared in water. (The carrier is necessary to prevent surface adsorption during dilution and storage of low concentrations of calibrator RNA oligos.) 50 µl of a calibrator cocktail containing 1 µM of each

calibrator RNA oligo was prepared. The calibrator cocktail was diluted 1:10 in carrier solution resulting in a calibrator concentration of 0.1 µM each (50 µl calibrator solution plus 450 µl of carrier). The calibrator cocktail was diluted further 1:100 in carrier solution resulting in a final RNA oligo calibrator concentration of 1 nM each. Following pre-adenylated 3' adapter sequences were used (L, 3' aminolinker blocking group; rApp, 5' terminal adenosine residue connected via a 5', 5'-diphosphate bridge to the 5'OH of the 5' nucleotide, which activates the adapter for ligation): *3' pre-adenylated RNA adapter:* rApp**NNNNN**TCGTATGCCGTCTTCTGCTTG-L, where **NNNNN**: *26.75,* TCACT; *26.76,* TCATC; *26.77,* TCCAC; *26.78,* TCCGT; *26.79,* TCCTA; *26.80,* TCGAT; *26.81,* TCGCG; *26.82,* TCTAG; *26.83,* TCTCC; *26.84,* TCTGA; *26.85,* TTAAG; *26.89,* TAACG; *26.90,* TAATA; *26.91,* TAGAG; *26.92,* TAGGA; *26.93,* TATCA; *26.94,* TGATG; *26.95,* TGTGT; *26.96,* TTACA; *26.98,* TTGGT. The adapters each contain a unique pentamer barcode sequence at the 5' end (bold and underlined) and were preadenlyated according to a preadenylation protocol established in the laboratory. The 5' RNA adapter *26.68* had following sequence: 5'-GUUCAGAGUUCUACAGUCCGACGAUC. The primers for amplification of the barcoded cDNA library used were the following: 5' adapter primer *44.32:* 5'-AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA; 3' adapter primer *21.929*: 4'-CAAGCAGAAGACGGCATACGA.

Alkaline Hydrolysis: 0.3 µg of total RNA was subjected to partial alkaline hydrolysis in a mixture of 10 mM Na_2CO_3 and 10 mM $NaHCO_3$ for 5 min at 90°C.

Dephosphorylation*:* Fragmented RNA was dephosphorylated with 10 U calf intestinal alkaline phosphatase (CIP) in dephosphorylation buffer (20 mM Tris-OAc pH 7.9, 50 mM KOAc, 10 mM MgOAc, 100 µg/ml BSA) for 1 hr at 37°C, phenol chloroform extracted and ethanol precipitated.

Re-phosphorylation: RNA was re-phosphorylated with T4 polynucleotide kinase (PNK) in reaction buffer (70 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT) for 1 hr at 37°C, phenol chloroform extracted and ethanol precipitated. The RNA was redissolved in 8.5 µl water.

Subsequent steps were identical to a small RNA library preparation protocol previously described in Hafner *et al.* (Hafner et al., 2012) and are in detail described below.

3' adapter ligation: Re-phosphorylated RNA (in 8.5 µl water) was incubated with 8.6 µl 3'-ligation mix [2 µl of 10x RNA ligase buffer without ATP (0.5 M Tris-HCl, pH 7.6, 0.1 M $MgCl_2$, 0.1 M 2-mercaptoethanol, 1 mg/ml acetylated BSA (Sigma)], 6 µl 50% aqueous DMSO, 0.5 µl of 1 nM of each calibrator cocktail and 0.1 µl of 1:100 dilution of the 5'-^{32}P-labeled length marker oligoribonucleotide mix), and 2 µl 3'adapter (different adapters for different samples), and incubated at 90ºC for 1 min to denature any RNA secondary structures. The samples were cooled down on ice, 1 µl Rnl2(1-249)K227Q (1 µg/µl) added and incubated on ice at 4ºC overnight. As ligation control 2 µl of the 1:100 diluted length marker was mixed with 6.5 µl of water and processed separately. To precipitate and stop the 3'ligation reaction 3x the total volume of the combined 3'-adapter ligation reactions of absolute ethanol was added sequentially to each tube and the reaction transferred until all samples were pooled. Ethanol addition ensured stopping the reaction by deactivation of Rnl2(1-249)K227Q. To the pooled samples a 3 M NaCl solution was added in a 1/10 of the total volume to achieve a final 0.3 M NaCl concentration. The reaction was precipitated at -20ºC for 1 hr or overnight, and the RNA pellet collected by centrifugation in a tabletop centrifuge at 4ºC at maximum speed (approx. 14,000 g) for 30 min. The supernatant was discarded and the pellet dried by vacuum centrifugation. The RNA pellet was dissolved in 15 µl water per pooled library, mixed with 15 µl formamide stop mix (98.8% formamide, 1% (v/v) 0.5 M Na_2 H_2EDTA, pH 8.0, 0.2% Bromophenol blue), and heated at 90ºC for 1 min to denature RNA secondary structures. The samples were loaded on a 15% polyacrylamide/8M urea gel and separated at 28 W for 45 min. The gel was exposed on a phosphorimaging screen to visualize the ligation product. The ligated product (at ligated size marker control length) was excised from the gel, the ligated RNAs eluted from the gel in 400 µl 0.3 M NaCl solution by incubating the tube overnight at 4ºC under constant agitation (on the thermomixer shaking at 11,000 rpm). The RNA

was ethanol precipitated as previously described. The RNA pellet was vacuum dried and dissolved in 9 µl water.

5' adapter ligation: To the ligated product 9 µl of 5' ligation mix was added (1 µl of 100 µM 5'-adapter 26.68, 2 µl of 10x RNA ligase buffer with ATP and 6 µl 50% aqueous DMSO), the RNA mixture was incubated at 90ºC for 1 min to denature RNA secondary structures and the reaction immediately cooled on ice for 2 min. 2 µl of T4 RNA ligase 1 (Rnl1) (1 µg/µl, Fermentas) was added, the reaction mixed gently and incubated for 1 hr at 37 ºC. To stop the reaction 20 µl of denaturing formamide stop mix was added and the samples incubated for 1 min at 90ºC. Samples were directly loaded onto a 12% polyacrylamide/8M urea gel and separated at 28W for 45 min. The gel was exposed on a phosphorimaging screen and gel pieces of the size of the 5' ligation product (as assessed by the ligation control) were excised. The RNA was eluted in 400 µl 0.3 M NaCl solution (approximately 3x the gel piece volume), 1 µl of 100 µM 3'-primer 21.929 was added as carrier and the ligated RNAs eluted from the gel by incubating the solution overnight at 4ºC under constant agitation (on the thermomixer shaking at 11000 rpm). The RNA was ethanol precipitated, the pellet collected by centrifugation and air dried.

Reverse transcription: Reverse transcription was performed using the SuperScript III reverse transcriptase (Invitrogen) reaction kit. The RNA pellet was dissolved in 5.6 µl water. The control length marker ligation product was carried forward as control. The RNA was denatured by incubating the tube for 30 sec at 90ºC and transferring the tube to a 50ºC incubator. 8.7 µl of the RT reaction mix were added (1.5 µl 0.1 M DTT, 3 µl 5x first-strand buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM $MgCl_2$, 100 mM DTT, buffer provided by the manufacturer), and 4.2 µl 10 mM each dNTPs) to each sample and incubated for 3 min at 50ºC. 0.75 µl of Superscript III reverse transcriptase (Invitrogen) was added and the reaction incubated for 30 min at 42ºC. After the RT reaction, the RNA template was hydrolyzed by adding 40 µl of 150 mM KOH/20 mM Tris base and incubated for 10 min at 90ºC. The solution was neutralized by addition of 40

µl of 150 mM HCl and the pH adjusted to a range of 7.5 (monitored on pH paper). The pH solution should be slightly alkaline to not inhibit the subsequent PCR amplification.

PCR amplification: 10 µl of the cDNA solution, 0.5 µM of each primer (0.5 µl of each 100 µM primer 21.929 and 44.32), 10 µl 10x dNTP mix (2 mM each dNTP), 10 µl 10x PCR buffer (100 mM Tris-HCl, pH 8.0, 500 mM KCl, 1% Triton-X100, 20 mM $MgCl_2$, 10 mM 2-Mercapthoethanol), 68 µl water and 1 µl of Taq DNA polymerase (5 U/µl) were mixed to perform a standard 100 µl PCR with Taq polymerase. As control a no-template control PCR reaction was performed with H_2O to check for DNA contamination in the reaction mixture. Enough master mix was prepared to have four 100 µl PCR reactions per sample (pilot PCR and large-scale PCR). For the cDNA amplification following PCR conditions were chosen: 45 sec at 94ºC denaturation step, 85 sec at 50ºC annealing, 60 s at 72ºC polymerase extension. To determine the number of cycles for the final library amplification a 100 µl pilot PCR was performed. 10 µl aliquots were removed every other cycle following cycle number 10 until cycle 28 by temporarily putting the PCR cycler on hold at the end of the 72ºC step. The PCR aliquots were mixed with 5x DNA loading dye (0.2% bromophenol blue, 0.2% xylene cyanol FF, 50mM EDTA pH 8, 20% Ficoll type 400) and analyzed on a 2.5% agarose gel in 0.5x TBE buffer for 2 hrs at 180 V. The optimal cycle number was determined by choosing the cycle number approximately 5 cycles below the PCR saturation threshold. Limiting the amplification cycles to the exponential phase was important to minimize sequence-specific distortions of some sequences amplifying better than others and leading to distortions in the RNA sequence profiles, commonly referred to as clonal amplification. After determining the optimal cycle, 3 large scale PCR reactions were carried out per sample, the PCR products pooled, and a 5 µl aliquot removed for analysis on a 2.5% agarose gel. The remaining product was phenol/chloroform extracted and ethanol precipitated by adding 30 µl 3 M NaCl solution, adding 1 volume of neutral Phenol/chloroform/IAA and vortexing the mixture for 30 sec. The phases were separated by

centrifugation at 14,000 g for 2 min. The aqueous supernatant was transferred to a new tube and re-extracted with 1 volume of chloroform to remove residual phenol. The mixture was vortexed for 30 sec and separated by centrifugation at 14,000 g for 2 min. The aqueous supernatant was transferred to a new tube and ethanol precipitated by adding 3 volumes of 100% ethanol and incubating the solution at -20ºC for 1 hr or overnight. The DNA pellet was collected by centrifugation at 14,000 g for 30 min at 4ºC. The ethanol supernatant was removed. The DNA pellet was not air dried to prevent denaturation of DNA and directly dissolved in 20 µl PmeI reaction mix. Denaturation and subsequent re-annealing of a complex sequence pool will result in imperfect rehybridization and formation of DNA duplexes with internal bulges that might compromise PmeI digestion.

PmeI digest of size marker sequences: In a final step the 19-nt and 24-nt size markers in the samples were PmeI digested to not lead to overrepresentation of size markers in the sequencing reaction. Per reaction following PmeI digest mix was prepared: 2 µl 10x PmeI buffer (NEB Cutsmart buffer: 20 mM Tris-acetate pH 7.9, 50 mM Potassium Acetate, 10 mM Magnesium Acetate, 100 µg/ml BSA, 17.5 µl of water and 0.5 µl (5 U) of PmeI (NEB). The DNA pellet was directly dissolved in PmeI reaction mix and incubated at 37ºC for 2 hrs. The reaction mix was mixed with 20 µl 5x DNA loading dye, loaded onto a 2.5% agarose gel and run in 0.5x TBE buffer for 2 hrs at 180 V until 5'-adapter-3'-adpater ligation product could be sufficiently separated from the cDNA library. The cDNA library product was excised from the gel and gel purified using the QiaQuick gel purification kit (Qiagen). The final cDNA library was eluted in 30 µl water and submitted for Illumina sequencing.

3.2.27 RNA-seq data analysis

RNA-seq libraries were adapter extracted and aligned against the *Drosophila melanogaster* genome dm3 with BWA or tophat2 (for poly(A) and total RNA-seq) in standard settings (Li and

Durbin, 2009). Hydro-seq and total RNA-seq libraries were hierarchically aligned against separately created indexes of different RNA classes, according to their relative abundance in the cell: (1) rRNA, (2) tRNA, (3) snRNA/snRNA, (4) transposons, (5) miRNA and other short ncRNAs, (6) mRNAs. Library and transcript normalized rpkm count values were calculated separately for each alignment category and their correlation presented as scatterplots in R. Alignments against the full 47S rRNA precursor were done using the NCBI GenBank *M21017.1* reference transcript. Further downstream analysis was performed in R. Rpkm values for poly(A)-mRNA –seq were calculated using cufflinks2 (Trapnell et al., 2012a) and HTSeq (Anders et al., 2015). Differential gene expression was calculated with DEseq (Anders and Huber, 2010) and further downstream analysis was carried out in R. Gene Ontology [using the DAVID functional annotation database (Ashburner et al., 2000; Huang et al., 2008)] and GOrilla (Eden et al., 2009) analyses were performed using the cuffdiff output data, selecting genes with adjusted p-values of ≤ 0.05.

3.3 Results

3.3.1 The NEF-sp nuclease family has a unique structural domain organization in vertebrates

The NEF-sp nuclease family possesses one distant *S. cerevisiae* homolog, the RNA exonuclease Rex1p (RNH70). However, the homologous relationship between them is not close and the human REXO1 family represents the main homolog by sequence conservation and evolutionary distance. Examining the phylogenetic tree of DEDDh RNase T nucleases, NEF-sp proteins branch off as a distinct, newly evolved protein family (Figure 3.6). NEF-sp proteins contain one RNase T exonuclease domain, which is conserved across eukaryotes and gained at the vertebrate level two RNA-recognition motif domains (RRMs), which are absent in invertebrates and lower eukaryotes (Figure 3.8).

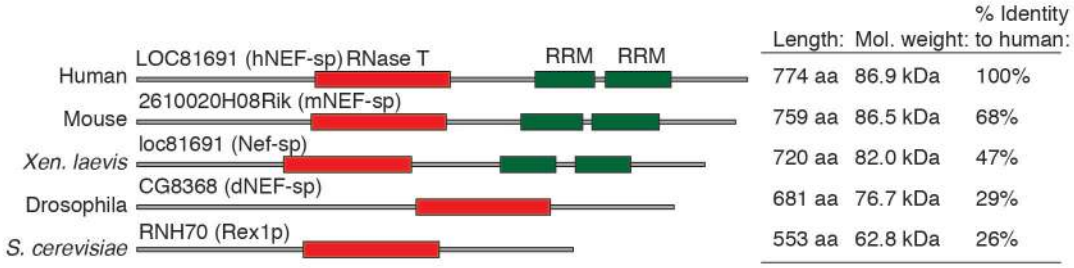


Figure 3.8 NEF-sp domain conservation across eukaryotes. Schematic domain organization of NEF-sp proteins scaled according to protein length for human NEF-sp, mouse, *Xenopus*, *Drosophila* and the *S. cerevisiae* homolog Rex1p. Protein length [in amino acids (aa)], predicted molecular weight (in kDa), and percentage identity to the human NEF-sp protein are shown.

This structural domain combination of a catalytic RNA exonuclease domain and two single-stranded RNA binding domains is unique across the genome (Gerstberger et al., 2014a). The recent evolutionary acquisition of two RRM domains in NEF-sp homologs in vertebrates might suggest an increasing requirement for affinity or specificity to recognize target RNAs.

To characterize the role of NEF-sp in RNA processing, we chose to investigate its function biochemically and genetically in *Drosophila melanogaster* as this provided a tractable

and well-established model system for gonad development. Due to the structural differences in NEF-sp proteins at the vertebrate level we also started initial characterizations of human NEF-sp in HEK293 cells and generated a monoclonal antibody against the mouse NEF-sp homolog for future genetic and biochemical studies. In the following chapters I will first describe the work performed in human cell lines on hNEF-sp and the generation of a monoclonal antibody against the mouse homolog mNEF-sp, before I describe the full characterization of *Drosophila* dNEF-sp and its *in vivo* phenotypes and targets.

3.3.2 Recombinant protein purification of NEF-sp and generation of stable cell lines

In order to study the *in vitro* RNase activities and to raise antibodies against NEF-sp proteins several overexpression and purification strategies were pursued to generate recombinant human, mouse and *Drosophila* NEF-sp proteins. We used bacterial and baculoviral expression in Sf9 insect cells (Army worm). The human isoform 1 of NEF-sp (NM_030941.2), mouse isoform 2 NEF-sp (NM_028129.2), and the *Drosophila* CG8368 ORFs were cloned into the bacterial expression vector pET28a-HisSUMO with an N-terminal cleavable HisSUMO tag and into the baculoviral pDEST10 N-terminal His6 vector using the Bac-to-Bac expression system (Invitrogen). Bacterial expression and purification of NEF-sp proteins was not successful for either human and *Drosophila* NEF-sp proteins and they were already degraded before lysis. Expression constructs were switched to Glutathione S-transferase (GST) N-terminal tags, and lengths of expression constructs varied, but none of the tested conditions produced stable protein, monitored by Western blot analysis and Commassie staining. Only mNEF-sp protein was stably expressed in bacteria, although degraded fragments of mNEF-sp protein were already observed before lysis and co-purified (Figure 3.9 A). mNEF-sp protein was insoluble and could only be purified under denaturing conditions. We switched to baculoviral expression of NEF-sp proteins in Sf9 cells. In this expression system mNEF-sp protein was expressed at full length, but also

showed unstable fragments before lysis (Figure 3.9 B). Insoluble in Sf9 cells, mNEF-sp could only be purified under denaturing conditions. The dialyzed mNEF-sp protein purified from bacterial and baculoviral expression was further used for injection into Armenian hamsters and screening for monoclonal antibodies against the mNEF-sp protein. Human NEF-sp protein was unstable in Sf9 cells *in vivo* and its purification was not further pursued (Figure 3.9 C).

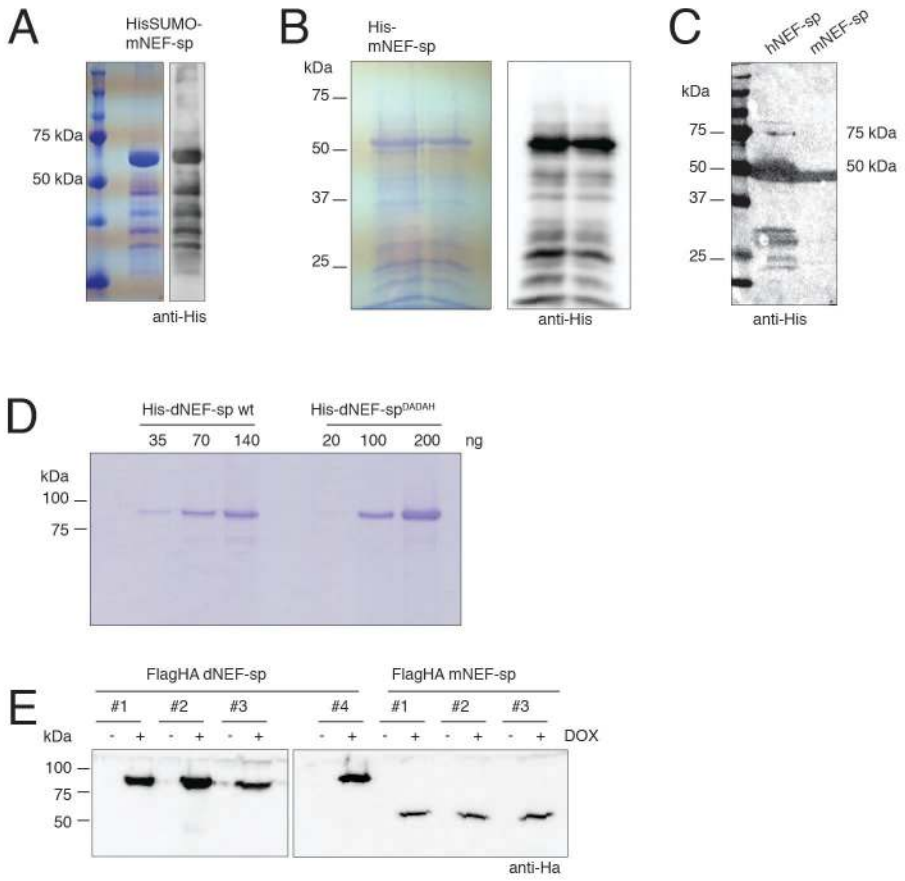


Figure 3.9 Recombinant expression of human, mouse and *Drosophila* NEF-sp. (A) Commassie gel and anti-His Western blot (anti-His) of recombinant HisSUMO-mNEF-sp protein expressed in BL21(DE3)-RIL *E. coli* strain. (B) Same as in (A) for recombinant His-mNEF-sp protein expressed with baculoviral expression in Sf9 cells. (C) Western blot analysis of Sf9 lysates with baculoviral expression of human NEF-sp and mNEF-sp protein. (D) Commassie gel of recombinant His-dNEF-sp protein purified from Sf9 cells. (E) Western blot (anti-HA) of HEK293 stable cell line clones expressing doxycycline-induced mNEF-sp and dNEF-sp protein.

Drosophila NEF-sp protein could be successfully expressed and purified under native conditions in Sf9 cells and displayed nuclease activity *in vitro* (Figure 3.9 D, Figure 3.23-3.24). Lastly, using the FlpIn system, stable HEK293 cell lines expressing doxycycline inducible N-terminally FLAG/HA tagged hNEF-sp, mNEF-sp, and dNEF-sp protein were successfully generated (Figure 3.9 E). These cell lines served as tool for testing and validating antibodies in the hybridoma screen, for subcellular localization studies, and for PAR-CLIP RNA-crosslinking experiments.

3.3.3 Pilot PAR-CLIP of hNEF-sp in HEK293 cells

Predominantly expressed in gonads, human NEF-sp is expressed at low levels in HEK293 cells. Using doxycycline-inducible overexpression of FLAG/HA protein we investigated the localization and RNA-binding activity of hNEF-sp. The human FLAG/HA-NEF-sp protein localized to nucleoli (Figure 3.10 A). Applying 4-SU crosslinking *in vivo* (PAR-CLIP), I assessed the RNA-crosslinking efficiency of hNEF-sp, visualizing the crosslinked and radioactively labeled RNA fragments by phosphorimaging. While a crosslinking band at the expected ~100 kDa could be detected, the intensity was very low and not high enough above background to pursue with a full cDNA library preparation (Figure 3.10 B).

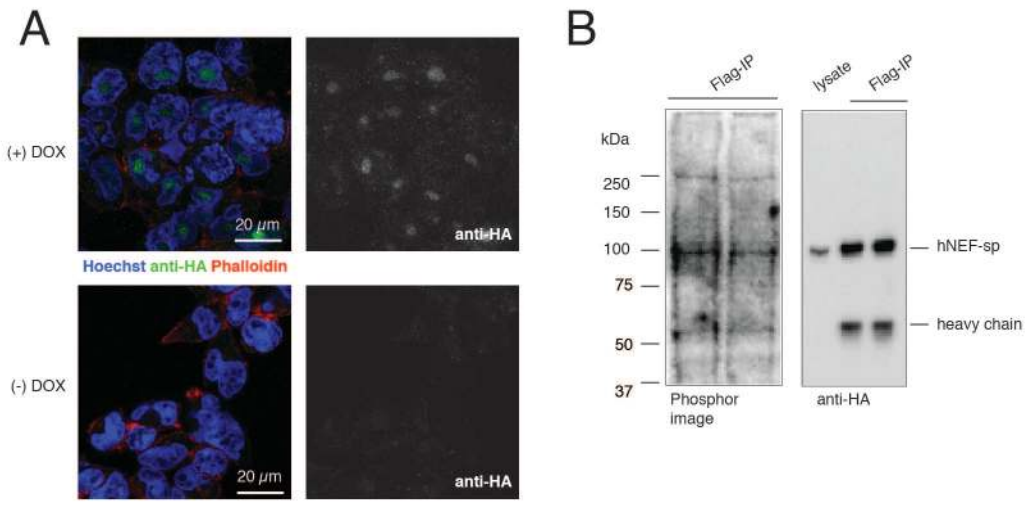


Figure 3.10 Localization and RNA-crosslinking of hNEF-sp. (A) Immunofluorescence of doxycycline-induced hNEF-sp, and non-induced HEK293 cells as control. hNEF-sp shows nucleolar localization. (B) Phosphor image and Western blot analysis of crosslinked, radiolabeled, and immunoprecipitated hNEF-sp in a small-scale PAR-CLIP.

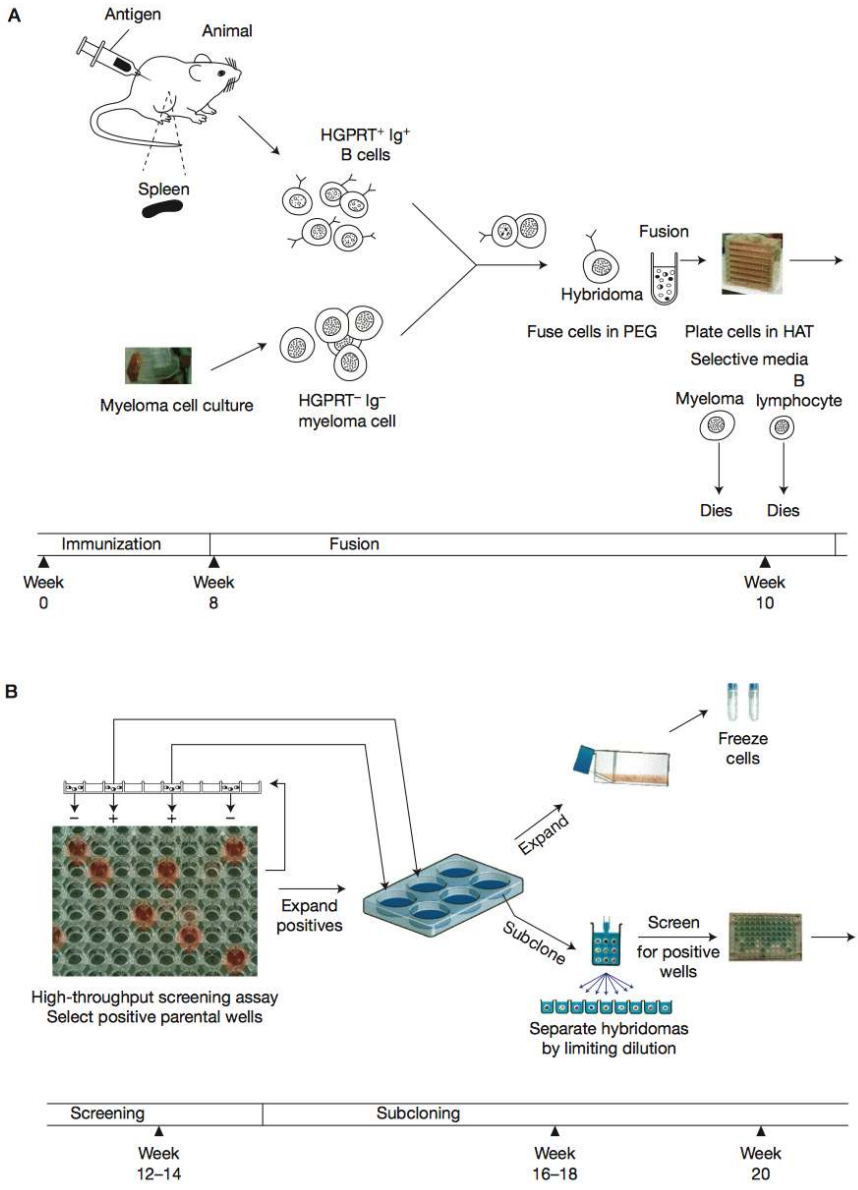
We reasoned that the observed low crosslinking efficiency may be due to either low abundance of the *in vivo* targets of NEF-sp in HEK293 cells, or because RNA interactions of hNEF-sp nuclease with its targets were highly transient or had low affinity and did not allow sufficient crosslinking intensity.

3.3.4 Generation and characterization of a monoclonal antibody against mNEF-sp

To study the role of mammalian NEF-sp proteins in gonad development one needs a suitable *in vivo* system in which knockouts can be generated and RNA targets easily isolated. Knockout mouse embryonic stem cells for mNEF-sp exist, generated by genome-wide gene knockout consortiums, but we needed high affinity biochemical reagents to isolate and characterize mNEF-sp *in vivo*. Thus, we decided to raise a monoclonal antibody against the mNEF-sp protein isoform 2 (RefSeq transcript ID NM_028129.2), which could immunoprecipitate native RNA-protein complexes of mNEF-sp. Monoclonal antibodies allow, in contrast to polyclonal antibodies,

selective, and epitope-specific immunoprecipitation of protein-RNA complexes *in vivo* from cells and tissues. Immunoprecipitation-grade antibodies are difficult to make and often researchers resort to high affinity tags of proteins to isolate their protein of interest. The advantage of an immunoprecipitation-grade monoclonal antibody is that it enables isolation of RBP complexes *in vivo* without relying on laborious knock-in affinity tags, which may affect target and protein complex affinity. This project was done in collaboration with the monoclonal antibody core facility at Memorial Sloan Kettering Research Center, who isolated antibody-secreting B cells from immunized animals and fused them with a myeloma cell line, a type of B-cell tumor. These hybrid cells or hybridomas can be maintained *in vitro* and will continue to secrete antibodies with a defined specificity. An overview of the general procedure for monoclonal antibody generation is shown in Figure 3.11 (Greenfield, 2014).

Figure 3.11 Outline of the stages in hybridoma production. (A) Animals are injected with an antigen preparation. Once a good humoral response (titer) has appeared in the immunized animal a screening procedure is developed and fusion started. The sera from test bleeds are used to develop and validate the screening procedure. Several days before the fusion, animals are boosted with a sample of the antigen. For the fusion, antibody-secreting cells are prepared from the immunized animal, mixed with the myeloma cells, and fused. After the fusion, cells are diluted in selective medium (hypoxanthine-aminopterin-thymidine medium (HAT), selection medium usually used for selection of immortal, fused hybridomas) and plated in multiwell tissue culture dishes. Hybridomas are ready to test beginning 1 week after the fusion. Cells from positive wells are expanded and then single-cell-cloned. (B) Hybridoma cells grow at approximately the same rate and tissue culture supernatants from all fusions are usually ready to screen within a few days of one another. This makes screening the most labor-intensive part of hybridoma production. Approximately one week after the fusion, colonies of hybrid cells are ready to screen. During the screening supernatants of growing hybridomas are tested for the presence of the desired antibodies. Successful fusions generally produce between 200-20,000 hybridoma colonies, on average ~1000 colonies. Typically, the first wells are ready to screen on day 7-8 and most of the wells need to be screened within 4-5 days. In our hybridoma generation we devised two primary screens per week, which were followed by secondary validations each subsequent day. Positive wells are expanded and frozen; selected colonies are chosen for further subcloning. Figure adapted from (Greenfield).


A
Antigen
Animal
Spleen
HGPRT+ Ig+
B cells
Myeloma cell culture
HGPRT− Ig−
myeloma cell
Fusion
Hybridoma
Fuse cells in PEG
Plate cells in HAT
Selective media
Myeloma
B
lymphocyte
Dies
Dies
Immunization
Fusion
Week
0
Week
8
Week
10
B
Freeze
cells
Expand
Expand
positives
Subclone
Screen
for positive
wells
High-throughput screening assay
Select positive parental wells
Separate hybridomas
by limiting dilution
Screening
Subcloning
Week
12–14
Week
16–18
Week
20

HisSUMO- and His-mNEF-sp protein purified from BL21(DE-3)RIL and baculoviral Sf9 cells was injected in four injections into Armenian hamsters to raise an immune response. Before starting the hybridoma fusions, I tested humoral responses of each hamster against mNEF-sp by standard and sandwich ELISA. In standard ELISA, response of all test bleeds was screened for baculoviral purified His-NEF-sp protein, indicating the highest response titer for hamster 3 (Figure 3.12 A). As negative control baculoviral purified human FMRP protein was used [a gift of M. Ascano (Ascano et al., 2012b)]. Since mNEF-sp protein could only be purified under denaturing conditions, we reasoned that standard ELISA may predominantly over-represent clones that recognized epitopes of the denatured protein. However, for IP-grade hybridoma antibodies, we needed to isolate clones, which also recognized the folded, native protein in an immunoprecipitation-like assay. Thus, as second screen we decided to use sandwich ELISAs to screen the test bleeds and select clones, which recognized FLAG-bound, HEK293 doxycycline-induced, native mNEF-sp protein. This assay also had the advantage that mNEF-sp had a different affinity tag than the injected protein, thus clones recognizing the His tag were excluded. By sandwich ELISAs hamster 2 test bleeds showed the highest response for mNEF-sp, and we decided to use hamster 2 for clonal fusion (Figure 3.12 B).

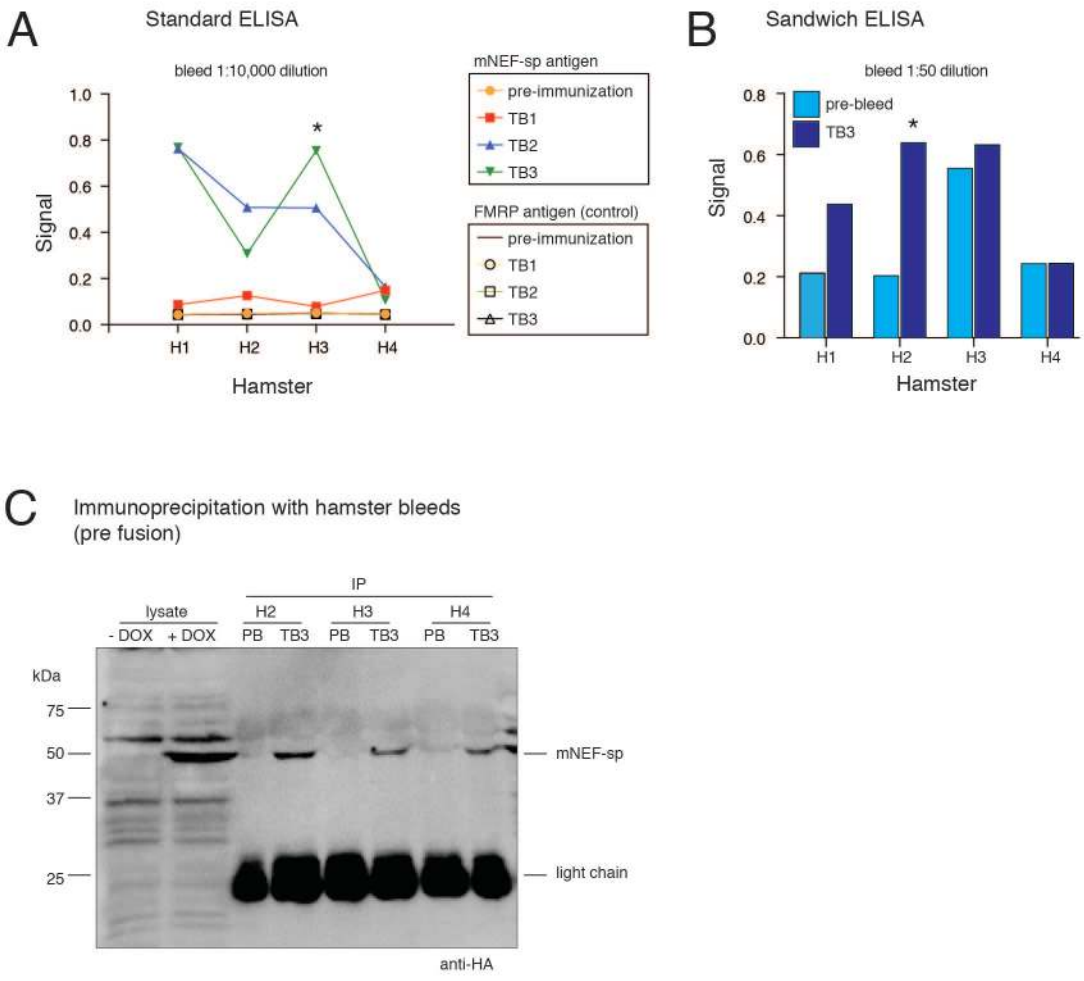


Figure 3.12 Results from different screens for mNEF-sp immunized hamsters. (A) Standard ELISA using Sf9 purified recombinant mNEF-sp and FMRP antigen as negative control. Bleeds were tested in serial dilution, shown here 1:10,000 dilution. Tested in ELISA with mNEF-sp: pre-bleed (orange), test bleed 1 (TB1, red), test bleed 2 (TB2, blue), test bleed 3 (TB3, green). Control with FMRP: pre-bleed (brown), TB1 (yellow), TB2 (light green), TB3 (black). (B) Sandwich ELISA: coated with rabbit monoclonal anti-FLAG, incubated with doxycycline induced mNEF-sp HEK293 cell lysate and incubated with 1:50 dilution test bleeds: TB3 (dark blue) and pre-bleed (light blue). (C) Western blot analysis (anti-HA) on immunoprecipitated FLAG/HA-mNEF-sp inducibly expressed in HEK293 cells. Shown are total lysate control (-) and (+) doxycycline induction in lane 1 and 2. Lane 3-8 show immunoprecipitations of FLAG/HA-mNEF-sp protein from doxycycline induced mNEF-sp HEK293 cells using test bleed TB3 and pre-bleed in 1:50 dilution. mNEF-sp isoform NM_028129.2 runs at ~50 kDa.

Test bleeds of all hamsters were able to immunoprecipitate FLAG/HA-mNEF-sp from HEK293 cells (Figure 3.12 C, shown here are hamsters 2-4).

For our screening strategy of hybridoma clones we developed a triage-screen (Figure 3.13). First, supernatants of all clones were tested in parallel in standard. Second, positive clones were selected for further validation for their immunoprecipitation quality by sandwich ELISAs and immunoprecipitation and supernatants were used to immunoprecipitate doxycycline-induced FLAG/HA-mNEF-sp protein from HEK293 cells. Immunoprecipitated FLAG/HA-mNEF-sp protein was detected by Western blot analysis against the HA peptide epitope. The hybridoma fusion cultures were tested from three separate pulls for their performance in the triage screen. Only the clones that tested positive by standard ELISA were selected for sandwich ELISA and immunoprecipitation.

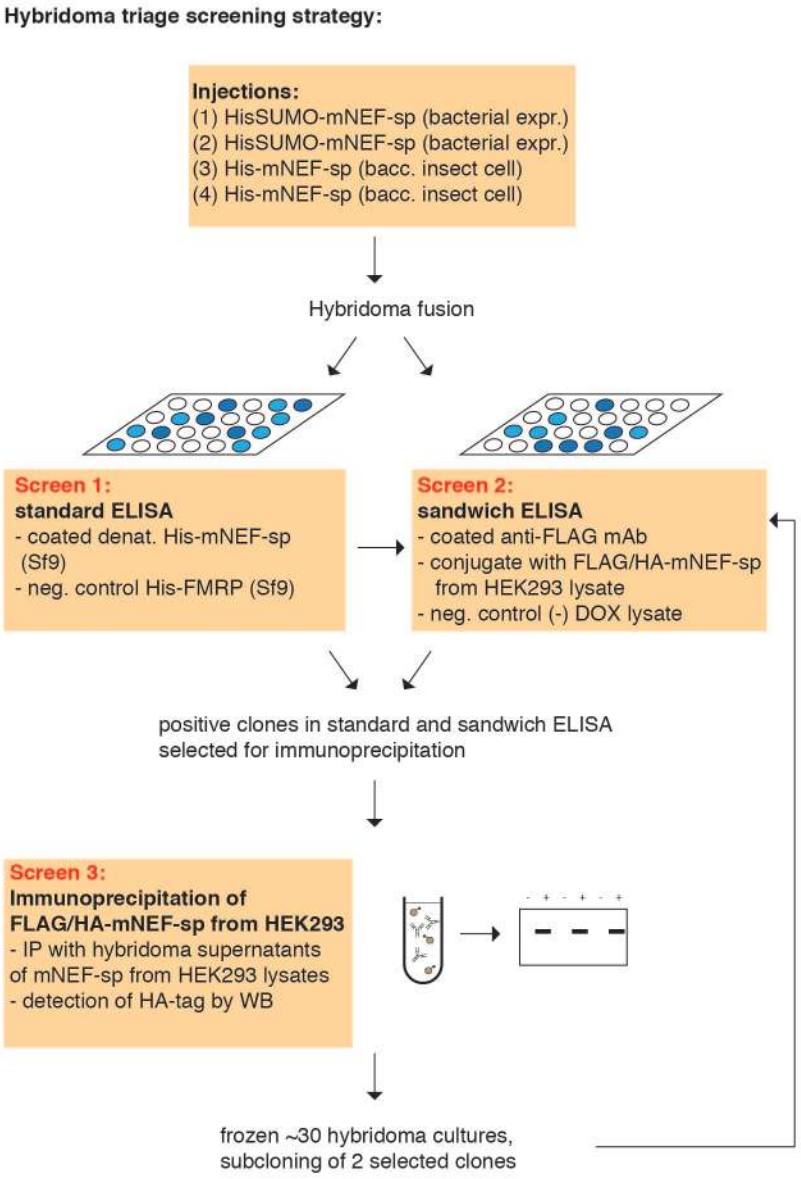


Figure 3.13 Hybridoma screening strategy. Bacterial and baculoviral purified mNEF-sp antigen was injected into four hamsters. In the pre-screening the hamster with highest titer response was selected. After a final boost with Sf9 purified mNEF-sp protein B cells were fused with myeloma cells to create stable hybridoma colonies. In screen 1 hybridoma cultures were tested by standard ELISA with denatured Sf9 purified mNEF-sp antigen. All positive hits were tested in screen 2 by sandwich ELISA of FLAG/HA-mNEF-sp protein and in screen 3 by immunoprecipitation with hybridoma supernatants conjugated to agarose-Protein A/G beads. After three repeats of the triage screen positive hybridoma cultures were frozen in liquid nitrogen and two hybridoma cultures further subcloned to isolate single clones. The subcloned colonies went through the same triage screen and were also validated by standard Western blot analysis.

Of the 2,376 hybridoma cultures (approximately 6,000 hybridomas), 161 tested positive in standard ELISA, 74 of those were only positive in standard ELISA and showed no signal in sandwich ELISA or immunoprecipitation, 23 tested positive in all three assays, 61 in standard ELISA and immunoprecipitation, and 3 in standard and sandwich ELISA but were negative in immunoprecipitations. Based on these results, we froze 27 hybridoma cultures and selected 2 for further subcloning (Table 3.1, Figure 3.14).

Table 3.2: Summary of screened hybridoma cultures.

Screen	Total tested
Cultures screened	2,376
Hybridomas	~6,000 (~2.5 per well)
Standard ELISA (+)	161
Standard ELISA (+), sandwich ELISA (+), IP (+)	23
Standard ELISA (+), sandwich ELISA (-), IP (-)	74
Standard ELISA (+), sandwich ELISA (+), IP (-)	3
Standard ELISA (+), sandwich ELISA (-), IP (+)	61
Frozen cultures/ pools	27
Subcloned	2

03D01 and 23F05 were selected for further subcloning, as these showed the highest immunoprecipitation quality, assessed by semi-quantitative Western blot analysis (Figure 3.14).

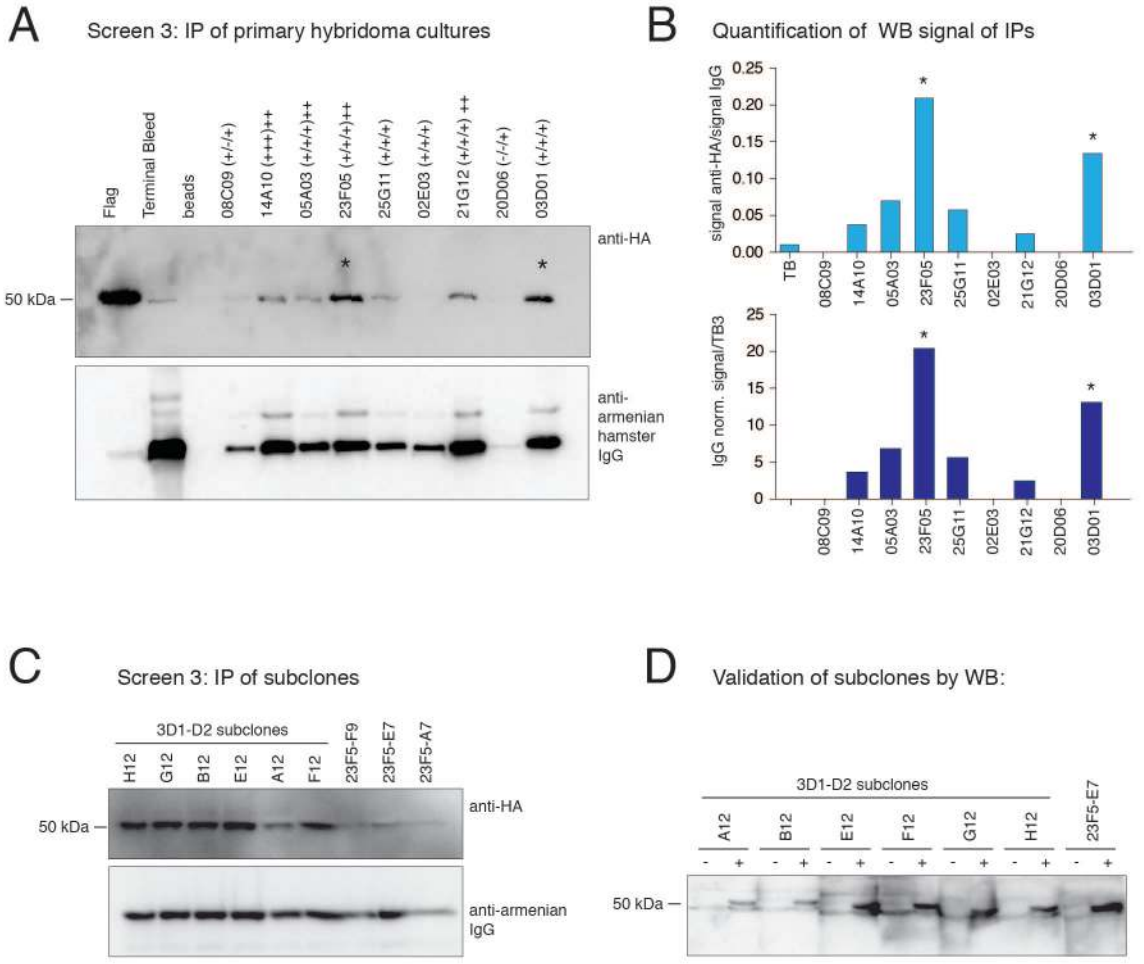


Figure 3.14 Selected results from hybridoma screening. (A) Immunoprecipitation screening results for a selection of hybridoma cultures from screen 3. Hybridoma supernatants are incubated with FLAG/HA mNEF-sp lysates in 1:10 dilution and conjugated to Protein A/G agarose beads. Shown are Western blots probing against HA peptide, to assess loading blots are reprobed for Protein A/G bound Armenian hamster IgG. The stars indicate the two clones (23F05 and 03D01), which were selected for subcloning. (+/-, +/-, +/-) indicates whether the clone tested previously positive in (1) standard ELISA, (2) sandwich ELISA, and (3) immunoprecipitation assay. The plus behind the brackets indicates how strong the overall response was. (B) (i) Semi-quantification of Western blot signal from (A) for immunoprecipitated mNEF-sp protein normalized over anti-Armenian hamster IgG signal (light blue), (ii) further normalization of the signal in (i) over the final test bleed response. Selected colonies for subcloning are marked with a star. (C) Same as (A) for a selection of subclones. (D) Validation of subclones recognizing FLAG/HA-mNEF-sp in Western blot analysis (hybridoma supernatants are used in dilution 1:10).

3.3.5 Generation and characterization of a polyclonal antibody against dNEF-sp

We decided to raise a polyclonal antibody against the *Drosophila melanogaster* dNEF-sp protein. This allowed quantification of the protein *in vivo*, but was more time efficient and since I had generated an *in vivo* FLAG/HA-dNEF-sp transgenic fly (see section 3.3.7) we did not rely on a dNEF-sp-specific immunoprecipitation-grade antibody. Recombinant dNEF-sp protein, purified from Sf9 cells, was sent for injection into two rabbits (Covance). Anti-sera responses against dNEF-sp were tested in HEK293 cells expressing doxycycline inducible FLAG/HA-dNEF-sp (Figure 3.15). Sera were highly potent and could be used in 1:20,000 dilution to specifically detect the antigen in 40 µg lysates (shown here are dilutions 1:5,000). To minimize background detection of other insect proteins, the polyclonal sera were further affinity-purified with recombinant dNEF-sp protein and the resulting eluate used in 1:1000 dilution for Western blot analysis of *Drosophila* lysates.

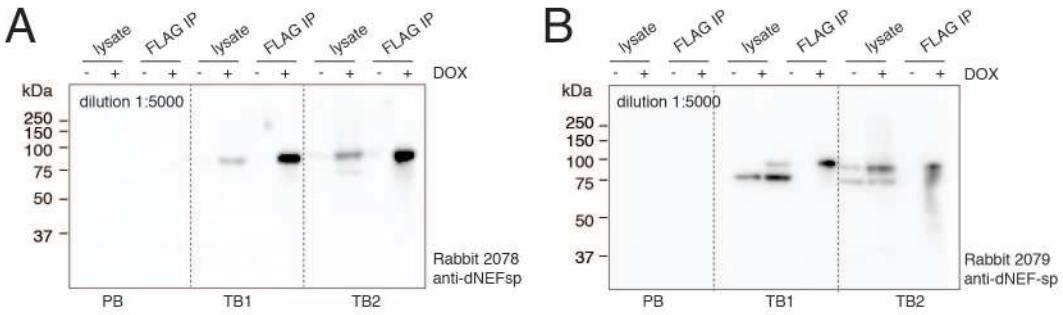


Figure 3.15 Characterization of a polyclonal antibody against dNEF-sp. Recombinant baculoviral Sf9 purified His-dNEF-sp protein was injected into two rabbits to generate a humoral response. Specificity of bleeds was validated with doxycycline-inducible dNEF-sp HEK293 cell lines. (A) Response of pre-bleed, test bleed 1 (TB1), and test bleed 2 (TB2) for rabbit 2078 tested on doxycycline-induced FLAG/HA-dNEF-sp HEK293 total cell lysates and FLAG-immunoprecipitated protein. (B) Same as in (A) for rabbit 2079.

3.3.6 Immunoprecipitation of the mNEF-sp homolog in testis

I tested the supernatants of the subcloned for immunoprecipitation of mNEF-sp protein from mouse liver and testis lysates. Unlike the known RefSeq isoform predictions, I detected a specific protein band at ~80-90 kDa in testis lysates. This putative testis-specific isoform was in the

expected size range of NEF-sp proteins from other organisms. mNEF-sp (official gene name 2610020H08Rik) has currently two annotated isoforms in the RefSeq NCBI annotation. Isoform 2 is predicted to be 479 amino acids long and did not have a predicted RRM domain by Pfam and SMART structural domain predictions, while isoform 1, 445 amino acids long, possessed one RRM domain instead of two (Figure 3.16 A).

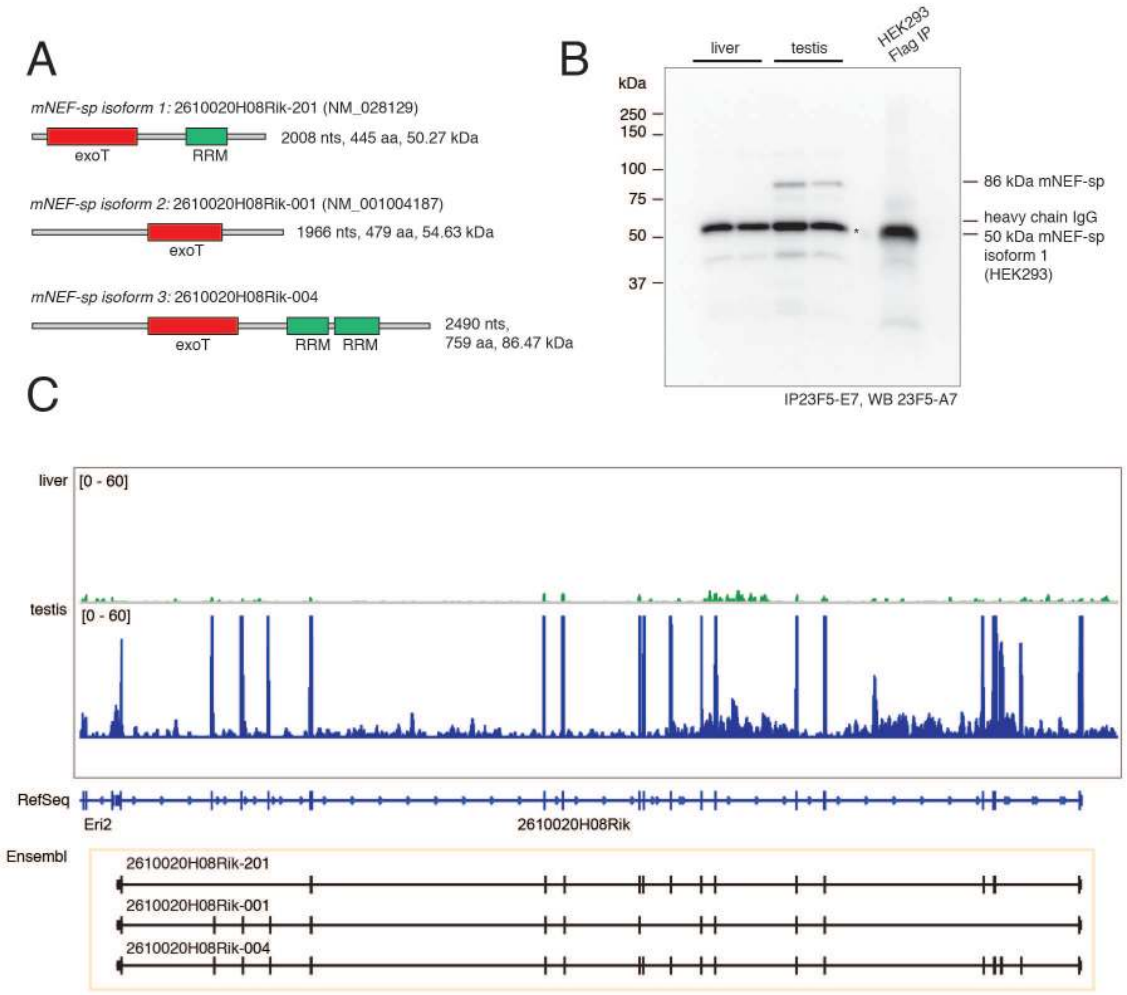


Figure 3.16 Immunoprecipitation of mNEF-sp from testis and liver lysates. (A) mNEF-sp isoforms defined by Ensembl and RefSeq. RNase T domains (red), and RRM domains (green) are highlighted. (B) Hybridoma supernatant clone 23F5-E7 was used for immunoprecipitation (IP) of mNEF-sp from liver and testis. Immunoprecipitated mNEF-sp was probed detected by Western blot (WB) with clone 23F5-A7. (C) Genome Browser representation of RNA-seq read alignments from liver and testis at the mNEF-sp locus (2610020H08Rik).

NEF-sp proteins across mammals and vertebrates are usually ~700 amino acids long and have a molecular weight of ~90 kDa (including rat). Hence, the deviance in size of the mouse homolog predicted by RefSeq seemed surprising, but we decided at the start of the project to work with isoform 1 as this appeared to be the closest related homolog (Figure 3.16 A). Given the unexpected immunoprecipitation results, we re-examined other genome annotations and detected recent updates of the Ensembl genome consortium, which included a new, longer isoform of mNEF-sp that was consistent with our biochemical results. This additional isoform (isoform 3) was 759 amino acids long, contained two predicted RRM domains C-terminal to the RNase T domain and had a predicted molecular weight of 86.5 kDa (Figure 3.16 B). By domain structure, weight and protein length, it was closest to other mammalian NEF-sp proteins. This isoform remains currently not listed in the RefSeq genome annotation, but it is supported by Uniprot data. To confirm whether the observed protein band could correspond to the longest predicted isoform, I analyzed published adult mRNA-seq data from mouse testis and liver and assessed read abundances for the different isoforms (You et al., 2015) (Figure 3.16 C). Isoform 3 was highly abundant in adult testis but not expressed in the liver. Together, given the RNA-seq and immunoprecipitation data, we concluded that isoform 3 is the dominant isoform of mNEF-sp in testis and based on our results we revised our homology annotation for NEF-sp proteins in Figure 3.8 to include the longest isoform of mNEF-sp as the representative isoform.

This section concludes the studies on the mammalian homologs of NEF-sp. Much of this work was done in parallel to the work on *Drosophila melanogaster* dNEF-sp. Because of the complexity and time needed for each organism, I decided to shift my focus entirely on the full characterization of the physiological function and targets of dNEF-sp for the remaining part of the project.

3.3.7 Characterization of *Drosophila* dNEF-sp genetic mutants

The dNEF-sp (CG8368) gene is located on the 3rd chromosome in *Drosophila melanogaster* (Figure 3.17 A). One genetic mutant, *C04255*, contained a PiggyBac (PB) transposon insertion into the 5' UTR of isoform CG8368-a and the first intron of isoform CG8368-b and was previously generated by the Harvard Exelixis collection (location of the PB insertion is shown in Figure 3.17 A). I confirmed the correct location of the PB by PCR using a transposon-specific primer in the forward and a gene-specific primer in the reverse direction (Figure 3.17 B). Homozygous mutants were lethal during second instar larval development. We confirmed that lethality was due to mutations in the cytogenetic region 65B3-65C1, (~135 kb) by crossing *C04255* to the deficiency lines *Df(3L)BSC410* (deletion of 64E7;65B3, 3L:5770673;3L:6490185) and *Df(3L)BSC411* (deletion of 65A2;65C1, 3L:5975960;3L:6625626-6625629). Crosses of *C04255* to *Df(3L)BSC411*, which had a 650 kb segment deleted encompassing the dNEF-sp gene locus, were lethal, while complementation of *C04255* to the Df(3L)BSC410 deficiency line, containing a deletion overlapping with *Df(3L)BSC411* except for a 135 kb region, did not affect viability. Next, we screened a number of previously generated ethyl methanesulfonate chemical mutagenesis lines deposited in the Drosophila Bloomington Stock center (DBSC), which had mutations in the 65C-65D cytogenetic band, but had previously not been further genetically mapped (Anderson et al., 1995). Complementation of *C04255* with one of the mutants, *M100*, was lethal, suggesting that this mutant also caused loss-of-function of dNEF-sp. Similar to *C04255*, *M100* homozygous mutants were lethal at the second instar larval stage. Crosses to the genetic deficiency lines gave the same results as above. Sequencing of *M100* within the coding region of dNEF-sp identified two missense mutations in the ORF; one of the mutations, a glutamate to lysine substitution (E497K) located four amino residues away from the conserved catalytic histidine in the RNase T domain presumably causes the loss-of-function of dNEF-sp (locations of nucleotide substitutions shown in Figure 3.17 A).

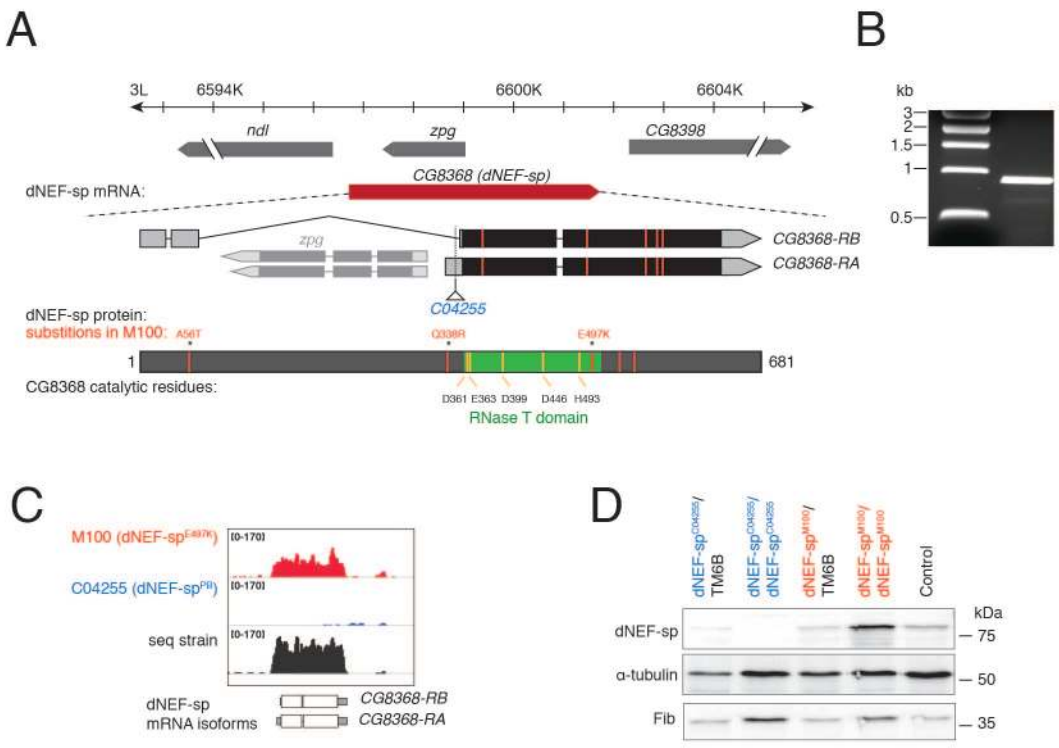


Figure 3.17 Characterization of dNEF-sp mutants. (A) Genomic map of the dNEF-sp gene, mRNA isoforms, and protein. The RNase T domain is highlighted in green, catalytic DEDDh residues are shown in orange, the location of the nucleotide substitutions in mutant *M100* and the resulting amino acid changes are shown in red. The PiggyBac insertion is shown for dNEF-sp mutant *C04255* (blue). (B) Verification of PiggyBac transposon insertion in mutant *C04255* into the predicted site at the dNEF-sp locus. (C) Expression of dNEF-sp mRNA in mutants *M100*, *C04255*, and wild type 2nd instar larvae. (D) dNEF-sp protein levels in mutants and wild type 2nd instar larvae. Loading controls are tubulin and the nucleolar snoRNP component Fibrillarin (Fib). Genotypes are highlighted in blue (C04255, dNEF-sp^{PB}) and red (M100, dNEF-sp^{E497K}). dNEF-sp Genotypes: heterozygotes: dNEF-sp^{M100}/TM6B and dNEF-sp^{C04255}/TM6B; homozygotes: dNEF-sp^{M100}/dNEF-sp^{M100} and dNEF-sp^{C04255}/dNEF-sp^{C04255}, wild type: sequencing strain.

Finally, to generate transgenic flies, I cloned the full-length ORF of dNEF-sp into the pUAS-attB plasmid and, in addition, generated four genomic pattB rescue constructs. Two of these genomic rescue constructs contained the genomic dNEF-sp region including either a 3 or 1.7 kb promoter region and the genomic region 1 kb region downstream of the 3'UTR. In addition, I created two genomic, N-terminal tagged dNEF-sp rescue constructs by inserting GFP or FLAG/HA-tags into the 3 kb genomic construct at the start of the dNEF-sp ORF, the N-

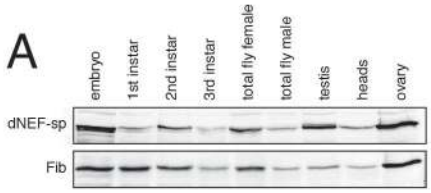
terminus of dNEF-sp protein. The pUAS-attB and pattB constructs were send for injection and were specifically inserted at 25C7 cytogenetic site on the second chromosome using the phiC31 integrase transgenesis system (Bischof et al., 2007). All genomic dNEF-sp transgenic lines rescued lethality of dNEF-sp mutants and produced adult flies in both mutants. Expression of the dNEF-sp ORF alone (pUAS-attB-dNEF-sp) under the actinGal4 driver also rescued lethality. Poly(A) mRNA-seq of homozygous mutants and wild type 2nd instar larvae showed that in *C04255* mutants mRNA levels were close to depleted, while in *M100* mutants mRNA levels were similar to wild type (Figure 3.17 C). Western blot analysis of protein levels in homo- and heterozygous mutants and wild type controls showed that heterozygous mutants had protein levels similar to wild type larvae. In contrast, homozygous *C04255* mutant protein levels were undetectable by Western blot analysis, while *M100* mutants had consistently higher protein levels, suggesting some posttranscriptional compensation mechanism of *M100* homozygous mutants to produce more dNEF-sp protein, which nevertheless is not functional (Figure 3.17 D).

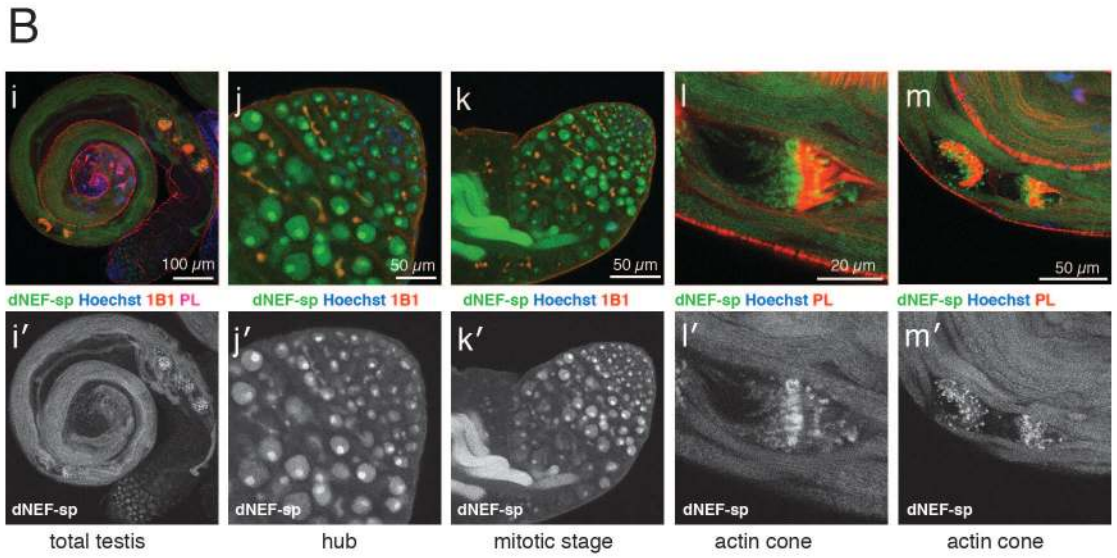
3.3.8 dNEF-sp is a nucleolar/nuclear protein and translocates to the cytoplasm during terminal differentiation in testis development

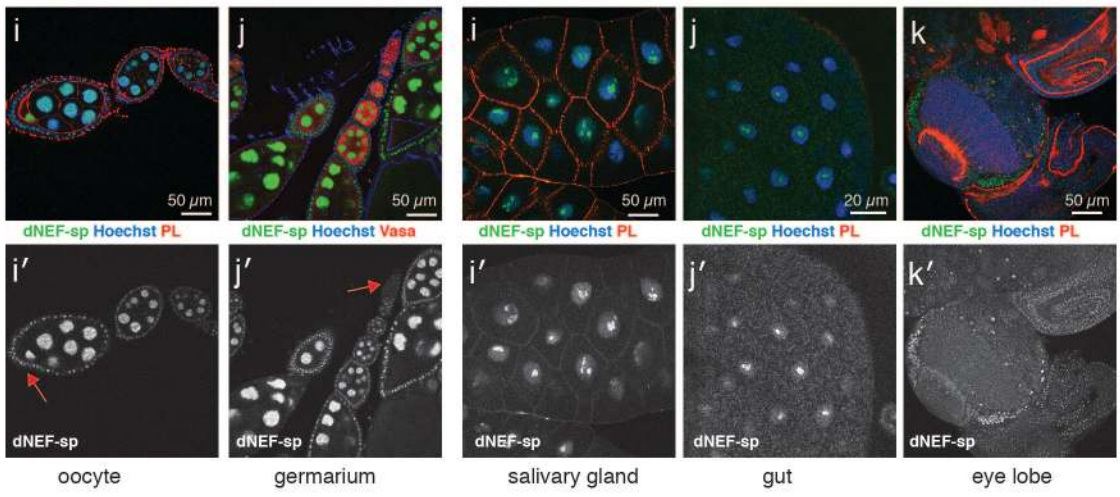
In contrast to the mRNA abundances, dNEF-sp protein was predominantly expressed in both male and female gonads in adult flies, but also expressed highly during early embryogenesis. To investigate the localization of dNEF-sp, I studied the *in vivo* localization of GFP-dNEF-sp in the transgenic GFP-dNEF-sp fly rescue, expressing GFP-dNEF-sp under its endogenous promoter. In *Drosophila* testis GFP-dNEF-sp was expressed in GSCs and mitotic cells, as well as during terminal spermatocyte differentiation (Figure 3.18 B). In mitotic gonialblasts and GSCs dNEF-sp was localized to the nucleolus. In contrast, dNEF-sp abundance was highest during terminal spermatocyte differentiation, where it localized in the cytoplasm and was concentrated predominantly to the individualization complexes, localizing in a distinct pattern in opposite

polarity to the actin cones (Figure 3.18 B). This pattern was unique to dNEF-sp and unlike other rRNA biogenesis and ribosomal proteins, and may suggest an active role for dNEF-sp during terminal differentiation. In ovaries dNEF-sp was highly expressed throughout the tissue in the germarium in GSCs and cystoblasts, and highest expressed in the nuclei of oocytes, nurse cells, and follicle cells (Figure 3.18 C). dNEF-sp expression was not restricted to gonads and was ubiquitously expressed at low level in the nucleolus and nucleus of somatic tissues (Figure 3.18 D).

Figure 3.18 dNEF-sp is predominantly expressed in the gonads. (A) Expression of dNEF-sp and the nucleolar marker Fibrillarin (Fib) assessed by Western blot analysis on total 40 µg tissue lysates. (B-D) *In vivo* localization of GFP-dNEF-sp. Shown are GFP-dNEF-sp (green), phalloidin (PL, magenta or red), 1B1 in testes (red) and Vasa in ovaries (red), and DNA (Hoechst, blue). Localization of GFP-dNEF-sp is also shown in gray scale images. (B) Localization of dNEF-sp in testis. Shown are (i,i') whole testis, (j,j') hub region with GSCs, and (k,k') mitotically dividing gonialblasts, (l,l',m,m') the individualization complex during terminal differentiation with punctate localization of dNEF-sp in the leading edge of the actin cones. (C) *In vivo* localization of GFP-dNEF-sp in ovaries. (i,i') dNEF-sp is localized to the nuclei of the oocyte, nurse cells and somatic follicle cells. (j,j') Localization of dNEF-sp in nuclei of GSCs and gonialblasts in the germarium (red arrow in j'). (D) GFP-dNEF-sp is predominantly expressed in nucleoli in somatic tissues. Expression for (i,i') salivary glands, (j,j') foregut, (k,k') eye lobe and imaginal discs. Genotypes: GFP-dNEF-sp/Cyo; Sb/TM6B.


A
embryo
1st instar
2nd instar
3rd instar
total fly female
total fly male
testis
heads
ovary
dNEF-sp
Fib


B
i
j
k
l
m
100 μm
50 μm
50 μm
20 μm
50 μm
dNEF-sp Hoechst 1B1 PL
dNEF-sp Hoechst 1B1
dNEF-sp Hoechst 1B1
dNEF-sp Hoechst PL
dNEF-sp Hoechst PL
i′
j′
k′
l′
m′
dNEF-sp
dNEF-sp
dNEF-sp
dNEF-sp
dNEF-sp
total testis
hub
mitotic stage
actin cone
actin cone


C
D
i
j
i
j
k
50 μm
50 μm
50 μm
20 μm
50 μm
dNEF-sp Hoechst PL
dNEF-sp Hoechst Vasa
dNEF-sp Hoechst PL
dNEF-sp Hoechst PL
dNEF-sp Hoechst PL
i′
j′
i′
j′
k′
dNEF-sp
dNEF-sp
dNEF-sp
dNEF-sp
dNEF-sp
oocyte
germarium
salivary gland
gut
eye lobe

3.3.9 Homozygous mutants show gonad developmental defects

dNEF-sp homozygous mutant larvae were developmentally delayed by 24-48 days, but organs looked generally morphologically normal in somatic tissues compared to wild type larval tissues at the same developmental stage. We reasoned that defects in dNEF-sp mutants would most strongly accumulate in the tissues of highest expression, i.e. during testis development, and therefore decided to investigate early phenotypes in homozygous mutant testes during 2nd instar larval development, their latest developmental stage before death. By confocal microscopy mutant gonads were smaller in size and arrested in the first or second mitotic divisions (as assessed by the level branching of connected fusomes in mitotic cells) and they also had fewer cells compared to wild type 2nd instar male gonads (Figure 3.19 A). These results demonstrated that dNEF-sp was required for testis development and that dNEF-sp mutants were arrested in early mitosis.

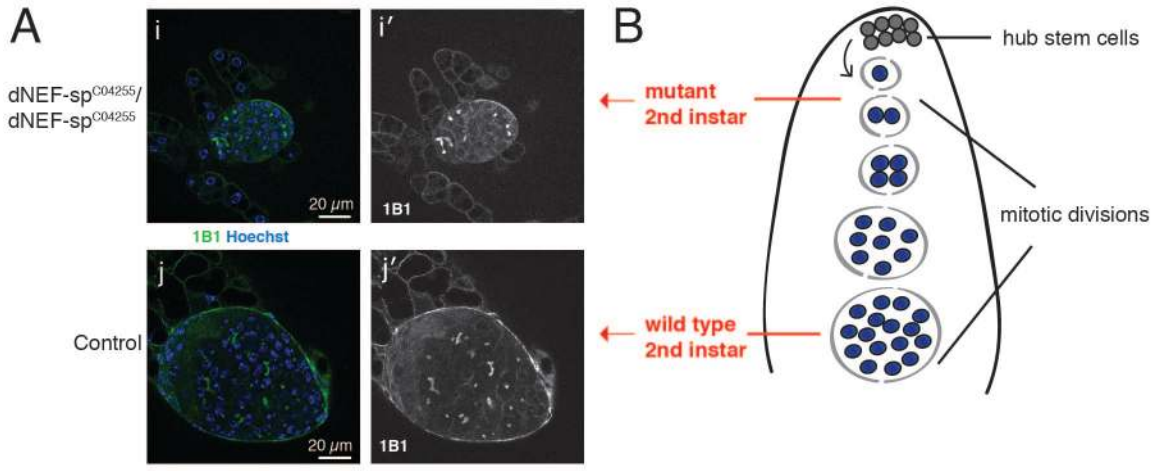
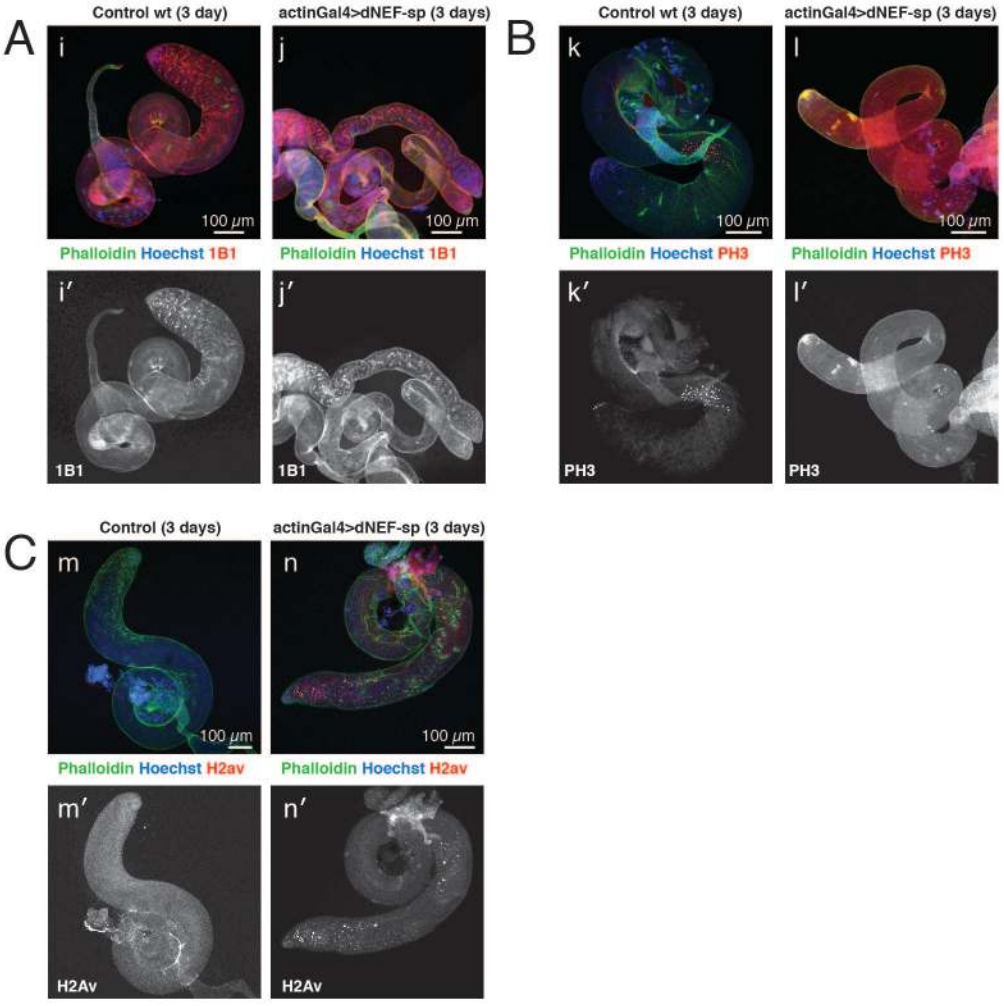


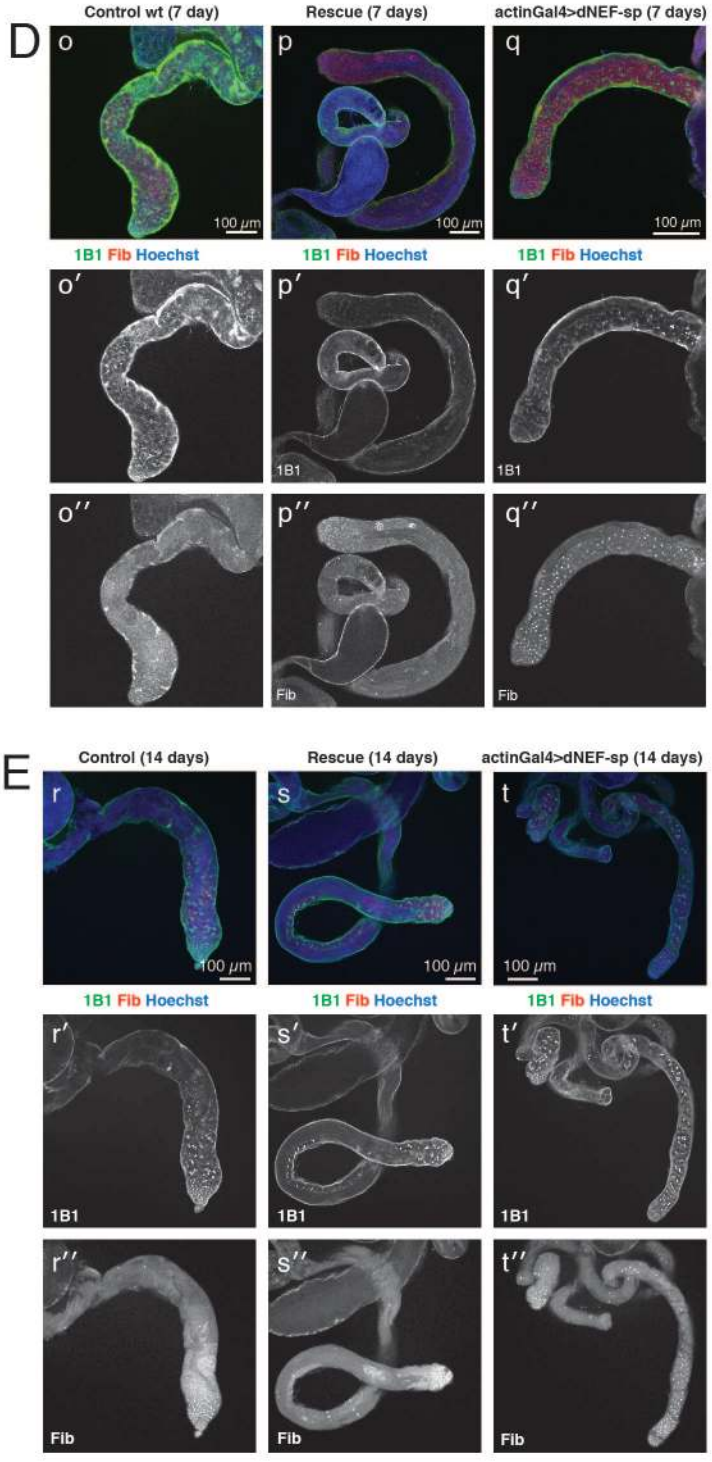
Figure 3.19 dNEF-sp mutant gonads are developmentally arrested. (A) Second instar larval male gonads. (i,i') dNEF-sp^{C04255}/dNEF-sp^{C04255} homozygous mutant (j,j') wild type. DNA is shown with Hoechst dye in blue, the 1B1 fusome marker in green. 1B1 staining is separately shown in grayscale images. (B) Schematic representation of the mitotic developmental program in *Drosophila* testis development. GSCs asymmetrically divide to produce one stem cell daughter cell and one gonialblast cell. The gonialblast migrates from the hub and undergoes 4 mitotic divisions. Stages of mutant and wild type 2nd instar larval gonad development are highlighted in red. Genotypes: dNEF-sp^{C04255}/dNEF-sp^{C04255}, wild type: sequencing strain.

3.3.10 Developmental arrest in testes and ovaries of actinGal4 rescued dNEF-sp adult mutants

To investigate gonad developmental defects in adult flies, I rescued the lethality in somatic cells by transgenic expression of UAS-dNEF-sp under the actinGal4 driver in both mutants. The UAS promoter works inefficiently in the germline (Brand and Perrimon, 1993), thus actinGal4 UAS-dNEF-sp rescued flies effectively lack dNEF-sp expression in germ cells. I examined mutant adult testes and ovaries after 3, 7, and 14 days (Figure 3.20). Homozygous adult mutant testes showed phenotypic variation. Severely affected testes were arrested in mitosis, and instead of being confined to the hub, mitotic cells were spread throughout the entire organ (Figure 3.20 A). Phosphorylation of histone H3 levels showed that mutant testes did not enter meiosis (Figure 3.20 B). Furthermore, H2Av phosphorylation levels, a marker of DNA double-stranded breaks, indicated that also genomic stability was affected in mutant testes (Figure 3.20 C). In *Drosophila* males do not undergo meiotic recombination, thus double-stranded DNA breaks are entirely due to genomic instability. Over time, fewer cysts went through mitosis and testes became morphologically thinner (Figure 3.20 D-E).

Figure 3.20 dNEF-sp mutant adult testes arrest in mitosis. (A-E) dNEF-sp mutants are lethal, but transgenic expression of dNEF-sp in the soma rescues this lethality and permits the recovery of adult flies that lack dNEF-sp in germ cells. These flies display defects in growth and the cell-cycle progression gonads. Shown are Phalloidin (green), Hoechst (blue), individual marker (red). Individual developmental markers are shown separately in grayscale images. Genotypes of rescued mutants: actin-Gal4/UAS-dNEF-sp, dNEF-sp^{C02455}/dNEF-sp^{M100}. (A) dNEF-sp mutant testes are thinner and show continuous mitotic divisions with the fusome branching throughout the entire testicular tube (1B1). (B) Serine 10 phosphorylated histone H3 (PH3) marks meiotic divisions. Mutant testes show fewer meiotic nuclei. Wild type testes enter meiosis as detected by the 32 PH3-stained meiotic nuclei. (C) Histone H2Av phosphorylation (H2Av) is a marker for double-stranded DNA breaks, genomic instability and apoptosis (Fernandez-Capetillo et al., 2004; Kotova et al., 2011). H2Av phosphorylation levels are increased in mutant testis compared to wild type, indicating an increase in genomic instability in dNEF-sp mutants. (D-E) 1B1 and Fibrillarin (Fib) stainings after 7 and 14 days in wild type, genomic rescue of dNEF-sp mutants (full rescue) and actinGal4 UAS-dNEF-sp rescued mutants (germ cell loss). dNEF-sp mutants show progression in mitosis beyond the hub throughout the entire testis (1B1, green) and the nucleolar marker Fib (red) shows continuous staining, also marking mitotic cells. Genotypes: mutant: actinGal4/UAS-dNEF-sp, dNEF-sp^{M100}/dNEF-sp^{C04255}; genomic rescue: 1.5kb-dNEF-sp, *dNEF-sp*^{M100}; wild type: sequencing strain.


A
Control wt (3 day)
actinGal4>dNEF-sp (3 days)
i
j
100 μm
Phalloidin Hoechst 1B1
i′
j′
1B1
B
Control wt (3 days)
actinGal4>dNEF-sp (3 days)
k
l
Phalloidin Hoechst PH3
k′
l′
PH3
C
Control (3 days)
actinGal4>dNEF-sp (3 days)
m
n
Phalloidin Hoechst H2av
m′
n′
H2Av


D
Control wt (7 day)
Rescue (7 days)
actinGal4>dNEF-sp (7 days)
o
p
q
100 μm
1B1 Fib Hoechst
o′
p′
q′
1B1
o″
p″
q″
Fib
E
Control (14 days)
Rescue (14 days)
actinGal4>dNEF-sp (14 days)
r
s
t
r′
s′
t′
r″
s″
t″

Homozygous mutant females were sterile and had rudimentarily developed ovaries. Morphologically, the oocyte was placed correctly and expression of the germ cell marker Vasa in mutants was the same as in control wild type ovaries. H2Av phosphorylation, a marker of dsDNA breaks, did not show any difference to wild type, indicating that genomic integrity of mutant ovaries was not compromised as observed for mutant testes. However, mutant oocytes did not fully develop, leading to homozygous females having >10 fold smaller, rudimentarily developed ovaries. Ovaries displayed egg-chamber degeneration and arrested at stage 4-6 of oocyte development. Moreover, complete GSCs loss occurred over time, indicating that dNEF-sp was essential for germ cell development (Figure 3.21 A-B). We recognized that the presented ovarian phenotypes resembled previously described ovarian defects for loss-of-function of the ribosomal protein RpS2 (Cramton and Laski, 1994) (Figure 3.21 C), suggesting that dNEF-sp might also be involved in a ribosome-related process.

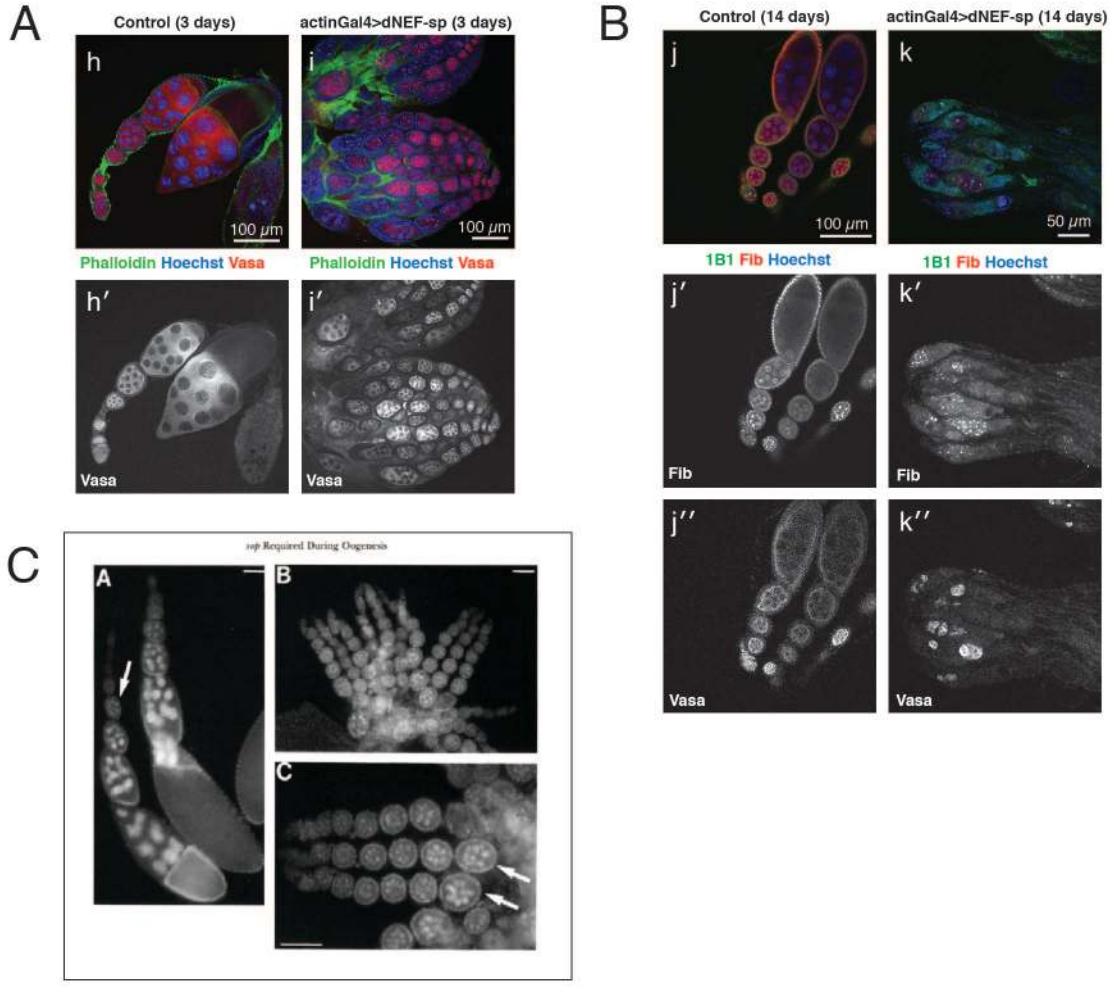


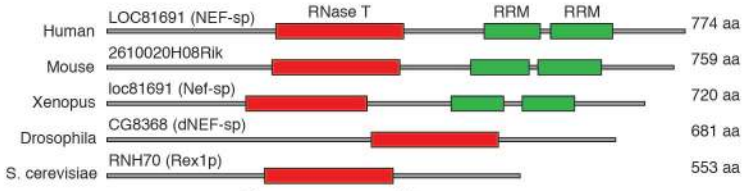
Figure 3.21 dNEF-sp mutant adult ovaries are growth arrested. (A, B) Mutant ovaries are smaller than wild type and arrest at developmental stage 4-6. Organization of oocytes and nurse cells are positioned normally, expression of the germ cell marker Vasa (red) is normal, no physical deformities are observed, but oocyte and nurse cell growth is impaired and cells become apoptotic at stage 4-6. (A) Mutant and wild type ovaries after 3 days. Mutant ovaries are 10-100-fold smaller and degenerate, entering apoptosis at stage 4-6. Vasa (red) expression and position of oocyte, as well as the number of nurse cells are normal compared to wild type. (B) Loss of egg chambers becomes more severe over time; ovarioles contain only the germarium and stage 1 egg chambers. (C) Ovarian phenotype for the string-of-pearls (RpS2) gene characterized by Cramton *et al.* (Cramton and Laski, 1994) shows similarity to the dNEF-sp loss-of-function phenotype. Genotypes: mutant: actinGal4/UAS-dNEF-sp, dNEF-sp^{M100}/dNEF-sp^{C04255}; wild type: sequencing strain.

In summary, the mutant phenotypes suggested a pivotal role of dNEF-sp in early gonad development, but they did not address why high dNEF-sp protein levels were observed in later stages post meiosis during terminal differentiation of spermatocytes. Since neither the biological role nor the molecular mechanisms were known for dNEF-sp, I focused my attention for the remaining study on the early gonadal and somatic roles of dNEF-sp in order to elucidate the essential function of dNEF-sp in general RNA metabolism.

3.3.11 The catalytic RNase T domain of dNEF-sp is required for viability *in vivo*

The predicted DEDDh RNase T exonuclease domain in the NEF-sp protein family is highly conserved across organisms and requires a divalent metal cation for activity (Zuo and Deutscher, 2001) (Figure 3.22 A). To test whether the catalytic activity of NEF-sp was essential for viability, I genetically complemented *C04255* and *M100* mutants with UAS-dNEF-sp transgenic flies, for which different catalytic residues were mutated to alanine in the RNase T domain. All catalytic residues were required for survival and none of the tested constructs, which contained a single or double alanine mutation, could rescue lethality (Figure 3.22 B). Substitution of the catalytic histidine residue (H) to an arginine (R), an amino acid change present in the mouse NEF-sp homolog, rescued homozygous larvae to 3rd instar, suggesting that substitution to arginine gave still partially active dNEF-sp protein. Therefore, we concluded that mNEF-sp is most likely a functional RNA exonuclease, despite a substitution of the catalytic histidine to arginine. The results demonstrated that the RNA exonuclease activity of dNEF-sp, and not a secondary protein association, was essential for survival.

A RNase T domain multiple sequence alignment:




```
RNH70 (Rex1p)              FALDCEMCLSEQG-LVLTRISLVNFDNEVIYEELVKPDVPIVDYLTRYSGITEEKLTVGA
dNEF-sp (CG8368)           FGVDCEMCHTEAGCNELTRISIVNENYETVYETLVLPNNRITDYLTQYSGITAEIME-QV
X.laevis NEF-sp (loc81691) FGLDCEMCLTDKG-SELTRISLVDASGSCIMDELVKPDNTIRDYMTRYSGITRKLLL-PV
hNEF-sp (LOC81691)         FGLDCEMCLTSKG-RELTRISLVAEGGCCVMDELVKPENKILDYLTSFSGITKKILN-PV
mNEF-sp (2610020H08Rik)    FGLDCEVCLTSMG-KELTRISLVTEGGYCLIDELVKPDLKILDYLTSFTGITKEILN-PV
                           *.:***:* :. *   *****:*      : : ** *:  * **:* ::*** : :   .

RNH70 (Rex1p)              KKTLREVQKDLLKIISRSDILIGHSLQNDLKVMKLKHPLVVDTAIIYHHKAGDPFKPSLK
dNEF-sp (CG8368)           TKRLDVVQKEVSELLPPDAILVGQSLNSDLNAMKMMHPYVIDTSVCFNTSGVRRRKTKLK
X.laevis NEF-sp (loc81691) KTKLKDVQQKLKSVLPPDAVLVGHSLDNDLRALQMIHTSVIDTALLFAREYGR--KFRLK
hNEF-sp (LOC81691)         TTKLKDVQRQLKALLPPDAVLVGHSLDLDLRALKMIHPYVIDTSLLYVREQGR--RFKLK
mNEF-sp (2610020H08Rik)    TTKLKDVQKLLRELLPPDAVLVGHCLDLDLRVLKMIHPYVIDTSLLYAGKQGR--RFKLT
                           .. *  **: :  ::  . :*:*:.*: **..::: *  *:**:: :  .    :  *.

RNH70 (Rex1p)              YLSETFLNKSIQNG---EHDSVEDARACLELT (226-373)
dNEF-sp (CG8368)           DLAKTFLQ-EIIQENIDGHDSIEDSRATLKLV (357-506)
X.laevis NEF-sp (loc81691) FLAQAVLGREIQTDDVMGHCPAEDARAALNLA (190-337)
hNEF-sp (LOC81691)         FLAKVILGKDIQCPDRLGHDATEDARTILELA (229-376)
mNEF-sp (2610020H08Rik)    FLARVILGKDIQCPNKLGRDGIEDARAALELL (221-317)
                            *:...*  .*       :   **:*: *:*
```

B actinGal4>UAS-dNEF-sp mutant rescues:

(1)	DEDDH	adult full rescue
(2)	DEDDR	3rd instar
(2)	DEDDA	2nd instar- no rescue
(3)	AEDDH	
(4)	DADDH	
(5)	DEDAH	
(6)	AEDAH	
(7)	DADAH	

Genotype:
dNEF-sp$^{\text{DEDDH mutants}}$/actinGal4;dNEF-sp$^{\text{C04255}}$/dNEF-sp$^{\text{M100}}$

Figure 3.22 Multiple sequence alignment of conserved NEF-sp active site residues. (A) Scheme of the structural domain organization of NEF-sp proteins across yeast, *D. melanogaster*, *Xen. laevis*, human and mouse and Clustal Omega alignment of the conserved RNase T domain. Conserved catalytic acidic residues (DEDD) are highlighted in orange, the general base histidine residue is highlighted in yellow. (B) Viability test rescuing lethality of mutants with actin-Gal4 UAS-dNEF-sp constructs with substituted residues in the catalytic side.

3.3.12 dNEF-sp displays single-stranded RNase activity *in vitro*

To verify the catalytic activity of dNEF-sp *in vitro*, I purified the recombinant full-length dNEF-sp protein and the catalytic dead mutant (dNEF-sp^{DADAH}, section 3.3.2) and carried out DNA and RNA oligomer cleavage assays. Cleavage assays were carried out for both wild type and mutant full-length proteins with 18mer sequences. dNEF-sp showed selective 3'-5' exonucleolytic degradation of RNA, while DNA was not degraded (Figure 3.23 A). RNA cleavage of dNEF-sp displayed preference for pyrimidine bases (C and U) over purine bases (A) (Figure 3.23 D-F). As previously characterized for other DEDDh RNase T nucleases, RNA cleavage activity of dNEF-sp was dependent on the presence of divalent cations (Mg^{2+}) in the active site and the reaction was inhibited in the presence of EDTA (Figure 3.23). Cleavage assays with circularized RNAs demonstrated that dNEF-sp could not endonucleolytically process circularized ssRNA (Figure 3.24 A-B).

Figure 3.23 RNase *in vitro* assays show specificity of dNEF-sp for ssRNA. *In vitro* RNA exonuclease assays of recombinant wild type and DADAH mutant dNEF-sp protein. RNase activity of the recombinant nucleases was tested for a selection of 5'-^{32}P-labeled RNA and DNA 18-mer oligo(deoxy)ribonucleotides at 20-25 nM enzyme concentration in a time series (0-30/60 min). Control = no enzyme added, control EDTA = dNEF-sp wt protein in presence of EDTA (50 mM). (A) Activity of dNEF-sp wild type and mutant protein for poly-adenosine RNA and DNA substrate at 20 nM enzyme concentration. (B) Same as in (A) for poly-cytosine RNA and DNA substrates. (C) Same as in (A) for poly-uridine and poly-thymidine. (D) Quantification of the degradation of substrates at 25 nM protein concentration measured in (E) and (F). Intensity measurements are normalized to the intensity of the 1 min time point for each experimental condition. (E) Comparison of nucleolytic activity of dNEF-sp wild type and mutant protein at 25 nM concentration for poly-cytosines versus poly-adenosine-cytosine oligomer. (F) Exonucleolytic activity of dNEF-sp wild type and mutant protein for poly-uridine and poly-adenosine RNA and DNA oligomers. DNA is not a processed substrate, dNEF-sp DADAH mutant is inactive, addition of EDTA inhibits enzymatic activity of wild type dNEF-sp protein (control EDTA).

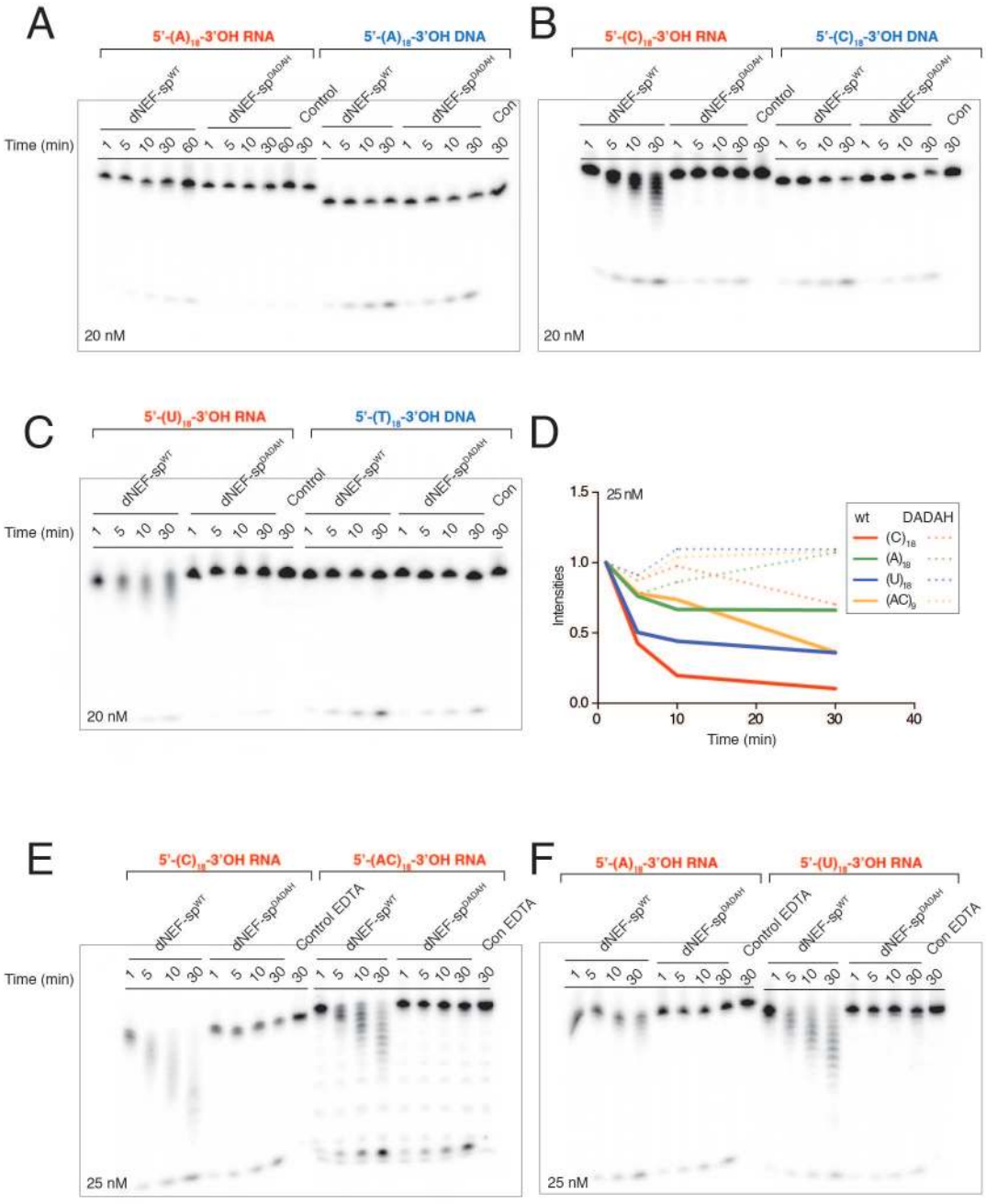
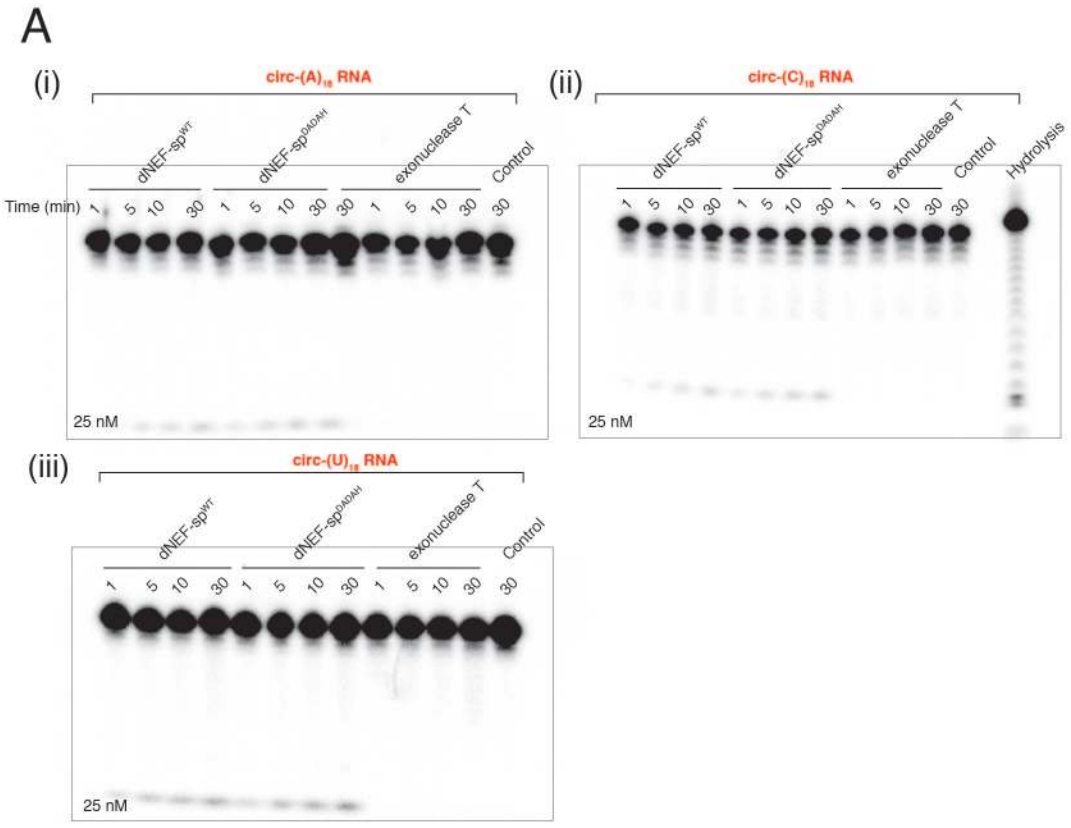
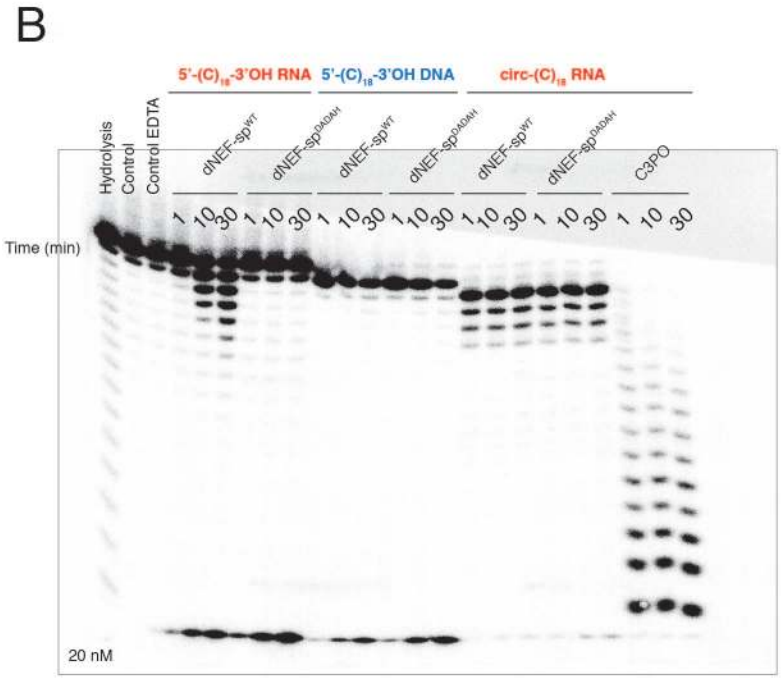

A
5'-(A)18-3'OH RNA
5'-(A)18-3'OH DNA
dNEF-spWT
dNEF-spDADAH
Control
Con
Time (min)
20 nM
B
5'-(C)18-3'OH RNA
5'-(C)18-3'OH DNA
C
5'-(U)18-3'OH RNA
5'-(T)18-3'OH DNA
D
25 nM
wt
DADAH
(C)18
(A)18
(U)18
(AC)9
Intensities
Time (min)
E
5'-(C)18-3'OH RNA
5'-(AC)18-3'OH RNA
Control EDTA
Con EDTA
25 nM
F
5'-(A)18-3'OH RNA
5'-(U)18-3'OH RNA

Figure 3.24 dNEF-sp has no endonucleolytic activity. (A-B) *In vitro* RNA nuclease assays of recombinant wild type and DADAH mutant dNEF-sp protein and commercial RNA exonuclease T (ExoT, NEB). RNase activity of the recombinant nucleases was tested for a selection of 5'-^{32}P-labeled RNA and DNA 18-mer single-stranded and circularized oligo(deoxy)ribonucleotides at 20-25 nM enzyme concentration in a time series of 0-30 min. (A) (i) Activity of dNEF-sp wild type and mutant protein for circularized poly-adenosine RNA and at 25 nM enzyme concentration. (ii) Same as in (i) for circularized poly-cytosine RNA substrates. (iii) Same as in (i) for circularized poly-uridine oligomers. (B) High-resolution sequencing gel for the comparison of nucleolytic activity of dNEF-sp wild type and mutant protein, as well as the RNA endonuclease C3PO as control (gift of M. Ascano) at 20 nM concentration for poly(C) ssRNA, ssDNA and circularized RNA 18mers. Final concentration of substrate is 100 nM per reaction.


A
(i)
circ-(A)18 RNA
dNEF-spWT
dNEF-spDADH
exonuclease T
Control
Time (min)
1 5 10 30 1 5 10 30 30 1 5 10 30 30
25 nM
(ii)
circ-(C)18 RNA
dNEF-spWT
dNEF-spDADH
exonuclease T
Control
Hydrolysis
1 5 10 30 1 5 10 30 1 5 10 30 30
25 nM
(iii)
circ-(U)18 RNA
dNEF-spWT
dNEF-spDADH
exonuclease T
Control
1 5 10 30 1 5 10 30 1 5 10 30 30
25 nM


B
5'-(C)18-3'OH RNA
5'-(C)18-3'OH DNA
circ-(C)18 RNA
Hydrolysis
Control
Control EDTA
dNEF-spWT
dNEF-spDADH
dNEF-spWT
dNEF-spDADH
dNEF-spWT
dNEF-spDADH
C3PO
Time (min)
1 10 30 1 10 30 1 10 30 1 10 30 1 10 30 1 10 30 1 10 30
20 nM

3.3.13 The 3' ETS of the 47S rRNA precursor accumulates in dNEF-sp mutants

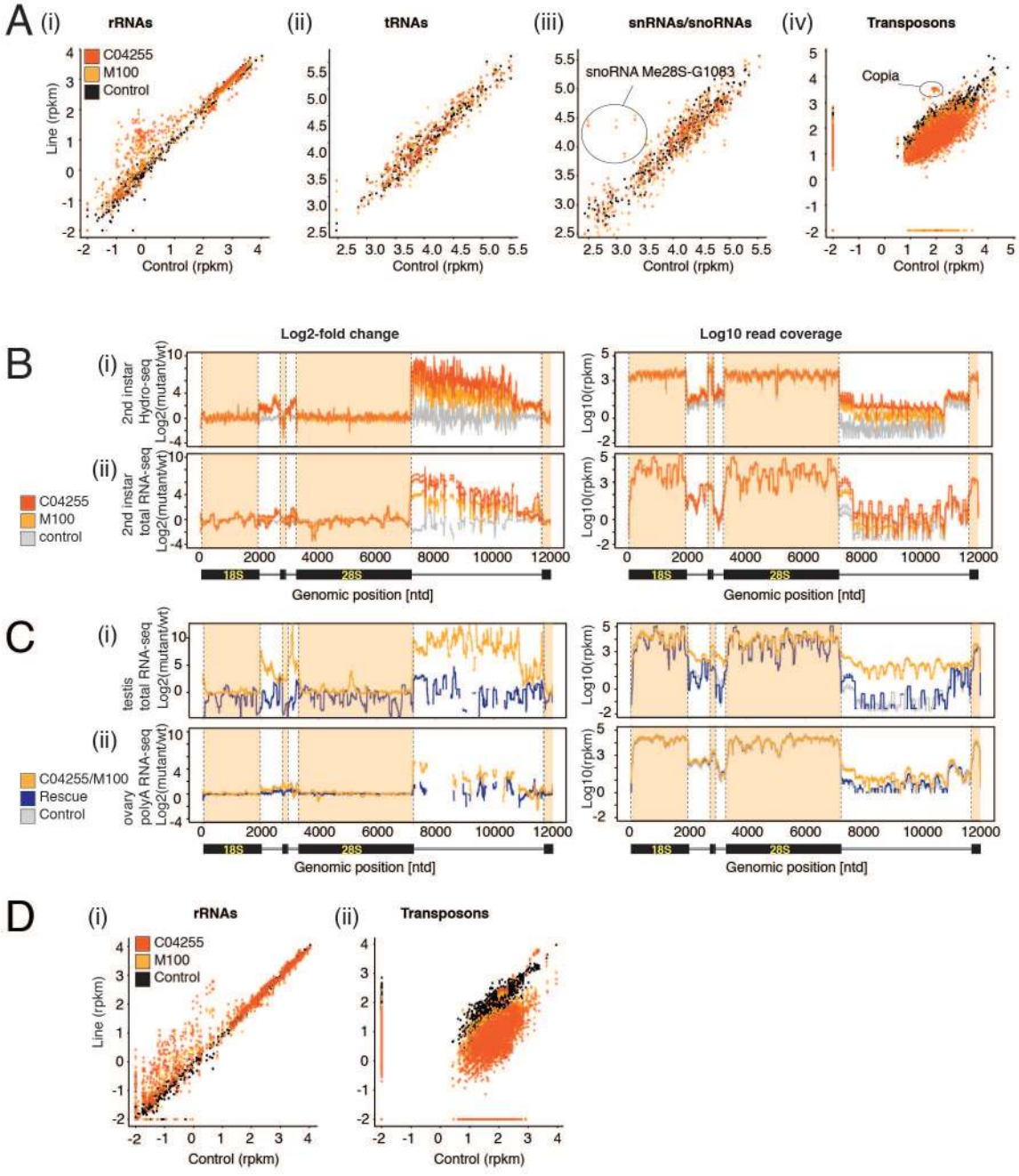
To identify processing defects of dNEF-sp targets in the homozygous mutants *in vivo* we employed next generation sequencing of total RNA isolated from wild type and dNEF-sp genetic mutant larvae. The localization of dNEF-sp to the nucleolus, as well as the mutant phenotypes observed in adult ovaries, suggested a role in rRNA biogenesis. Previous literature on the family member RNH70 (Rex1p) indicated that the *S. cerevisiae* nuclease was involved in the processing of a variety of noncoding RNA substrates, including tRNAs, snRNAs, and snoRNAs (Copela et al., 2008; Ozanick et al., 2009; Piper and Stråby, 1989; Piper et al., 1983; 1987; van Hoof et al., 2000). Thus we considered that, if dNEF-sp was a functional homolog of Rex1p, dNEF-sp could process small nuclear noncoding RNAs. In addition, the selective upregulation of dNEF-sp in gonads indicated a gonad-specific specific role, putatively in germline-specific noncoding RNA metabolism such as the piRNA pathway (Siomi et al., 2011).

To analyze differential expression of long and short coding and non-coding RNAs, and also to investigate highly structured transcripts, we combined the poly(A) mRNA-seq and total RNA sequencing (Illumina TruSeq) protocols with an established small RNA cloning protocol (Hafner et al., 2012), for which we included an alkaline hydrolysis step to fragment total RNA prior to library preparation ("Hydro-Seq") (developed by D. Briskin, M. Brown, T. Farazi, T. Tuschl, Methods 3.3). The latter RNA-sequencing method allowed us to investigate nuclear noncoding RNAs not covered by conventional RNA-seq protocols. tRNAs, snRNAs and snoRNAs have a size range between ~100-150 nt length, but current RNA sequencing methodologies, such as the Illumina TruSeq protocol, are optimized for size ranges between 200-300 nts and do not appropriately cover RNAs of smaller sizes. Therefore, to capture small noncoding RNAs between 20-100 nts, we needed read coverage of shorter RNAs. Using alkaline hydrolysis, we include a fragmentation step of total RNA, selecting RNA of 20-30 nt-long

segments, before proceeding with adapter ligation and library preparation of small RNAs as previously described (Hafner et al., 2012).

I analyzed gene expression of abundant noncoding RNAs by annotating all sequence reads in a hierarchical fashion, starting from the most abundant RNAs (rRNAs) to the less abundant ones (in descending order: 1= rRNA, 2=tRNA, 3=snRNA/snoRNA, 4=transposon/repeats, 5=miscRNA (miRNAs, other ncRNAs). Detecting differential expression of rRNA genes can be challenging, as rRNA sequences are highly heterogeneous and lie within repetitive regions of the genome (Kominami et al., 1981; Sylvester et al., 1986; Tautz et al., 1988). As a result current genome annotation databases, such as Flybase, RefSeq or Ensembl, are poorly curated and incomplete in their rRNA annotations. In addition, expression changes within regions of the large rRNA precursor transcript can be difficult to detect in the presence of the highly abundant, mature rRNAs. Failing initially to detect differential expression changes in rRNAs, we decided to create our own rRNA alignment index by selecting annotated *Drosophila melanogaster* rRNA sequences on NCBI and Flybase and blast them against the NCBI nucleotide database, thereby adding sequence-related insect rRNA sequences from the NCBI nucleotide database (Altschul et al., 1990). Hierarchical alignment against the most abundant RNAs showed specific upregulation of dNEF-sp mutants for rRNAs, one snoRNA cluster, and few transposons (Figure 3.25 A).

Figure 3.25 RNA-sequencing reveals accumulation of the ribosomal precursor 3'ETS in dNEF-sp mutants. (A) Hydro-seq scatterplots for different abundant noncoding RNAs. (i) NCBI and BLAST created rRNA sequences (ii) annotated tRNAs with +/- 20 nt extensions, (iii) Flybase annotated snRNAs and snoRNAs, (iv) Flybase annotated transposons. Mutant *C04255* (dark orange), mutant *M100* (bright orange), wild type control (black). (B-C) Log2 fold changes of read coverage and log10 read coverage along the 47S rRNA precursor. Mutant *C04255* (dark orange), mutant *M100* (bright orange), wild type control (grey), dNEF-sp genomic rescue (dark blue). (B) (i) Hydro-seq fold changes and coverage. (ii) Total RNA-seq fold changes and coverage. (C) (i) Total RNA-seq fold changes and coverage for adult actinGal4 UAS-dNEF-sp rescued mutant testes. (ii) poly(A) RNA-seq fold changes and coverage for adult actinGal4 UAS-dNEF-sp rescued mutant ovaries. (D) Total RNA-seq scatterplots for long abundant noncoding RNAs. Mutant *C04255* (dark orange), mutant *M100* (bright orange), wild type control (black). (i) rRNA reference, (ii) Flybase transposon annotation. Genotypes: heterozygotes: dNEF-sp^{M100}/TM6B and dNEF-sp^{C04255}/TM6B; homozygotes: dNEF-sp^{M100}/dNEF-sp^{M100} and dNEF-sp^{C04255}/dNEF-sp^{C04255}, wild type: sequencing strain.


A
(i) rRNAs
(ii) tRNAs
(iii) snRNAs/snoRNAs
(iv) Transposons
C04255
M100
Control
Line (rpkm)
Control (rpkm)
snoRNA Me28S-G1083
Copia
B
Log2-fold change
Log10 read coverage
2nd instar Hydro-seq Log2(mutant/wt)
2nd instar total RNA-seq Log2(mutant/wt)
Log10(rpkm)
C04255
M100
control
18S
28S
Genomic position [ntd]
C
testis total RNA-seq Log2(mutant/wt)
ovary polyA RNA-seq Log2(mutant/wt)
C04255/M100
Rescue
Control
D
(i) rRNAs
(ii) Transposons

Using the new reference index I could detect more than 10-fold upregulation of specific rRNA regions within the 3'ETS of the 47S precursor [Figure 3.25 A(i)]. When visualizing the read coverage along the full 47S rRNA precursor transcript, I detected, reproducibly across datasets, a significant 200-1000 upregulation of the 3' ETS in mutants (Figure 3.25 B). Visualization of the fold changes along the complete precursor at each nucleotide position revealed that fold change differences were highest closest to the mature 3' end of 28S (200-1000 fold), and smallest at the 5' end of the 47S precursor (4-fold). In wild type 2nd instar larvae the intergenic region between the 28S rRNA 3' end and 18S rRNA 5' end was close to undetectable with 0-5 reads on average along the entire region. In mutants the intergenic region was fully covered with several hundred reads directly downstream of the 3' end of 28S, which slowly decreased towards the 5'ETS [Figure 3.25 B (i)]. The continuous decrease in 3'ETS coverage downstream of the 28S mature rRNA end supported an exonucleolytic trimming process instead of a precise cleavage event within the 3' ETS intergenic region. We also noticed a smaller increase (4-fold) in the ITS1 and 5'ETS regions, which we attributed to an accumulation of the 47S precursor. As expected, the processing defects were also observable by Illumina total RNA sequencing, however read coverages were less contiguous, most likely due to the presence of extensive secondary structures not resolved by the Illumina protocol, a higher number mismatches against the reference sequence due to sequencing errors in longer reads and nucleotide polymorphisms [Figure 3.25 B (ii)]. Despite a role of the yeast REX1 family member in 5S rRNA exonucleolytic trimming (Piper et al., 1983), we did not detect any differences in abundance or 3' end length of 5S rRNAs for dNEF-sp mutants.

Gene expression of other abundant noncoding RNAs, such as tRNAs (extended by +/-20 nts), snoRNAs/snRNAs, transposon transcripts and annotated short noncoding RNAs (ncRNAs. e.g. miRNAs, piRNAs) did not show any significant differences in the relative abundance, except for one snoRNA cluster Me28S-G1083 and a few transposon groups (Copia, Cr1a, R1A1), which

were upregulated in mutants. The Me28S-G1083a-d C/D box snoRNA family of four snoRNAs is a conserved snoRNA family in humans (SNORD80) and *S. cerevisae* (snR60) (Yoshihama et al., 2013). The *Drosophila* Me28S-G1083a-d (FBgn0082935) cluster is located in the Nop60b gene and predicted to methylate 28S rRNA at position G1083 (Huang et al., 2005; Riccardo et al., 2007; Yoshihama et al., 2013). While the Me28S-G1083a-d cluster was expressed at low levels in wild type larvae (~100 library normalized read counts), it was reproducibly ~15-30 fold higher expressed (~1,500-3,000 library normalized read counts) in dNEF-sp mutants. However, although this was consistently observed for both mutants, we did not observe a similar upregulation after RNAi knockdown of dNEF-sp (next section). We also detected upregulation of a number of transposons, most notably Copia, which we also found to be upregulated in RNAi knockdowns of other rRNA processing factors (next section). We conclude that these are upregulated in response to cellular stress or rRNA processing defects.

3.3.14 *In vivo* RNAi knockdown of dNEF-sp and Ddx51 accumulate extended 3'ETS precursors

To test whether 3' ETS extension was specific for dNEF-sp or whether upregulation could also be observed for other rRNA biogenesis factors, I selected a number of transgenic RNAi fly lines with genomically integrated short hairpin RNAs (UAS-shRNAs) and expressed these constitutively under the TubGal4 driver. Only transgenic lines for which knockdown was lethal at the 1st or 2nd instar larval stage were selected. This served as an internal control that shRNA knockdown was strong enough to cause a physiological effect on rRNA processing. The choice of this *in vivo* system was important, as in a previous dsRNA-mediated knockdowns of dNEF-sp in S2 cells (>10 fold depletion) I could not detect the accumulation of the 3'ETS in the rRNA precursors after 3 days incubation and cell viability was not affected (Figure 3.26 A). Tin contrast to the cell culture system, *in vivo* shRNA knockdown of dNEF-sp and other known rRNA

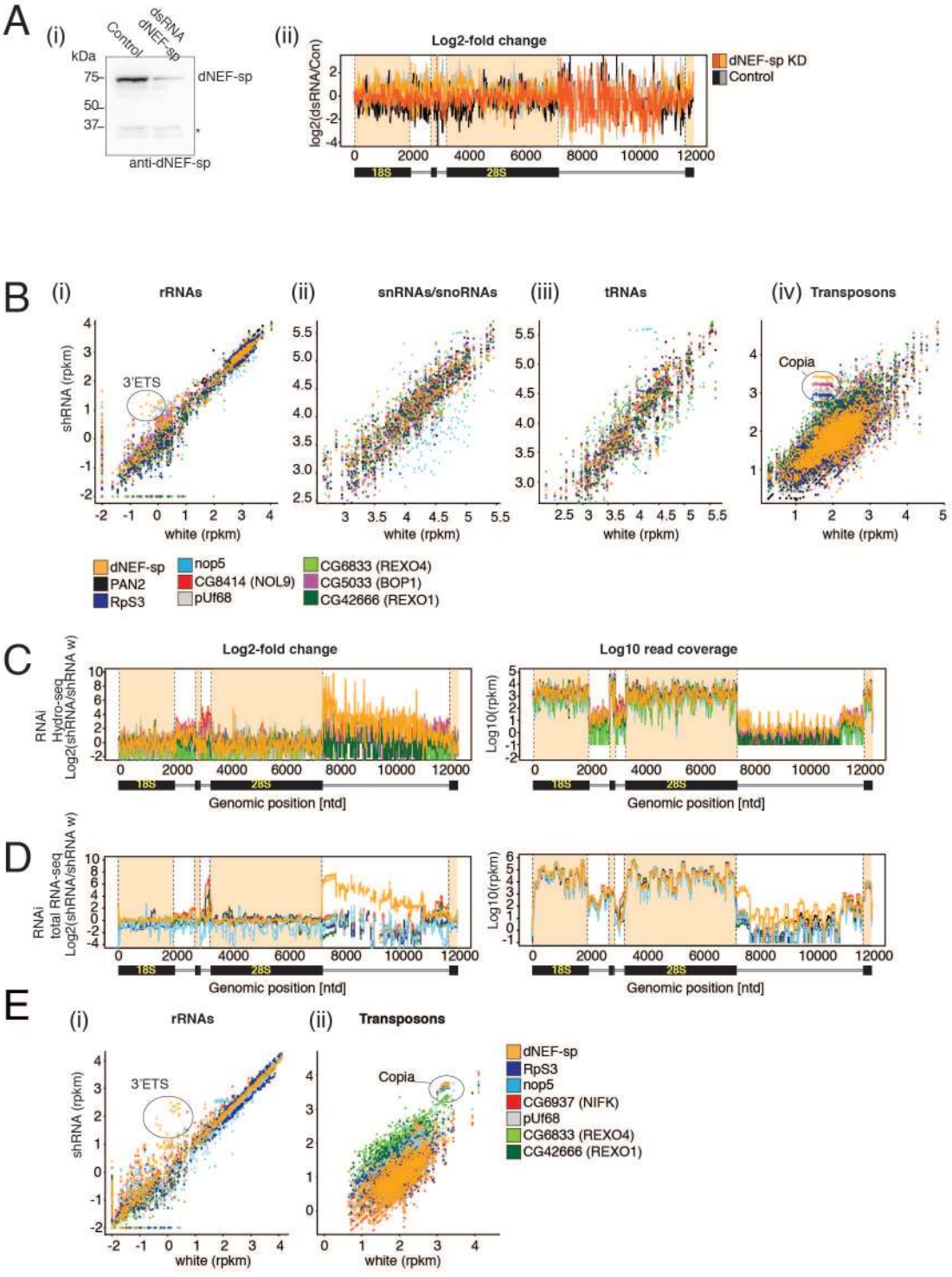
biogenesis factors resulted in 1st or 2nd instar larval death, confirming their essential function *in vivo*. Thus, screening for loss of viability ensured that the shRNA knockdown was strong enough to accumulate pre-RNA processing defects.

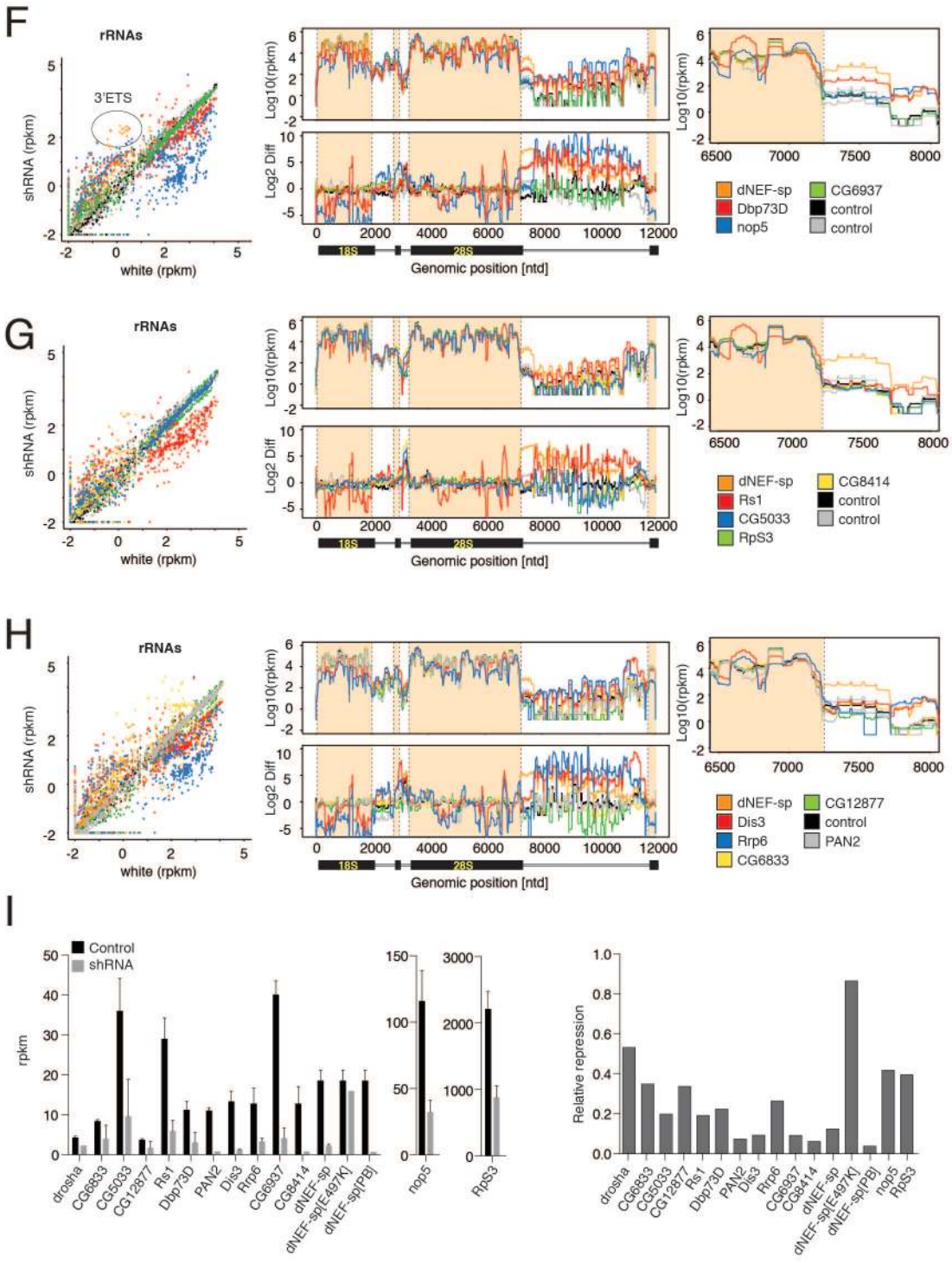
To compare pre-rRNA processing defects resulting from loss-of-function dNEF-sp with the loss of other rRNA biogenesis components, I selected rRNA biogenesis factors characterized to be involved in the processing at different sites of the rRNA precursor (e.g. pre-18S, -5.8, -28S maturation). The following pre-rRNA processing factors were chosen, which did not possess RNA nuclease activity: CG5033 (homolog of mammalian BOP1 and yeast ERB1, required for processing of the 32S precursor to give mature 28S rRNA (Strezoska et al., 2000), CG8414 [homolog of mammalian NOL9, involved in 5'-3' 5.8S rRNA trimming (Heindl and Martinez, 2010)], CG6937 [homolog of mammalian NIFK and yeast NOP15, involved in ITS1 processing (Tafforeau et al., 2013)], RpS3 (ribosomal protein, associates with early pre-90S ribosomal complexes), nop5 [homolog of mammalian NOP58, the box C/D snoRNP methylase, required for 18S rRNA processing, accumulates the large 47S, and the 18S 34S and 30S precursors (Dragon et al., 2002; Tafforeau et al., 2013; Wu et al., 1998)]. In addition, we chose a number of RNA nucleases, such as the RNA exosome catalytic 3'-5' exonucleolytic subunits, Dis3 and Rrp6, known to trim 5.8S precursor 3' ends (Briggs et al., 1998), and two members of the DEDDh RNA nuclease family CG8414 (homolog of mammalian REXO4), and CG42666 (homolog of mammalian REXO1 and closest homolog to yeast Rex1) (Figure 3.26 B-F). *In vivo* shRNA knockdowns of the other two Rex family members in *Drosophila melanogaster*, CG12877 (homolog of yeast Rex1p) and CG10214 (mammalian REXO2 homolog, yeast Rex2p) were viable (either because of insufficiently strong knockdowns or nonessentiality of the gene) and excluded from the analysis. Furthermore, the homologs of the two characterized protein factors involved in mammalian 28S rRNA 3'ETS end processing, DEAD box helicases DDX51 (*Drosophila* homolog Dbp73D) and DDX27 (*Drosophila* homolog Rs1) were also chosen

(Kellner et al., 2015; Srivastava et al., 2010). For both genes shRNA knockdowns were lethal. As negative controls the non-essential RNase T-class PAN2 3'-5' exonuclease, involved in poly(A) deadenylation, and the *Drosophila* non-essential *white* gene were chosen.

Accumulation of the 3' ETS was specific for dNEF-sp and Dbp73D (Figure 3.26 F). We did not detect processing defects for Rs1 under the tested conditions (Figure 3.26 G). None of the other rRNA biogenesis factors showed extended 3'ETS, but they did display accumulation at different sites or changes in overall abundance, which in some cases were easier to detect by Northern Blot than RNA-seq (Figure 3.26 B-G, Figure 3.28). None of the tested RNA exonucleases, including the RNA exosome components Dis3 and Rrp6 had an effect on 3'ETS removal (3.26 H). Under the tested conditions we also could not detect any involvement of the paralogous DEDDh RNA exonucleases (CG42666, CG12877) in 3'ETS trimming (Figure 3.26 H). For all tested factors, including dNEF-sp, we could not detect significant differential expression of tRNAs, snRNAs, snoRNAs, or small noncoding RNAs (Figure 3.26 B).

Figure 3.26 Total RNA-seq and Hydro-seq of RNAi knockdowns of rRNA processing factors show a unique function of dNEF-sp in 3'ETS removal. (A) dsRNA-mediated knockdown of dNEF-sp in S2 cells. (i) Western blot of dsRNA knockdown in S2 cells. dNEF-sp is depleted >10 fold, the star marks an unspecific protein band which serves as internal loading control. (ii) Hydro-seq log2 fold changes of the 47S precursor for dNEF-sp knockdown and control. (B-E) *In vivo* RNAi mediated by UAS-shRNA expression under the TubGal4 driver for dNEF-sp and selected ribosomal biogenesis genes. (B) Log10 Hydro-seq scatterplots for different abundant noncoding RNAs for dNEF-sp and selected ribosomal processing and control genes. (i) NCBI and BLAST created rRNA sequences (ii) annotated tRNAs with +/- 20 nt extensions, (iii) Flybase annotated snRNAs and snoRNAs, (iv) Flybase annotated transposons. shRNA knockdown of dNEF-sp (orange), mutant *M100* (bright orange), PAN2 control (black), RpS2 (dark blue), nop5 (light blue), CG81414 (homolog of human NOL9, red), splicing factor pUf68 (control, grey), CG6833 (homolog of human REXO4, light green), CG5033 (homolog of human BOP1, violet), CG42666 (homolog of human REXO1), plotted in correlation to shRNA knockdown of the *white* gene as control (white). (C-D) Log2 fold changes of read coverage over shRNA knockdown of *white* (control) across the 47S ribosomal precursor and log10 read coverage. (C) Hydro-seq fold changes and coverage for the same genes and color code as in (B). (D) Total RNA-seq fold changes and coverage for dNEF-sp (orange), RpS3 (dark blue), nop5 (light blue), CG6937 (NIFK), pUF68 (grey), CG6833 (light green), CG42666 (dark green). (E) Total RNA-seq scatterplots for long abundant noncoding RNAs for shRNA knockdowns shown in (D). (i) rRNA reference, (ii) Flybase transposon annotation, plotted in correlation to shRNA knockdown of *white* (shRNA w). (F-H) Total RNA-seq scatterplots for rRNAs (in log10 scale), log10 coverage along the 47S rRNA precursor with magnified 28S-3'ETS border, and log2 fold change difference plots along the 47S rRNA precursor for a number of shRNA *in vivo* knockdowns under the TubGal4 driver. For all *D. melanogaster* genes the human homolog is given in apprentices. (F) rRNA biogenesis factors: dNEF-sp (orange), Dbp73D (DDX51) (red), nop5 (NOP58) (blue), CG6937 (NIFK) (green), control white (black), control sequencing strain (grey). (G) dNEF-sp (orange), Rs1 (DDX27) (red), CG5033 (BOP1) (blue), RpS3 (green), CG8414 (NOL9), control white (black), control sequencing strain (grey). (H) dNEF-sp (orange), Dis3 (red), Rrp6 (blue), CG6833 (REXO4), CG12877 (REXO1), control white (black), PAN2 (grey). Genotypes: UAS-shRNA/Cyo, TubGal4/+ or +/Cyo, TubGal4/UAS-shRNA. (I) Left panel: mRNA-seq levels (in rpkm) of white and other gene knockdowns under the TubulinGal4 driver. Known genes involved in rRNA biogenesis or belonging to RNA exonuclease families were selected and compared to mRNA expression levels of $dNEF\text{-}sp^{E497K}$ and $dNEF\text{-}sp^{PB}$ mutants. All selected knockdowns (except for white, PAN2, selected controls, and drosha) caused death at the first or second developmental larval stage. Right panel: Relative repression of genes in the respective knockdowns compared to control white shRNA knockdown larvae.


A
(i)
Control
dsRNA dNEF-sp
kDa
75
50
37
dNEF-sp
anti-dNEF-sp
(ii)
Log2-fold change
log2(dsRNA/Con)
dNEF-sp KD
Control
18S
28S
B
(i) rRNAs
(ii) snRNAs/snoRNAs
(iii) tRNAs
(iv) Transposons
shRNA (rpkm)
3'ETS
Copia
white (rpkm)
dNEF-sp
PAN2
RpS3
nop5
CG8414 (NOL9)
pUf68
CG6833 (REXO4)
CG5033 (BOP1)
CG42666 (REXO1)
C
Log2-fold change
Log10 read coverage
RNAi Hydro-seq Log2(shRNA/shRNA w)
Log10(rpkm)
Genomic position [ntd]
D
RNAi total RNA-seq Log2(shRNA/shRNA w)
Log10(rpkm)
Genomic position [ntd]
E
(i) rRNAs
(ii) Transposons
shRNA (rpkm)
3'ETS
Copia
white (rpkm)
dNEF-sp
RpS3
nop5
CG6937 (NIFK)
pUf68
CG6833 (REXO4)
CG42666 (REXO1)


F
rRNAs
3'ETS
shRNA (rpkm)
white (rpkm)
Log10(rpkm)
Log2 Diff
18S
28S
Genomic position [ntd]
dNEF-sp
Dbp73D
nop5
CG6937
control
control
G
rRNAs
dNEF-sp
Rs1
CG5033
RpS3
CG8414
control
control
H
rRNAs
dNEF-sp
Dis3
Rrp6
CG6833
CG12877
control
PAN2
I
Control
shRNA
rpkm
Relative repression
drosha
CG6833
CG5033
CG12877
Rs1
Dbp73D
PAN2
Dis3
Rrp6
CG6937
CG8414
dNEF-sp
dNEF-sp[E497K]
dNEF-sp[PB]
nop5
RpS3

Upregulation of a number of transposons was observed in knockdowns of dNEF-sp and other rRNA biogenesis factors. Most upregulated were the transposons Copia, INE1 and accord. Environmental stress is known to activate transposons such as Copia, and the expression may be a consequence of cellular stress or increased genomic instability induced by defects in rRNA biogenesis (Capy et al., 2000; Strand and McDonald, 1985).

3.3.15 Loss-of-function of dNEF-sp does not affect mRNA expression

To investigate changes in mRNA metabolic pathways in dNEF-sp loss-of-function mutants I conducted differential gene expression analysis from mRNA-seq data of knockdowns of dNEF-sp (3 biological replicates) and the *white* gene (2 biological replicates). This had the advantage that, in contrast to mRNA-seq comparisons of the genetic mutants *C04255* and *M100*, these had the same genetic background in all experiments and therefore less expression noise compared to control conditions. dNEF-sp expression was downregulated 10-fold in dNEF-sp knockdowns in all three biological replicates (control ~20 rpkm, dNEF-sp RNAi ~2 rpkm). dNEF-sp knockdowns displayed less than 250 genes differentially expressed genes as assessed by different differential gene expression analysis methods (cuffdiff and DEseq/HTSeq) (Anders and Huber, 2010; Trapnell et al., 2012b). Only 223 genes showed significant differential expression (q value ≤ 0.05); of these 121 were upregulated while 102 were downregulated compared to *white* controls (Figure 3.27 A). Most enriched Gene Ontology and REACTOME pathways for upregulated genes were involved in cellular stress, immune and humoral response, and DNA repair (Ashburner et al., 2000; Croft et al., 2014). Most significantly, the components of the nonhomologous end joining (NHEJ) dsDNA repair pathway, the MRN complex [including Rad50, meiotic recombination 11 (mre11), and Nibrin (Nbs) (Ciapponi et al., 2006; van den Bosch et al., 2003)], as well as replication protein A 70 (RpA-70/Ku70), and Ku80 (XRCC5 homolog) (Fell and Schild-Poulter, 2015) were upregulated ~6-fold in mutants (Figure 3.27 D). Of the 102

downregulated genes most annotated ones were involved in cuticle/chitin production pathways, lipid digestions, as well as a number of (metallo-)peptidases were enriched.

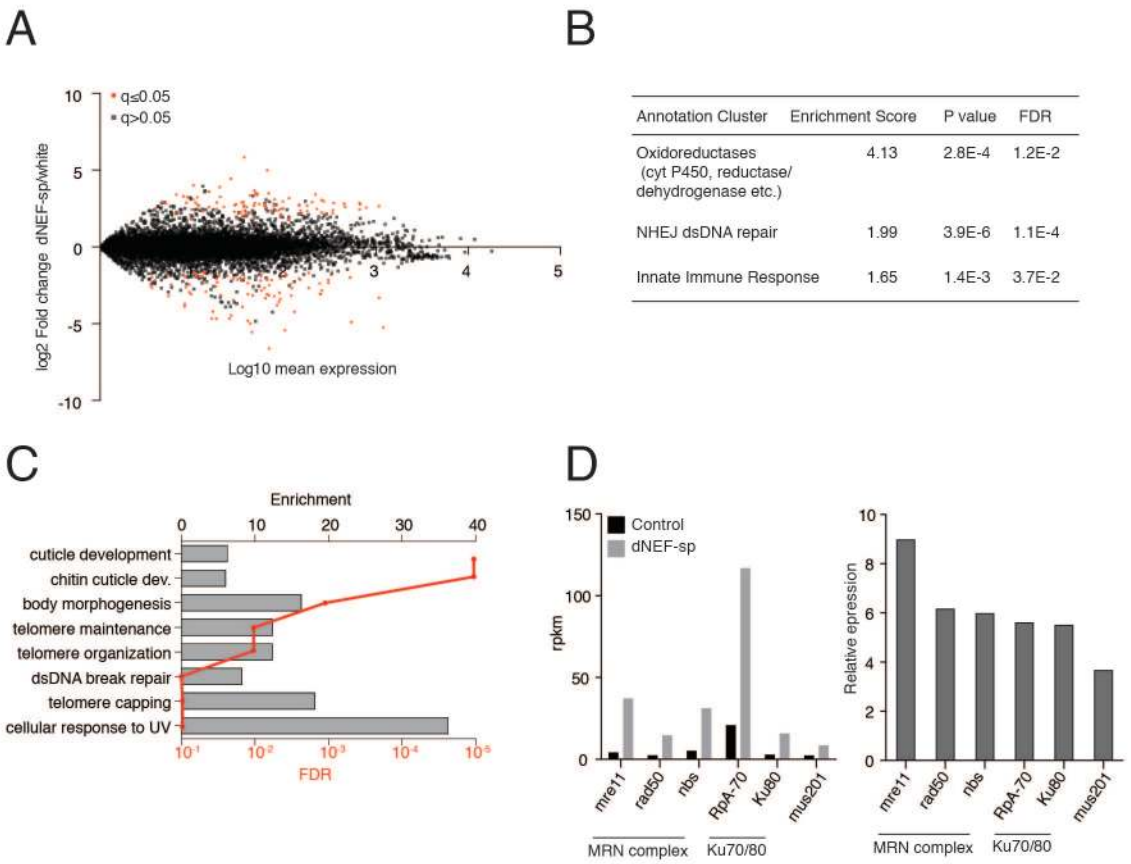


Figure 3.27 Differential mRNA expression upon dNEF-sp knockdown in *Drosophila* second instar larvae. (A) MA plot (log2 ratios over log10 mean average between the groups) of mRNA gene expression for dNEF-sp over *white* knockdown calculated with cuffdiff (Trapnell et al., 2012a). (B) Upregulated in mutants, enriched pathways in Gene Ontology pathway analysis. (C) GOrilla analysis of differentially expressed genes in dNEF-sp knockdown 2nd instar larvae (Eden et al., 2009). (D) Components of the nonhomologous healing and end joining (NHEJ) dsDNA damage repair pathway are upregulated by mRNA-seq more than 4-fold in dNEF-sp knockdown 2nd instar larvae. Shown are the mRNA expression of NHEJ components in rpkm for the control white (black) and dNEF-sp (grey) shRNA knockdowns (left panel), and their relative fold change expression in mutants over wild type (right panel).

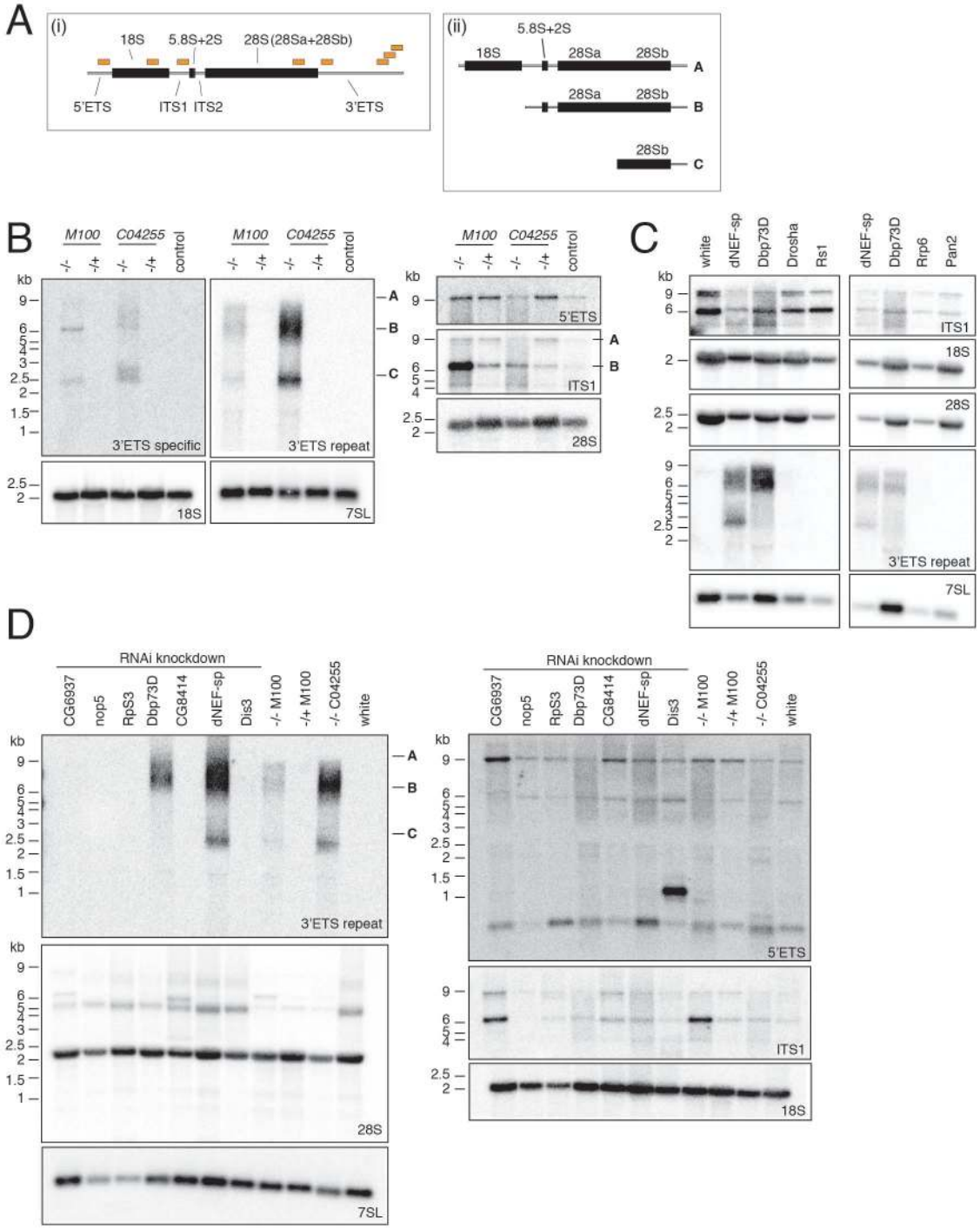
In conclusion, surprisingly few protein-coding genes were differentially expressed in mutants versus wild type larvae. These were involved in pathways related to energy supply, cytotoxic stress response and dsDNA repair (Figure 3.27 B). This data agreed with our previous

finding that dsDNA breaks increased in mutant testes (Figure 3.20). The majority of protein-coding genes, including genes involved in ribosomal biogenesis, did not change in their expression. Thus, we concluded that loss-of-function of dNEF-sp does not globally influence mRNA expression pathways and interpret the observed expression changes as a result of a general cell stress in response to deficient rRNA biogenesis, increased genomic instability, and deficient energy supply.

3.3.16 Northern Blot analysis shows rRNA precursor intermediates accumulate with a distribution of sizes

RNA-sequencing gave an unbiased view of genomic targets and allowed the detection of processing defects. However, in its two-dimensional information content it is difficult to estimate the accumulation of different intermediates and precursors of different sizes as seen in rRNA biogenesis. To address the size distribution of the 3'ETS misprocessed transcripts I carried out Northern blot analysis on total RNA isolated from 2nd instar larvae to (1) confirm the defects detected by RNA-seq and (2) to gain further insights into the size distribution of the rRNA precursors. Probes specific to the 3'ETS showed a specific signal for homozygous larvae in both mutants. Misprocessed rRNA precursors showed a range of products between ~6 and 9 kb in size (by ssRNA size ladder), and around 2.5 kb. These corresponded to the full size 7.3 kb precursor and the 4.6 kb precursor intermediate b (containing ITS1, 5.8S, ITS2, and 28S) and the mature 28Sb rRNA (2.3 kb) (Long and Dawid, 1980) (Figure 3.28 A,B). Insect rRNAs contain a hydrolytic cleavage site in the center of 28S rRNA that generates two 28S fragments, 28Sa and 28Sb, which run at similar electrophoretic mobility as 18S rRNA (Jordan, 1975) (Figure 3.2). Confirming the RNA-seq results, misprocessed intermediates did not have a defined size, but displayed a continuous size range. Fragments with length of the ~28Sb rRNA were longer in length, above the correctly processed mature 28Sb rRNA.

Figure 3.28 dNEF-sp mutants accumulate a distribution of rRNA precursor intermediates. Northern Blot analysis of rRNA precursors in mutant 2nd instar larvae. (A) (i) Location of Northern blot probes are shown on the 47S pre-rRNA precursor in the 5'ETS, ITS1, mature 18A and 28S rRNA, 3'ETS (unique sequence, labeled 3'ETS specific), 3'ETS (repetitive sequence within the 3'ETS, labeled 3'ETS repeat); (ii) sizes of different 28S rRNA precursors (A to C). (B) Northern Blot of both genetic mutants of dNEF-sp, RNA is loaded from homozygotes (-/-), heterozygotes (-/+) and wild type control 2nd instar larvae. Blot was stripped and reprobed for all probes. (C-D) Northern Blots of shRNA-mediated knockdowns of known ribosomal RNA processing genes [CG6937 (MKI67IP), nop5, RpS3, CG8414 (NOL9)], known factors involved in 28S pre-RNA 3' maturation: Dbp73D (DDX51) and Rs1 (DDX27). Also tested is total RNA isolated from shRNA knockdowns of the RNA exosome nucleases Rrp6 and Dis3. In (C) shRNA knockdowns are also compared with the genetic loss-of-function dNEF-sp mutants (homozygous M100 and C04255). shRNA knockdown of the *white* gene and the polyA mRNA-specific DEDDh RNA exonuclease Pan2 serve as controls. Genotypes: heterozygotes: dNEF-sp^{M100}/TM6B and dNEF-sp^{C04255}/TM6B; homozygotes: dNEF-sp^{M100}/dNEF-sp^{M100} and dNEF-sp^{C04255}/dNEF-sp^{C04255}, wild type: sequencing strain. shRNA knockdowns: UAS-shRNA/Cyo, TubGal4/+ or +/Cyo; TubGal4/UAS-shRNA.


A
(i)
18S
5.8S+2S
28S (28Sa+28Sb)
5'ETS
ITS1
ITS2
3'ETS
(ii)
5.8S+2S
18S
28Sa
28Sb
A
28Sa
28Sb
B
28Sb
C
B
M100
C04255
control
-/-
-/+
kb
3'ETS specific
3'ETS repeat
18S
7SL
5'ETS
ITS1
28S
C
white
dNEF-sp
Dbp73D
Drosha
Rs1
Rrp6
Pan2
ITS1
18S
28S
3'ETS repeat
7SL
D
RNAi knockdown
CG6937
nop5
RpS3
Dbp73D
CG8414
dNEF-sp
Dis3
-/- M100
-/+ M100
-/- C04255
white
3'ETS repeat
28S
7SL
5'ETS
ITS1
18S

From the processing patterns it remains unclear whether 28S rRNA processing continuous at the 5' end keeping the extended 3'ETS trailers until the final step or whether the 2.3 kb fragments of 28Sb rRNA is the result of an early hydrolytic cleavage of the misprocessed precursor. Because of the hydrolytic cleavage of insect 28S rRNA one cannot unequivocally exclude either, however, we found that this effect was specific for dNEF-sp and not observed for Dbp73D RNA helicase, the Ddx51 homolog involved in 28S rRNA processing (Figure 3.28 C), which may suggest that 28S rRNA processing can partially continue to the 28Sb product.

As detected by RNA-seq, processing defects were several fold stronger for the transposon insertion line *C04255* compared to the EMS point mutant *M100*. Furthermore, probes specific to ITS1 revealed the accumulation of 4.6 kb precursors. In mammals and yeast, maturation of the 5′ end of the 5.8S rRNA is coordinated with formation of the 3′ end of the 25S/28S rRNA and both events are coupled (Henras et al., 2014). The accumulation of the ITS1 (Figure 3.28 B-C) and ITS2 suggest a partial coupling of 5' end 5.8S and 3' end 28S formation in *Drosophila* similar to U8 snoRNA-mediated processing described in *Xen. laevis* (Peculis and Steitz, 1993).

Comparison of pre-rRNA processing defects in shRNAs knockdowns for the panel of rRNA biogenesis factors and dNEF-sp showed a similar accumulation of extended 3'ETS precursors in the shRNA mediated knockdown of dNEF-sp. Notably, by Northern Blot analysis some defects on pre-rRNA processing were easier to detect than in a two-dimensional representation resulting from the RNA-seq analysis. Perhaps in a tiling array-like method, selecting short regions within the rRNA precursors and mapping RNA-seq read abundances against these will allow comparisons of the relative rRNA precursor ratios. Accumulation of intermediates was observed for all tested biogenesis factors, confirming that knockdown of the selected proteins led to pre-rRNA processing defects (Figure 3.28 C). Furthermore, Dbp73D also displayed a range of accumulated 3'ETS extended precursors at 9 kb, confirming an early role of Dbp73D in 3'ETS removal (Figure 3.28 C). In contrast to dNEF-sp, accumulation of

intermediates of 6 and 2.5 kb were not observed for Dbp73D knockdowns, suggesting that dNEF-sp and Dbp73D action are not coupled and some defective precursors can be further processed in dNEF-sp mutants.

3.3.17 dNEF-sp crosslinks to RNA *in vivo*

To investigate whether dNEF-sp binds RNA *in vivo*, I designed several transient transfection constructs of dNEF-sp with a mutated catalytic domain and an N-terminal HA/FLAG-tag, which allowed immunoprecipitation of the tagged protein using a monoclonal anti-FLAG antibody. Colloidal blue staining could detect immunoprecipitated and FLAG-eluted HA/FLAG-dNEF-sp protein (Figure 3.29 A). Immunoprecipitation of 4-SU labeled RNA-protein complexes followed by RNA radiolabeling (PAR-CLIP) showed that dNEF-sp protein immunoprecipitated crosslinked RNA (Figure 3.29 B). A second endogenous and highly abundant protein in S2 cells, recognized by FLAG monoclonal antibodies, was also immunoprecipitated and displayed an RNA crosslinking band (Figure 3.29 B). Currently, large-scale PAR-CLIP and protein mass spectrometry experiments are underway to identify bound RNA fragments and protein complexes by next generation sequencing and quantitative mass spectrometry. In addition to confirming the direct binding of NEF-sp to the 3'ETS and identifying additional RNA targets transcriptome-wide, PAR-CLIP analysis will elucidate the consensus binding motif required for dNEF-sp to recognize its targets. Overall, our analysis will characterize the interactions of dNEF-sp with its targets and protein partners and define the molecular complex during ribosome biogenesis.

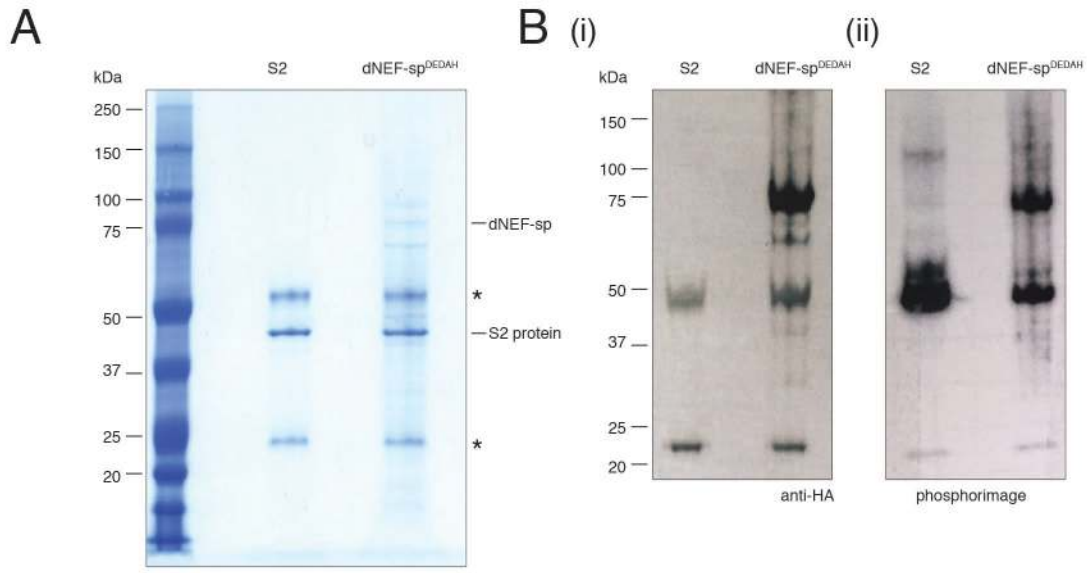


Figure 3.29 dNEF-sp crosslinks RNA *in vivo*. Immunoprecipitation of HA/FLAG-dNEF-sp mutants and untransfected S2 cells as control. (A) Colloidal Blue staining of FLAG-immunoprecipitated and FLAG peptide- eluted dNEF-sp^{DEDAH} protein and control untransfected S2 cells. Residual antibody eluted is marked with a star. An unknown protein of ~50 kDa is known to immunoprecipitate with FLAG antibody from S2 cells (Miyoshi et al., 2013). (B) (i) Western blot of FLAG immunoprecipitates crosslinked S2 cell lysates expressing dNEF-sp^{DADAH} mutant protein and untransfected S2 cells. (ii) Phosphorimage of RNA-radiolabeled FLAG immunoprecipitates from (i).

3.3.18 dNEF-sp mutant cells accumulate unprocessed 3'ETS in the nucleolus and display increased nucleolar size

While examining second instar larval tissues we noticed an increased nucleolar size in homozygous mutant gut cells. In agreement with this observation a previous screen for regulators of nucleolar size in S2 cells had found that dNEF-sp knockdown increased nucleolar size 2-fold (Neumuller et al., 2013). To measure the extent to which nucleolar structure was affected in dNEF-sp deficient cells *in vivo*, we used clonal expression of RNAi mediated by shRNA knockdowns. shRNA knockdown of dNEF-sp under the hsFlp-tubGal4 driver resulted in 4-fold increase of the nucleolus in NEF-sp knockdown cells (Figure 3.30 A).

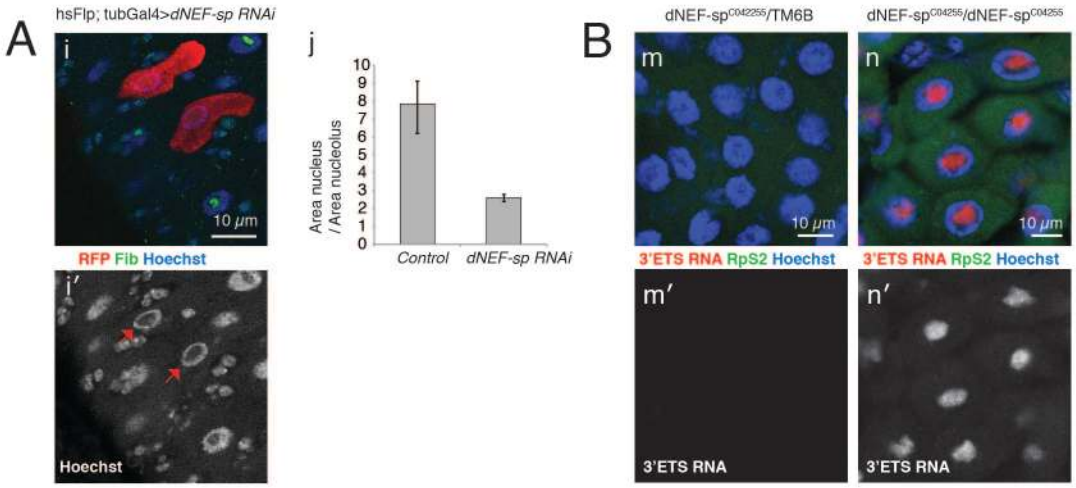


Figure 3.30 Loss-of-function of dNEF-sp increases nucleolar size in *Drosophila* 2nd instar larvae. (A) hsFlp-mediated shRNA knockdown of dNEF-sp in mid gut cells of 2nd instar larvae. Cells expressing UAS-shRNA against dNEF-sp also express UAS-Red47. (i) Confocal 3-color image of mutant and wild type mid gut cells. Mutant cells are red (RFP), the nucleolus is marked with Fibrillarin (Fib, green), the nucleus stained with Hoechst dye (blue). (i') Nuclear staining with Hoechst shows increased nucleolar size (absence of Hoechst) for shRNA mediated RNAi knockdown cells (marked with red arrows). (j) Quantification of the ratio of nuclear/nucleolar area size for wild type and mutant cells in mid gut cells. Wild type cells usually have a ratio Area (nucleus)/ Area (nucleolus) 8:1, while mutant cells display a ratio of 2:1, i.e. in mutant cells the nucleolus occupies half of the nucleus and is ~4 times larger than in wild type cells. (B) Fluorescent RNA *in situ* hybridization probing against the repeat sequence in the 3'ETS (red) in a dNEF-sp mutant background expressing RpS2-GFP (green). Hoechst dye is shown in blue. In homozygous mutants the unprocessed 3'ETS is retained in the nucleolus, which increases in size. In heterozygous mutants (viable, no phenotypes), no unprocessed 3'ETS can be detected and nucleolar size is unchanged. Genotypes: (A) hsFlp; UAS-shRNA dNEF-sp (P{KK101144}VIE-260B)/Cyo, UAS#Red47a#1 tub<+GFP<Gal4/TM6B. (B) homozygotes: RpS2-GFP/Cyo, dNEF-sp^{C04255}/dNEF-sp^{C04255}; heterozygotes: RpS2-GFP/Cyo, dNEF-sp^{C04255}/TM6B.

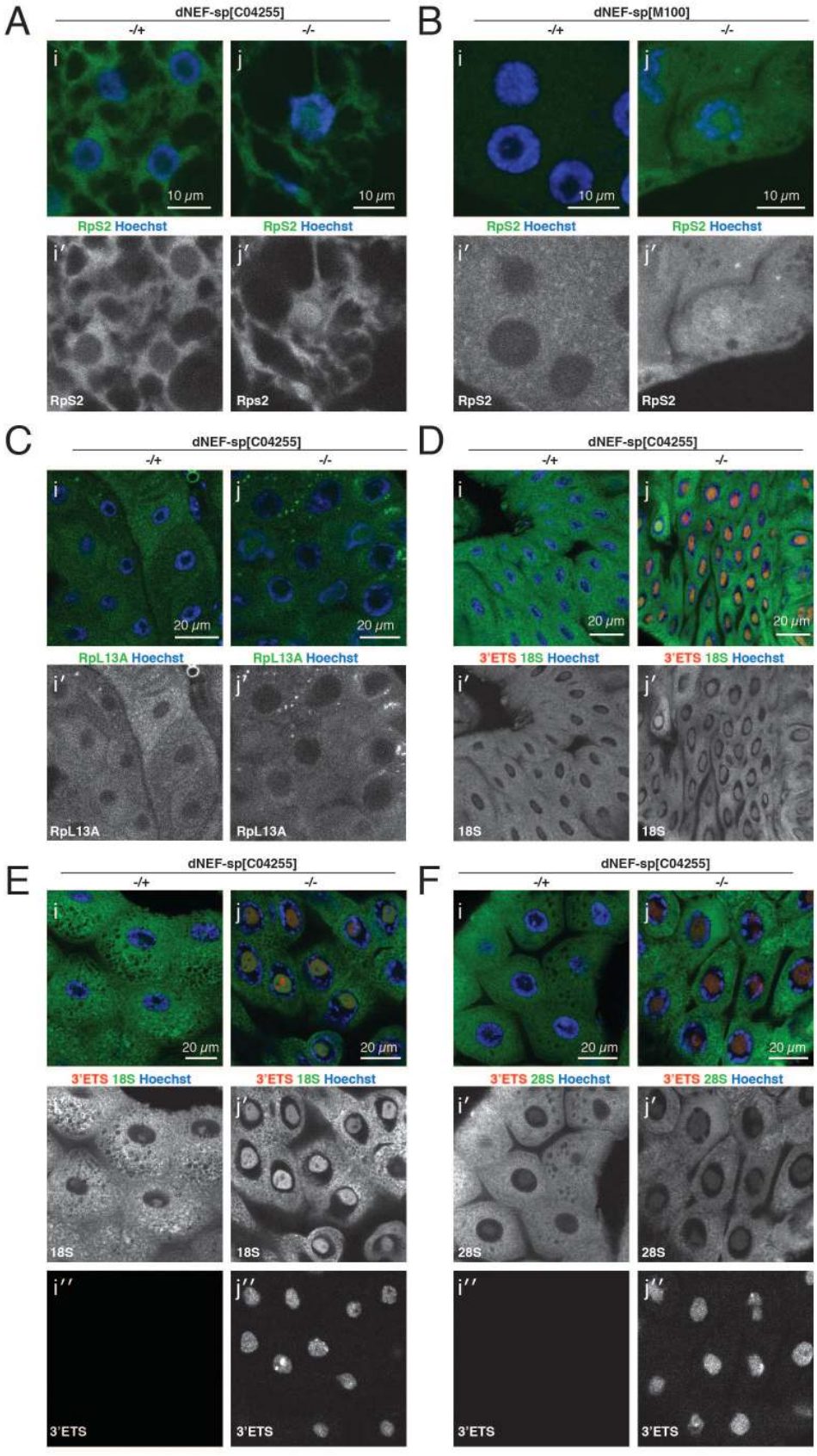
Since the Northern Blot results indicated aberrant 28S fragments close to the size of mature 28S rRNA, we were curious to see whether loss-of-function of dNEF-sp resulted in misprocessed 28S rRNAs being exported to the cytoplasm or whether these aberrant 28S precursor transcripts were retained in the nucleolus. Therefore, to assess 3'ETS export we synthesized *in situ* LNA probes specific to a repeated sequence within 3'ETS, which allowed a natural amplification of the signal. RNA fluorescent *in situ* hybridization (RNA FISH) experiments showed specific signal of the extended 3'ETS region only in homozygous mutants,

where it was exclusively confined to the nucleolus, while no signal was detected in heterozygous larvae. Hence, aberrant 3'ETS-extended 28S rRNA precursors were not exported from the nucleolus. The increase in nucleolar size may therefore be the result of the accumulation of incorrectly processed 3'ETS fragments (Figure 3.30 B).

3.3.19 Loss-of-function of dNEF-sp impairs ribosome export

In order to determine whether misprocessing also affects ribosome export we combined a RpS2-GFP reporter line with our mutant backgrounds (Buszczak et al., 2007). Heterozygote flies, which were essentially wild type in viability and phenotype, displayed predominantly cytoplasmic localization of the GFP-tagged ribosomal proteins (Figure 3.31). In contrast, homozygous mutants showed nuclear/nucleolar retention of RpS2-GFP and retention of RpL13A-GFP, although nucleolar retention of Rpl13A was only found in few mutant cells. Nucleolar retention of RpS2-GFP most strongly accumulated in foregut cells, but was also observed in other tissues, such as salivary glands and fat cells (Figure 3.31 A-C). RNA FISH against the 3'ETS showed that the 3'ETS strongly accumulated in these tissues (Figure 3.31 D-F). In addition to an increased nucleolus, overall cell morphology was also deformed and nuclei were larger relative to the cell body (Figure 3.31 D). In all tissues, processing defects of the 3'ETS affected export of the small ribosomal subunits (RpS2-GFP) and showed stronger nucleolar staining of 18S rRNA (Figure 3.31 D-E). RNA FISH against mature 28S rRNA did not show, apart from a larger nucleolar size in homozygous mutants, significant nucleolar accumulation in mutants (Figure 3.31 F). Together with the data on RpL13A, these results suggested to us that the observed phenotypes of 3'ETS misprocessing took place early in rRNA biogenesis, thereby affecting the export of the small ribosomal subunit first.

Figure 3.31 Ribosome export is impaired in dNEF-sp mutants. (A-B) RpS2-GFP protein trap in the dNEF-sp mutant backgrounds shows nucleolar/nuclear retention of RpS2-GFP in homozygous mutants. In heterozygous mutants (no phenotypes, viable) RpS2-GFP is exported to the cytoplasm. Shown are RpS2-GFP (green) and DNA staining (Hoechst, blue). (A) (i,i') dNEF-sp^{C04255}/dNEF-sp^{C04255}, fat body; (j,j') dNEF-sp^{C04255}/TM6B, fat body. (B) (i,i') dNEF-sp^{M100}/TM6B, salivary glands; (j,j') dNEF-sp^{M100}/dNEF-sp^{M100}, salivary glands. (C) *In vivo* localization of RpL13A-GFP in the C04255 dNEF-sp mutant background shows nucleolar/nuclear retention of RpL13A-GFP in homozygous mutants for some cells, foregut. (D-E) RNA FISH of the 3'ETS (red) and mature 18S rRNA (green) in foregut cells in homozygous and heterozygous *C04255* mutants. 18S rRNA shows nucleolar accumulation, the misprocessed 3'ETS RNA is not exported. Cells show increased nucleolar size and altered nuclear and overall cell morphology. (F) RNA FISH of the 3'ETS (red) and mature 28S rRNA (green) shows nuclear retention of the 3'ETS, but unaltered distribution of mature 28S rRNA.


A
dNEF-sp[C04255]
-/+
-/-
i
j
10 μm
RpS2 Hoechst
i′
j′
RpS2
Rps2
B
dNEF-sp[M100]
C
RpL13A Hoechst
RpL13A
D
3′ETS 18S Hoechst
18S
E
20 μm
3′ETS
F
3′ETS 28S Hoechst
28S
i′′
j′′

28S rRNA is incorporated into the large ribosomal subunit and the small and large subunits are exported in separate pathways, with the small subunit undergoing export more rapidly (Figure 3.1). In light of our data, we interpret the observed nucleolar retention of the small ribosomal subunit that either removal of 3'ETS takes place early on in rRNA biogenesis and thus faulty processing affects early biogenesis or, alternatively, that accumulation of defective 3'ETS precursors sends a negative feedback to halt export of the small ribosomal subunit.

3.4 Discussion

3.4.1 3'-5' exonucleolytic trimming of the 3'ETS of 47S pre-rRNA is a conserved mechanism across lower and higher eukaryotes

In this chapter I identified and characterized the DEDDh 3'-5' RNA exonuclease dNEF-sp and showed that it is involved in the exonucleolytic trimming of the 3'ETS rRNA precursor. I characterized the function of dNEF-sp *in vivo* and *in vitro* and showed that 3'ETS removal is an essential process in *Drosophila melanogaster*. In absence of dNEF-sp and other factors required for 3'ETS removal, such as Dbp73D (DDX51), misprocessed rRNA precursor transcripts accumulate in the nucleolus and impair ribosome export to the cytoplasm. Through an unbiased RNA sequencing approach, profiling small and large noncoding RNAs, we investigated abundant noncoding RNAs transcriptome-wide and studied targets and regulatory effects of dNEF-sp and other ribosomal RNA processing factors. I found that dNEF-sp is unique in its role of removing the 3'ETS in *Drosophila melanogaster*. The conserved expression and nucleolar localization of the *Drosophila* and human NEF-sp homologs may imply a conserved function of NEF-sp proteins in higher eukaryotes. Contrary to previous models of 28S rRNA 3' end maturation, we find that 3'-5' exonucleolytic trimming is an essential process during 28S rRNA biogenesis in higher eukaryotes. dNEF-sp is the enzyme responsible for 28S rRNA 3' end trimming and generation of mature 28S rRNA ends in *Drosophila melanogaster*. In *S. cerevisiae* the distant

homolog of NEF-sp, Rex1p, generates the mature 28S rRNA 3' end by exonucleolytically trimming the 28S rRNA precursors after co-transcriptional cleavage of Rnt1p that releases the 37S rRNA precursor [Figure 3.31 A (i)]. Future studies of the human and mouse NEF-sp homologs, which have been briefly started here, will be needed to answer the role of NEF-sp in mammalian rRNA biogenesis, its tissue-specific expression, and how its domain divergence from lower to higher eukaryotes has evolved to recognize its RNA targets. Furthermore, biochemical purifications of the protein complex of NEF-sp are the next steps to yield a molecular understanding of the recruitment and processing factors involved in 3'ETS removal and to delineate the precise mechanism of release of the 47S rRNA precursor.

In higher eukaryotes, a direct endonucleolytic cleavage is thought to occur, in which cleavage is guided by the C/D box snoRNA U8 (Peculis, 1997). The enzymatic factor catalyzing such cleavage has not been characterized so far and the mammalian and the *Drosophila* homologs of Rnt1p, Drosha, do not have a role in 3'ETS cleavage (Chong et al., 2008; Smibert et al., 2011). The DEAD box helicase Ddx51 has previously been shown to be required for correct 3' end processing by unwinding and releasing U8 snoRNA from 5.8S-28S-binding (Srivastava et al., 2010) [Figure 3.31 A (ii)]. In agreement with the mammalian Ddx51 function, we find that the function of the *Drosophila* homolog Dbp73D is conserved and knockdown of Dbp73D displays 3'ETS processing defects similar to dNEF-sp [Figure 3.31 A (iii)].

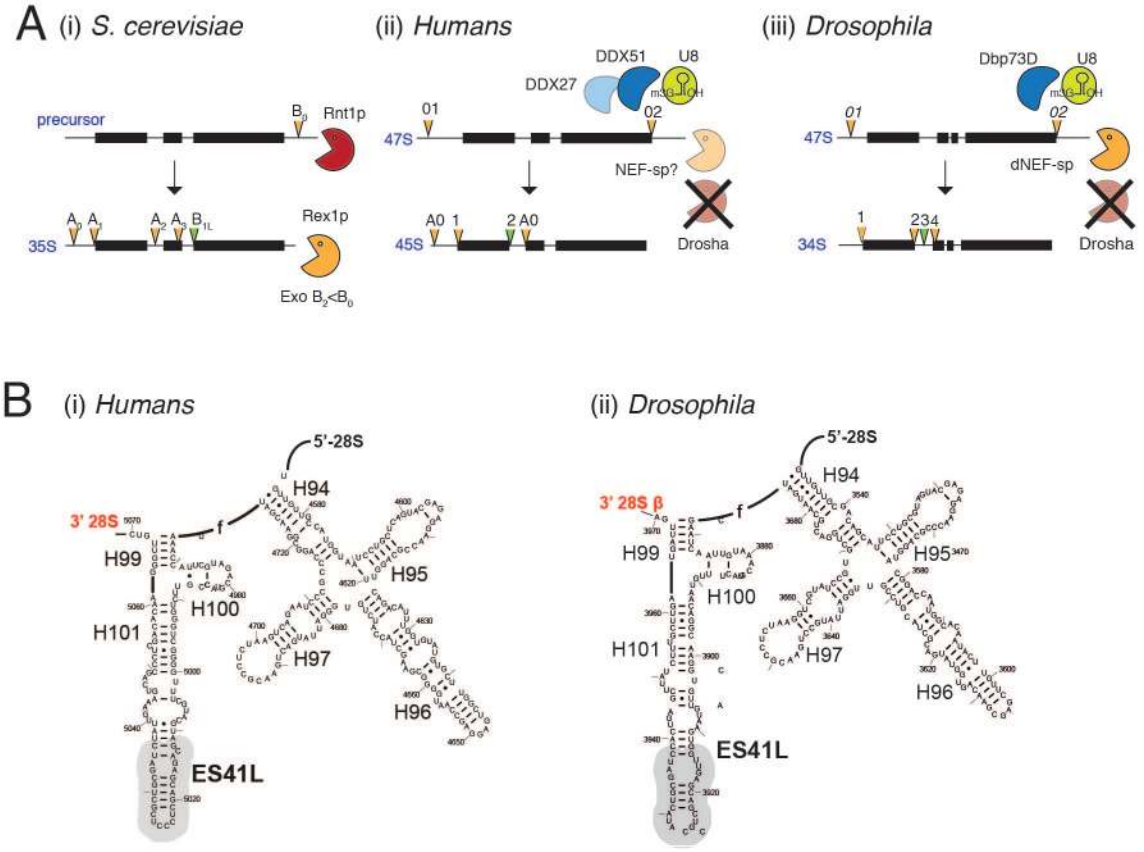


Figure 3.32 Current view of 28S rRNA 3' end maturation in *S. cerevisiae*, humans, and *Drosophila*. (A) Schematic comparison of 28S rRNA 3' end maturation in *S. cerevisiae*, human and *Drosophila*. (i) In *S. cerevisiae* Rnt1p endonucleolytically cleaves the 3'ETS and releases the 37S precursor transcript. Rex1p trims the residual 3'ETS overhangs to generate the mature 3' end of 28S rRNA. (ii) In humans and other mammalians, as well as in *Xenopus laevis*, a homologous endonucleolytic cleavage reaction by the Rnt1p homolog DROSHA has not been identified, and cleavage or processing factors maturing the 3'end of 28S rRNA remain uncharacterized. U8 snoRNA is known to basepair to 28S rRNA and guides a proposed direct endonucleolytic cleavage reaction at the 3' end of 28S rRNA. The DEAD box helicase DDX51 is required for unwinding of U8 snoRNA and release of the rRNA precursor. DDX27 has been implicated in 3'ETS end definition and knockdowns of DDX27 display precursors, which extend beyond the defined end of the 3'ETS. (iii) In *Drosophila* melanogaster, a conserved U8 snoRNA homolog has been identified. This study shows that the DDX51 homolog Dbp73D is functionally conserved and knockdowns display, like the mammalian DDX51 protein, extended 3'ETS precursors in *D. melanogaster*. Loss-of-function dNEF-sp mutants accumulate extended 3'ETS rRNA precursors, showing that dNEF-sp exonucleolytic processing is a required step in 28S 3'end maturation in *D. melanogaster*. (B) RNA secondary structure of the mature 3' end of human and *Drosophila* 28S rRNA show that the nucleotides at the mature 3' end differ, however secondary structure remains conserved. Figure adapted from (Anger et al., 2013).

Our study provides a fresh perspective on 28S rRNA biogenesis in higher eukaryotes. Previous works suggested a model in which a precise endonucleolytic cleavage event at the 28S-3'ETS transition released the mature 3' end of 28S rRNA, guided by U8 snoRNA. However, our data clearly supports a model that requires exonucleolytic trimming to generate mature 3'ends of 28S rRNAs, similar to *S. cerevisae*. The conserved base-pairing interactions of U8 snoRNA may ensure the required specificity to define the mature 28S rRNA end. In support of such a mechanism, *in vivo* transcripts have been found to readily extend beyond the T1 site in *Xenopus laevis* and previous work found that endonucleolytic cleavage was followed by rapid exonucleolytic trimming to generate the 40S precursors (Labhart and Reeder, 1986). Processing at the cleavage site of mature 28S 3' ends (T1) in *Xenopus laevis* (the mature 28S 3' end site) also takes place *in vitro* in fractions of impure albumin serum, suggesting that the nucleolytic reaction requires little or none sequence specificity to generate the mature 28S 3' ends (Labhart and Reeder, 1986). Given the heterogeneity and lack of conservation of 3'ETS sequences across organisms (Figure 3.31 B), an exonucleolytic mechanism may also be preferred over an endonucleolytic cleavage, which requires specific base pair recognition and conservation.

3.4.2 The role of tissue-specific levels of rRNA biogenesis factors

dNEF-sp exhibits high tissue specificity, which is conserved across metazoans. Every cell needs to synthesize rRNAs abundantly and general pre-rRNA processing factors involved in cleavage and maturation of rRNAs usually do not show regulated gene expression across tissues. Thus the restricted expression of one of its essential rRNA processing factor is mechanistically unexpected and points towards tissue-specific regulation of rRNA biogenesis.

Posttranscriptional gene regulation of rRNA modifications in different cell types and tissues is known to lead to cell-type specific variations in processing pathways and rRNA intermediates. Heterogeneity of rRNA modifications and changing ratios of $5.8S_{S/L}$ rRNA in

different cell types are among the reported examples (Lafontaine, 2015; Mullineux and Lafontaine, 2012). Furthermore, tailoring translational efficiency of ribosomes to specific mRNA targets by tissue-specific incorporation of ribosomal proteins (or in response to cytosolic stress) has also recently been reported (Kondrashov et al., 2011; Vesper et al., 2011; Zhang et al., 2015b). However, tissue-specific regulation of rRNA biogenesis remains currently unknown.

The tissue-specific expression of dNEF-sp may suggest (1) a tissue-specific regulation/function or targets of dNEF-sp or (2) 28S rRNA 3' end maturation requires the involvement of other nucleases, which have overlapping specificities. In our study we have not found tissue-specific RNA targets of dNEF-sp. Maturation of the 3'ETS through several RNA exonuclease with some overlapping specificity would allow tissue-specific regulation of 28S rRNA maturation through e.g. signaling pathways that modulate activity of rRNA biogenesis factors. Thereby they can directly influence posttranscriptional processing steps in rRNA biogenesis according to the demands of the specific cell type or tissue. However, in this study, we could not identify any overlapping specificities of NEF-sp and its family paralogs, or other 3'-5' RNA exonucleases, such as those of the RNA exosome Rrp6 and Dis3, which have been described in 5.8S 3' end maturation. In conclusion, we find that, at present, the role of dNEF-sp in 3'ETS maturation of 28S rRNA is unique.

4 Conclusion

In this short chapter I aim to provide a brief summary and highlight unanswered questions this work raises. In chapter 2, I studied characteristics of PTGR from a genome-wide perspective. The central role of PTGR in cellular metabolism can be appreciated by considering the large number of proteins interacting with RNA. Over 1,500 of the 20,500 unique human proteins are directly contributing to PTGR. RBPs form many distinct families with few members and human RBPs can be grouped into ~1,100 paralogous families related by greater than 20% sequence identity.

The complexity of PTGR was established early in evolutionary time scales. The earliest common ancestor of metazoans had a set of ~200 RBPs (Kerner et al., 2011), and of the ~1,100 human RBP families ~600 families have homologs in yeast. Consistent with their high degree of conservation, most RBPs (98%) do not display highly tissue-specific expression, but they are abundant and make up to 25% of the expressed transcripts encoding for cellular proteins. Interestingly, dysregulation of ubiquitous and general components in PTGR often shows highly tissue-specific phenotypes; for instance defects involving mRBPs are most frequently associated with neurological diseases, especially of the peripheral nervous system.

Given that the common RBD folds have been characterized and the majority of RBPs do not fall into large families, novel RBPs are most probably singular or have recently evolved RNA-binding activity independent of their family. This makes RBP prediction challenging and leaves experimental approaches as the most suitable strategy for their identification. Genome-wide experimental methods such as covalent RNA-protein cross-linking coupled with high-throughput sequencing to identify RNA target sites, or combined with mass spectrometric approaches to identify proteome-wide RBPs cross-linked to RNAs, will advance these efforts towards the elucidation of novel RBP and with increasing sensitivity of these experimental approaches, the number of RBPs is likely to grow. Based on our current collective data and recent experimental genome-wide studies, which added less than <50 novel RBPs to our census, we

believe that most RBPs or their homologs (>95%) have been identified and we will see fewer novel discoveries in the future. Even at present though, the precise biological functions and RNA targets for the majority (>50%) of the known human RBPs have not been characterized. For example, rRNA and tRNA-biogenesis components and their targets are often inferred by homology, as well as many of the proteins with canonical RBDs have not been characterized in humans. Some of the RNA regulatory processes, such as noncoding RNA maturation pathways, control of RNA transport, the role of RNA modifications, sensing of intracellular RNA, RNA as structural component for genome organization (Ding et al., 2012) or regulation (Cech and Steitz, 2014), as well as mechanisms of RNA/RNP clearance, remain poorly understood. Thus, the main challenges in the field are the characterization of these processes and the mechanisms leading to human disease.

Much of the dynamics and the integration of co-regulatory functions of RBPs in PTGR regulation remain to be elucidated. We are still in the process of profiling the targets and binding affinities for most RBPs. In the future, with a more detailed knowledge of targets and co-regulated expression of RBPs in developmental processes, we can begin to study their integration into different regulatory pathways, and ask questions on how for example signaling pathways and innate immunity are connected to RNA metabolism and neuroplasticity (Figure 2.11). Stoichiometries and time scales strongly determine regulatory outcome, thus studying dynamic assembly processes in RBP regulation will be important to understand the final outcome in gene expression control.

Through *in silico* investigation of novel PTGR factors I identified and characterized two unknown proteins, the RG/RGG-rich RBP FAM98A and the conserved 3'-5' RNA exonuclease NEF-sp. Our findings shed light on one of the members of the poorly characterized DEDDh RNase T exonuclease class and investigated in detail the mechanism of 3'ETS 28S rRNA maturation in higher eukaryotes. We show here that 28S rRNA undergoes exonucleolytic

trimming, a process conserved from yeast to higher eukaryotes. Still the question remains how the pre-rRNA precursor transcript is released from RNA Pol I in higher eukaryotes: is a separate endonucleolytic cleavage required or is transcription termination, followed by rapid exonucleolytic trimming, sufficient to release the precursor transcript? The latter would explain the absence of a homologous enzymatic function of the Rnt1p homolog Drosha. The evolution of a distinct RNA endonuclease, which carries out 3'ETS nucleolytic cleavage in higher eukaryotes, but not in *S. cerevisiae*, seems unlikely for a highly conserved mechanism such as rRNA biogenesis. Perhaps an intrinsic nucleolytic mechanism of the Pol I complex, in combination with other protein factors, could carry out an equivalent cleavage reaction.

Our work also raises the question of the tissue-specific requirement for high levels of NEF-sp in germ cells. NEF-sp could either function in gonad-specific regulation of rRNA biogenesis or serve an additional gonad-specific function in different noncoding RNA regulatory pathways. In addition, the distinct subcellular localization of dNEF-sp during terminal differentiation points to a specific role of NEF-sp in the final stages of spermatid differentiation.

Given the structural differences between the *Drosophila* and vertebrate NEF-sp proteins, the characterization of NEF-sp in higher organisms will be important to understand its function and conservation. NEF-sp is a relatively newly expanded RNA nuclease family that evolved from an old RNase T family. What is the role of the newly evolved RRM domains? None of the other DEDDh RNA exonuclease members or any other RNA exonucleases share structural similarities with NEF-sp proteins, making it a unique enzymatic protein family. Furthermore, how are the domain differences between invertebrates and vertebrates explained and how does that alter target recognition? It remains to be asked whether substrates are directly recognized through the RBDs of NEF-sp or whether NEF-sp is recruited by additional regulatory factors.

Finally, it is likely that 3'ETS exonucleolytic trimming is executed by several RNA exonucleases, as redundancies in exonucleolytic trimming have been observed for several pre-

rRNA substrates, such as e.g. 5'exonucleolytic trimming of 5.8S pre-rRNA through Rrp17 and Xrn1 (Oeffinger et al., 2007).

Another interesting aspect of this project has been to understand the role of RNA exonucleases regulating pre-rRNA maturation. As one of the earliest steps in rRNA maturation, 3'ETS exonucleolytic trimming regulates rRNA production. In contrast to one specific RNA endonucleolytic event, the utilization of a number of RNA exonucleases allows for a complex posttranscriptional control of rRNA processing. The control of a number of exonucleases to carry out this processing step allows for fine-tuning of rRNA biogenesis in response to cellular needs. This may also have implications for dysregulated rRNA biogenesis in human diseases. Hence, the study of tissue-specific processing in rRNA biogenesis and its regulation in different developmental processes opens up an interesting field of PTGR in the future.

5 Appendix of Tables

Table 5.1 Canonical RNA-binding domains

Domain	Description
RRM	RNA recognition motif, single-strand RNA (ssRNA)-binding (Cléry et al., 2008)
KH	KH-homology domain, ssRNA-binding (Grishin, 2001; Lunde et al., 2007)
DEAD	DEAD and DEAH box helicase motif, unwinds RNA (and DNA) (Rocak and Linder, 2004)
dsrm	Double-stranded dsRNA binding motif (Lunde et al., 2007)
zf-CCCH	Zinc finger motif type C-x8-C-x5-C-x3-H, ssRNA-binding (Krishna et al., 2003; Lunde et al., 2007)
zf-C2H2_jaz	JAZ dsRNA-binding protein zinc-fingers, dsRNA-binding (Krishna et al., 2003; Lunde et al., 2007)
zf-CCHC	Zinc knuckle, C-x2-C-x4-H-x4-C, ssRNA-binding (Krishna et al., 2003; Lunde et al., 2007)
zf-RanBP	RNA-binding Ran-binding-protein-like zinc finger (Krishna et al., 2003; Lunde et al., 2007)
zf-nanos	Zinc finger domains found in the eukaryotic proteins RBP families Nanos and Xcat-2 (Krishna et al., 2003; Lunde et al., 2007)
Zf-U1	Zinc finger motif found in several U1 small nuclear ribonucleoprotein C (U1-C) proteins (Krishna et al., 2003; Lunde et al., 2007)
LSM	LSM (Like Sm) domain is found in snRNP complexes, bind A/U rich regions (Wilusz and Wilusz, 2005)
SAP	(SAF-A/B, Acinus and PIAS) motif, RNA/DNA-binding domain (Aravind and Koonin, 2000)
YTH	YTH (YT521 homology) domain, ssRNA-binding, conserved aromatic residues within the ß-stands of the YTH domain similar to RRM, PUA, and OB-fold structures (Zhang et al., 2010)
SAM	Sterile alpha motif, 4-5-helical bundle of two orthogonally packed alpha-hairpins, ssRNA-binding (Kim and Bowie, 2003)
La	La protein RBD; alpha/beta fold that comprises a winged-helix motif, ssRNA-binding (Kenan and Keene, 2004)
PWI	~80 amino acid module with PWI tri-peptide located in N-terminal region, found at the N or C terminus of RBPs, generally found in association with RRM and RS RBDs (Szymczyna et al., 2003)
PABP	C-terminal domain in polyadenylate-binding protein, involved in homodimerization (Derry et al., 2006)
S1, S1-like	S1 ssRNA-binding domain, found in ribosomal proteins, similar to cold shock domain (Lunde et al., 2007)
SMN	Survival of motor neuron gene (SMN) contains a Tudor domain (SMN domain), which facilitates SMN-Sm protein interaction in the assembly of spliceosomal uridine-rich small nuclear ribonucleoprotein (Selenko et al., 2001)
PUA	Pseudouridine synthase and archaeosine transglycosylase domain often found in RNA modification enzymes and ribonucleoprotein complexes, ssRNA-binding (Pérez-Arellano et al., 2007)
Surp	Commonly found in splicing factors, ssRNA-binding domain, alpha-helical (Kuwasako et al., 2006)
PAZ	Piwi Argonaut and Zwille (PAZ) domain, posttranscriptional silencing domain, binds siRNAs (Meister, 2013)
Piwi	Piwi domain (P-element induced wimpy testis), posttranscriptional silencing domain, dsRNA guide hydrolysis of ssRNA (Meister, 2013)
CSD	Cold-shock domain, ssRNA/ssDNA binding (Mihailovich et al., 2010)
Agenet	Tudor-like domain found in FMRP and other RBPs, putative RBD, tandem Agenet-like domains preferentially recognize trimethylated peptides in a sequence-specific manner, four-stranded antiparallel beta sheet (Adinolfi et al., 2003; Maurer-Stroh et al., 2003)
TUDOR	Tudor domain, found in Tudor proteins, Tudor proteins are in complexes with RBPs (Siomi et al., 2010)
PUF	Tandem repeat of eight domains, ssRNA-binding domain (Wang et al., 2001; Zamore et al., 1997)
TROVE	TROVE (Telomerase, Ro and Vault) domain is a module of ~300-500 residues that is found in TEP1 and Ro60 the protein components of three ribonucleoprotein particles (Bateman and Kickhoefer, 2003)
THUMP	Thiouridine synthases, methylases and pseudouridine synthases, α/β fold, RNA-binding (Aravind and Koonin, 2001b)

Table 5.2 Pfam RNA-binding and RNA-related domains

Index	Pfam Domain	Count	Index	Pfam Domain	Count
1	RRM_1	178	401	PRP1_N	1
2	DEAD	62	402	PRP21_like_P	1
3	zf-CCCH	45	403	PRP3	1
4	RRM_6	43	404	PRP8_domainIV	1
5	KH_1	38	405	PTS_2-RNA	1
6	RRM_5	35	406	Paf67	1
7	GTP_EFTU	21	407	Pept_tRNA_hydro	1
8	dsrm	21	408	Pescadillo_N	1
9	zf-CCHC	20	409	PolyA_pol	1
10	LSM	19	410	Pox_MCEL	1
11	GTP_EFTU_D2	17	411	ProRS-C_1	1
12	HA2	17	412	Prp18	1
13	G-patch	15	413	Prp19	1
14	IBN_N	14	414	Prp31_C	1
15	RnaseA	13	415	RAI1	1
16	SAP	13	416	RAM	1
17	TUDOR	13	417	RNA_GG_bind	1
18	KOW	12	418	RNA_POL_M_15KD	1
19	MMR_HSR1	12	419	RNA_bind	1
20	zf-C2H2_jaz	12	420	RNA_pol	1
21	RNase_T	11	421	RNA_poll_A34	1
22	MIF4G	10	422	RNA_pol_I_A49	1
23	NTF2	9	423	RNA_pol_N	1
24	PAM2	9	424	RNA_pol_Rpb1_1	1
25	PAZ	9	425	RNA_pol_Rpb1_2	1
26	RBM1CTR	9	426	RNA_pol_Rpb1_3	1
27	zf-RanBP	9	427	RNA_pol_Rpb1_4	1
28	Anticodon_1	8	428	RNA_pol_Rpb1_5	1
29	CSD	8	429	RNA_pol_Rpb1_R	1
30	GTP_EFTU_D3	8	430	RNA_pol_Rpb2_1	1
31	HGTP_anticodon	8	431	RNA_pol_Rpb2_3	1
32	Piwi	8	432	RNA_pol_Rpb2_4	1
33	R3H	8	433	RNA_pol_Rpb2_5	1
34	RNase_Zc3h12a	8	434	RNA_pol_Rpb2_6	1
35	Ribosomal_L7Ae	8	435	RNA_pol_Rpb2_7	1
36	S1	8	436	RNA_pol_Rpb4	1
37	Xpo1	8	437	RNA_pol_Rpb5_C	1
38	tRNA-synt_2b	8	438	RNA_pol_Rpb5_N	1
39	Exo_endo_phos	7	439	RNA_pol_Rpb6	1
40	La	7	440	RNA_pol_Rpb8	1
41	Nol1_Nop2_Fmu	7	441	RNase_H	1
42	PAP_assoc	7	442	RNase_H2-Ydr279	1
43	RNase_PH	7	443	RNase_H2_suC	1
44	W2	7	444	RNase_HII	1
45	tRNA-synt_1	7	445	RNase_P_p30	1
46	tRNA_SAD	7	446	ROKNT	1
47	zf-C2H2	7	447	RRF	1
48	zf-met	7	448	RRM_4	1
49	APOBEC_N	6	449	RRP7	1
50	A_deamin	6	450	RRS1	1
51	MA3	6	451	RTC	1
52	RAP	6	452	RTC_insert	1
53	RNase_PH_C	6	453	RVT_1	1
54	Surp	6	454	Ribonuc_L-PSP	1
55	tRNA_anti-codon	6	455	Ribonuc_P_40	1
56	Brix	5	456	Ribonucleas_3_3	1
57	DZF	5	457	Ribonuclease_T2	1
58	FtsJ	5	458	Ribosomal_L12	1
59	PARP	5	459	Ribosomal_L13e	1
60	PWI	5	460	Ribosomal_L14e	1
61	Tap-RNA_bind	5	461	Ribosomal_L15e	1
62	YTH	5	462	Ribosomal_L17	1

63	tRNA-synt_2	5	463	Ribosomal_L18ae	1
64	AF-4	4	464	Ribosomal_L19	1
65	Beta-Casp	4	465	Ribosomal_L19e	1
66	CAF1	4	466	Ribosomal_L20	1
67	CRM1_C	4	467	Ribosomal_L21e	1
68	DND1_DSRM	4	468	Ribosomal_L21p	1
69	Dus	4	469	Ribosomal_L23eN	1
70	EFG_IV	4	470	Ribosomal_L27	1
71	IF4E	4	471	Ribosomal_L27e	1
72	Lactamase_B	4	472	Ribosomal_L28	1
73	OAS1_C	4	473	Ribosomal_L29	1
74	PABP	4	474	Ribosomal_L29e	1
75	PIH1	4	475	Ribosomal_L31e	1
76	PIN_4	4	476	Ribosomal_L32e	1
77	PUA	4	477	Ribosomal_L32p	1
78	PseudoU_synth_2	4	478	Ribosomal_L33	1
79	RF-1	4	479	Ribosomal_L34	1
80	Ribosomal_L30	4	480	Ribosomal_L34e	1
81	S4	4	481	Ribosomal_L35Ae	1
82	SUI1	4	482	Ribosomal_L35p	1
83	THRAP3_BCLAF1	4	483	Ribosomal_L36	1
84	tRNA-synt_1b	4	484	Ribosomal_L36e	1
85	tRNA-synt_2c	4	485	Ribosomal_L37	1
86	2OG-FeII_Oxy_2	3	486	Ribosomal_L37ae	1
87	AKAP95	3	487	Ribosomal_L37e	1
88	Ago_hook	3	488	Ribosomal_L38e	1
89	CBF	3	489	Ribosomal_L40e	1
90	CBFNT	3	490	Ribosomal_L41	1
91	Calreticulin	3	491	Ribosomal_L5	1
92	DUF2414	3	492	Ribosomal_L50	1
93	DUF2465	3	493	Ribosomal_L5_C	1
94	DcpS_C	3	494	Ribosomal_L6	1
95	EST1_DNA_bind	3	495	Ribosomal_L6e	1
96	FXR1P_C	3	496	Ribosomal_L6e_N	1
97	FYTT	3	497	Ribosomal_L9_N	1
98	Fox-1_C	3	498	Ribosomal_S13	1
99	IF-2B	3	499	Ribosomal_S13_N	1
100	LSM14	3	500	Ribosomal_S16	1
101	LUC7	3	501	Ribosomal_S19	1
102	MBD	3	502	Ribosomal_S19e	1
103	Methyltransf_31	3	503	Ribosomal_S21	1
104	Nop	3	504	Ribosomal_S21e	1
105	Nucleoplasmin	3	505	Ribosomal_S24e	1
106	PAP_RNA-bind	3	506	Ribosomal_S25	1
107	PseudoU_synth_1	3	507	Ribosomal_S26e	1
108	PurA	3	508	Ribosomal_S27	1
109	RIG-I_C-RD	3	509	Ribosomal_S28e	1
110	RIO1	3	510	Ribosomal_S30	1
111	RMMBL	3	511	Ribosomal_S3Ae	1
112	RNB	3	512	Ribosomal_S3_C	1
113	RS4NT	3	513	Ribosomal_S6	1
114	Ribonuc_2-5A	3	514	Ribosomal_S6e	1
115	Ribosomal_60s	3	515	Ribosomal_S7e	1
116	Ribosomal_L1	3	516	Ribosomal_S8	1
117	Ribosomal_L10	3	517	Rpp20	1
118	Ribosomal_L16	3	518	Rpr2	1
119	Ribosomal_L18e	3	519	Rrp15p	1
120	Ribosomal_L3	3	520	Rsm1	1
121	Ribosomal_S10	3	521	RtcB	1
122	Ribosomal_S18	3	522	S10_plectin	1
123	Ribosomal_S4e	3	523	SAP18	1
124	RrnaAD	3	524	SART-1	1
125	S1-like	3	525	SBDS	1
126	SMN	3	526	SBDS_C	1
127	Sas10_Utp3	3	527	SF3a60_bindingd	1

128	SpoU_methylase	3	528	SF3b1	1
129	Suf	3	529	SF3b10	1
130	TAP_C	3	530	SHQ1	1
131	THUMP	3	531	SHS2_Rpb7-N	1
132	TruB_N	3	532	SIP1	1
133	UPF0020	3	533	SLBP_RNA_bind	1
134	Utp12	3	534	SNase	1
135	cwf21	3	535	SR-25	1
136	eIF-1a	3	536	SRA1	1
137	eIF-5a	3	537	SRP14	1
138	tRNA-synt_1c	3	538	SRP19	1
139	tRNA-synt_His	3	539	SRP40_C	1
140	tRNA_m1G_MT	3	540	SRP72	1
141	zf-RNPHF	3	541	SRP9-21	1
142	zf-U1	3	542	SURF6	1
143	zf-nanos	3	543	SUV3_C	1
144	AARP2CN	2	544	Sen15	1
145	AKAP7_NLS	2	545	Seryl_tRNA_N	1
146	AXH	2	546	Slu7	1
147	Aconitase	2	547	Snurportin1	1
148	Aconitase_C	2	548	Sof1	1
149	Alba	2	549	Spt4	1
150	B3_4	2	550	Ssu72	1
151	Bin3	2	551	Sua5_yciO_yrdC	1
152	CPSF_A	2	552	Symplekin_C	1
153	Caprin-1_C	2	553	THOC7	1
154	Clp1	2	554	TRM13	1
155	CwfJ_C_1	2	555	TYW3	1
156	CwfJ_C_2	2	556	Telomerase_RBD	1
157	Cwf_Cwc_15	2	557	Thg1	1
158	DALR_1	2	558	Tho2	1
159	DCP1	2	559	Thoc2	1
160	DRY_EERY	2	560	Transformer	1
161	DUF1387	2	561	Transposase_22	1
162	DUF1897	2	562	Trm112p	1
163	DUF2051	2	563	U1snRNP70_N	1
164	ECR1_N	2	564	U3_assoc_6	1
165	EF1_GNE	2	565	U3snoRNP10	1
166	EST1	2	566	U5_2-snRNA_bdg	1
167	Endonuclease_NS	2	567	U6-snRNA_bdg	1
168	FDF	2	568	UPF0086	1
169	FDX-ACB	2	569	UPF0113	1
170	FRG1	2	570	UPF1_Zn_bind	1
171	Fcf1	2	571	UTP15_C	1
172	Fibrillarin	2	572	Upf2	1
173	GCD14	2	573	Urb2	1
174	GUCT	2	574	Urm1	1
175	Gar1	2	575	Utp11	1
176	HABP4_PAI-RBP1	2	576	Utp13	1
177	HEXIM	2	577	Utp21	1
178	HnRNPA1	2	578	Vault	1
179	IF-2	2	579	WGG	1
180	INT_SG_DDX_CT_C	2	580	Wyosine_form	1
181	L51_S25_CI-B8	2	581	XendoU	1
182	LsmAD	2	582	eIF-3_zeta	1
183	MRP-S27	2	583	eIF3_N	1
184	MT-A70	2	584	eIF3_subunit	1
185	Mago_nashi	2	585	eIF3g	1
186	Met_10	2	586	efThoc1	1
187	NOP5NT	2	587	mRNA_cap_C	1
188	NOT2_3_5	2	588	mRNA_cap_enzyme	1
189	NUC153	2	589	mTERF	1
190	Nop25	2	590	rRNA_processing	1
191	Nop52	2	591	tRNA_Me_trans	1
192	PAT1	2	592	tRNA_U5-meth_tr	1

193	PDCD9	2	593	tRNA_bind_2	1
194	PRP38	2	594	tRNA_int_end_N2	1
195	PRP4	2	595	tRNA_int_endo_N	1
196	PTH2	2	596	tRNA_synt_1c_R1	1
197	PUF	2	597	tRNA_synt_1c_R2	1
198	RNA_pol_L_2	2	598	z-alpha	1
199	RNase_P_Rpp14	2	599	zf-C3H1	1
200	RRM_3	2	600	zf-C3HC	1
201	ResIII	2	601	zf-CCHC_2	1
202	Ribonuclease_3	2	602	zf-FPG_IleRS	1
203	Ribosom_S12_S23	2	603	zf-TRM13_CCCH	1
204	Ribosomal_L11	2	604	2OG-FeII_Oxy	0
205	Ribosomal_L11_N	2	605	2_5_RNA_ligase	0
206	Ribosomal_L13	2	606	3_5_exonuc	0
207	Ribosomal_L14	2	607	5_3_exonuc	0
208	Ribosomal_L18p	2	608	5_3_exonuc_N	0
209	Ribosomal_L2	2	609	ANTAR	0
210	Ribosomal_L22	2	610	APOBEC_C	0
211	Ribosomal_L22e	2	611	APO_RNA-bind	0
212	Ribosomal_L23	2	612	ASCH	0
213	Ribosomal_L24e	2	613	Agenet	0
214	Ribosomal_L28e	2	614	Arb1	0
215	Ribosomal_L2_C	2	615	Arb2	0
216	Ribosomal_L30_N	2	616	AviRa	0
217	Ribosomal_L39	2	617	B2	0
218	Ribosomal_L4	2	618	BDHCT	0
219	Ribosomal_L44	2	619	BRK	0
220	Ribosomal_S11	2	620	BTV_NS2	0
221	Ribosomal_S14	2	621	CAT_RBD	0
222	Ribosomal_S15	2	622	CM1	0
223	Ribosomal_S17	2	623	CM2	0
224	Ribosomal_S17e	2	624	CPDase	0
225	Ribosomal_S2	2	625	CRS1_YhbY	0
226	Ribosomal_S27e	2	626	DALR_2	0
227	Ribosomal_S4	2	627	DEAD_assoc	0
228	Ribosomal_S5	2	628	DUF1325	0
229	Ribosomal_S5_C	2	629	DUF1669	0
230	Ribosomal_S7	2	630	DUF1866	0
231	Ribosomal_S8e	2	631	DUF446	0
232	Ribosomal_S9	2	632	DbpA	0
233	SID-1_RNA_chan	2	633	EIAV_Rev	0
234	SM-ATX	2	634	Ebola_NP	0
235	SRP54	2	635	EndoU_bacteria	0
236	SRP54_N	2	636	Endonuclea_NS_2	0
237	SRRM_C	2	637	Fibrillarin_2	0
238	SYF2	2	638	FlbT	0
239	Smg4_UPF3	2	639	FmrO	0
240	SpoU_sub_bind	2	640	GAD	0
241	TGT	2	641	GIIM	0
242	TRM	2	642	GidB	0
243	TROVE	2	643	HA	0
244	Translin	2	644	Helicase_Sgs1	0
245	TruD	2	645	IF2_N	0
246	Tudor-knot	2	646	Init_tRNA_PT	0
247	Utp14	2	647	KH_3	0
248	XRN_N	2	648	KH_4	0
249	eIF-3c_N	2	649	KH_5	0
250	eIF-5_eIF-2B	2	650	L31	0
251	eIF2A	2	651	Lactamase_B2	0
252	eIF2_C	2	652	Leu_Phe_trans	0
253	eRF1_1	2	653	MKT1_C	0
254	eRF1_2	2	654	MKT1_N	0
255	eRF1_3	2	655	MRL1	0
256	rRNA_proc-arch	2	656	MRP	0
257	tRNA-synt_1c_C	2	657	MetRS-N	0

258	tRNA-synt_1d	2	658	Methyltrans_RNA	0
259	tRNA-synt_1e	2	659	Mtr2	0
260	tRNA-synt_1g	2	660	N-Term_TEN	0
261	tRNA-synt_2d	2	661	NRDE-2	0
262	tRNA_bind	2	662	NSP10	0
263	tRNA_int_endo	2	663	NSP13	0
264	zf-CCCH_2	2	664	Nab2	0
265	zf-U11-48K	2	665	Nab6_mRNP_bdg	0
266	AAR2	1	666	Nol1_Nop2_Fmu_2	0
267	AD	1	667	Nsp1	0
268	ARS2	1	668	Nup35_RRM	0
269	AdoMet_MTase	1	669	Nup35_RRM_2	0
270	Arg_tRNA_synt_N	1	670	PAZ_siRNAbind	0
271	B5	1	671	PIN	0
272	BOP1NT	1	672	PIN_2	0
273	BUD22	1	673	PIN_3	0
274	Btz	1	674	PNPase_C	0
275	Bud13	1	675	PORR	0
276	Bystin	1	676	PP_M1	0
277	CMS1	1	677	Pet127	0
278	CNPase	1	678	Phe_tRNA-synt_N	0
279	CPL	1	679	PolyG_pol	0
280	CPSF73-100_C	1	680	Pox_ATPase-GT	0
281	CWC25	1	681	Pox_Rap94	0
282	Cgr1	1	682	Pox_mRNA-cap	0
283	DAP3	1	683	Pox_polyA_pol	0
284	DBP10CT	1	684	ProRS-C_2	0
285	DBR1	1	685	ProSAAS	0
286	DCP2	1	686	Queuosine_synth	0
287	DHHA1	1	687	RIX1	0
288	DKCLD	1	688	RMF	0
289	DNA_RNApol_7kD	1	689	RNA12	0
290	DRIM	1	690	RNA_Me_trans	0
291	DUF1604	1	691	RNA_bind_2	0
292	DUF1693	1	692	RNA_helicase	0
293	DUF1917	1	693	RNA_lig_T4_1	0
294	DUF2040	1	694	RNA_ligase	0
295	DUF2356	1	695	RNA_polI_A14	0
296	DUF2363	1	696	RNA_pol_3_Rpc31	0
297	DUF2638	1	697	RNA_pol_A_CTD	0
298	DUF2650	1	698	RNA_pol_A_bac	0
299	DUF382	1	699	RNA_pol_I_TF	0
300	DUF947	1	700	RNA_pol_L	0
301	DcpS	1	701	RNA_pol_Rbc25	0
302	EF1G	1	702	RNA_pol_Rpa2_4	0
303	EF_TS	1	703	RNA_pol_Rpb1_6	0
304	EIF4E-T	1	704	RNA_pol_Rpb1_7	0
305	EIF_2_alpha	1	705	RNA_pol_Rpb2_2	0
306	EMG1	1	706	RNA_pol_Rpb2_45	0
307	EXOSC1	1	707	RNA_pol_Rpc34	0
308	Ebp2	1	708	RNA_pol_Rpc4	0
309	Endonuclease_5	1	709	RNA_pol_Rpc82	0
310	Es2	1	710	RNA_pol_Rpo13	0
311	FTO_CTD	1	711	RNA_pol_delta	0
312	FTO_NTD	1	712	RNA_replicase_B	0
313	Fcf2	1	713	RNaseH_C	0
314	Fip1	1	714	RNase_E_G	0
315	GEMIN8	1	715	RNase_P_pop3	0
316	GIDA_assoc_3	1	716	RPM2	0
317	GLE1	1	717	RRM	0
318	GN3L_Grn1	1	718	RRM_2	0
319	GatB_Yqey	1	719	RSD-2	0
320	Gcd10p	1	720	RSS_P20	0
321	Gemin6	1	721	RVT_2	0
322	Gemin7	1	722	RdRP	0

323	Glu-tRNAGln	1	723	RdRP_1	0
324	HAT	1	724	RdRP_2	0
325	HGTP_anticodon2	1	725	RdRP_3	0
326	HRDC	1	726	RdRP_4	0
327	HVSL	1	727	RdRP_5	0
328	HnRNP_M	1	728	RdgC	0
329	INTS2	1	729	Rho_N	0
330	INTS5_C	1	730	Rho_RNA_bind	0
331	INTS5_N	1	731	Ribonuclease	0
332	IPPT	1	732	Ribonuclease_BN	0
333	Img2	1	733	Ribonuclease_P	0
334	Isy1	1	734	Ribosomal_L25p	0
335	KH_2	1	735	Ribosomal_L31	0
336	Kin17_mid	1	736	Ribosomal_L9_C	0
337	Lactamase_B_4	1	737	Ribosomal_S20p	0
338	Las1	1	738	Ribosomal_S22	0
339	Lsm_interact	1	739	Ribosomal_S23p	0
340	MIF4G_like	1	740	Ribosomal_S30AE	0
341	MIF4G_like_2	1	741	Ribosomal_S3_N	0
342	MPP6	1	742	Ribosomal_S4Pg	0
343	MRP-63	1	743	Rif1_N	0
344	MRP-L27	1	744	RimK	0
345	MRP-L28	1	745	RimM	0
346	MRP-L46	1	746	RnaseH	0
347	MRP-L47	1	747	Rsm22	0
348	MRP-L51	1	748	S1-P1_nuclease	0
349	MRP-S22	1	749	SEN1_N	0
350	MRP-S23	1	750	SPOUT_MTase	0
351	MRP-S24	1	751	Se-cys_synth_N	0
352	MRP-S26	1	752	SelA	0
353	MRP-S28	1	753	SelB-wing_1	0
354	MRP-S31	1	754	SelB-wing_2	0
355	MRP-S32	1	755	SelB-wing_3	0
356	MRP-S33	1	756	She2p	0
357	MRP-S35	1	757	SnAPC_2_like	0
358	MRP_L53	1	758	SpoU_methylas_C	0
359	MTS	1	759	Stb3	0
360	Maelstrom	1	760	Suppressor_P21	0
361	Mago-bind	1	761	THP2	0
362	Methyltransf_15	1	762	TPP1	0
363	Methyltransf_8	1	763	TSNR_N	0
364	Methyltrn_RNA_3	1	764	TilS	0
365	Mitoc_L55	1	765	TilS_C	0
366	Mpp10	1	766	Trm56	0
367	Myb_Cef	1	767	TruB-C_2	0
368	NMD3	1	768	TruB_C	0
369	NOB1_Zn_bind	1	769	U3_snoRNA_assoc	0
370	NOC3p	1	770	Utp8	0
371	NOG1	1	771	UvrD-helicase	0
372	NOGCT	1	772	VAR1	0
373	NUC129	1	773	Val_tRNA-synt_C	0
374	NUC173	1	774	Vmethyltransf	0
375	NUC205	1	775	Vmethyltransf_C	0
376	NUDIX_2	1	776	WT1	0
377	NUFIP1	1	777	XS	0
378	NUFIP2	1	778	cwf18	0
379	Noc2	1	779	dsRNA_bind	0
380	Nop10p	1	780	eIF3_p135	0
381	Nop14	1	781	eIF_4EBP	0
382	Nop16	1	782	eIF_4G1	0
383	Nop53	1	783	mRNA_triPase	0
384	Not1	1	784	nsp7	0
385	Not3	1	785	nsp8	0
386	Npa1	1	786	nsp9	0
387	Nrap	1	787	rRNA_methylase	0

388	OB_RNB	1	788	tRNA-Thr_ED	0
389	P68HR	1	789	tRNA-synt_1f	0
390	PC4	1	790	tRNA-synt_2e	0
391	PHAT	1	791	tRNA_NucTran2_2	0
392	PHF5	1	792	tRNA_NucTransf2	0
393	PMC2NT	1	793	tRNA_anti	0
394	PNPase	1	794	tRNA_anti-like	0
395	POP1	1	795	tRNA_deacylase	0
396	POPLD	1	796	tRNA_lig_CPD	0
397	PPR	1	797	tRNA_lig_kinase	0
398	PRO8NT	1	798	tRNA_synt_2b	0
399	PROCN	1	799	tRNA_synt_2f	0
400	PROCT	1			

Table 5.3 RBP census

Supplementary table S3 at:

http://www.nature.com/nrg/journal/v15/n12/full/nrg3813.html#supplementary-information

Table 5.4 Most abundant RBDs in the human proteome

Domain	Description
RRM	RNA recognition motif, single-strand RNA (ssRNA)-binding
RG/RGG	RG/RGG box repeats are arginine glycine rich low complexity regions, may bind RNA or act as protein-protein interaction domains in shuttling
DEAD	DEAD and DEAH box helicase motif, unwinds RNA (and DNA)
zf-CCCH	Zinc finger motif type C-x8-C-x5-C-x3-H, ssRNA-binding
KH	KH-homology domain, ssRNA-binding
GTP_EFTU, GTP_EFTU_D2, GTP_EFTU_D3	GTP-elongation factor family, proteins usually consist of 3 structural domains, 2 oligonucleotide binding domains (D2 and D3) and a GTP-binding domain
dsrm	Double-stranded RNA binding motif
zf-CCHC	Zinc knuckle, C-x2-C-x4-H-x4-C, ssRNA-binding
LSM	LSM (Like Sm) domain is found in snRNP complexes, bind A/U rich regions
OB_NTP_bind	Oligonucleotide/oligosaccharide-binding (OB)-fold, found in DEAD-box helicases in association with HA2 domain, regulates helicase activity through RNA binding
HA2	Helicase-associated domain, found in RNA helicases
G-patch	G-patch domain, ~48 amino acids with 6 conserved glycines, found in RBPs
IBN_N	Importin-beta N-terminal domain, RNA transport or RBP transport proteins
SAP	(SAF-A/B, Acinus and PIAS) motif, RNA/DNA-binding domain
TUDOR	Tudor domain, found in Tudor proteins, Tudor proteins are in complexes with RBPs
RnaseA	RNase A domain, ssRNA endonuclease
zf-C2H2_jaz	JAZ dsRNA-binding protein zinc-fingers, dsRNA-binding
MMR_HSR1	50S ribosome-binding GTPase domain, found in RBPs
KOW	KOW (Kyprides, Ouzounis, Woese) motif, found in a variety of ribosomal proteins
RNase_T	RNase T ssRNA exonuclease domain
MIF4G	MIF4G [Middle domain of eukaryotic initiation factor 4G (eIF4G)], RNA- (and DNA-) binding
zf-RanBP	RNA-binding Ran-binding-protein-like zinc finger
NTF2	Nuclear transport factor 2 (NTF2) domain, found in RNA export factors
PAZ	Piwi Argonaut and Zwille (PAZ) domain, posttranscriptional silencing domain, binds siRNAs
RBM1CTR	C-terminal region found in hnRNPs
PAM2	PABP-interacting motif PAM2, found in RBPs
Xpo1	exportin 1 domain, RNA transport or RBP transport proteins
S1	S1 ssRNA-binding domain
HGTP_anticodon	Anticodon binding domain, found in aminoacyl-tRNA synthetases
tRNA-synt_2b	tRNA synthetase class II core domain (G, H, P, S and T), core catalytic domain of tRNA synthetases
Piwi	Piwi domain (P-element induced wimpy testis), posttranscriptional silencing domain, dsRNA guide hydrolysis of ssRNA
CSD	cold-shock domain, ssRNA/ssDNA binding
Ribosomal_L7Ae	domain found in ribosomal proteins L7Ae/L30e/S12e/Gadd45
RNase_Zc3h12a	ssRNA endonuclease domain found in Zc3h12a proteins, member of the NYN domain family
Anticodon_1	tRNA anticodon-binding domain, found in tRNA synthetases
R3H	R3H domain, R-x3-H conserved core, binds ssRNA/ssDNA

Table 5.5 TF census

Supplementary Table S4 at:
http://www.nature.com/nrg/journal/v15/n12/full/nrg3813.html#supplementary-information

6 References

Abbasi-Moheb, L., Mertel, S., Gonsior, M., Nouri-Vahid, L., Kahrizi, K., Cirak, S., Wieczorek, D., Motazacker, M.M., Esmaeeli-Nieh, S., Cremer, K., et al. (2012). Mutations in NSUN2 cause autosomal-recessive intellectual disability. Am. J. Hum. Genet. *90*, 847–855.

Adinolfi, S., Ramos, A., Martin, S.R., Dal Piaz, F., Pucci, P., Bardoni, B., Mandel, J.L., and Pastore, A. (2003). The N-terminus of the fragile X mental retardation protein contains a novel domain involved in dimerization and RNA binding. Biochemistry *42*, 10437–10444.

Agamy, O., Ben Zeev, B., Lev, D., Marcus, B., Fine, D., Su, D., Narkis, G., Ofir, R., Hoffmann, C., Leshinsky-Silver, E., et al. (2010). Mutations disrupting selenocysteine formation cause progressive cerebello-cerebral atrophy. Am. J. Hum. Genet. *87*, 538–544.

Ahmad, Y., Boisvert, F.-M., Gregor, P., Cobley, A., and Lamond, A.I. (2009). NOPdb: Nucleolar Proteome Database--2008 update. Nucleic Acids Res. *37*, D181–D184.

Akopian, D., Shen, K., Zhang, X., and Shan, S.-O. (2013). Signal recognition particle: an essential protein-targeting machine. Annu. Rev. Biochem. *82*, 693–721.

Al-Sukhni, W., Rothenmund, H., Borgida, A.E., Zogopoulos, G., O'Shea, A.-M., Pollett, A., and Gallinger, S. (2008). Germline BRCA1 mutations predispose to pancreatic adenocarcinoma. Hum. Genet. *124*, 271–278.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat. Genet. *23*, 185–188.

Amouri, R., Moreira, M.-C., Zouari, M., Euch, El, G., Barhoumi, C., Kefi, M., Belal, S., Koenig, M., and Hentati, F. (2004). Aprataxin gene mutations in Tunisian families. Neurology *63*, 928–929.

Anantharaman, V., Koonin, E.V., and Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res. *30*, 1427–1464.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Anderson, M.G., Perkins, G.L., Chittick, P., Shrigley, R.J., and Johnson, W.A. (1995). drifter, a Drosophila POU-domain transcription factor, is required for correct differentiation and migration of tracheal cells and midline glia. Genes Dev. *9*, 123–137.

Anderson, P., and Kedersha, N. (2009). RNA granules: post-transcriptional and epigenetic modulators of gene expression. Nat. Rev. Mol. Cell Biol. *10*, 430–436.

Anger, A.M., Armache, J.-P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., and Beckmann, R. (2013). Structures of the human and Drosophila 80S ribosome. Nature *497*, 80–85.

Ansel, K.M., Pastor, W.A., Rath, N., Lapan, A.D., Glasmacher, E., Wolf, C., Smith, L.C., Papadopoulou, N., Lamperti, E.D., Tahiliani, M., et al. (2008). Mouse Eri1 interacts with the ribosome and catalyzes 5.8S rRNA processing. Nat. Struct. Mol. Biol. *15*, 523–530.

Anthony, K., and Gallo, J.M. (2010). Aberrant RNA processing events in neurological disorders. Brain Res. *1338*, 67–77.

Antonellis, A., and Green, E.D. (2008). The role of aminoacyl-tRNA synthetases in genetic diseases. Annu Rev Genomics Hum Genet *9*, 87–107.

Antonellis, A., Ellsworth, R.E., Sambuughin, N., Puls, I., Abel, A., Lee-Lin, S.-Q., Jordanova, A., Kremensky, I., Christodoulou, K., Middleton, L.T., et al. (2003). Glycyl tRNA synthetase mutations in Charcot-Marie-Tooth disease type 2D and distal spinal muscular atrophy type V. Am. J. Hum. Genet. *72*, 1293–1299.

Antonicka, H., Ostergaard, E., Sasarman, F., Weraarpachai, W., Wibrand, F., Pedersen, A.M.B., Rodenburg, R.J., van der Knaap, M.S., Smeitink, J.A.M., Chrzanowska-Lightowlers, Z.M., et al. (2010). Mutations in C12orf65 in patients with encephalomyopathy and a mitochondrial translation defect. Am. J. Hum. Genet. *87*, 115–122.

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D., et al. (2001). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res. *29*, 37–40.

Aravind, L., and Koonin, E.V. (2000). SAP - a putative DNA-binding motif involved in chromosomal organization. Trends Biochem. Sci. *25*, 112–114.

Aravind, L., and Koonin, E.V. (2001a). A natural classification of ribonucleases. Meth. Enzymol. *341*, 3–28.

Aravind, L., and Koonin, E.V. (2001b). THUMP--a predicted RNA-binding domain shared by 4-thiouridine, pseudouridine synthases and RNA methylases. Trends Biochem. Sci. *26*, 215–217.

Arcus, V. (2002). OB-fold domains: a snapshot of the evolution of sequence, structure and function. Curr. Opin. Struct. Biol. *12*, 794–801.

Armanios, M.Y., Chen, J.J.-L., Cogan, J.D., Alder, J.K., Ingersoll, R.G., Markin, C., Lawson, W.E., Xie, M., Vulto, I., Phillips, J.A., et al. (2007). Telomerase mutations in families with idiopathic pulmonary fibrosis. N. Engl. J. Med. *356*, 1317–1326.

Armistead, J., Khatkar, S., Meyer, B., Mark, B.L., Patel, N., Coghlan, G., Lamont, R.E., Liu, S., Wiechert, J., Cattini, P.A., et al. (2009). Mutation of a gene essential for ribosome biogenesis, EMG1, causes Bowen-Conradi syndrome. Am. J. Hum. Genet. *84*, 728–739.

Arnold, S.E., and Trojanowski, J.Q. (1996). Human fetal hippocampal development: I. Cytoarchitecture, myeloarchitecture, and neuronal morphologic features. J. Comp. Neurol. *367*, 274–292.

Ascano, M., Gerstberger, S., and Tuschl, T. (2013). Multi-disciplinary methods to define RNA-protein interactions and regulatory networks. Curr. Opin. Genet. Dev. *23*, 20–28.

Ascano, M., Hafner, M., Cekan, P., Gerstberger, S., and Tuschl, T. (2012a). Identification of RNA-protein interaction networks using PAR-CLIP. WIREs RNA *3*, 159–177.

Ascano, M., Mukherjee, N., Bandaru, P., Miller, J.B., Nusbaum, J.D., Corcoran, D.L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M., et al. (2012b). FMRP targets distinct mRNA sequence elements to regulate protein expression. Nature *492*, 382–386.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Astuti, D., Morris, M.R., Cooper, W.N., Staals, R.H.J., Wake, N.C., Fews, G.A., Gill, H., Gentle, D., Shuib, S., Ricketts, C.J., et al. (2012). Germline mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility. Nat. Genet. *44*, 277–284.

Atianand, M.K., and Fitzgerald, K.A. (2013). Molecular basis of DNA recognition in the immune system. J. Immunol. *190*, 1911–1918.

Auweter, S.D., Oberstrass, F.C., and Allain, F.H.T. (2006). Sequence-specific binding of single-stranded RNA: is there a code for recognition? Nucleic Acids Res. *34*, 4943–4959.

Bachellerie, J.-P., Cavaillé, J., and Hüttenhofer, A. (2002). The expanding snoRNA world. Biochimie *84*, 775–790.

Baltz, A.G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. Mol. Cell *46*, 674–690.

Banfi, S., Servadio, A., Chung, M.Y., Kwiatkowski, T.J., McCall, A.E., Duvick, L.A., Shen, Y., Roth, E.J., Orr, H.T., and Zoghbi, H.Y. (1994). Identification and characterization of the gene causing type 1 spinocerebellar ataxia. Nat. Genet. *7*, 513–520.

Bartel, D.P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. Cell *136*, 215–233.

Bartsch, I., Schoneberg, C., and Grummt, I. (1987). Evolutionary changes of sequences and factors that direct transcription termination of human and mouse ribsomal genes. Mol. Cell. Biol. *7*, 2521–2529.

Bateman, A., and Kickhoefer, V. (2003). The TROVE module: a common element in Telomerase, Ro and Vault ribonucleoproteins. BMC Bioinformatics *4*, 49.

Batista, P.J., and Chang, H.Y. (2013). Long Noncoding RNAs: Cellular Address Codes in

Development and Disease. Cell *152*, 1298–1307.

Battle, D.J., KASIM, M., Yong, J., LOTTI, F., LAU, C.K., Mouaikel, J., Zhang, Z., HAN, K., Wan, L., and Dreyfuss, G. (2006). The SMN complex: an assembly machine for RNPs. Cold Spring Harb. Symp. Quant. Biol. *71*, 313–320.

Beese, L.S., and Steitz, T.A. (1991a). Structural basis for the 3“-5” exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. Embo J. *10*, 25–33.

Beese, L.S., and Steitz, T.A. (1991b). STRUCTURAL BASIS FOR THE 3“-5” EXONUCLEASE ACTIVITY OF ESCHERICHIA-COLI DNA-POLYMERASE-I - A 2 METAL-ION MECHANISM. Embo J. *10*, 25–33.

Bell, J.L., Wächter, K., Mühleck, B., Pazaitis, N., Köhn, M., Lederer, M., and Hüttelmaier, S. (2013). Insulin-like growth factor 2 mRNA-binding proteins (IGF2BPs): post-transcriptional drivers of cancer progression? Cell. Mol. Life Sci. *70*, 2657–2675.

Beloglazova, N., Flick, R., Tchigvintsev, A., Brown, G., Popovic, A., Nocek, B., and Yakunin, A.F. (2013). Nuclease Activity of the Human SAMHD1 Protein Implicated in the Aicardi-Goutieres Syndrome and HIV-1 Restriction. J. Biol. Chem. *288*, 8101–8110.

Belostotsky, R., Ben-Shalom, E., Rinat, C., Becker-Cohen, R., Feinstein, S., Zeligson, S., Segel, R., Elpeleg, O., Nassar, S., and Frishberg, Y. (2011). Mutations in the mitochondrial seryl-tRNA synthetase cause hyperuricemia, pulmonary hypertension, renal failure in infancy and alkalosis, HUPRA syndrome. Am. J. Hum. Genet. *88*, 193–200.

Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L., et al. (2012). U1 snRNP Determines mRNA Length and Regulates Isoform Expression. Cell *150*, 53–64.

Berger, W., Steiner, E., Grusch, M., Elbling, L., and Micksche, M. (2008). Vaults and the major vault protein: Novel roles in signal pathway regulation and immunity. Cell. Mol. Life Sci. *66*, 43–61.

Bernier, F.P., Caluseriu, O., Ng, S., Schwartzentruber, J., Buckingham, K.J., Innes, A.M., Jabs, E.W., Innis, J.W., Schuette, J.L., Gorski, J.L., et al. (2012). Haploinsufficiency of SF3B4, a component of the pre-mRNA spliceosomal complex, causes Nager syndrome. Am. J. Hum. Genet. *90*, 925–933.

Bettegowda, A., and Wilkinson, M.F. (2010). Transcription and post-transcriptional regulation of spermatogenesis. Philosophical Transactions of the Royal Society B: Biological Sciences *365*, 1637–1651.

Bhalla, K., Phillips, H.A., Crawford, J., McKenzie, O.L.D., Mulley, J.C., Eyre, H., Gardner, A.E., Kremmidiotis, G., and Callen, D.F. (2004). The de novo chromosome 16 translocations of two patients with abnormal phenotypes (mental retardation and epilepsy) disrupt the A2BP1 gene. J. Hum. Genet. *49*, 308–311.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc. Natl. Acad.

Sci. U. S. a. *104*, 3312–3317.

Boissel, S., Reish, O., Proulx, K., Kawagoe-Takaki, H., Sedgwick, B., Yeo, G.S.H., Meyre, D., Golzio, C., Molinari, F., Kadhom, N., et al. (2009). Loss-of-function mutation in the dioxygenase-encoding FTO gene causes severe growth retardation and multiple malformations. Am. J. Hum. Genet. *85*, 106–111.

Boisvert, F.-M., Ahmad, Y., Gierlinski, M., Charriere, F., Lamont, D., Scott, M., Barton, G., and Lamond, A.I. (2012). A quantitative spatial proteomics analysis of proteome turnover in human cells. Mol. Cell. Proteomics *11*, M111.011429.

Boisvert, F.-M., van Koningsbruggen, S., Navascués, J., and Lamond, A.I. (2007). The multifunctional nucleolus. Nat. Rev. Mol. Cell Biol. *8*, 574–585.

Bonifati, V., Rizzu, P., van Baren, M.J., Schaap, O., Breedveld, G.J., Krieger, E., Dekker, M.C.J., Squitieri, F., Ibanez, P., Joosse, M., et al. (2003). Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science *299*, 256–259.

Boocock, G.R.B., Morrison, J.A., Popovic, M., Richards, N., Ellis, L., Durie, P.R., and Rommens, J.M. (2003). Mutations in SBDS are associated with Shwachman-Diamond syndrome. Nat. Genet. *33*, 97–101.

Boulanger, L.M. (2009). Immune proteins in brain development and synaptic plasticity. Neuron *64*, 93–109.

Boulon, S., Westman, B.J., Hutten, S., Boisvert, F.-M., and Lamond, A.I. (2010). The Nucleolus under Stress. Mol. Cell *40*, 216–227.

Brais, B., Bouchard, J.-P., Xie, Y.-G., Rochefort, D.L., Chrétien, N., Tomé, F.M.S., Lafrentére, R.G., Rommens, J.M., Uyama, E., Nohira, O., et al. (1998). Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. Nat. Genet. *18*, 164–167.

Bramham, C.R., and Wells, D.G. (2007). Dendritic mRNA: transport, translation and function. Nat. Rev. Neurosci. *8*, 776–789.

Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401–415.

Bratkovič, T., and Rogelj, B. (2014). The many faces of small nucleolar RNAs. Bba *1839*, 438–443.

Brauch, K.M., Karst, M.L., Herron, K.J., de Andrade, M., Pellikka, P.A., Rodeheffer, R.J., Michels, V.V., and Olson, T.M. (2009). Mutations in ribonucleic acid binding protein gene cause familial dilated cardiomyopathy. J. Am. Coll. Cardiol. *54*, 930–941.

Brickwood, S., Bonthron, D.T., Al-Gazali, L.I., Piper, K., Hearn, T., Wilson, D.I., and Hanley, N.A. (2003). Wolcott-Rallison syndrome: pathogenic insights into neonatal diabetes from new mutation and expression studies of EIF2AK3. J. Med. Genet. *40*, 685–689.

Briggs, M.W., Burkard, K.T., and Butler, J.S. (1998). Rrp6p, the yeast homologue of the human

PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. J. Biol. Chem. *273*, 13255–13263.

Broderick, P., Carvajal-Carmona, L., Pittman, A.M., Webb, E., Howarth, K., Rowan, A., Lubbe, S., Spain, S., Sullivan, K., Fielding, S., et al. (2007). A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat. Genet. *39*, 1315–1317.

Brook, M., Smith, J.W.S., and Gray, N.K. (2009). The DAZL and PABP families: RNA-binding proteins with interrelated roles in translational control in oocytes. Reproduction *137*, 595–617.

Brooks, S.A., and Blackshear, P.J. (2013). Tristetraprolin (TTP): Interactions with mRNA and proteins, and current thoughts on mechanisms of action. BBA - Gene Regul. Mech. *1829*, 666–679.

Bruni, F., Gramegna, P., Oliveira, J.M.A., Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M.A. (2013). REXO2 Is an Oligoribonuclease Active in Human Mitochondria. PLoS ONE *8*, e64670.

Buchan, J.R., and Parker, R. (2009). Eukaryotic Stress Granules: The Ins and Outs of Translation. Mol. Cell *36*, 932–941.

Buchan, J.R., Kolaitis, R.-M., Taylor, J.P., and Parker, R. (2013). Eukaryotic stress granules are cleared by autophagy and Cdc48/VCP function. Cell *153*, 1461–1474.

Buckanovich, R.J., Yang, Y.Y., and Darnell, R.B. (1996). The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. J. Neurosci. *16*, 1114–1122.

Budde, B.S., Namavar, Y., Barth, P.G., Poll-The, B.T., Nuernberg, G., Becker, C., van Ruissen, F., Weterman, M.A.J., Fluiter, K., Beek, E.T.T., et al. (2008). tRNA splicing endonuclease mutations cause pontocerebellar hypoplasia. Nat. Genet. *40*, 1113–1118.

Buhler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell *125*, 873–886.

Burd, C.G., and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. Science *265*, 615–621.

Buszczak, M., Paterno, S., Lighthouse, D., Bachman, J., Planck, J., Owen, S., Skora, A.D., Nystul, T.G., Ohlstein, B., Allen, A., et al. (2007). The carnegie protein trap library: a versatile tool for Drosophila developmental studies. Genetics *175*, 1505–1531.

Buszczak, M., Signer, R.A.J., and Morrison, S.J. (2014). Cellular differences in protein synthesis regulate tissue homeostasis. Cell *159*, 242–251.

Bykhovskaya, Y., Casas, K., Mengesha, E., Inbal, A., and Fischel-Ghodsian, N. (2004). Missense mutation in pseudouridine synthase 1 (PUS1) causes mitochondrial myopathy and sideroblastic anemia (MLASA). Am. J. Hum. Genet. *74*, 1303–1308.

Camargos, S., Scholz, S., Simón-Sánchez, J., Paisán-Ruiz, C., Lewis, P., Hernandez, D., Ding, J.,

Gibbs, J.R., Cookson, M.R., Bras, J., et al. (2008). DYT16, a novel young-onset dystonia-parkinsonism disorder: identification of a segregating mutation in the stress-response protein PRKRA. Lancet Neurol. *7*, 207–215.

Cancel, G., Dürr, A., Didierjean, O., Imbert, G., Bürk, K., Lezin, A., Belal, S., Benomar, A., Abada-Bendib, M., Vial, C., et al. (1997). Molecular and clinical correlations in spinocerebellar ataxia 2: a study of 32 families. Hum. Mol. Genet. *6*, 709–715.

Capy, P., Gasperi, G., Biémont, C., and Bazin, C. (2000). Stress and transposable elements: co-evolution or useful parasites? Heredity (Edinb) *85 (Pt 2)*, 101–106.

Carmo-Fonseca, M., Mendes-Soares, L., and Campos, I. (2000). To be or not to be in the nucleolus. Nat. Cell Biol. *2*, E107–E112.

Carpten, J., Nupponen, N., Isaacs, S., Sood, R., Robbins, C., Xu, J., Faruque, M., Moses, T., Ewing, C., Gillanders, E., et al. (2002). Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. Nat. Genet. *30*, 181–184.

Casey, G., Neville, P.J., Plummer, S.J., Xiang, Y., Krumroy, L.M., Klein, E.A., Catalona, W.J., Nupponen, N., Carpten, J.D., Trent, J.M., et al. (2002). RNASEL Arg462Gln variant is implicated in up to 13% of prostate cancer cases. Nat. Genet. *32*, 581–583.

Cassandrini, D., Biancheri, R., Tessa, A., Di Rocco, M., Di Capua, M., Bruno, C., Denora, P.S., Sartori, S., Rossi, A., Nozza, P., et al. (2010). Pontocerebellar hypoplasia: clinical, pathologic, and genetic studies. Neurology *75*, 1459–1464.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell *149*, 1393–1406.

Castilla, L.H., Couch, F.J., Erdos, M.R., Hoskins, K.F., Calzone, K., Garber, J.E., Boyd, J., Lubin, M.B., Deshano, M.L., and Brody, L.C. (1994). Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. Nat. Genet. *8*, 387–391.

Castle, J.C., Armour, C.D., Löwer, M., Haynor, D., Biery, M., Bouzek, H., Chen, R., Jackson, S., Johnson, J.M., Rohl, C.A., et al. (2010). Digital genome-wide ncRNA expression, including SnoRNAs, across 11 human tissues using polyA-neutral amplification. PLoS ONE *5*, e11779.

Cavdar Koc, E., Burkhart, W., Blackburn, K., Moseley, A., and Spremulli, L.L. (2001). The small subunit of the mammalian mitochondrial ribosome. Identification of the full complement of ribosomal proteins present. J. Biol. Chem. *276*, 19363–19374.

Cech, T.R., and Steitz, J.A. (2014). The Noncoding RNA Revolution-Trashing Old Rules to Forge New Ones. Cell *157*, 77–94.

Cerritelli, S.M., and Crouch, R.J. (2009). Ribonuclease H: the enzymes in eukaryotes. Febs J. *276*, 1494–1505.

Chakarova, C.F., Hims, M.M., Bolz, H., Abu-Safieh, L., Patel, R.J., Papaioannou, M.G., Inglehearn, C.F., Keen, T.J., Willis, C., Moore, A.T., et al. (2002). Mutations in HPRP3, a third

member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. Hum. Mol. Genet. *11*, 87–92.

Chang, H., Lim, J., Ha, M., and Kim, V.N. (2014). TAIL-seq: Genome-wide Determination of Poly(A) Tail Length and 3' End Modifications. Mol. Cell *53*, 1044–1052.

Chang, K.-Y., and Ramos, A. (2005). The double-stranded RNA-binding motif, a versatile macromolecular docking platform. Febs J. *272*, 2109–2117.

Chartier-Harlin, M.-C., Dachsel, J.C., Vilariño-Güell, C., Lincoln, S.J., Leprêtre, F., Hulihan, M.M., Kachergus, J., Milnerwood, A.J., Tapia, L., Song, M.-S., et al. (2011). Translation initiator EIF4G1 mutations in familial Parkinson disease. Am. J. Hum. Genet. *89*, 398–406.

Cheadle, J.P., Gill, H., Fleming, N., Maynard, J., Kerr, A., Leonard, H., Krawczak, M., Cooper, D.N., Lynch, S., Thomas, N., et al. (2000). Long-read sequence analysis of the MECP2 gene in Rett syndrome patients: correlation of disease severity with mutation type and location. Hum. Mol. Genet. *9*, 1119–1129.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat. Rev. Mol. Cell Biol. *10*, 741–754.

Chen, X., and Wolin, S.L. (2004). The Ro 60 kDa autoantigen: insights into cellular function and role in autoimmunity. J. Mol. Med. *82*, 232–239.

Chiu, C., Tebo, M., Ingles, J., Yeates, L., Arthur, J.W., Lind, J.M., and Semsarian, C. (2007). Genetic screening of calcium regulation genes in familial hypertrophic cardiomyopathy. J. Mol. Cell. Cardiol. *43*, 337–343.

Cho, J., Chang, H., Kwon, S.C., Kim, B., Kim, Y., Choe, J., Ha, M., Kim, Y.K., and Kim, V.N. (2012). LIN28A is a suppressor of ER-associated translation in embryonic stem cells. Cell *151*, 765–777.

Chong, M.M.W., Rasmussen, J.P., Rudensky, A.Y., Rundensky, A.Y., and Littman, D.R. (2008). The RNAseIII enzyme Drosha is critical in T cells for preventing lethal inflammatory disease. J. Exp. Med. *205*, 2005–2017.

Ciapponi, L., Cenci, G., and Gatti, M. (2006). The Drosophila Nbs protein functions in multiple pathways for the maintenance of genome stability. Genetics *173*, 1447–1454.

Ciganda, M., and Williams, N. (2011). Eukaryotic 5S rRNA biogenesis. WIREs RNA *2*, 523–533.

Cirillo, D., Marchese, D., Agostini, F., Livi, C.M., Botta-Orfila, T., and Tartaglia, G.G. (2014). Constitutive patterns of gene expression regulated by RNA-binding proteins. Genome Biol. *15*, R13.

Cléry, A., Blatter, M., and Allain, F.H.T. (2008). RNA recognition motifs: boring? Not quite. Curr. Opin. Struct. Biol. *18*, 290–298.

Cmejla, R., Cmejlova, J., Handrkova, H., Petrak, J., and Pospisilova, D. (2007). Ribosomal

protein S17 gene (RPS17) is mutated in Diamond-Blackfan anemia. Hum. Mutat. *28*, 1178–1182.

Cobben, J.M., van der Steege, G., Grootscholten, P., de Visser, M., Scheffer, H., and Buys, C.H. (1995). Deletions of the survival motor neuron gene in unaffected siblings of patients with spinal muscular atrophy. Am. J. Hum. Genet. *57*, 805–808.

Coenen, M.J.H., Antonicka, H., Ugalde, C., Sasarman, F., Rossi, R., Heister, J.G.A.M.A., Newbold, R.F., Trijbels, F.J.M.F., van den Heuvel, L.P., Shoubridge, E.A., et al. (2004). Mutant mitochondrial elongation factor G1 and combined oxidative phosphorylation deficiency. N. Engl. J. Med. *351*, 2080–2086.

Cook, K.B., Kazan, H., Zuberi, K., Morris, Q., and Hughes, T.R. (2011). RBPDB: a database of RNA-binding specificities. Nucleic Acids Res. *39*, D301–D308.

Cools, J., DeAngelo, D.J., Gotlib, J., Stover, E.H., Legare, R.D., Cortes, J., Kutok, J., Clark, J., Galinsky, I., Griffin, J.D., et al. (2003). A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. N. Engl. J. Med. *348*, 1201–1214.

Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and Disease. Cell *136*, 777–793.

Copela, L.A., Fernandez, C.F., Sherrer, R.L., and Wolin, S.L. (2008). Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation. Rna *14*, 1214–1227.

Corcoran, D.L., Georgiev, S., Mukherjee, N., Gottwein, E., Skalsky, R.L., Keene, J.D., and Ohler, U. (2011). PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol. *12*, R79.

Coute, Y., Kindbeiter, K., Belin, S., Dieckmann, R., Duret, L., Bezin, L., Sanchez, J.C., and Diaz, J.J. (2007). ISG20L2, a Novel Vertebrate Nucleolar Exoribonuclease Involved in Ribosome Biogenesis. Mol. Cell. Proteomics *7*, 546–559.

Cramton, S.E., and Laski, F.A. (1994). string of pearls encodes Drosophila ribosomal protein S2, has Minute-like characteristics, and is required during oogenesis. Genetics *137*, 1039–1048.

Criscuolo, C., Mancini, P., Menchise, V., Saccà, F., De Michele, G., Banfi, S., and Filla, A. (2005). Very late onset in ataxia oculomotor apraxia type I. Ann. Neurol. *57*, 777.

Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., et al. (2014). The Reactome pathway knowledgebase. Nucleic Acids Res. *42*, D472–D477.

Crosby, A.H., Patel, H., Chioza, B.A., Proukakis, C., Gurtz, K., Patton, M.A., Sharifi, R., Harlalka, G., Simpson, M.A., Dick, K., et al. (2010). Defective mitochondrial mRNA maturation is associated with spastic ataxia. Am. J. Hum. Genet. *87*, 655–660.

Crow, Y.J., Hayward, B.E., Parmar, R., Robins, P., Leitch, A., Ali, M., Black, D.N., van Bokhoven, H., Brunner, H.G., Hamel, B.C., et al. (2006a). Mutations in the gene encoding the 3 "-5 " DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. Nat.

Genet. *38*, 917–920.

Crow, Y.J., Leitch, A., Hayward, B.E., Garner, A., Parmar, R., Griffith, E., Ali, M., Semple, C., Aicardi, J., Babul-Hirji, R., et al. (2006b). Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection. Nat. Genet. *38*, 910–916.

Curry, S., Kotik-Kogan, O., Conte, M.R., and Brick, P. (2009). Getting to the end of RNA: Structural analysis of protein recognition of 5′ and 3′ termini. BBA - Gene Regul. Mech. *1789*, 653–666.

Czeschik, J.C., Voigt, C., Alanay, Y., Albrecht, B., Avci, S., Fitzpatrick, D., Goudie, D.R., Hehr, U., Hoogeboom, A.J., Kayserili, H., et al. (2013). Clinical and mutation data in 12 patients with the clinical diagnosis of Nager syndrome. Hum. Genet. *132*, 885–898.

Daiger, S.P., Sullivan, L.S., and Bowne, S.J. (2013). Genes and mutations causing retinitis pigmentosa. Clin. Genet. *84*, 132–141.

Darnell, J.E. (2011). **RNA: Life's Indispensable Molecule** CHSL press (New York)

Dauwerse, J.G., Dixon, J., Seland, S., Ruivenkamp, C.A.L., Van Haeringen, A., Hoefsloot, L.H., Peters, D.J.M., Boers, A.C.-D., Daumer-Haas, C., Maiwald, R., et al. (2011). Mutations in genes encoding subunits of RNA polymerases I and III cause Treacher Collins syndrome. Nat. Genet. *43*, 20–22.

de Cuevas, M., and Matunis, E.L. (2011). The stem cell niche: lessons from the Drosophila testis. Development *138*, 2861–2869.

Dean, M., Park, M., and Vande Woude, G.F. (1987). Characterization of the rearranged tpr-met oncogene breakpoint. Mol. Cell. Biol. *7*, 921–924.

DeHoratius, R.J., Pillarisetty, R., Messner, R.P., and Talal, N. (1975). Anti-nucleic acid antibodies in systemic lupus erythematosus patients and their families. Incidence and correlation with lymphocytotoxic antibodies. J. Clin. Invest. *56*, 1149–1154.

Delépine, M., Nicolino, M., Barrett, T., Golamaully, M., Lathrop, G.M., and Julier, C. (2000). EIF2AK3, encoding translation initiation factor 2-alpha kinase 3, is mutated in patients with Wolcott-Rallison syndrome. Nat. Genet. *25*, 406–409.

Derry, M.C., Yanagiya, A., Martineau, Y., and Sonenberg, N. (2006). Regulation of poly(A)-binding protein through PABP-interacting proteins. Cold Spring Harb. Symp. Quant. Biol. *71*, 537–543.

Dever, T.E., and Green, R. (2012). The elongation, termination, and recycling phases of translation in eukaryotes. Cold Spring Harb. Perspect. Biol. *4*, a013706.

Deverman, B.E., and Patterson, P.H. (2009). Cytokines and CNS development. Neuron *64*, 61–78.

Devys, D., Biancalana, V., Rousseau, F., Boué, J., Mandel, J.L., and Oberlé, I. (1992). Analysis

of full fragile X mutations in fetal tissues and monozygotic twins indicate that abnormal methylation and somatic heterogeneity are established early in development. Am. J. Med. Genet. *43*, 208–216.

Dezso, Z., Nikolsky, Y., Sviridov, E., Shi, W., Serebriyskaya, T., Dosymbekov, D., Bugrim, A., Rakhmatulin, E., Brennan, R.J., Guryanov, A., et al. (2008). A comprehensive functional analysis of tissue specificity of human gene expression. BMC Biol. *6*, 49.

Di Giammartino, D.C., Nishida, K., and Manley, J.L. (2011). Mechanisms and Consequences of Alternative Polyadenylation. Mol. Cell *43*, 853–866.

Dieci, G., Preti, M., and Montanini, B. (2009). Eukaryotic snoRNAs: a paradigm for gene expression flexibility. Genomics *94*, 83–88.

Ding, D.Q., Okamasa, K., Yamane, M., Tsutsumi, C., Haraguchi, T., Yamamoto, M., and Hiraoka, Y. (2012). Meiosis-Specific Noncoding RNA Mediates Robust Pairing of Homologous Chromosomes in Meiosis. Science *336*, 732–736.

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature *505*, 696–700.

Dittmar, K.A., Goodenbour, J.M., and Pan, T. (2006). Tissue-Specific Differences in Human Transfer RNA Expression. PLoS Genet. *2*, e221.

Doherty, L., Sheen, M.R., Vlachos, A., Choesmel, V., O'Donohue, M.-F., Clinton, C., Schneider, H.E., Sieff, C.A., Newburger, P.E., Ball, S.E., et al. (2010). Ribosomal protein genes RPS10 and RPS26 are commonly mutated in Diamond-Blackfan anemia. Am. J. Hum. Genet. *86*, 222–228.

Doma, M.K., and Parker, R. (2007). RNA quality control in eukaryotes. Cell *131*, 660–668.

Dominski, Z., Yang, X.-C., Kaygun, H., Dadlez, M., and Marzluff, W.F. (2003). A 3“ exonuclease that specifically interacts with the 3” end of histone mRNA. Mol. Cell *12*, 295–305.

Dragon, F., Gallagher, J.E.G., Compagnone-Post, P.A., Mitchell, B.M., Porwancher, K.A., Wehner, K.A., Wormsley, S., Settlage, R.E., Shabanowitz, J., Osheim, Y., et al. (2002). A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. Nature *417*, 967–970.

Drake, K.M., Zygmunt, D., Mavrakis, L., Harbor, P., Wang, L., Comhair, S.A., Erzurum, S.C., and Aldred, M.A. (2011). Altered MicroRNA processing in heritable pulmonary arterial hypertension: an important role for Smad-8. Am. J. Respir. Crit. Care Med. *184*, 1400–1408.

Draper, D.E., and Reynaldo, L.P. (1999). RNA binding strategies of ribosomal proteins. Nucleic Acids Res. *27*, 381–388.

Draptchinskaia, N., Gustavsson, P., Andersson, B., Pettersson, M., Willig, T.N., Dianzani, I., Ball, S., Tchernia, G., Klar, J., Matsson, H., et al. (1999). The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. Nat. Genet. *21*, 169–175.

Dreyfuss, G., Choi, Y.D., and Adam, S.A. (1984). Characterization of heterogeneous nuclear RNA-protein complexes in vivo with monoclonal antibodies. Mol. Cell. Biol. *4*, 1104–1114.

Dreyfuss, G., Kim, V.N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. Nat. Rev. Mol. Cell Biol. *3*, 195–205.

Dufau, M.L., and Tsai-Morris, C.-H. (2007). Gonadotropin-regulated testicular helicase (GRTH/DDX25): an essential regulator of spermatogenesis. Trends Endocrinol. Metab. *18*, 314–320.

Dumitrescu, A.M., Liao, X.-H., Abdullah, M.S.Y., Lado-Abeal, J., Majed, F.A., Moeller, L.C., Boran, G., Schomburg, L., Weiss, R.E., and Refetoff, S. (2005). Mutations in SECISBP2 result in abnormal thyroid hormone metabolism. Nat. Genet. *37*, 1247–1252.

Durocher, F., Faure, R., Labrie, Y., Pelletier, L., Bouchard, I., and Laframboise, R. (2006). A novel mutation in the EIF2AK3 gene with variable expressivity in two patients with Wolcott-Rallison syndrome. Clin. Genet. *70*, 34–38.

Echeverria, G.V., and Cooper, T.A. (2012). RNA-binding proteins in microsatellite expansion disorders: Mediators of RNA toxicity. Brain Res. *1462*, 100–111.

Eddy, S.R. (1998). Profile hidden Markov models. Bioinformatics *14*, 755–763.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics *10*, 48.

Edvardson, S., Shaag, A., Kolesnikova, O., Gomori, J.M., Tarassov, I., Einbinder, T., Saada, A., and Elpeleg, O. (2007). Deleterious mutation in the mitochondrial arginyl-transfer RNA synthetase gene is associated with pontocerebellar hypoplasia. Am. J. Hum. Genet. *81*, 857–862.

Edwards, S.J., Gladwin, A.J., and Dixon, M.J. (1997). The mutational spectrum in Treacher Collins syndrome reveals a predominance of mutations that create a premature-termination codon. Am. J. Human Genet. *60*, 515–524.

Egan, E.D., and Collins, K. (2012). Biogenesis of telomerase ribonucleoproteins. Rna *18*, 1747–1759.

Elden, A.C., Kim, H.-J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M.M., et al. (2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature *466*, 1069–1075.

Eliazer, S., and Buszczak, M. (2011). Finding a niche: studies from the Drosophila ovary. Stem Cell Res. Ther. *2*, 45.

Ellis, J.C., and Brown, J.W. (2009). The RNase P family. RNA Biol. *6*, 362–369.

Elo, J.M., Yadavalli, S.S., Euro, L., Isohanni, P., Götz, A., Carroll, C.J., Valanne, L., Alkuraya, F.S., Uusimaa, J., Paetau, A., et al. (2012). Mitochondrial phenylalanyl-tRNA synthetase mutations underlie fatal infantile Alpers encephalopathy. Hum. Mol. Genet. *21*, 4521–4529.

Esakova, O., and Krasilnikov, A.S. (2010). Of proteins and RNA: The RNase P/MRP family. Rna *16*, 1725–1747.

Espert, L., Degols, G., Gongora, C., Blondel, D., Williams, B.R., Silverman, R.H., and Mechti, N. (2003). ISG20, a new interferon-induced RNase specific for single-stranded RNA, defines an alternative antiviral pathway against RNA genomic viruses. J. Biol. Chem. *278*, 16151–16158.

Espert, L., Eldin, P., Gongora, C., Bayard, B., Harper, F., Chelbi-Alix, M.K., Bertrand, E., Degols, G., and Mechti, N. (2006). The exonuclease ISG20 mainly localizes in the nucleolus and the Cajal (Coiled) bodies and is associated with nuclear SMN protein-containing complexes. J. Cell. Biochem. *98*, 1320–1333.

Espert, L., Rey, C., Gonzalez, L., Degols, G., Chelbi-Alix, M.K., Mechti, N., and Gongora, C. (2004). The exonuclease ISG20 is directly induced by synthetic dsRNA via NF-kappaB and IRF1 activation. Oncogene *23*, 4636–4640.

Faber, A.W., Vos, J.C., Vos, H.R., Ghazal, G., Elela, S.A., and Raué, H.A. (2004). The RNA catabolic enzymes Rex4p, Rnt1p, and Dbr1p show genetic interaction with trans-acting factors involved in processing of ITS1 in Saccharomyces cerevisiae pre-rRNA. Rna *10*, 1946–1956.

Fabian, M.R., Cieplak, M.K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T.F., et al. (2011). miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4–NOT. Nat. Struct. Mol. Biol. *18*, 1211–1217.

Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T., Svitkin, Y.V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J.A., et al. (2009). Mammalian miRNA RISC Recruits CAF1 and PABP to Affect PABP-Dependent Deadenylation. Mol. Cell *35*, 868–880.

Fabre, A., Charroux, B., Martinez-Vinson, C., Roquelaure, B., Odul, E., Sayar, E., Smith, H., Colomb, V., Andre, N., Hugot, J.-P., et al. (2012). SKIV2L mutations cause syndromic diarrhea, or trichohepatoenteric syndrome. Am. J. Hum. Genet. *90*, 689–692.

Fardaei, M. (2002). Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. Hum. Mol. Genet. *11*, 805–814.

Farrar, J.E., Nater, M., Caywood, E., McDevitt, M.A., Kowalski, J., Takemoto, C.M., Talbot, C.C., Meltzer, P., Esposito, D., Beggs, A.H., et al. (2008). Abnormalities of the large ribosomal subunit protein, Rpl35a, in Diamond-Blackfan anemia. Blood *112*, 1582–1592.

Fatica, A., and Tollervey, D. (2003). Insights into the structure and function of a guide RNP. Nat. Struct. Mol. Biol. *10*, 237–239.

Feinstein, M., Markus, B., Noyman, I., Shalev, H., Flusser, H., Shelef, I., Liani-Leibson, K., Shorer, Z., Cohen, I., Khateeb, S., et al. (2010). Pelizaeus-Merzbacher-like disease caused by AIMP1/p43 homozygous mutation. Am. J. Hum. Genet. *87*, 820–828.

Fell, V.L., and Schild-Poulter, C. (2015). The Ku heterodimer: function in DNA repair and beyond. Rev. Mutat. Res. *763*, 15–29.

Fernandez-Capetillo, O., Lee, A., Nussenzweig, M., and Nussenzweig, A. (2004). H2AX: the histone guardian of the genome. DNA Repair *3*, 959–967.

Fernandez-Vizarra, E., Berardinelli, A., Valente, L., Tiranti, V., and Zeviani, M. (2007). Nonsense mutation in pseudouridylate synthase 1 (PUS1) in two brothers affected by myopathy, lactic acidosis and sideroblastic anaemia (MLASA). J. Med. Genet. *44*, 173–180.

Filipowicz, W., and Pogacić, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. Curr. Opin. Cell Biol. *14*, 319–327.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., et al. (2010). The Pfam protein families database. Nucleic Acids Res. *38*, D211–D222.

Fogli, A., Wong, K., Eymard-Pierre, E., Wenger, J., Bouffard, J.-P., Goldin, E., Black, D.N., Boespflug-Tanguy, O., and Schiffmann, R. (2002). Cree leukoencephalopathy and CACH/VWM disease are allelic at the EIF2B5 locus. Ann. Neurol. *52*, 506–510.

Foulkes, W.D., Bahubeshi, A., Hamel, N., Pasini, B., Asioli, S., Baynam, G., Choong, C.S., Charles, A., Frieder, R.P., Dishop, M.K., et al. (2011). Extending the phenotypes associated with DICER1 mutations. Hum. Mutat. *32*, 1381–1384.

Fox, A.H., and Lamond, A.I. (2010). Paraspeckles. Cold Spring Harb. Perspect. Biol. *2*, a000687.

Freilich, S., Massingham, T., Bhattacharyya, S., Ponsting, H., Lyons, P.A., Freeman, T.C., and Thornton, J.M. (2005). Relationship between the tissue-specificity of mouse gene expression and the evolutionary origin and function of the proteins. Genome Biol. *6*, R56.

Freude, K., Hoffmann, K., Jensen, L.-R., Delatycki, M.B., Portes, des, V., Moser, B., Hamel, B., van Bokhoven, H., Moraine, C., Fryns, J.-P., et al. (2004). Mutations in the FTSJ1 gene coding for a novel S-adenosylmethionine-binding protein cause nonsyndromic X-linked mental retardation. Am. J. Hum. Genet. *75*, 305–309.

Frost, R.J.A., Hamra, F.K., Richardson, J.A., Qi, X., Bassel-Duby, R., and Olson, E.N. (2010). MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. Proc. Natl. Acad. Sci. U. S. a. *107*, 11847–11852.

Fuller, M.T. (1998). Genetic control of cell proliferation and differentiation in Drosophila spermatogenesis. Sem. Cell & Dev. Biol. *9*, 433–444.

Gabel, H.W., and Ruvkun, G. (2008). The exonuclease ERI-1 has a conserved dual role in 5.8S rRNA processing and RNAi. Nat. Struct. Mol. Biol. *15*, 531–533.

Gaipl, U.S., Voll, R.E., Sheriff, A., Franz, S., Kalden, J.R., and Herrmann, M. (2005). Impaired clearance of dying cells in systemic lupus erythematosus. Autoimmun. Rev. *4*, 189–194.

Galante, P.A.F., Sandhu, D., de Sousa Abreu, R., Gradassi, M., Slager, N., Vogel, C., de Souza, S.J., and Penalva, L.O.F. (2009). A comprehensive in silico expression analysis of RNA binding proteins in normal and tumor tissue: Identification of potential players in tumor formation. RNA Biol. *6*, 426–433.

Galmiche, L., Serre, V., Beinat, M., Assouline, Z., Lebre, A.-S., Chrétien, D., Nietschke, P., Benes, V., Boddaert, N., Sidi, D., et al. (2011). Exome sequencing identifies MRPL3 mutation in mitochondrial cardiomyopathy. Hum. Mutat. *32*, 1225–1231.

Gambardella, A., Mazzei, R., Toscano, A., Annesi, G., Pasqua, A., Annesi, F., Quattrone, F., Oliveri, R.L., Valentino, P., Bono, F., et al. (1998). Spinal muscular atrophy due to an isolated deletion of exon 8 of the telomeric survival motor neuron gene. Ann. Neurol. *44*, 836–839.

Gamberi, C., Johnstone, O., and Lasko, P. (2006). Drosophila RNA binding proteins. Int. Rev. Cytol. *248*, 43–139.

Gazda, H.T., Grabowska, A., Merida-Long, L.B., Latawiec, E., Schneider, H.E., Lipton, J.M., Vlachos, A., Atsidaftos, E., Ball, S.E., Orfali, K.A., et al. (2006). Ribosomal protein S24 gene is mutated in Diamond-Blackfan anemia. Am. J. Hum. Genet. *79*, 1110–1118.

Gazda, H.T., Sheen, M.R., Vlachos, A., Choesmel, V., O'Donohue, M.-F., Schneider, H., Darras, N., Hasman, C., Sieff, C.A., Newburger, P.E., et al. (2008). Ribosomal protein L5 and L11 mutations are associated with cleft palate and abnormal thumbs in Diamond-Blackfan anemia patients. Am. J. Hum. Genet. *83*, 769–780.

Gedeon, A.K., Baker, E., Robinson, H., Partington, M.W., Gross, B., Manca, A., Korn, B., Poustka, A., Yu, S., and Sutherland, G.R. (1992). Fragile X syndrome without CCG amplification has an FMR1 deletion. Nat. Genet. *1*, 341–344.

Gehman, L.T., Stoilov, P., Maguire, J., Damianov, A., Lin, C.-H., Shiue, L., Ares, M., Mody, I., and Black, D.L. (2011). The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. Nat. Genet. *43*, 706–711.

Gelpi, C., Sontheimer, E.J., and Rodriguez-Sanchez, J.L. (1992). Autoantibodies against a serine tRNA-protein complex implicated in cotranslational selenocysteine insertion. Proc. Natl. Acad. Sci. U. S. a. *89*, 9739–9743.

Georgiev, S., Boyle, A.P., Jayasurya, K., Ding, X., Mukherjee, S., and Ohler, U. (2010). Evidence-ranked motif identification. Genome Biol. *11*, R19.

Gerstberger, S., Hafner, M., and Tuschl, T. (2013). Learning the language of post-transcriptional gene regulation. Genome Biol. *14*, 130.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014a). A census of human RNA-binding proteins. Nat. Rev. Genet. *15*, 829–845.

Gerstberger, S., Hafner, M., Ascano, M., and Tuschl, T. (2014b). Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. Adv. Exp. Med. Biol. *825*, 1–55.

Ghezzi, D., Baruffini, E., Haack, T.B., Invernizzi, F., Melchionda, L., Dallabona, C., Strom, T.M., Parini, R., Burlina, A.B., Meitinger, T., et al. (2012). Mutations of the mitochondrial-tRNA modifier MTO1 cause hypertrophic cardiomyopathy and lactic acidosis. Am. J. Hum. Genet. *90*, 1079–1087.

Gibson, W.T., Hood, R.L., Zhan, S.H., Bulman, D.E., Fejes, A.P., Moore, R., Mungall, A.J., Eydoux, P., Babul-Hirji, R., An, J., et al. (2012). Mutations in EZH2 cause Weaver syndrome. Am. J. Hum. Genet. *90*, 110–118.

Gill, S., McManus, A.P., Crew, A.J., Benjamin, H., Sheer, D., Gusterson, B.A., Pinkerton, C.R., Patel, K., Cooper, C.S., and Shipley, J.M. (1995). Fusion of the EWS gene to a DNA segment from 9q22-31 in a human myxoid chondrosarcoma. Genes Chromosom. Cancer *12*, 307–310.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. *582*, 1977–1986.

Gold, H.A., Craft, J., Hardin, J.A., Bartkiewicz, M., and Altman, S. (1988). Antibodies in human serum that precipitate ribonuclease P. Proc. Natl. Acad. Sci. U. S. a. *85*, 5483–5487.

Gongora, C., David, G., Pintard, L., Tissot, C., Hua, T.D., Dejean, A., and Mechti, N. (1997). Molecular cloning of a new interferon-induced PML nuclear body-associated protein. J. Biol. Chem. *272*, 19457–19463.

Gonzatti-Haces, M., Seth, A., Park, M., Copeland, T., Oroszlan, S., and Vande Woude, G.F. (1988). Characterization of the TPR-MET oncogene p65 and the MET protooncogene p140 protein-tyrosine kinases. Proc. Natl. Acad. Sci. U. S. a. *85*, 21–25.

Gordon, C.T., Petit, F., Oufadem, M., Decaestecker, C., Jourdain, A.-S., Andrieux, J., Malan, V., Alessandri, J.-L., Baujat, G., Baumann, C., et al. (2012). EFTUD2 haploinsufficiency leads to syndromic oesophageal atresia. J. Med. Genet. *49*, 737–746.

Götz, A., Tyynismaa, H., Euro, L., Ellonen, P., Hyötyläinen, T., Ojala, T., Hämäläinen, R.H., Tommiska, J., Raivio, T., Oresic, M., et al. (2011). Exome sequencing identifies mitochondrial alanyl-tRNA synthetase mutations in infantile mitochondrial cardiomyopathy. Am. J. Hum. Genet. *88*, 635–642.

Granneman, S., and Baserga, S.J. (2004). Ribosome biogenesis: of knobs and RNA processing. Exp. Cell Res. *296*, 43–50.

Graubert, T.A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris, C.C., Koboldt, D.C., Larson, D.E., et al. (2012). Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat. Genet. *44*, 53–57.

Greenfield, E. (2014). Antibodies: A Laboratory Manual. CHSL press (New York)

Greenway, M.J., Andersen, P.M., Russ, C., Ennis, S., Cashman, S., Donaghy, C., Patterson, V., Swingler, R., Kieran, D., Prehn, J., et al. (2006). ANG mutations segregate with familial and "sporadic" amyotrophic lateral sclerosis. Nat. Genet. *38*, 411–413.

Griffin, J.H., Leung, J., Bruner, R.J., Caligiuri, M.A., and Briesewitz, R. (2003). Discovery of a fusion kinase in EOL-1 cells and idiopathic hypereosinophilic syndrome. Proc. Natl. Acad. Sci. U. S. a. *100*, 7830–7835.

Grishin, N.V. (2001). KH domain: one motif, two folds. Nucleic Acids Res. *29*, 638–643.

Grohmann, K., Schuelke, M., Diers, A., HOFFMANN, K., Lucke, B., Adams, C., Bertini, E., Leonhardt-Horti, H., Muntoni, F., Ouvrier, R., et al. (2001). Mutations in the gene encoding immunoglobulin mu-binding protein 2 cause spinal muscular atrophy with respiratory distress type 1. Nat. Genet. *29*, 75–77.

Grummt, I., Kuhn, A., Bartsch, I., and Rosenbauer, H. (1986). A transcription terminator located upstream of the mouse rDNA initiation site affects rRNA synthesis. Cell *47*, 901–911.

Grzmil, M., and Hemmings, B.A. (2012). Translation regulation as a therapeutic target in cancer. Cancer Res. *72*, 3891–3900.

Guan, M.-X., Yan, Q., Li, X., Bykhovskaya, Y., Gallo-Teran, J., Hajek, P., Umeda, N., Zhao, H., Garrido, G., Mengesha, E., et al. (2006). Mutation in TRMU related to transfer RNA modification modulates the phenotypic expression of the deafness-associated mitochondrial 12S ribosomal RNA mutations. Am. J. Hum. Genet. *79*, 291–302.

Gubitz, A.K., Feng, W., and Dreyfuss, G. (2004). The SMN complex. Exp. Cell Res. *296*, 51–56.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Guo, W., Schafer, S., Greaser, M.L., Radke, M.H., Liss, M., Govindarajan, T., Maatz, H., Schulz, H., Li, S., Parrish, A.M., et al. (2012). RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. Nat. Med. *18*, 766–773.

Gurney, T. (1985). Characterization of mouse 45S ribosomal RNA subspecies suggests that the first processing cleavage occurs 600 +/- 100 nucleotides from the 5“ end and the second 500 +/- 100 nucleotides from the 3” end of a 13.9 kb precursor. Nucleic Acids Res. *13*, 4905–4919.

Gwinn-Hardy, K., Chen, J.Y., Liu, H.C., Liu, T.Y., Boss, M., Seltzer, W., Adam, A., Singleton, A., Koroshetz, W., Waters, C., et al. (2000). Spinocerebellar ataxia type 2 with parkinsonism in ethnic Chinese. Neurology *55*, 800–805.

Haack, T.B., Kopajtich, R., Freisinger, P., Wieland, T., Rorbach, J., Nicholls, T.J., Baruffini, E., Walther, A., Danhauser, K., Zimmermann, F.A., et al. (2013). ELAC2 Mutations Cause a Mitochondrial RNA Processing Defect Associated with Hypertrophic Cardiomyopathy. Am. J. Hum. Genet. *93*, 211–223.

Hackman, P., Sarparanta, J., Lehtinen, S., Vihola, A., Evilä, A., Jonson, P.H., Luque, H., Kere, J., Screen, M., Chinnery, P.F., et al. (2012). Welander distal myopathy is caused by a mutation in the RNA-binding protein TIA1. Ann. Neurol. *73*, 500–509.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.C., Munschauer, M., et al. (2010a). Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. Cell *141*, 129–141.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., et al. (2010b). PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. J. Vis. Exp. e2034–e2034.

Hafner, M., Max, K.E.A., Bandaru, P., Morozov, P., Gerstberger, S., Brown, M., Molina, H., and Tuschl, T. (2013). Identification of mRNAs bound and regulated by human LIN28 proteins and molecular requirements for RNA recognition. Rna *19*, 613–626.

Hafner, M., Renwick, N., Farazi, T.A., Mihailovic, A., Pena, J.T.G., and Tuschl, T. (2012). Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. Methods *58*, 164–170.

Haft, D.H., Loftus, B.J., Richardson, D.L., Yang, F., Eisen, J.A., Paulsen, I.T., and White, O. (2001). TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res. *29*, 41–43.

Hage, El, A., Koper, M., Kufel, J., and Tollervey, D. (2008). Efficient termination of transcription by RNA polymerase I requires the 5' exonuclease Rat1 in yeast. Genes Dev. *22*, 1069–1081.

Hagerman, P. (2013). Fragile X-associated tremor/ataxia syndrome (FXTAS): pathology and mechanisms. Acta Neuropathol *126*, 1–19.

Hagerman, R.J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B., and Hagerman, P.J. (2001). Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. Neurology *57*, 127–130.

Hahnen, E., Schönling, J., Rudnik-Schöneborn, S., Raschke, H., Zerres, K., and Wirth, B. (1997). Missense mutations in exon 6 of the survival motor neuron gene in patients with spinal muscular atrophy (SMA). Hum. Mol. Genet. *6*, 821–825.

Hall, A.E., Turnbull, C., and Dalmay, T. (2013). Y RNAs: recent developments. BioMol. Concepts *4*, 103–110.

Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., and McKusick, V.A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. *33*, D514–D517.

Hanna, J.H., Saha, K., and Jaenisch, R. (2010). Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. Cell *143*, 508–525.

Hannan, K.M., Sanij, E., Rothblum, L.I., Hannan, R.D., and Pearson, R.B. (2013). Dysregulation of RNA polymerase I transcription during disease. Bba *1829*, 342–360.

Hanson, K.A., Kim, S.H., and Tibbetts, R.S. (2011). RNA-binding proteins in neurodegenerative disease: TDP-43 and beyond. WIREs RNA *3*, 265–285.

Harms, M.B., Ori-McKenney, K.M., Scoto, M., Tuck, E.P., Bell, S., Ma, D., Masi, S., Allred, P., Al-Lozi, M., Reilly, M.M., et al. (2012). Mutations in the tail domain of DYNC1H1 cause dominant spinal muscular atrophy. Neurology *78*, 1714–1720.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., van de Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. Nature *489*, 391–399.

Hayashi, Y.K., Matsuda, C., Ogawa, M., Goto, K., Tominaga, K., Mitsuhashi, S., Park, Y.-E., Nonaka, I., Hino-Fukuyo, N., Haginoya, K., et al. (2009). Human PTRF mutations cause secondary deficiency of caveolins resulting in muscular dystrophy with generalized lipodystrophy. J. Clin. Invest. *119*, 2623–2633.

Hein, N., Hannan, K.M., George, A.J., Sanij, E., and Hannan, R.D. (2013). The nucleolus: an emerging target for cancer therapy. Trends Mol. Med. *19*, 643–654.

Heindl, K., and Martinez, J. (2010). Nol9 is a novel polynucleotide 5'-kinase involved in ribosomal RNA processing. Embo J. *29*, 4161–4171.

Heiss, N.S., Knight, S.W., Vulliamy, T.J., Klauck, S.M., Wiemann, S., Mason, P.J., Poustka, A., and Dokal, I. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nat. Genet. *19*, 32–38.

Hekman, K.E., Yu, G.-Y., Brown, C.D., Zhu, H., Du, X., Gervin, K., Undlien, D.E., Peterson, A., Stevanin, G., Clark, H.B., et al. (2012). A conserved eEF2 coding variant in SCA26 leads to loss of translational fidelity and increased susceptibility to proteostatic insult. Hum. Mol. Genet. *21*, 5472–5483.

Henderson, S., and Sollner-Webb, B. (1986). A transcriptional terminator is a novel element of the promoter of the mouse ribosomal RNA gene. Cell *47*, 891–900.

Hendrick, J.P., Wolin, S.L., Rinke, J., Lerner, M.R., and Steitz, J.A. (1981). Ro small cytoplasmic ribonucleoproteins are a subclass of La ribonucleoproteins: further characterization of the Ro and La small ribonucleoproteins from uninfected mammalian cells. Mol. Cell. Biol. *1*, 1138–1149.

Henneke, M., Diekmann, S., Ohlenbusch, A., Kaiser, J., Engelbrecht, V., Kohlschuetter, A., Kraetzner, R., Madruga-Garrido, M., Mayer, M., Opitz, L., et al. (2009). RNASET2-deficient cystic leukoencephalopathy resembles congenital cytomegalovirus brain infection. Nat. Genet. *41*, 773–775.

Henras, A.K., Soudet, J., Gérus, M., Lebaron, S., Caizergues-Ferrer, M., Mougin, A., and Henry, Y. (2008). The post-transcriptional steps of eukaryotic ribosome biogenesis. Cell. Mol. Life Sci. *65*, 2334–2359.

Henras, A.K., Plisson-Chastang, C., O'Donohue, M.-F., Chakraborty, A., and Gleizes, P.-E. (2014). An overview of pre-ribosomal RNA processing in eukaryotes. WIREs RNA *6*, 225–242.

Hill, D.A., Ivanovich, J., Priest, J.R., Gurnett, C.A., Dehner, L.P., Desruisseau, D., Jarzembowski, J.A., Wikenheiser-Brokamp, K.A., Suarez, B.K., Whelan, A.J., et al. (2009). DICER1 mutations in familial pleuropulmonary blastoma. Science *325*, 965.

Hoefig, K.P., Rath, N., Heinz, G.A., Wolf, C., Dameris, J., Schepers, A., Kremmer, E., Ansel, K.M., and Heissmeyer, V. (2013). Eri1 degrades the stem-loop of oligouridylated histone mRNAs to induce replication-dependent decay. Nat. Struct. Mol. Biol. *20*, 73–81.

Hopper, A.K., Pai, D.A., and Engelke, D.R. (2010). Cellular dynamics of tRNAs and their genes. FEBS Lett. *584*, 310–317.

Houmard, B., Small, C., Yang, L.Z., Naluai-Cecchini, T., Cheng, E., Hassold, T., and Griswold, M. (2009). Global Gene Expression in the Human Fetal Testis and Ovary. Biol. Reprod. *81*, 438–443.

Houseley, J., LaCava, J., and Tollervey, D. (2006). RNA-quality control by the exosome. Nat. Rev. Mol. Cell Biol. *7*, 529–539.

Howe, J.R., Roth, S., Ringold, J.C., Summers, R.W., Järvinen, H.J., Sistonen, P., Tomlinson, I.P., Houlston, R.S., Bevan, S., Mitros, F.A., et al. (1998). Mutations in the SMAD4/DPC4 gene in juvenile polyposis. Science *280*, 1086–1088.

Hsiao, Y.-Y., Yang, C.-C., Lin, C.L., Lin, J.L.J., Duh, Y., and Yuan, H.S. (2011). Structural basis for RNA trimming by RNase T in stable RNA 3 '-end maturation. Nature Chem. Biol. *7*, 236–243.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Huang, Z.-P., Zhou, H., He, H.-L., Chen, C.-L., Liang, D., and Qu, L.-H. (2005). Genome-wide analyses of two families of snoRNA genes from Drosophila melanogaster, demonstrating the extensive utilization of introns for coding of snoRNAs. Rna *11*, 1303–1316.

Huppke, P., Laccone, F., Krämer, N., Engel, W., and Hanefeld, F. (2000). Rett syndrome: analysis of MECP2 and clinical characterization of 31 patients. Hum. Mol. Genet. *9*, 1369–1375.

Hussain, S., Sajini, A.A., Blanco, S., Dietmann, S., Lombard, P., Sugimoto, Y., Paramor, M., Gleeson, J.G., Odom, D.T., Ule, J., et al. (2013). NSun2-Mediated Cytosine-5 Methylation of Vault Noncoding RNA Determines Its Processing into Regulatory Small RNAs. Cell Rep. *4*, 255–261.

Ichikawa, H., Shimizu, K., Hayashi, Y., and Ohki, M. (1994). An RNA-binding protein gene, TLS/FUS, is fused to ERG in human myeloid leukemia with t(16;21) chromosomal translocation. Cancer Res. *54*, 2865–2868.

Iida, T., Kawaguchi, R., and Nakayama, J.-I. (2006). Conserved Ribonuclease, Eri1, Negatively Regulates Heterochromatin Assembly in Fission Yeast. Curr. Biol. *16*, 1459–1464.

IJlst, L., Loupatty, F.J., Ruiter, J.P.N., Duran, M., Lehnert, W., and Wanders, R.J.A. (2002). 3-Methylglutaconic aciduria type I is caused by mutations in AUH. Am. J. Hum. Genet. *71*, 1463–1466.

Ingolia, N.T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. Nat. Rev. Genet. *15*, 205–213.

Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. *11*, 113–127.

Jan, C.H., Friedman, R.C., Ruby, J.G., and Bartel, D.P. (2011). Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs. Nature *469*, 97–101.

Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. Trends Biochem. Sci. *36*, 19–29.

Jankowsky, E., and Fairman-Williams, M.E. (2010). An introduction to RNA helicases: superfamilies, families, and major themes. RNA Helicases *19*, 1–31.

Jarrous, N. (2002). Human ribonuclease P: subunits, function, and intranuclear localization. Rna *8*, 1–7.

Johanson, T.M., Lew, A.M., and Chong, M.M.W. (2013). MicroRNA-independent roles of the RNase III enzymes Drosha and Dicer. Open Biol. *3*, 130144.

Johnston, J.J., Teer, J.K., Cherukuri, P.F., Hansen, N.F., Loftus, S.K., NIH Intramural Sequencing Center (NISC), Chong, K., Mullikin, J.C., and Biesecker, L.G. (2010). Massively parallel sequencing of exons on the X chromosome identifies RBM10 as the gene that causes a syndromic form of cleft palate. Am. J. Hum. Genet. *86*, 743–748.

Jordan, B.R. (1975). Demonstration of intact 26 S ribosomal RNA molecules in Drosophila cells. J. Mol. Biol. *98*, 277–280.

Jordanova, A., Irobi, J., Thomas, F.P., Van Dijck, P., Meerschaert, K., Dewil, M., Dierick, I., Jacobs, A., De Vriendt, E., Guergueltcheva, V., et al. (2006). Disrupted function and axonal distribution of mutant tyrosyl-tRNA synthetase in dominant intermediate Charcot-Marie-Tooth neuropathy. Nat. Genet. *38*, 197–202.

Jung, H., Gkogkas, C.G., Sonenberg, N., and Holt, C.E. (2014). Remote Control of Gene Function by Local Translation. Cell *157*, 26–40.

Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature *468*, 664–668.

Kalscheuer, V.M., Freude, K., Musante, L., Jensen, L.R., Yntema, H.G., Gecz, J., Sefiani, A., Hoffmann, K., Moser, B., Haas, S., et al. (2003). Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation. Nat. Genet. *35*, 313–315.

Kalsotra, A., and Cooper, T.A. (2011). Functional consequences of developmentally regulated alternative splicing. Nat. Rev. Genet. *12*, 715–729.

Kandel, E.R., Dudai, Y., and Mayford, M.R. (2014). The Molecular and Systems Biology of Memory. Cell *157*, 163–186.

Kanegane, H., Kasahara, Y., Okamura, J., Hongo, T., Tanaka, R., Nomura, K., Kojima, S., and Miyawaki, T. (2005). Identification of DKC1 gene mutations in Japanese patients with X-linked dyskeratosis congenita. Br. J. Haematol. *129*, 432–434.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

Kaneko, H., Dridi, S., Tarallo, V., Gelfand, B.D., Fowler, B.J., Cho, W.G., Kleinman, M.E., Ponicsan, S.L., Hauswirth, W.W., Chiodo, V.A., et al. (2011). DICER1 deficit induces Alu RNA

toxicity in age-related macular degeneration. Nature *471*, 325–330.

Kang, M.K., and Han, S.J. (2011). Post-transcriptional and post-translational regulation during mouse oocyte maturation. BMB Rep. *44*, 147–157.

Kapeli, K., and Yeo, G.W. (2012). Genome-wide approaches to dissect the roles of RNA binding proteins in translational control: implications for neurological diseases. Front. Neurosci. *6*, 144.

Karaca, E., Weitzer, S., Pehlivan, D., Shiraishi, H., Gogakos, T., Hanada, T., Jhangiani, S.N., Wiszniewski, W., Withers, M., Campbell, I.M., et al. (2014). Human CLP1 mutations alter tRNA biogenesis, affecting both peripheral and central nervous system function. Cell *157*, 636–650.

Kass, S., and Sollner-Webb, B. (1990). The first pre-rRNA-processing event occurs in a large complex: analysis by gel retardation, sedimentation, and UV cross-linking. Mol. Cell. Biol. *10*, 4920–4931.

Kawase, T., Ichikawa, H., Ohta, T., Nozaki, N., Tashiro, F., Ohki, R., and Taya, Y. (2008). p53 target gene AEN is a nuclear exonuclease required for p53-dependent apoptosis. Oncogene *27*, 3797–3810.

Kechavarzi, B., and Janga, S.C. (2014). Dissecting the expression landscape of RNA-binding proteins in human cancers. Genome Biol. *15*, R14.

Kedde, M., van Kouwenhove, M., Zwart, W., Oude Vrielink, J.A.F., Elkon, R., and Agami, R. (2010). A Pumilio-induced RNA structure switch in p27-3′ UTR controls miR-221 and miR-222 accessibility. Nat. Cell Biol. *12*, 1014–1020.

Kedersha, N., and Anderson, P. (2007). Mammalian stress granules and processing bodies. Meth. Enzymol. *431*, 61–81.

Kee, K., Angeles, V.T., Flores, M., Nguyen, H.N., and Reijo Pera, R.A. (2009). Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation. Nature *462*, 222–225.

Keen, T.J., Hims, M.M., McKie, A.B., Moore, A.T., Doran, R.M., Mackey, D.A., Mansfield, D.C., Mueller, R.F., Bhattacharya, S.S., Bird, A.C., et al. (2002). Mutations in a protein target of the Pim-1 kinase associated with the RP9 form of autosomal dominant retinitis pigmentosa. Eur. J. Hum. Genet. *10*, 245–249.

Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. Nat. Rev. Genet. *8*, 533–543.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. *10*, 845–858.

Kellner, M., Rohrmoser, M., Forné, I., Voss, K., Burger, K., Mühl, B., Gruber-Eber, A., Kremmer, E., Imhof, A., and Eick, D. (2015). DEAD-box helicase DDX27 regulates 3' end formation of ribosomal 47S RNA and stably associates with the PeBoW-complex. Exp. Cell Res. *334*, 146–159.

Kempers-Veenstra, A.E., Oliemans, J., Offenberg, H., Dekker, A.F., Piper, P.W., Planta, R.J., and Klootwijk, J. (1986). 3'-End formation of transcripts from the yeast rRNA operon. Embo J. *5*, 2703–2710.

Kenan, D.J., and Keene, J.D. (2004). La gets its wings. Nat. Struct. Mol. Biol. *11*, 303–305.

Kennedy, S., Wang, D., and Ruvkun, G. (2004). A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans. Nature *427*, 645–649.

Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. Nat. Rev. Genet. *11*, 345–355.

Kerkis, J. (1931). The Growth of the Gonads in DROSOPHILA MELANOGASTER. Genetics *16*, 212–224.

Kerner, P., Degnan, S.M., Marchand, L., Degnan, B.M., and Vervoort, M. (2011). Evolution of RNA-binding proteins in animals: insights from genome-wide analysis in the sponge Amphimedon queenslandica. Mol. Biol. Evol. *28*, 2289–2303.

Khan, M.A., Rafiq, M.A., Noor, A., Hussain, S., Flores, J.V., Rupp, V., Vincent, A.K., Malli, R., Ali, G., Khan, F.S., et al. (2012). Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability. Am. J. Hum. Genet. *90*, 856–863.

Khatter, H., Myasnikov, A.G., Natchiar, S.K., and Klaholz, B.P. (2015). Structure of the human 80S ribosome. Nature *520*, 640–645.

Kiledjian, M., and Dreyfuss, G. (1992). Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. Embo J. *11*, 2655–2664.

Kim, C.A., and Bowie, J.U. (2003). SAM domains: uniform structure, diversity of function. Trends Biochem. Sci. *28*, 625–628.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013a). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kim, H.J., Kim, N.C., Wang, Y.-D., Scarborough, E.A., Moore, J., Diaz, Z., MacLea, K.S., Freibaum, B., Li, S., Molliex, A., et al. (2013b). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature *495*, 467–473.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. *10*, 126–139.

Kirilly, D., and Xie, T. (2007). The Drosophila ovary: an active stem cell community. Cell Res. *17*, 15–25.

Kirwan, M., Walne, A.J., Plagnol, V., Velangi, M., Ho, A., Hossain, U., Vulliamy, T., and Dokal, I. (2012). Exome sequencing identifies autosomal-dominant SRP72 mutations associated with familial aplasia and myelodysplasia. Am. J. Hum. Genet. *90*, 888–892.

Kiss, T., Fayet, E., Jády, B.E., Richard, P., and Weber, M. (2006). Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. Cold Spring Harb. Symp. Quant. Biol. *71*, 407–417.

Kiss, T. (2004). Biogenesis of small nuclear RNPs. J. Cell Sci. *117*, 5949–5951.

Klauck, S.M., Felder, B., Kolb-Kokocinski, A., Schuster, C., Chiocchetti, A., Schupp, I., Wellenreuther, R., Schmötzer, G., Poustka, F., Breitenbach-Koller, L., et al. (2006). Mutations in the ribosomal protein gene RPL10 suggest a novel modulating disease mechanism for autism. Mol Psychiatry *11*, 1073–1084.

Klein, C., and Westenberger, A. (2012). Genetics of Parkinson's disease. Cold Spring Harb. Perspect. Med. *2*, a008888.

Knight, S.J.L., Flannery, A.V., Hirst, M.C., Campbell, L., Christodoulou, Z., Phelps, S.R., Pointon, J., Middleton-Price, H.R., Barnicoat, A., Pembrey, M.E., et al. (1993). Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell *74*, 127–134.

Knight, S.W., Heiss, N.S., Vulliamy, T.J., Greschner, S., Stavrides, G., Pai, G.S., Lestringant, G., Varma, N., Mason, P.J., Dokal, I., et al. (1999). X-linked dyskeratosis congenita is predominantly caused by missense mutations in the DKC1 gene. Am. J. Hum. Genet. *65*, 50–58.

Knight, S.W., Vulliamy, T.J., Morgan, B., Devriendt, K., Mason, P.J., and Dokal, I. (2001). Identification of novel DKC1 mutations in patients with dyskeratosis congenita: implications for pathophysiology and diagnosis. Hum. Genet. *108*, 299–303.

Kobayashi, H., Abe, K., Matsuura, T., Ikeda, Y., Hitomi, T., Akechi, Y., Habu, T., Liu, W., Okuda, H., and Koizumi, A. (2011). Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. Am. J. Hum. Genet. *89*, 121–130.

Kominami, R., Urano, Y., Mishima, Y., and Muramatsu, M. (1981). Organization of ribosomal RNA gene repeats of the mouse. Nucleic Acids Res. *9*, 3219–3233.

Kondrashov, N., Pusic, A., Stumpf, C.R., Shimizu, K., Hsieh, A.C., Xue, S., Ishijima, J., Shiroishi, T., and Barna, M. (2011). Ribosome-Mediated Specificity in Hox mRNA Translation and Vertebrate Tissue Patterning. Cell *145*, 383–397.

Konig, J., Zarnack, K., Luscombe, N.M., and Ule, J. (2011). Protein-RNA interactions: new genomic technologies and perspectives. Nat. Rev. Genet. *13*, 77–83.

Kornblihtt, A.R., Schor, I.E., Allo, M., Dujardin, G., Petrillo, E., and Muñoz, M.J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat. Rev. Mol. Cell Biol. *14*, 153–165.

Korobeinikova, A.V., Garber, M.B., and Gongadze, G.M. (2012). Ribosomal proteins: Structure, function, and evolution. Biochemistry Moscow *77*, 562–574.

Koss-Harnes, D., Høyheim, B., Anton-Lamprecht, I., Gjesti, A., Jørgensen, R.S., Jahnsen, F.L.,

Olaisen, B., Wiche, G., and Gedde-Dahl, T. (2002). A site-specific plectin mutation causes dominant epidermolysis bullosa simplex Ogna: two identical de novo mutations. J. Invest. Dermatol. *118*, 87–93.

Kotaja, N., and Sassone-Corsi, P. (2007). The chromatoid body: a germ-cell-specific RNA-processing centre. Nat. Rev. Mol. Cell Biol. *8*, 85–90.

Kotova, E., Lodhi, N., Jarnik, M., Pinnola, A.D., Ji, Y., and Tulin, A.V. (2011). Drosophila histone H2A variant (H2Av) controls poly(ADP-ribose) polymerase 1 (PARP1) activation in chromatin. Proc. Natl. Acad. Sci. U.S.a. *108*, 6205–6210.

Köhn, M., Pazaitis, N., and Hüttelmaier, S. (2013). Why YRNAs? About Versatile RNAs and Their Functions. Biomolecules *3*, 143–156.

Kraus, M.R.-C., Clauin, S., Pfister, Y., Di Maïo, M., Ulinski, T., Constam, D., Bellanné-Chantelot, C., and Grapin-Botton, A. (2012). Two mutations in human BICC1 resulting in Wnt pathway hyperactivity associated with cystic renal dysplasia. Hum. Mutat. *33*, 86–90.

Kremer, E.J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R., and Richards, R.I. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science *252*, 1711–1714.

Kressler, D., Linder, P., and la Cruz, de, J. (1999). Protein trans-acting factors involved in ribosome biogenesis in Saccharomyces cerevisiae. Mol. Cell. Biol. *19*, 7897–7912.

Krishna, S.S., Majumdar, I., and Grishin, N.V. (2003). Structural classification of zinc fingers: survey and summary. Nucleic Acids Res. *31*, 532–550.

Kuchta, K., Knizewski, L., Wyrwicz, L.S., Rychlewski, L., and Ginalski, K. (2009). Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. Nucleic Acids Res. *37*, 7701–7714.

Kufel, J., Dichtl, B., and Tollervey, D. (1999). Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3“ ETS but not the 5” ETS. Rna *5*, 909–917.

Kuhn, A., and Grummt, I. (1989). 3'-end formation of mouse pre-rRNA involves both transcription termination and a specific processing reaction. Genes Dev. *3*, 224–231.

Kuwasako, K., He, F., Inoue, M., Tanaka, A., Sugano, S., Güntert, P., Muto, Y., and Yokoyama, S. (2006). Solution structures of the SURP domains and the subunit-assembly mechanism within the splicing factor SF3a complex in 17S U2 snRNP. Structure *14*, 1677–1689.

Kwiatkowski, T.J., Bosco, D.A., Leclerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T., et al. (2009). Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science *323*, 1205–1208.

Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1122–1130.

Labhart, P., and Reeder, R.H. (1986). Characterization of three sites of RNA 3' end formation in the Xenopus ribosomal gene spacer. Cell *45*, 431–443.

Labhart, P., and Reeder, R.H. (1987). A 12-base-pair sequence is an essential element of the ribosomal gene terminator in Xenopus laevis. Mol. Cell. Biol. *7*, 1900–1905.

Lachke, S.A., Alkuraya, F.S., Kneeland, S.C., Ohn, T., Aboukhalil, A., Howell, G.R., Saadi, I., Cavallesco, R., Yue, Y., Tsai, A.C.-H., et al. (2011). Mutations in the RNA granule component TDRD7 cause cataract and glaucoma. Science *331*, 1571–1576.

Lafontaine, D.L., and Tollervey, D. (2001). The function and synthesis of ribosomes. Nat. Rev. Mol. Cell Biol. *2*, 514–520.

Lafontaine, D.L.J. (2015). Noncoding RNAs in eukaryotic ribosome biogenesis and function. Nat. Struct. Mol. Biol. *22*, 11–19.

Lagier-Tourenne, C., Polymenidou, M., and Cleveland, D.W. (2010). TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. Hum. Mol. Genet. *19*, R46–R64.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lasko, P. (2000). The Drosophila melanogaster genome: Translation factors and RNA binding proteins. J. Cell Biol. *150*, F51–F56.

Lasko, P. (2013). The DEAD-box helicase Vasa: evidence for a multiplicity of functions in RNA processes and developmental biology. Bba *1829*, 810–816.

Latour, P., Thauvin-Robinet, C., Baudelet-Méry, C., Soichot, P., Cusin, V., Faivre, L., Locatelli, M.-C., Mayençon, M., Sarcey, A., Broussolle, E., et al. (2010). A major determinant for binding and aminoacylation of tRNA(Ala) in cytoplasmic Alanyl-tRNA synthetase is mutated in dominant axonal Charcot-Marie-Tooth disease. Am. J. Hum. Genet. *86*, 77–82.

Lee, E.B., Lee, V.M.Y., and Trojanowski, J.Q. (2012). Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. Nat. Rev. Neurosci. *13*, 38–50.

Lee, M.-H., and Schedl, T. (2006). RNA-binding proteins. WormBook 1–13.

Lee, Y., Ahn, C., Han, J.J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415–419.

Leegwater, P.A., Vermeulen, G., Könst, A.A., Naidu, S., Mulders, J., Visser, A., Kersbergen, P., Mobach, D., Fonds, D., van Berkel, C.G., et al. (2001). Subunits of the translation initiation factor eIF2B are mutant in leukoencephalopathy with vanishing white matter. Nat. Genet. *29*, 383–388.

Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., and Zeviani, M. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. Cell *80*, 155–165.

Lerner, M.R., Boyle, J.A., Hardin, J.A., and Steitz, J.A. (1981). Two novel classes of small

ribonucleoproteins detected by antibodies associated with lupus erythematosus. Science *211*, 400–402.

Letunic, I., Doerks, T., and Bork, P. (2009). SMART 6: recent updates and new developments. Nucleic Acids Res. *37*, D229–D232.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, Q., Lee, J.-A., and Black, D.L. (2007). Neuronal regulation of alternative pre-mRNA splicing. Nat. Rev. Neurosci. *8*, 819–831.

Li, Y.R., King, O.D., Shorter, J., and Gitler, A.D. (2013). Stress granules as crucibles of ALS pathogenesis. J. Cell Biol. *201*, 361–372.

Li, Z., and Deutscher, M.P. (1995). The tRNA processing enzyme RNase T is essential for maturation of 5S RNA. Proc. Natl. Acad. Sci. U. S. a. *92*, 6883–6886.

Li, Z., Pandit, S., and Deutscher, M.P. (1998). 3' exoribonucleolytic trimming is a common feature of the maturation of small, stable RNAs in Escherichia coli. Proc. Natl. Acad. Sci. U. S. a. *95*, 2856–2861.

Li, Z., Pandit, S., and Deutscher, M.P. (1999). Maturation of 23S ribosomal RNA requires the exoribonuclease RNase T. Rna *5*, 139–146.

Lianoglou, S., Garg, V., Yang, J.L., Leslie, C.S., and Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev. *27*, 2380–2396.

Lin, H., Yue, L., and Spradling, A.C. (1994). The Drosophila fusome, a germline-specific organelle, contains membrane skeletal proteins and functions in cyst formation. Development *120*, 947–956.

Lin, K.-P., Soong, B.-W., Yang, C.-C., Huang, L.-W., Chang, M.-H., Lee, I.-H., Antonellis, A., Antonellis, A., and Lee, Y.-C. (2011). The mutational spectrum in a cohort of Charcot-Marie-Tooth disease type 2 among the Han Chinese in Taiwan. PLoS ONE *6*, e29393.

Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. Nat. Rev. Mol. Cell Biol. *12*, 505–516.

Lindsley, R.C., and Ebert, B.L. (2013). Molecular pathophysiology of myelodysplastic syndromes. Annu Rev Pathol *8*, 21–47.

Lines, M.A., Huang, L., Schwartzentruber, J., Douglas, S.L., Lynch, D.C., Beaulieu, C., Guion-Almeida, M.L., Zechi-Ceide, R.M., Gener, B., Gillessen-Kaesbach, G., et al. (2012). Haploinsufficiency of a spliceosomal GTPase encoded by EFTUD2 causes mandibulofacial dysostosis with microcephaly. Am. J. Hum. Genet. *90*, 369–377.

Ling, S.-C., Polymenidou, M., and Cleveland, D.W. (2013). Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. Neuron *79*, 416–438.

Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., and Ranum, L.P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science *293*, 864–867.

Liu, J.M. (2006). Ribosomes and marrow failure: coincidental association or molecular paradigm? Blood *107*, 4583–4588.

Liu-Yesucevitz, L., Bassell, G.J., Gitler, A.D., Hart, A.C., Klann, E., Richter, J.D., Warren, S.T., and Wolozin, B. (2011). Local RNA translation at the synapse and in disease. J. Neurosci. *31*, 16086–16093.

Long, E.O., and Dawid, I.B. (1980). Alternative pathways in the processing of ribosomal RNA precursor in Drosophila melanogaster. J. Mol. Biol. *138*, 873–878.

Lovci, M.T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T.Y., Stark, T.J., Gehman, L.T., Hoon, S., et al. (2013). Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. Nat. Struct. Mol. Biol. *20*, 1434–1442.

Lukong, K.E., Chang, K.W., Khandjian, E.W., and Richard, S. (2008). RNA-binding proteins in human genetic disease. Trends Genet. *24*, 416–425.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479–490.

Luteijn, M.J., and Ketting, R.F. (2013). PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat. Rev. Genet. *14*, 523–534.

Ly, T.B.N., Peters, V., Gibson, K.M., Liesert, M., Buckel, W., Wilcken, B., Carpenter, K., Ensenauer, R., Hoffmann, G.F., Mack, M., et al. (2003). Mutations in the AUH gene cause 3-methylglutaconic aciduria type I. Hum. Mutat. *21*, 401–407.

Ma, T., Trinh, M.A., Wexler, A.J., Bourbon, C., Gatti, E., Pierre, P., Cavener, D.R., and Klann, E. (2013). Suppression of eIF2α kinases alleviates Alzheimer's disease-related plasticity and memory deficits. Nature Neuroscience *16*, 1299–1305.

MacDonald, C.C., and McMahon, K.W. (2010). Tissue-specific mechanisms of alternative polyadenylation: testis, brain, and beyond. WIREs RNA *1*, 494–501.

Macias, S., Plass, M., Stajuda, A., Michlewski, G., Eyras, E., and Cáceres, J.F. (2012). DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. Nat. Struct. Mol. Biol. *19*, 760–766.

Mandal, R.K., and Dawid, I.B. (1981). The nucleotide sequence at the transcription termination site of ribosomal RNA in Drosophila melanogaster. Nucleic Acids Res. *9*, 1801–1811.

Mankodi, A., Urbinati, C.R., Yuan, Q.P., Moxley, R.T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M.S., and Thornton, C.A. (2001). Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. Hum. Mol. Genet. *10*, 2165–2170.

Mann, M. (2006). Functional and quantitative proteomics using SILAC. Nat. Rev. Mol. Cell Biol. *7*, 952–958.

Maraia, R.J., and Lamichhane, T.N. (2011). 3 ' processing of eukaryotic precursor tRNAs. WIREs RNA *2*, 362–375.

Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M.K., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., et al. (2013). CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res. *41*, D348–D352.

Marrone, A., Walne, A., Tamary, H., Masunari, Y., Kirwan, M., Beswick, R., Vulliamy, T., and Dokal, I. (2007). Telomerase reverse-transcriptase homozygous mutations in autosomal recessive dyskeratosis congenita and Hoyeraal-Hreidarsson syndrome. Blood *110*, 4198–4205.

Martin, C.L., Duvall, J.A., Ilkin, Y., Simon, J.S., Arreaza, M.G., Wilkes, K., Alvarez-Retuerto, A., Whichello, A., Powell, C.M., Rao, K., et al. (2007). Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism. Am. J. Med. Genet. B Neuropsychiatr. Genet. *144B*, 869–876.

Martin, I., Kim, J.W., Lee, B.D., Kang, H.C., Xu, J.-C., Jia, H., Stankowski, J., Kim, M.-S., Zhong, J., Kumar, M., et al. (2014). Ribosomal Protein s15 Phosphorylation Mediates LRRK2 Neurodegeneration in Parkinson's Disease. Cell *157*, 472–485.

Martin-Lluesma, S., Stucke, V.M., and Nigg, E.A. (2002). Role of Hec1 in spindle checkpoint signaling and kinetochore recruitment of Mad1/Mad2. Science *297*, 2267–2270.

Masliah, G., Barraud, P., and Allain, F.H.T. (2013). RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. Cell. Mol. Life Sci. *70*, 1875–1895.

Matera, A.G., Terns, R.M., and Terns, M.P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat. Rev. Mol. Cell Biol. *8*, 209–220.

Matsson, H., Klar, J., Draptchinskaia, N., Gustavsson, P., Carlsson, B., Bowers, D., de Bont, E., and Dahl, N. (1999). Truncating ribosomal protein S19 mutations and variable clinical expression in Diamond-Blackfan anemia. Hum. Genet. *105*, 496–500.

Matthews, D.E., Hessler, R.A., Denslow, N.D., Edwards, J.S., and O'Brien, T.W. (1982). Protein composition of the bovine mitochondrial ribosome. J. Biol. Chem. *257*, 8788–8794.

Maurer-Stroh, S., Dickens, N.J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F., and Ponting, C.P. (2003). The Tudor domain "Royal Family": Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends Biochem. Sci. *28*, 69–74.

Mauro, V.P., and Edelman, G.M. (2002). The ribosome filter hypothesis. Proc. Natl. Acad. Sci. U. S. a. *99*, 12031–12036.

May, W.A., Gishizky, M.L., Lessnick, S.L., Lunsford, L.B., Lewis, B.C., Delattre, O., Zucman, J., Thomas, G., and Denny, C.T. (1993). Ewing sarcoma 11;22 translocation produces a chimeric transcription factor that requires the DNA-binding domain encoded by FLI1 for transformation. Proc. Natl. Acad. Sci. U. S. a. *90*, 5752–5756.

McCann, K.L., and Baserga, S.J. (2013). Genetics. Mysterious ribosomopathies. Science *341*, 849–850.

McHugh, C.A., Russell, P., and Guttman, M. (2014). Methods for comprehensive experimental identification of RNA-protein interactions. Genome Biol. *15*, 203.

McKee, A.E., Minet, E., Stern, C., Riahi, S., Stiles, C.D., and Silver, P.A. (2005). A genome-wide in situ hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. BMC Dev. Biol. *5*, 14.

McKie, A.B., McHale, J.C., Keen, T.J., Tarttelin, E.E., Goliath, R., van Lith-Verhoeven, J.J., Greenberg, J., Ramesar, R.S., Hoyng, C.B., Cremers, F.P., et al. (2001). Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). Hum. Mol. Genet. *10*, 1555–1562.

McLaughlin, H.M., Sakaguchi, R., Liu, C., Igarashi, T., Pehlivan, D., Chu, K., Iyer, R., Cruz, P., Cherukuri, P.F., Hansen, N.F., et al. (2010). Compound heterozygosity for loss-of-function lysyl-tRNA synthetase mutations in a patient with peripheral neuropathy. Am. J. Hum. Genet. *87*, 560–566.

McLean, W.H., Pulkkinen, L., Smith, F.J., Rugg, E.L., Lane, E.B., Bullrich, F., Burgeson, R.E., Amano, S., Hudson, D.L., Owaribe, K., et al. (1996). Loss of plectin causes epidermolysis bullosa with muscular dystrophy: cDNA cloning and genomic organization. Genes Dev. *10*, 1724–1735.

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. Nat. Rev. Genet. *14*, 447–459.

Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., and Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. Nat. Struct. Mol. Biol. *19*, 560–567.

Melo, S.A., Moutinho, C., Ropero, S., Calin, G.A., Rossi, S., Spizzo, R., Fernandez, A.F., Davalos, V., Villanueva, A., Montoya, G., et al. (2010). A Genetic Defect in Exportin-5 Traps Precursor MicroRNAs in the Nucleus of Cancer Cells. Cancer Cell *18*, 303–315.

Melo, S.A., Ropero, S., Moutinho, C., Aaltonen, L.A., Yamamoto, H., Calin, G.A., Rossi, S., Fernandez, A.F., Carneiro, F., Oliveira, C., et al. (2009). A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. Nat. Genet. *41*, 365–370.

Merritt, W.M., Lin, Y.G., Han, L.Y., Kamat, A.A., Spannuth, W.A., Schmandt, R., Urbauer, D., Pennacchio, L.A., Cheng, J.-F., Nick, A.M., et al. (2008). Dicer, Drosha, and outcomes in patients with ovarian cancer. N. Engl. J. Med. *359*, 2641–2650.

Meyer, K.D., and Jaffrey, S.R. (2014). The dynamic epitranscriptome: N(6)-methyladenosine and gene expression control. Nat. Rev. Mol. Cell Biol. *15*, 313–326.

Mientjes, E.J., Willemsen, R., Kirkpatrick, L.L., Nieuwenhuizen, I.M., Hoogeveen-Westerveld, M., Verweij, M., Reis, S., Bardoni, B., Hoogeveen, A.T., Oostra, B.A., et al. (2004). Fxr1 knockout mice show a striated muscle phenotype: implications for Fxr1p function in vivo. Hum. Mol. Genet. *13*, 1291–1302.

Mihailovich, M., Militti, C., Gabaldón, T., and Gebauer, F. (2010). Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. Bioessays *32*, 109–118.

Miller, C., Saada, A., Shaul, N., Shabtai, N., Ben-Shalom, E., Shaag, A., Hershkovitz, E., and Elpeleg, O. (2004). Defective mitochondrial translation caused by a ribosomal protein (MRPS16) mutation. Ann. Neurol. *56*, 734–738.

Miller, J.W. (2000). Recruitment of human muscleblind proteins to (CUG)n expansions associated with myotonic dystrophy. Embo J. *19*, 4439–4448.

Miller, J.A., Ding, S.-L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., et al. (2014). Transcriptional landscape of the prenatal human brain. Nature *508*, 199–206.

Miller, O.L., and Beatty, B.R. (1969). Visualization of nucleolar genes. Science *164*, 955–957.

Mitchell, S.F., and Parker, R. (2014). Principles and Properties of Eukaryotic mRNPs. Mol. Cell *54*, 547–558.

Mitchell, S.F., Jain, S., She, M., and Parker, R. (2013). Global analysis of yeast mRNPs. Nat. Struct. Mol. Biol. *20*, 127–133.

Mittal, N., Roy, N., Babu, M.M., and Janga, S.C. (2009). Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks. Proc. Natl. Acad. Sci. U. S. a. *106*, 20300–20305.

Miyamura, Y., Suzuki, T., Kono, M., Inagaki, K., Ito, S., Suzuki, N., and Tomita, Y. (2003). Mutations of the RNA-specific adenosine deaminase gene (DSRAD) are involved in dyschromatosis symmetrica hereditaria. Am. J. Hum. Genet. *73*, 693–699.

Miyoshi, K., Ogino, A., Siomi, M.C., and Siomi, H. (2013). Purification of dFMR1-containing complexes using tandem affinity purification. Methods Mol. Biol. *1010*, 111–121.

Mody, M., Cao, Y., Cui, Z., Tay, K.Y., Shyong, A., Shimizu, E., Pham, K., Schultz, P., Welsh, D., and Tsien, J.Z. (2001). Genome-wide gene expression profiles of the developing mouse hippocampus. Proc. Natl. Acad. Sci. U. S. a. *98*, 8862–8867.

Monemi, S., Spaeth, G., DaSilva, A., Popinchalk, S., Ilitchev, E., Liebmann, J., Ritch, R., Héon, E., Crick, R.P., Child, A., et al. (2005). Identification of a novel adult-onset primary open-angle glaucoma (POAG) gene on 5q22.1. Hum. Mol. Genet. *14*, 725–733.

Montanaro, L., Treré, D., and Derenzini, M. (2008). Nucleolus, ribosomes, and cancer. Am. J. Pathol. *173*, 301–310.

Montell, D.J. (2003). Border-cell migration: the race is on. Nat. Rev. Mol. Cell Biol. *4*, 13–24.

Monti, L., Rodolfo, M., Rudolfo, M., Russo, Lo, G., Noonan, D., Acquati, F., and Taramelli, R. (2008). RNASET2 as a tumor antagonizing gene in a melanoma cancer model. Oncol. Res. *17*, 69–74.

Mootha, V.K., Lepage, P., Miller, K., Bunkenborg, J., Reich, M., Hjerrild, M., Delmonte, T., Villeneuve, A., Sladek, R., Xu, F., et al. (2003). Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. Proc. Natl. Acad. Sci. U. S. a. *100*,

605–610.

Mougey, E.B., O'Reilly, M., Osheim, Y., Miller, O.L., Beyer, A., and Sollner-Webb, B. (1993). The terminal balls characteristic of eukaryotic rRNA transcription units in chromatin spreads are rRNA processing complexes. Genes Dev. *7*, 1609–1619.

Mroczek, S., and Dziembowski, A. (2013). U6 RNA biogenesis and disease association. WIREs RNA *4*, 581–592.

Mukherjee, N., Corcoran, D.L., Nusbaum, J.D., Reid, D.W., Georgiev, S., Hafner, M., Ascano, M., Tuschl, T., Ohler, U., and Keene, J.D. (2011). Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. Mol. Cell *43*, 327–339.

Mukherjee, N., Jacobs, N.C., Hafner, M., Kennington, E.A., Nusbaum, J.D., Tuschl, T., Blackshear, P.J., and Ohler, U. (2014). Global target mRNA specification and regulation by the RNA-binding protein ZFP36. Genome Biol. *15*, R12.

Mullineux, S.-T., and Lafontaine, D.L.J. (2012). Mapping the cleavage sites on mammalian pre-rRNAs: where do we stand? Biochimie *94*, 1521–1532.

Muñoz, L.E., Lauber, K., Schiller, M., Manfredi, A.A., and Herrmann, M. (2010). The role of defective clearance of apoptotic cells in systemic autoimmunity. Nat. Rev. Rheumatol. *6*, 280–289.

Murray, A., Webb, J., Grimley, S., Conway, G., and Jacobs, P. (1998). Studies of FRAXA and FRAXE in women with premature ovarian failure. J. Med. Genet. *35*, 637–640.

Murzin, A.G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. Embo J. *12*, 861–867.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. *247*, 536–540.

Muscatelli, F., Strom, T.M., Walker, A.P., Zanaria, E., Récan, D., Meindl, A., Bardoni, B., Guioli, S., Zehetner, G., and Rabl, W. (1994). Mutations in the DAX-1 gene give rise to both X-linked adrenal hypoplasia congenita and hypogonadotropic hypogonadism. Nature *372*, 672–676.

Müller-McNicoll, M., and Neugebauer, K.M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. Nat. Rev. Genet. *14*, 275–287.

Nakashima, E., Mabuchi, A., Makita, Y., Masuno, M., Ohashi, H., Nishimura, G., and Ikegawa, S. (2004). Novel SBDS mutations caused by gene conversion in Japanese patients with Shwachman-Diamond syndrome. Hum. Genet. *114*, 345–348.

Narla, A., and Ebert, B.L. (2010). Ribosomopathies: human disorders of ribosome dysfunction. Blood *115*, 3196–3205.

Nasim, M.T., Ogo, T., Ahmed, M., Randall, R., Chowdhury, H.M., Snape, K.M., Bradshaw,

T.Y., Southgate, L., Lee, G.J., Jackson, I., et al. (2011). Molecular genetic characterization of SMAD signaling molecules in pulmonary arterial hypertension. Hum. Mutat. *32*, 1385–1389.

Nawrot, B., Sochacka, E., and Düchler, M. (2011). tRNA structural and functional changes induced by oxidative stress. Cell. Mol. Life Sci. *68*, 4023–4032.

Neilson, D.E., Adams, M.D., Orr, C.M.D., Schelling, D.K., Eiben, R.M., Kerr, D.S., Anderson, J., Bassuk, A.G., Bye, A.M., Childs, A.-M., et al. (2009). Infection-triggered familial or recurrent cases of acute necrotizing encephalopathy caused by mutations in a component of the nuclear pore, RANBP2. Am. J. Hum. Genet. *84*, 44–51.

Nelson, D.L., Orr, H.T., and Warren, S.T. (2013). The Unstable Repeats—Three Evolving Faces of Neurological Disease. Neuron *77*, 825–843.

Nerurkar, P., Altvater, M., Gerhardy, S., Schütz, S., Fischer, U., Weirich, C., and Panse, V.G. (2015). Eukaryotic Ribosome Assembly and Nuclear Export. Int. Rev. Cell Mol. Biol. *319*, 107–140.

Neumuller, R.A., Gross, T., Samsonova, A.A., Vinayagam, A., Buckner, M., Founk, K., Hu, Y., Sharifpoor, S., Rosebrock, A.P., Andrews, B., et al. (2013). Conserved Regulators of Nucleolar Size Revealed by Global Phenotypic Analyses. Sci. Signaling *6*, ra70–ra70.

Nguyen, L.H., Erzberger, J.P., Root, J., and Wilson, D.M. (2000). The human homolog of Escherichia coli Orn degrades small single-stranded RNA and DNA oligomers. J. Biol. Chem. *275*, 25900–25906.

Norbury, C.J. (2013). Cytoplasmic RNA: a case of the tail wagging the dog. Nat. Rev. Cancer *13*, 643–653.

Nousbeck, J., Spiegel, R., Ishida-Yamamoto, A., Indelman, M., Shani-Adir, A., Adir, N., Lipkin, E., Bercovici, S., Geiger, D., van Steensel, M.A., et al. (2008). Alopecia, neurological defects, and endocrinopathy syndrome caused by decreased expression of RBM28, a nucleolar protein associated with ribosome biogenesis. Am. J. Hum. Genet. *82*, 1114–1121.

Nousiainen, H.O., Kestila, M., Pakkasjarvi, N., Honkala, H., Kuure, S., Tallila, J., Vuopala, K., Ignatius, J., Herva, R., and Peltonen, L. (2008). Mutations in mRNA export mediator GLE1 result in a fetal motoneuron disease. Nat. Genet. *40*, 155–157.

O'Brien, T.W. (2003). Properties of human mitochondrial ribosomes. IUBMB Life *55*, 505–513.

Oeffinger, M., Wei, K.E., Rogers, R., DeGrasse, J.A., Chait, B.T., Aitchison, J.D., and Rout, M.P. (2007). Comprehensive analysis of diverse ribonucleoprotein complexes. Nat. Methods *4*, 951–956.

Oeffinger, M., Zenklusen, D., Ferguson, A., Wei, K.E., Hage, El, A., Tollervey, D., Chait, B.T., Singer, R.H., and Rout, M.P. (2009). Rrp17p is a eukaryotic exonuclease required for 5' end processing of Pre-60S ribosomal RNA. Mol. Cell *36*, 768–781.

Oktem, O., and Urman, B. (2010). Understanding follicle growth in vivo. Hum. Reprod. *25*, 2944–2954.

Omran, H., Kobayashi, D., Olbrich, H., Tsukahara, T., Loges, N.T., Hagiwara, H., Zhang, Q., Leblond, G., O'Toole, E., Hara, C., et al. (2008). Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins. Nature *456*, 611–616.

Orloff, M., Peterson, C., He, X., Ganapathi, S., Heald, B., Yang, Y.-R., Bebek, G., Romigh, T., Song, J.H., Wu, W., et al. (2011). Germline mutations in MSR1, ASCC1, and CTHRC1 in patients with Barrett esophagus and esophageal adenocarcinoma. Jama *306*, 410–419.

Orr, H.T., Chung, M.Y., Banfi, S., Kwiatkowski, T.J., Servadio, A., Beaudet, A.L., McCall, A.E., Duvick, L.A., Ranum, L.P., and Zoghbi, H.Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat. Genet. *4*, 221–226.

Ozanick, S.G., Wang, X., Costanzo, M., Brost, R.L., Boone, C., and Anderson, J.T. (2009). Rex1p deficiency leads to accumulation of precursor initiator tRNAMet and polyadenylation of substrate RNAs in Saccharomyces cerevisiae. Nucleic Acids Res. *37*, 298–308.

Ozsolak, F., and Milos, P.M. (2011). RNA sequencing: advances, challenges and opportunities. Nat. Rev. Genet. *12*, 87–98.

Panagopoulos, I., Aman, P., Fioretos, T., Höglund, M., Johansson, B., Mandahl, N., Heim, S., Behrendtz, M., and Mitelman, F. (1994). Fusion of the FUS gene with ERG in acute myeloid leukemia with t(16;21)(p11;q22). Genes Chromosom. Cancer *11*, 256–262.

Panagopoulos, I., Mencinger, M., Dietrich, C.U., Bjerkehagen, B., Saeter, G., Mertens, F., Mandahl, N., and Heim, S. (1999). Fusion of the RBP56 and CHN genes in extraskeletal myxoid chondrosarcomas with translocation t(9;17)(q22;q11). Oncogene *18*, 7594–7598.

Panse, V.G., and Johnson, A.W. (2010). Maturation of eukaryotic ribosomes: acquisition of functionality. Trends Biochem. Sci. *35*, 260–266.

Parenteau, J., Durand, M., Morin, G., Gagnon, J., Lucier, J.-F., Wellinger, R.J., Chabot, B., and Elela, S.A. (2011). Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. Cell *147*, 320–331.

Parker, R., and Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. Mol. Cell *25*, 635–646.

Parsons, D.W., McAndrew, P.E., Monani, U.R., Mendell, J.R., Burghes, A.H., and Prior, T.W. (1996). An 11 base pair duplication in exon 6 of the SMN gene produces a type I spinal muscular atrophy (SMA) phenotype: further evidence for SMN as the primary SMA-determining gene. Hum. Mol. Genet. *5*, 1727–1732.

Paushkin, S., Gubitz, A.K., Massenet, S., and Dreyfuss, G. (2002). The SMN complex, an assemblyosome of ribonucleoproteins. Curr. Opin. Cell Biol. *14*, 305–312.

Pavlakis, G.N., Jordan, B.R., Wurst, R.M., and Vournakis, J.N. (1979). Sequence and secondary structure of Drosophila melanogaster 5.8S and 2S rRNAs and of the processing site between them. Nucleic Acids Res. *7*, 2213–2238.

Pearson, T., Curtis, F., Al-Eyadhy, A., Al-Tamemi, S., Mazer, B., Dror, Y., Abish, S., Bale, S.,

Compton, J., Ray, R., et al. (2008). An intronic mutation in DKC1 in an infant with Høyeraal-Hreidarsson syndrome. Am. J. Med. Genet. *146A*, 2159–2161.

Peculis, B.A. (1997). The sequence of the 5' end of the U8 small nucleolar RNA is critical for 5.8S and 28S rRNA maturation. Mol. Cell. Biol. *17*, 3702–3713.

Peculis, B.A., and Steitz, J.A. (1993). Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte. Cell *73*, 1233–1245.

Perron, M.P., and Provost, P. (2009). Protein components of the microRNA pathway and human diseases. Methods Mol. Biol. *487*, 369–385.

Peterlin, B.M., Brogie, J.E., and Price, D.H. (2011). 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. WIREs RNA *3*, 92–103.

Pettersson, I., Hinterberger, M., Mimori, T., Gottlieb, E., and Steitz, J.A. (1984). The structure of mammalian small nuclear ribonucleoproteins. Identification of multiple protein components reactive with anti-(U1)ribonucleoprotein and anti-Sm autoantibodies. J. Biol. Chem. *259*, 5907–5914.

Pérez-Arellano, I., Gallego, J., and Cervera, J. (2007). The PUA domain − a structural and functional overview. Febs J. *274*, 4972–4984.

Pérez-González, A., Pazo, A., Navajas, R., Ciordia, S., Rodriguez-Frandsen, A., and Nieto, A. (2014). hCLE/C14orf166 associates with DDX1-HSPC117-FAM98B in a novel transcription-dependent shuttling RNA-transporting complex. PLoS ONE *9*, e90957.

Phipps, K.R., Charette, J.M., and Baserga, S.J. (2011). The small subunit processome in ribosome biogenesis—progress and prospects. WIREs RNA *2*, 1–21.

Phizicky, E.M., and Hopper, A.K. (2010). tRNA biology charges to the front. Genes Dev. *24*, 1832–1860.

Pierce, S.B., Chisholm, K.M., Lynch, E.D., Lee, M.K., Walsh, T., Opitz, J.M., Li, W., Klevit, R.E., and King, M.-C. (2011). Mutations in mitochondrial histidyl tRNA synthetase HARS2 cause ovarian dysgenesis and sensorineural hearing loss of Perrault syndrome. Proc. Natl. Acad. Sci. U. S. a. *108*, 6543–6548.

Pinol-Roma, S., Choi, Y.D., Matunis, M.J., and Dreyfuss, G. (1988). Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. Genes Dev. *2*, 215–227.

Piper, P.W., and Stråby, K.B. (1989). Processing of transcripts of a dimeric tRNA gene in yeast uses the nuclease responsible for maturation of the 3' termini upon 5 S and 37 S precursor rRNAs. FEBS Lett. *250*, 311–316.

Piper, P.W., Bellatin, J.A., and Lockheart, A. (1983). Altered maturation of sequences at the 3' terminus of 5S gene transcripts in a Saccharomyces cerevisiae mutant that lacks a RNA processing endonuclease. Embo J. *2*, 353–359.

Piper, P.W., Lockheart, A., and Bellatin, J. (1987). The effects of protein synthesis inhibition, and of mutations rna1.1 and rna82.1, on the synthesis of small RNAs in yeast. FEBS Lett. *214*, 143–148.

Piper, P.W., Patel, N., and Lockheart, A. (1984). Processing of the 3' sequence extensions upon the 5S rRNA of a mutant yeast in Xenopus laevis germinal vesicle extract. Eur. J. Biochem. *141*, 115–118.

Plotkin, J.B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. *12*, 32–42.

Popow, J., Englert, M., Weitzer, S., Schleiffer, A., Mierzwa, B., Mechtler, K., Trowitzsch, S., Will, C.L., LUhrmann, R., Söll, D., et al. (2011). HSPC117 is the essential subunit of a human tRNA splicing ligase complex. Science *331*, 760–764.

Prokopowich, C.D., Gregory, T.R., and Crease, T.J. (2003). The correlation between rDNA copy number and genome size in eukaryotes. Genome *46*, 48–50.

Puffenberger, E.G., Jinks, R.N., Sougnez, C., Cibulskis, K., Willert, R.A., Achilly, N.P., Cassidy, R.P., Fiorentini, C.J., Heiken, K.F., Lawrence, J.J., et al. (2012). Genetic mapping and exome sequencing identify variants associated with five novel diseases. PLoS ONE *7*, e28936.

Pulkkinen, L., Smith, F.J., Shimizu, H., Murata, S., Yaoita, H., Hachisuka, H., Nishikawa, T., McLean, W.H., and Uitto, J. (1996). Homozygous deletion mutations in the plectin gene (PLEC1) in patients with epidermolysis bullosa simplex associated with late-onset muscular dystrophy. Hum. Mol. Genet. *5*, 1539–1546.

Pulst, S.M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X.N., Lopes-Cendes, I., Pearlman, S., Starkman, S., Orozco-Diaz, G., Lunkes, A., et al. (1996). Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nat. Genet. *14*, 269–276.

Rabe, B. (2013). Aicardi-Goutières syndrome: clues from the RNase H2 knock-out mouse. J. Mol. Med. *91*, 1235–1240.

Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., Jantsch, M.F., Konrat, R., Bläsi, U., and Schroeder, R. (2007). RNA chaperones, RNA annealers and RNA helicases. RNA Biol. *4*, 118–130.

Ramagopal, S. (1992). Are eukaryotic ribosomes heterogeneous? Affirmations on the horizon. Biochem. Cell Biol. *70*, 269–272.

Ramaswami, M., Taylor, J.P., and Parker, R. (2013). Altered Ribostasis: RNA-Protein Granules in Degenerative Disorders. Cell *154*, 727–736.

Ramser, J., Winnepenninckx, B., Lenski, C., Errijgers, V., Platzer, M., Schwartz, C.E., Meindl, A., and Kooy, R.F. (2004). A splice site mutation in the methyltransferase gene FTSJ1 in Xp11.23 is associated with non-syndromic mental retardation in a large Belgian family (MRX9). J. Med. Genet. *41*, 679–683.

Ramsköld, D., Wang, E.T., Burge, C.B., and Sandberg, R. (2009). An abundance of ubiquitously

expressed genes revealed by tissue transcriptome sequence data. PLoS Comput. Biol. *5*, e1000598.

Rankin, J., Brown, R., Dobyns, W.B., Harington, J., Patel, J., Quinn, M., and Brown, G. (2010). Pontocerebellar hypoplasia type 6: A British case with PEHO-like features. Am. J. Med. Genet. *152A*, 2079–2084.

Ravenscroft, J.C., Suri, M., Rice, G.I., Szynkiewicz, M., and Crow, Y.J. (2011). Autosomal dominant inheritance of a heterozygous mutation in SAMHD1 causing familial chilblain lupus. Am. J. Med. Genet. *155A*, 235–237.

Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature *499*, 172–177.

Regalado, E.S., Guo, D.-C., Villamizar, C., Avidan, N., Gilchrist, D., McGillivray, B., Clarke, L., Bernier, F., Santos-Cortez, R.L., Leal, S.M., et al. (2011). Exome sequencing identifies SMAD3 mutations as a cause of familial thoracic aortic aneurysm and dissection with intracranial and other arterial aneurysms. Circul. Res. *109*, 680–686.

Reijo, R., Lee, T.Y., Salo, P., Alagappan, R., Brown, L.G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., and Hovatta, O. (1995). Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat. Genet. *10*, 383–393.

Reyes-Turcu, F.E., and Grewal, S.I. (2012). Different means, same end—heterochromatin formation by RNAi and RNAi-independent RNA processing factors in fission yeast. Curr. Opin. Genet. Dev. *22*, 156–163.

Reynolds, N., and Cooke, H.J. (2005). Role of the DAZ genes in male fertility. Reprod. Biomed. Online *10*, 72–80.

Riccardo, S., Tortoriello, G., Giordano, E., Turano, M., and Furia, M. (2007). The coding/non-coding overlapping architecture of the gene encoding the Drosophila pseudouridine synthase. BMC Mol. Biol. *8*, 15.

Rice, G.I., Bond, J., Asipu, A., Brunette, R.L., Manfield, I.W., Carr, I.M., Fuller, J.C., Jackson, R.M., Lamb, T., Briggs, T.A., et al. (2009). Mutations involved in Aicardi-Goutières syndrome implicate SAMHD1 as regulator of the innate immune response. Nat. Genet. *41*, 829–832.

Rice, G.I., Kasher, P.R., Forte, G.M.A., Mannion, N.M., Greenwood, S.M., Szynkiewicz, M., Dickerson, J.E., Bhaskar, S.S., Zampini, M., Briggs, T.A., et al. (2012). Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. Nat. Genet. *44*, 1243–1248.

Rice, G.I., Reijns, M.A.M., Coffin, S.R., Forte, G.M.A., Anderson, B.H., Szynkiewicz, M., Gornall, H., Gent, D., Leitch, A., Botella, M.P., et al. (2013). Synonymous mutations in RNASEH2A create cryptic splice sites impairing RNase H2 enzyme function in Aicardi-Goutières syndrome. Hum. Mutat. *34*, 1066–1070.

Riley, L.G., Cooper, S., Hickey, P., Rudinger-Thirion, J., McKenzie, M., Compton, A., Lim, S.C., Thorburn, D., Ryan, M.T., Giegé, R., et al. (2010). Mutation of the mitochondrial tyrosyl-tRNA synthetase gene, YARS2, causes myopathy, lactic acidosis, and sideroblastic anemia--MLASA syndrome. Am. J. Hum. Genet. *87*, 52–59.

Rio Frio, T., Bahubeshi, A., Kanellopoulou, C., Hamel, N., Niedziela, M., Sabbaghian, N., Pouchet, C., Gilbert, L., O'Brien, P.K., Serfas, K., et al. (2011). DICER1 mutations in familial multinodular goiter with and without ovarian Sertoli-Leydig cell tumors. Jama *305*, 68–77.

Roberts, R., Timchenko, N.A., Miller, J.W., Reddy, S., Caskey, C.T., Swanson, M.S., and Timchenko, L.T. (1997). Altered phosphorylation and intracellular distribution of a (CUG)n triplet repeat RNA-binding protein in patients with myotonic dystrophy and in myotonin protein kinase knockout mice. Proc. Natl. Acad. Sci. U. S. a. *94*, 13221–13226.

Rocak, S., and Linder, P. (2004). DEAD-box proteins: the driving forces behind RNA metabolism. Nat. Rev. Mol. Cell Biol. *5*, 232–241.

Rosby, R., Cui, Z., Rogers, E., deLivron, M.A., Robinson, V.L., and DiMario, P.J. (2009). Knockdown of the Drosophila GTPase nucleostemin 1 impairs large ribosomal subunit biogenesis, cell growth, and midgut precursor cell maintenance. Mol Biol Cell *20*, 4424–4434.

Rosenberg, H.F. (2011). Vertebrate Secretory (RNase A) Ribonucleases and Host Defense. Ribonucleases 35–53.

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature *505*, 701–705.

Ruggero, D., and Pandolfi, P.P. (2003). Does the ribosome translate cancer? Nat. Rev. Cancer *3*, 179–192.

Ruggero, D., Grisendi, S., Piazza, F., Rego, E., Mari, F., Rao, P.H., Cordon-Cardo, C., and Pandolfi, P.P. (2003). Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification. Science *299*, 259–262.

Saada, A., Shaag, A., Arnon, S., Dolfin, T., Miller, C., Fuchs-Telem, D., Lombes, A., and Elpeleg, O. (2007). Antenatal mitochondrial disease caused by mitochondrial ribosomal protein (MRPS22) mutation. J. Med. Genet. *44*, 784–786.

Sakai, K., Gofuku, M., Kitagawa, Y., Ogasawara, T., Hirose, G., Yamazaki, M., Koh, C.S., Yanagisawa, N., and Steinman, L. (1994). A hippocampal protein associated with paraneoplastic neurologic syndrome and small cell lung carcinoma. Biochem. Biophys. Res. Commun. *199*, 1200–1208.

Santos-Cortez, R.L.P., Lee, K., Azeem, Z., Antonellis, P.J., Pollock, L.M., Khan, S., Irfanullah, Andrade-Elizondo, P.B., Chiu, I., Adams, M.D., et al. (2013). Mutations in KARS, encoding lysyl-tRNA synthetase, cause autosomal-recessive nonsyndromic hearing impairment DFNB89. Am. J. Hum. Genet. *93*, 132–140.

Scheper, G.C., van der Klok, T., van Andel, R.J., van Berkel, C.G.M., Sissler, M., Smet, J.,

Muravina, T.I., Serkov, S.V., Uziel, G., Bugiani, M., et al. (2007a). Mitochondrial aspartyl-tRNA synthetase deficiency causes leukoencephalopathy with brain stem and spinal cord involvement and lactate elevation. Nat. Genet. *39*, 534–539.

Scheper, G.C., van der Knaap, M.S., and Proud, C.G. (2007b). Translation matters: protein synthesis defects in inherited disease. Nat. Rev. Genet. *8*, 711–723.

Scherrer, T., Mittal, N., Janga, S.C., and Gerber, A.P. (2010). A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. PLoS ONE *5*, e15499.

Schneider, M.D., Bains, A.K., Rajendra, T.K., Dominski, Z., Matera, A.G., and Simmonds, A.J. (2010). Functional characterization of the Drosophila MRP (mitochondrial RNA processing) RNA gene. Rna *16*, 2120–2130.

Schou, K.B., Andersen, J.S., and Pedersen, L.B. (2014). A divergent calponin homology (NN-CH) domain defines a novel family: implications for evolution of ciliary IFT complex B proteins. Bioinformatics *30*, 899–902.

Schutte, M., Hruban, R.H., Hedrick, L., Cho, K.R., Nadasdy, G.M., Weinstein, C.L., Bova, G.S., Isaacs, W.B., Cairns, P., Nawroz, H., et al. (1996). DPC4 gene in various tumor types. Cancer Res. *56*, 2527–2530.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Seibler, P., Djarmati, A., Langpap, B., Hagenah, J., Schmidt, A., Brüggemann, N., Siebner, H., Jabusch, H.-C., Altenmüller, E., Münchau, A., et al. (2008). A heterozygous frameshift mutation in PRKRA (DYT16) associated with generalised dystonia in a German patient. Lancet Neurol. *7*, 380–381.

Selcen, D., Juel, V.C., Hobson-Webb, L.D., Smith, E.C., Stickler, D.E., Bite, A.V., Ohno, K., and Engel, A.G. (2011). Myasthenic syndrome caused by plectinopathy. Neurology *76*, 327–336.

Selenko, P., Sprangers, R., Stier, G., Bühler, D., Fischer, U., and Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nat. Struct. Mol. Biol. *8*, 27–31.

Senderek, J., Garvey, S.M., Krieger, M., Guergueltcheva, V., Urtizberea, A., Roos, A., Elbracht, M., Stendel, C., Tournev, I., Mihailova, V., et al. (2009). Autosomal-dominant distal myopathy associated with a recurrent missense mutation in the gene encoding the nuclear matrix protein, matrin 3. Am. J. Hum. Genet. *84*, 511–518.

Servadio, A., Koshy, B., Armstrong, D., Antalffy, B., Orr, H.T., and Zoghbi, H.Y. (1995). Expression analysis of the ataxin-1 protein in tissues from normal and spinocerebellar ataxia type 1 individuals. Nat. Genet. *10*, 94–98.

Seydoux, G., and Braun, R.E. (2006). Pathway to totipotency: Lessons from germ cells. Cell *127*, 891–904.

Shamseldin, H.E., Alshammari, M., Al-Sheddi, T., Salih, M.A., Alkhalidi, H., Kentab, A.,

Repetto, G.M., Hashem, M., and Alkuraya, F.S. (2012). Genomic analysis of mitochondrial diseases in a consanguineous population reveals novel candidate disease genes. J. Med. Genet. *49*, 234–241.

Sharma, S., and Lafontaine, D.L.J. (2015). "View From A Bridge": A New Perspective on Eukaryotic rRNA Base Modification. Trends Biochem. Sci. *40*, 560–575.

Shastry, S., Delgado, M.R., Dirik, E., Turkmen, M., Agarwal, A.K., and Garg, A. (2010). Congenital generalized lipodystrophy, type 4 (CGL4) associated with myopathy due to novel PTRF mutations. Am. J. Med. Genet. *152A*, 2245–2253.

Shen, E.C., Henry, M.F., Weiss, V.H., Valentini, S.R., Silver, P.A., and Lee, M.S. (1998). Arginine methylation facilitates the nuclear export of hnRNP proteins. Genes Dev. *12*, 679–691.

Sheng, X.R., Posenau, T., Gumulak-Smith, J.J., Matunis, E., Van Doren, M., and Wawersik, M. (2009). Jak-STAT regulation of male germline stem cell establishment during Drosophila embryogenesis. Dev. Biol. *334*, 335–344.

Shimazaki, H., Takiyama, Y., Ishiura, H., Sakai, C., Matsushima, Y., Hatakeyama, H., Honda, J., Sakoe, K., Naoi, T., Namekawa, M., et al. (2012). A homozygous mutation of C12orf65 causes spastic paraplegia with optic atrophy and neuropathy (SPG55). J. Med. Genet. *49*, 777–784.

Sibbritt, T., Patel, H.R., and Preiss, T. (2013). Mapping and significance of the mRNA methylome. WIREs RNA *4*, 397–422.

Silvera, D., Formenti, S.C., and Schneider, R.J. (2010). Translational control in cancer. Nat. Rev. Cancer *10*, 254–266.

Sim, S., and Wolin, S.L. (2011). Emerging roles for the Ro 60-kDa autoantigen in noncoding RNA metabolism. WIREs RNA *2*, 686–699.

Simard, J., Tonin, P., Durocher, F., Morgan, K., Rommens, J., Gingras, S., Samson, C., Leblanc, J.F., Bélanger, C., and Dion, F. (1994). Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families. Nat. Genet. *8*, 392–398.

Simone, L.E., and Keene, J.D. (2013). Mechanisms coordinating ELAV/Hu mRNA regulons. Curr. Opin. Genet. Dev. *23*, 35–43.

Simos, G., and Hurt, E. (1999). Transfer RNA biogenesis: A visa to leave the nucleus. Curr. Biol. *9*, R238–R241.

Singh, R.K., and Cooper, T.A. (2012). Pre-mRNA splicing in disease and therapeutics. Trends Mol. Med. *18*, 472–482.

Singh, R., and Valcarcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. Nat. Struct. Mol. Biol. *12*, 645–653.

Siomi, M.C., Mannen, T., and Siomi, H. (2010). How does the Royal Family of Tudor rule the PIWI-interacting RNA pathway? Genes Dev. *24*, 636–646.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat. Rev. Mol. Cell Biol. *12*, 246–258.

Siprashvili, Z., Webster, D.E., Kretz, M., Johnston, D., Rinn, J.L., Chang, H.Y., and Khavari, P.A. (2012). Identification of proteins binding coding and non-coding human RNAs using protein microarrays. BMC Genomics *13*, 633.

Skorupa, A., King, M.A., Aparicio, I.M., Dussmann, H., Coughlan, K., Breen, B., Kieran, D., Concannon, C.G., Marin, P., and Prehn, J.H.M. (2012). Motoneurons secrete angiogenin to induce RNA cleavage in astroglia. J. Neurosci. *32*, 5024–5038.

Skowronek, E., Grzechnik, P., Späth, B., Marchfelder, A., and Kufel, J. (2014). tRNA 3' processing in yeast involves tRNase Z, Rex1, and Rrp6. Rna *20*, 115–130.

Skrtić, M., Sriskanthadevan, S., Jhas, B., Gebbia, M., Wang, X., Wang, Z., Hurren, R., Jitkova, Y., Gronda, M., Maclean, N., et al. (2011). Inhibition of mitochondrial translation as a therapeutic strategy for human acute myeloid leukemia. Cancer Cell *20*, 674–688.

Smeitink, J.A.M., Elpeleg, O., Antonicka, H., Diepstra, H., Saada, A., Smits, P., Sasarman, F., Vriend, G., Jacob-Hirsch, J., Shaag, A., et al. (2006). Distinct clinical phenotypes associated with a mutation in the mitochondrial translation elongation factor EFTs. Am. J. Hum. Genet. *79*, 869–877.

Smibert, P., Bejarano, F., Wang, D., Garaulet, D.L., Yang, J.-S., Martin, R., Bortolamiol-Becet, D., Robine, N., Hiesinger, P.R., and Lai, E.C. (2011). A Drosophila genetic screen yields allelic series of core microRNA biogenesis factors and reveals post-developmental roles for microRNAs. Rna *17*, 1997–2010.

Smith, C.W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem. Sci. *25*, 381–388.

Smith, F.J., Eady, R.A., Leigh, I.M., McMillan, J.R., Rugg, E.L., Kelsell, D.P., Bryant, S.P., Spurr, N.K., Geddes, J.F., Kirtschig, G., et al. (1996). Plectin deficiency results in muscular dystrophy with epidermolysis bullosa. Nat. Genet. *13*, 450–457.

Smits, P., Smeitink, J., and van den Heuvel, L. (2010). Mitochondrial translation and beyond: processes implicated in combined oxidative phosphorylation deficiencies. J. Biomed. Biotech. *2010*, 737385.

Smogorzewska, A., and de Lange, T. (2004). Regulation of telomerase by telomeric proteins. Annu. Rev. Biochem. *73*, 177–208.

Sommerville, J. (1999). Activities of cold-shock domain proteins in translation control. Bioessays *21*, 319–325.

Sondalle, S.B., and Baserga, S.J. (2014). Human diseases of the SSU processome. Bba *1842*, 758–764.

Sontheimer, E.J. (1994). SITE-SPECIFIC RNA CROSS-LINKING WITH 4-THIOURIDINE. Mol. Biol. Rep. *20*, 35–44.

Sossi, V., Giuli, A., Vitali, T., Tiziano, F., Mirabella, M., Antonelli, A., Neri, G., and Brahe, C. (2001). Premature termination mutations in exon 3 of the SMN1 gene are associated with exon skipping and a relatively mild SMA phenotype. Eur. J. Hum. Genet. *9*, 113–120.

Spencer, C.M., Serysheva, E., Yuva-Paylor, L.A., Oostra, B.A., Nelson, D.L., and Paylor, R. (2006). Exaggerated behavioral phenotypes in Fmr1/Fxr2 double knockout mice reveal a functional genetic interaction between Fragile X-related proteins. Hum. Mol. Genet. *15*, 1984–1994.

Spitzer, J., Landthaler, M., and Tuschl, T. (2013). Rapid creation of stable mammalian cell lines for regulated expression of proteins using the Gateway® recombination cloning technology and Flp-In T-REx® lines. Meth. Enzymol. *529*, 99–124.

Sreedharan, J., Blair, I.P., Tripathi, V.B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J.C., Williams, K.L., Buratti, E., et al. (2008). TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science *319*, 1668–1672.

Srivastava, L., Lapik, Y.R., Wang, M., and Pestov, D.G. (2010). Mammalian DEAD box protein Ddx51 acts in 3' end maturation of 28S rRNA by promoting the release of U8 snoRNA. Mol. Cell. Biol. *30*, 2947–2956.

Steenweg, M.E., Ghezzi, D., Haack, T., Abbink, T.E.M., Martinelli, D., van Berkel, C.G.M., Bley, A., Diogo, L., Grillo, E., Water Naudé, Te, J., et al. (2012). Leukoencephalopathy with thalamus and brainstem involvement and high lactate “LTBL” caused by EARS2 mutations. Brain *135*, 1387–1394.

Steitz, T.A. (2008). A structural understanding of the dynamic ribosome machine. Nat. Rev. Mol. Cell Biol. *9*, 242–253.

Steller, H. (2008). Regulation of apoptosis in Drosophila. Cell Death Differ. *15*, 1132–1138.

Stetson, D.B., Ko, J.S., Heidmann, T., and Medzhitov, R. (2008). Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell *134*, 587–598.

Stettner, G.M., Shoukier, M., Höger, C., Brockmann, K., and Auber, B. (2011). Familial intellectual disability and autistic behavior caused by a small FMR2 gene deletion. Am. J. Med. Genet. *155*, 2003–2007.

Stoltenburg, R., Reinemann, C., and Strehlitz, B. (2007). SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. Biomol. Engineering *24*, 381–403.

Strand, D.J., and McDonald, J.F. (1985). Copia is transcriptionally responsive to environmental stress. Nucleic Acids Res. *13*, 4401–4410.

Strein, C., Alleaume, A.-M., Rothbauer, U., Hentze, M.W., and Castello, A. (2014). A versatile assay for RNA-binding proteins in living cells. Rna *20*, 721–731.

Strezoska, Z., Pestov, D.G., and Lau, L.F. (2000). Bop1 is a mouse WD40 repeat nucleolar protein involved in 28S and 5.8S rRNA processing and 60S ribosome biogenesis. Mol. Cell. Biol. *20*, 5516–5528.

Stults, D.M., Killen, M.W., Williamson, E.P., Hourigan, J.S., Vargas, H.D., Arnold, S.M., Moscow, J.A., and Pierce, A.J. (2009). Human rRNA gene clusters are recombinational hotspots in cancer. Cancer Res. *69*, 9096–9104.

Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature *508*, 66–71.

Sutton, M.A., and Schuman, E.M. (2006). Dendritic protein synthesis, synaptic plasticity, and memory. Cell *127*, 49–58.

Suzuki, T., Nagao, A., and Suzuki, T. (2011). Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. Annu. Rev. Genet. *45*, 299–329.

Sylvester, J.E., Whiteman, D.A., Podolsky, R., Pozsgay, J.M., Respess, J., and Schmickel, R.D. (1986). The human ribosomal RNA genes: structure and organization of the complete repeating unit. Hum. Genet. *73*, 193–198.

Szymczyna, B.R., Bowman, J., McCracken, S., Pineda-Lucena, A., Lu, Y., Cox, B., Lambermon, M., Graveley, B.R., Arrowsmith, C.H., and Blencowe, B.J. (2003). Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing. Genes Dev. *17*, 461–475.

Tafforeau, L., Zorbas, C., Langhendries, J.-L., Mullineux, S.-T., Stamatopoulou, V., Mullier, R., Wacheul, L., and Lafontaine, D.L.J. (2013). The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. Mol. Cell *51*, 539–551.

Talim, B., Pyle, A., Griffin, H., Topaloglu, H., Tokatli, A., Keogh, M.J., Santibanez-Koref, M., Chinnery, P.F., and Horvath, R. (2013). Multisystem fatal infantile disease caused by a novel homozygous EARS2 mutation. Brain *136*, e228.

Tamburino, A.M., Ryder, S.P., and Walhout, A.J.M. (2013). A Compendium of Caenorhabditis elegans RNA Binding Proteins Predicts Extensive Regulation at Multiple Levels. G3 (Bethesda) *3*, 297–304.

Tan, C.P., and Nakielny, S. (2006). Control of the DNA methylation system component MBD2 by protein arginine methylation. Mol. Cell. Biol. *26*, 7224–7235.

Tan, H.L., Glen, E., Töpf, A., Hall, D., O'Sullivan, J.J., Sneddon, L., Wren, C., Avery, P., Lewis, R.J., Dijke, ten, P., et al. (2012). Nonsynonymous variants in the SMAD6 gene predispose to congenital cardiovascular malformation. Hum. Mutat. *33*, 720–727.

Tanackovic, G., Ransijn, A., Ayuso, C., Harper, S., Berson, E.L., and Rivolta, C. (2011). A missense mutation in PRPF6 causes impairment of pre-mRNA splicing and autosomal-dominant retinitis pigmentosa. Am. J. Hum. Genet. *88*, 643–649.

Tanaka, A., Morice-Picard, F., Lacombe, D., Nagy, N., Hide, M., Taïeb, A., and McGrath, J. (2010). Identification of a homozygous deletion mutation in C16orf57 in a family with Clericuzio-type poikiloderma with neutropenia. Am. J. Med. Genet. *152A*, 1347–1348.

Tanner, N.K., and Linder, P. (2001). DExD/H box RNA helicases: from generic motors to specific dissociation functions. Mol. Cell *8*, 251–262.

Tarpey, P.S., Raymond, F.L., Nguyen, L.S., Rodriguez, J., Hackett, A., Vandeleur, L., Smith, R., Shoubridge, C., Edkins, S., Stevens, C., et al. (2007). Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. Nat. Genet. *39*, 1127–1133.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. *28*, 33–36.

Tautz, D., and Dover, G.A. (1986). Transcription of the tandem array of ribosomal DNA in Drosophila melanogaster does not terminate at any fixed point. Embo J. *5*, 1267–1273.

Tautz, D., Hancock, J.M., Webb, D.A., Tautz, C., and Dover, G.A. (1988). Complete sequences of the rRNA genes of Drosophila melanogaster. Mol. Biol. Evol. *5*, 366–376.

Tavtigian, S.V., Simard, J., Teng, D.H., Abtin, V., Baumgard, M., Beck, A., Camp, N.J., Carillo, A.R., Chen, Y., Dayananth, P., et al. (2001). A candidate prostate cancer susceptibility gene at chromosome 17p. Nat. Genet. *27*, 172–180.

Tcherkezian, J., Cargnello, M., Romeo, Y., Huttlin, E.L., Lavoie, G., Gygi, S.P., and Roux, P.P. (2014). Proteomic analysis of cap-dependent translation identifies LARP1 as a key regulator of 5'TOP mRNA translation. Genes Dev. *28*, 357–371.

Tenenbaum, S.A., Carson, C.C., Lager, P.J., and Keene, J.D. (2000). Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc. Natl. Acad. Sci. U. S. a. *97*, 14085–14090.

Tessier, M.-C., Qu, H.-Q., Fréchette, R., Bacot, F., Grabs, R., Taback, S.P., Lawson, M.L., Kirsch, S.E., Hudson, T.J., and Polychronakos, C. (2006). Type 1 diabetes and the OAS gene cluster: association with splicing polymorphism or haplotype? J. Med. Genet. *43*, 129–132.

Thandapani, P., O'Connor, T.R., Bailey, T.L., and Richard, S. (2013). Defining the RGG/RG motif. Mol. Cell *50*, 613–623.

Tharun, S. (2009). Roles of eukaryotic Lsm proteins in the regulation of mRNA function. Int. Rev. Cell Mol. Biol. *272*, 149–189.

Thiyagarajan, N., Ferguson, R., Subramanian, V., and Acharya, K.R. (2012). Structural and molecular insights into the mechanism of action of human angiogenin-ALS variants in neurons. Nat. Commun. *3*, 1121.

Thomas, M.F., Abdul-Wajid, S., Panduro, M., Babiarz, J.E., Rajaram, M., Woodruff, P., Lanier, L.L., Heissmeyer, V., and Ansel, K.M. (2012). Eri1 regulates microRNA homeostasis and mouse lymphocyte development and antiviral function. Blood *120*, 130–142.

Thomas, M.F., L'Etoile, N.D., and Ansel, K.M. (2014). Eri1: a conserved enzyme at the crossroads of multiple RNA-processing pathways. Trends Genet. *30*, 298–307.

Thomson, E., Ferreira-Cerca, S., and Hurt, E. (2013). Eukaryotic ribosome biogenesis at a glance. J. Cell Sci. *126*, 4815–4821.

Thomson, T., and Lin, H. (2009). The Biogenesis and Function of PIWI Proteins and piRNAs: Progress and Prospect. Annu. Rev. Cell Dev. Biol. *25*, 355–376.

Thornton, J.E., and Gregory, R.I. (2012). How does Lin28 let-7 control development and disease? Trends Cell Biol. *22*, 474–482.

Tian, Y., Simanshu, D.K., Ascano, M., Diaz-Avalos, R., Park, A.Y., Juranek, S.A., Rice, W.J., Yin, Q., Robinson, C.V., Tuschl, T., et al. (2011). Multimeric assembly and biochemical characterization of the Trax-translin endonuclease complex. Nat. Struct. Mol. Biol. *18*, 658–664.

Timchenko, L.T., Timchenko, N.A., Caskey, C.T., and Roberts, R. (1996). Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy. Hum. Mol. Genet. *5*, 115–121.

Todd, A.E., Orengo, C.A., and Thornton, J.M. (2001). Evolution of function in protein superfamilies, from a structural perspective. J. Mol. Biol. *307*, 1113–1143.

Toomey, M.E., Panaram, K., Fast, E.M., Beatty, C., and Frydman, H.M. (2013). Evolutionarily conserved Wolbachia-encoded factors control pattern of stem-cell niche tropism in Drosophila ovaries and favor infection. Proc. Natl. Acad. Sci. U.S.a. *110*, 10788–10793.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2012a). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 1–9.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012b). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Tsakiri, K.D., Cronkhite, J.T., Kuan, P.J., Xing, C., Raghu, G., Weissler, J.C., Rosenblatt, R.L., Shay, J.W., and Garcia, C.K. (2007). Adult-onset pulmonary fibrosis caused by mutations in telomerase. Proc. Natl. Acad. Sci. U. S. a. *104*, 7552–7557.

Tsvetanova, N.G., Klass, D.M., Salzman, J., and Brown, P.O. (2010). Proteome-wide search reveals unexpected RNA-binding proteins in Saccharomyces cerevisiae. PLoS ONE *5*, e12671.

Ule, J. (2008). Ribonucleoprotein complexes in neurologic diseases. Curr. Opin. Neurobiol. *18*, 516–523.

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: Genomics, Evolution, and Mechanisms. Cell *154*, 26–46.

Valente, L., Tiranti, V., Marsano, R.M., Malfatti, E., Fernandez-Vizarra, E., Donnini, C., Mereghetti, P., De Gioia, L., Burlina, A., Castellan, C., et al. (2007). Infantile encephalopathy and defective mitochondrial DNA translation in patients with mutations of mitochondrial elongation factors EFG1 and EFTu. Am. J. Hum. Genet. *80*, 44–58.

Valverde, R., Edwards, L., and Regan, L. (2008). Structure and function of KH domains. Febs J. *275*, 2712–2726.

van de Laar, I.M.B.H., Oldenburg, R.A., Pals, G., Roos-Hesselink, J.W., de Graaf, B.M., Verhagen, J.M.A., Hoedemaekers, Y.M., Willemsen, R., Severijnen, L.-A., Venselaar, H., et al. (2011). Mutations in SMAD3 cause a syndromic form of aortic aneurysms and dissections with early-onset osteoarthritis. Nat. Genet. *43*, 121–126.

van den Bosch, M., Bree, R.T., and Lowndes, N.F. (2003). The MRN complex: coordinating and mediating the response to broken chromosomes. EMBO Rep. *4*, 844–849.

van der Knaap, M.S., Leegwater, P.A.J., Könst, A.A.M., Visser, A., Naidu, S., Oudejans, C.B.M., Schutgens, R.B.H., and Pronk, J.C. (2002). Mutations in each of the five subunits of translation initiation factor eIF2B can cause leukoencephalopathy with vanishing white matter. Ann. Neurol. *51*, 264–270.

van Hoof, A., Lennertz, P., and Parker, R. (2000). Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. Embo J. *19*, 1357–1365.

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P., et al. (2009). Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science *323*, 1208–1211.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. *10*, 252–263.

Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A.C., Engelberg-Kulka, H., and Moll, I. (2011). Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in Escherichia coli. Cell *147*, 147–157.

Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. *19*, 327–335.

Vissers, L.E.L.M., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M., et al. (2010). A de novo paradigm for mental retardation. Nat. Genet. *42*, 1109–1112.

Vithana, E.N., Abu-Safieh, L., Allen, M.J., Carey, A., Papaioannou, M., Chakarova, C., Al-Maghtheh, M., Ebenezer, N.D., Willis, C., Moore, A.T., et al. (2001). A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). Mol. Cell *8*, 375–381.

Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M.,

Blencowe, B.J., and Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature *474*, 380–384.

Volpi, L., Roversi, G., Colombo, E.A., Leijsten, N., Concolino, D., Calabria, A., Mencarelli, M.A., Fimiani, M., Macciardi, F., Pfundt, R., et al. (2010). Targeted next-generation sequencing appoints c16orf57 as clericuzio-type poikiloderma with neutropenia gene. Am. J. Hum. Genet. *86*, 72–76.

Voronina, E., Seydoux, G., Sassone-Corsi, P., and Nagamori, I. (2011). RNA granules in germ cells. Cold Spring Harb. Perspect. Biol. *3*.

Vulliamy, T.J., Knight, S.W., Heiss, N.S., Smith, O.P., Poustka, A., Dokal, I., and Mason, P.J. (1999). Dyskeratosis congenita caused by a 3' deletion: germline and somatic mosaicism in a female carrier. Blood *94*, 1254–1260.

Vulliamy, T., Beswick, R., Kirwan, M., Marrone, A., Digweed, M., Walne, A., and Dokal, I. (2008). Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita. Proc. Natl. Acad. Sci. U. S. a. *105*, 8073–8078.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell *136*, 701–718.

Walne, A.J., Vulliamy, T., Marrone, A., Beswick, R., Kirwan, M., Masunari, Y., Al-Qurashi, F.-H., Aljurf, M., and Dokal, I. (2007). Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10. Hum. Mol. Genet. *16*, 1619–1629.

Wan, J., Yourshaw, M., Mamsa, H., Rudnik-Schöneborn, S., Menezes, M.P., Hong, J.E., Leong, D.W., Senderek, J., Salman, M.S., Chitayat, D., et al. (2012). Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration. Nat. Genet. *44*, 704–708.

Wan, M., Lee, S.S., Zhang, X., Houwink-Manville, I., Song, H.R., Amir, R.E., Budden, S., Naidu, S., Pereira, J.L., Lo, I.F., et al. (1999). Rett syndrome and beyond: recurrent spontaneous and familial MECP2 mutations at CpG hotspots. Am. J. Hum. Genet. *65*, 1520–1529.

Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E., et al. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. Nature *505*, 706–709.

Wang, G.-S., and Cooper, T.A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. Nat. Rev. Genet. *8*, 749–761.

Wang, M., and Pestov, D.G. (2011). 5'-end surveillance by Xrn2 acts as a shared mechanism for mammalian pre-rRNA maturation and decay. Nucleic Acids Res. *39*, 1811–1822.

Wang, T., Bray, S.M., and Warren, S.T. (2012). New perspectives on the biology of fragile X syndrome. Curr. Opin. Genet. Dev. *22*, 256–263.

Wang, X.Q., Zamore, P.D., and Hall, T.M.T. (2001). Crystal structure of a Pumilio homology

domain. Mol. Cell *7*, 855–865.

Wang, X., McLachlan, J., Zamore, P.D., and Hall, T.M.T. (2002). Modular recognition of RNA by a human pumilio-homology domain. Cell *110*, 501–512.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. Trends Biochem. Sci. *24*, 437–440.

Warner, J.R., and McIntosh, K.B. (2009). How common are extraribosomal functions of ribosomal proteins? Mol. Cell *34*, 3–11.

Watkins, N.J., and Bohnsack, M.T. (2012). The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. WIREs RNA *3*, 397–414.

Weedon, M.N., Hastings, R., Caswell, R., Xie, W., Paszkiewicz, K., Antoniadi, T., Williams, M., King, C., Greenhalgh, L., Newbury-Ecob, R., et al. (2011). Exome sequencing identifies a DYNC1H1 mutation in a large pedigree with dominant axonal Charcot-Marie-Tooth disease. Am. J. Hum. Genet. *89*, 308–312.

White-Cooper, H. (2010). Molecular mechanisms of gene regulation during Drosophila spermatogenesis. Reproduction *139*, 11–21.

Wilbert, M.L., Huelga, S.C., Kapeli, K., Stark, T.J., Liang, T.Y., Chen, S.X., Yan, B.Y., Nathanson, J.L., Hutt, K.R., Lovci, M.T., et al. (2012). LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. Mol. Cell *48*, 195–206.

Williamson, A., and Lehmann, R. (1996). Germ cell development in Drosophila. Annu. Rev. Cell Dev. Biol. *12*, 365–391.

Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C., and Gough, J. (2009). SUPERFAMILY--sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res. *37*, D380–D386.

Wilusz, C.J., and Wilusz, J. (2005). Eukaryotic Lsm proteins: lessons from bacteria. Nat. Struct. Mol. Biol. *12*, 1031–1036.

Winter, E.E., Goodstadt, L., and Ponting, C.P. (2004). Elevated rates of protein secretion, evolution, and disease among tissue-specific genes. Genome Res. *14*, 54–61.

Wolf, J., and Passmore, L.A. (2014). mRNA deadenylation by Pan2-Pan3. Biochem. Soc. Trans *42*, 184–187.

Wool, I.G. (1979). The Structure and Function of Eukaryotic Ribosomes. Annu. Rev. Biochem. *48*, 719–754.

Wool, I.G., Chan, Y.L., and Glück, A. (1995). Structure and evolution of mammalian ribosomal

proteins. Biochem. Cell Biol. *73*, 933–947.

Woolford, J.L., and Baserga, S.J. (2013). Ribosome Biogenesis in the Yeast Saccharomyces cerevisiae. Genetics *195*, 643–681.

Wöhrle, D., Kotzot, D., Hirst, M.C., Manca, A., Korn, B., Schmidt, A., Barbi, G., Rott, H.D., Poustka, A., and Davies, K.E. (1992). A microdeletion of less than 250 kb, including the proximal part of the FMR-I gene and the fragile-X site, in a male with the clinical phenotype of fragile-X syndrome. Am. J. Hum. Genet. *51*, 299–306.

Wu, H., Xu, H., Miraglia, L.J., and Crooke, S.T. (2000). Human RNase III is a 160-kDa protein involved in preribosomal RNA processing. J. Biol. Chem. *275*, 36957–36965.

Wu, P., Brockenbrough, J.S., Metcalfe, A.C., Chen, S., and Aris, J.P. (1998). Nop5p is a small nucleolar ribonucleoprotein component required for pre-18 S rRNA processing in yeast. J. Biol. Chem. *273*, 16453–16463.

Xiao, S., Scott, F., Fierke, C.A., and Engelke, D.R. (2002). Eukaryotic ribonuclease P: a plurality of ribonucleoprotein enzymes. Annu. Rev. Biochem. *71*, 165–189.

Xue, S., and Barna, M. (2012). Specialized ribosomes: a new frontier in gene regulation and organismal biology. Nat. Rev. Mol. Cell Biol. *13*, 355–369.

Yamaguchi, H., Calado, R.T., Ly, H., Kajigaya, S., Baerlocher, G.M., Chanock, S.J., Lansdorp, P.M., and Young, N.S. (2005). Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. N. Engl. J. Med. *352*, 1413–1424.

Yanase, T., Takayanagi, R., Oba, K., Nishi, Y., Ohe, K., and Nawata, H. (1996). New mutations of DAX-1 genes in two Japanese patients with X-linked congenital adrenal hypoplasia and hypogonadotropic hypogonadism. J. Clin. Endocrinol. Metab. *81*, 530–535.

Yao, P., and Fox, P.L. (2013). Aminoacyl-tRNA synthetases in medicine and disease. EMBO Mol. Med. *5*, 332–343.

Yin, Q.-F., Yang, L., Zhang, Y., Xiang, J.-F., Wu, Y.-W., Carmichael, G.G., and Chen, L.-L. (2012). Long Noncoding RNAs with snoRNA Ends. Mol. Cell *48*, 219–230.

Yisraeli, J.K. (2005). VICKZ proteins: a multi-talented family of regulatory RNA-binding proteins. Biol. Cell *97*, 87–96.

Yoon, J.-H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G., and Gorospe, M. (2012). LincRNA-p21 suppresses target mRNA translation. Mol. Cell *47*, 648–655.

Yoshihama, M., Nakao, A., and Kenmochi, N. (2013). snOPY: a small nucleolar RNA orthological gene database. BMC Res. Notes *6*, 426.

You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G., Akbalik, G., Wang, M., Glock, C., Quedenau, C., et al. (2015). Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. Nature Neuroscience *18*, 603–610.

Yuan, F., Dutta, T., Wang, L., Song, L., Gu, L., Qian, L., Benitez, A., Ning, S., Malhotra, A., Deutscher, M.P., et al. (2015). Human DNA Exonuclease TREX1 Is Also an Exoribonuclease That Acts on Single-stranded RNA. J. Biol. Chem. *290*, 13344–13353.

Zamore, P.D., Williamson, J.R., and Lehmann, R. (1997). The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. Rna *3*, 1421–1433.

Zeharia, A., Shaag, A., Pappo, O., Mager-Heckel, A.-M., Saada, A., Beinat, M., Karicheva, O., Mandel, H., Ofek, N., Segel, R., et al. (2009). Acute infantile liver failure due to mutations in the TRMU gene. Am. J. Hum. Genet. *85*, 401–407.

Zhang, A., Yu, H., He, Y., Shen, Y., Pan, N., Liu, J., Fu, B., Miao, F., and Zhang, J. (2013). The spatio-temporal expression of MHC class I molecules during human hippocampal formation development. Brain Res. *1529*, 26–38.

Zhang, L., Huang, J., Yang, N., Greshock, J., Megraw, M.S., Giannakakis, A., Liang, S., Naylor, T.L., Barchetti, A., Ward, M.R., et al. (2006). microRNAs exhibit high frequency genomic alterations in human cancer. Proc. Natl. Acad. Sci. U. S. a. *103*, 9136–9141.

Zhang, Q., Shalaby, N.A., and Buszczak, M. (2014). Changes in rRNA transcription influence proliferation and cell fate within a stem cell lineage. Science *343*, 298–301.

Zhang, X., Gao, X., Coots, R.A., Conn, C.S., Liu, B., and Qian, S.-B. (2015a). Translational control of the cytosolic stress response by mitochondrial ribosomal protein L18. Nat. Struct. Mol. Biol. *22*, 404–410.

Zhang, X., Gao, X., Coots, R.A., Conn, C.S., Liu, B., and Qian, S.-B. (2015b). Translational control of the cytosolic stress response by mitochondrial ribosomal protein L18. Nat. Struct. Mol. Biol. *22*, 404–410.

Zhang, Z.H., Niu, Z.M., Yuan, W.T., Zhao, J.J., Jiang, F.X., Zhang, J., Chai, B., Cui, F., Chen, W., Lian, C.H., et al. (2005). A mutation in SART3 gene in a Chinese pedigree with disseminated superficial actinic porokeratosis. Br. J. Dermatol. *152*, 658–663.

Zhang, Z., Theler, D., Kaminska, K.H., Hiller, M., la Grange, de, P., Pudimat, R., Rafalska, I., Heinrich, B., Bujnicki, J.M., Allain, F.H.T., et al. (2010). The YTH domain is a novel RNA binding domain. J. Biol. Chem. *285*, 14701–14710.

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., and Dreyfuss, G. (2008). SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. Cell *133*, 585–600.

Zhao, C., Bellur, D.L., Lu, S., Zhao, F., Grassi, M.A., Bowne, S.J., Sullivan, L.S., Daiger, S.P., Chen, L.J., Pang, C.P., et al. (2009). Autosomal-dominant retinitis pigmentosa caused by a mutation in SNRNP200, a gene required for unwinding of U4/U6 snRNAs. Am. J. Hum. Genet. *85*, 617–627.

Zheng, K., Xiol, J., Reuter, M., Eckardt, S., Leu, N.A., McLaughlin, K.J., Stark, A., Sachidanandam, R., Pillai, R.S., and Wang, P.J. (2010). Mouse MOV10L1 associates with Piwi

proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. Proc. Natl. Acad. Sci. U. S. a. *107*, 11841–11846.

Zhong, F., Savage, S.A., Shkreli, M., Giri, N., Jessop, L., Myers, T., Chen, R., Alter, B.P., and Artandi, S.E. (2011). Disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita. Genes Dev. *25*, 11–16.

Zuo, Y., and Deutscher, M.P. (2001). Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. Nucleic Acids Res. *29*, 1017–1026.

Zuo, Y.H., Zheng, H.P., Wang, Y., Chruszcz, M., Cymborowski, M., Skarina, T., Savchenko, A., Malhotra, A., and Minor, W. (2007). Crystal structure of RNase T, an exoribonuclease involved in tRNA maturation and end turnover. Structure *15*, 417–428.

BrainSpan: Atlas of the Developing Human Brain. http://developinghumanbrain.org.