The mRNA-bound Proteome of a human cell-line and in early Drosophila embryogenesis

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy

of Freie Universität Berlin

by

Dipl.-Biol. (t.o.) Alexander Georg Baltz

2013

1st Reviewer:

Prof. Dr. **Markus Wahl**

Institut für Chemie und Biochemie Freie Universität Berlin

Takustr. 6

14195 Berlin

Tel.: +49 30 838 53456

E-Mail: mwahl@chemie.fu-berlin.de

2nd Reviewer:

Prof. Dr. **Matthias Selbach**

Cell Signalling and Mass Spectrometry

Max Delbrück Center for Molecular Medicine

Robert-Rössle-Str. 10

13125 Berlin

Tel.: +49 30 9406 2464

Email: matthias.selbach@mdc-berlin.de

Disputation am 5. Mai 2014

Acknowledgement

My first thanks goes to my supervisors, *Markus Landthaler* and *Stephen J Small*.
I want to thank both for excellent supervision, giving me the opportunity to work in their lab and expanding my knowledge. By creating an inspiring, motivating and vivid research environment they helped making the four years an exciting, fruitful and beautiful time in my life.

I would like to thank all the colleagues in the lab, who were always helping and very supportive. You all contributed a lot to make this happen and I really appreciate it. It was such a pleasurable experience and I will miss you guys!
Also, merci vielmals an *Emanuel Wyler* for carefully proofreading the thesis.

This work would have never been achieved without my wonderful collaborators, I would therefore like to thank:

Matthias Selbach, Björn Schwanhäusser and *Koshi Imami* of the Flying Ion Circus for the mass spectrometry analysis and fruitful discussions.

Markus Schüler and *Christoph Dieterich* for computational analyses on this side of the atlantic.

Noah Youngs, Duncan Penfold-Brown, Kevin Drew and *Richard Bonneau* for computational analysis on the other side of the atlantic.

Claudia Langnick and Mirjam Feldkamp from the Wei Chen lab for sequencing.

Guido Mastrobuoni from the Stefan Kempa lab for the LC-MS.

I would also like to thank *Jennifer Stewart* and the whole PhD Graduate Office for the assistance helping me to find my way through bureaucracy.

I would like to thank all my friends for being who they are! Your support, humor, and everything else that certainly won't fit on one page made the last four years certainly one of the best times in my life. In particular, I would like to thank Fra, for keeping it all happy!

Vor allem moechte ich meiner Familie, ganz besonders meinen Eltern für die ständige Unterstützung und Liebe danken! Ohne Euch wäre ich nicht so weit gekommen. Ich kann Euch gar nicht genug danken!

Statement of contributions

Parts of the project have been published as:

Baltz, A.G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C. and Landthaler, M. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Molecular cell *46*, 674-690.

Parts of the methods section and figures are published in a similar or slightly modified way in *Baltz et al, 2012*.

The work presented here is the result of collaborative projects.
The first project, the mRNA-bound proteome of a human cell line was designed and supervised by Dr. Markus Landthaler. Experiments were designed, performed and analyzed by myself unless otherwise noted:
Björn Schwanhäusser performed proteomic analysis for the mRNA-bound proteome in HEK293 cells within the collaboration with the lab of Matthias Selbach. Björn Schwanhäusser, Matthias Selbach, Markus Landthaler and myself analyzed the proteomics data. Deep sequencing for the mRNA-bound proteome paper was performed in the BIMSB sequencing facility under supervision of Dr. Wei Chen. Deep Sequencing data was analyzed by Markus Schüler and Christoph Dieterich. The "mRNA-bound Proteome" part of the "The mRNA-bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts" paper was mainly written by Markus Landthaler and myself. Mathias Munschauer and Markus Landthaler wrote the "Global Occupancy Profile" part of the paper. Christoph Dieterich, Richard Bonneau and members of the respective labs performed the computational analysis in (Baltz et al., 2012). The PAR-CLIPseq experiments in the mRNA-bound proteome paper were performed by members of the Landthaler group (Mathias Munschauer, Alexandra Vasile, Yasuhiro Murakawa, Miha Milek and Emanuel Wyler). IFIT5 PAR-CLIPseq experiments were performed by myself and jointly analyzed by Markus Schüler, Yasuhiro Murakawa and myself.

Stephen J. Small supervised the "mRNA-bound proteome in *Drosophila melanogaster*" project. This research was conducted as part of the MDC-NYU PhD Exchange Program in

collaboration with the laboratory of Stephen J. Small at the Department of Biology, New York University. Experiments were designed, performed and analyzed by myself unless otherwise noted. Koshi Imami (Mathias Selbach lab) performed proteomic analysis for the mRNA-bound proteome in *Drosophila melanogaster* embryos. We thank Guido Mastrobuoni (Stefan Kempa lab) for the quantification of 4-thiouridine incorporation by LC-MS.

Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt habe.

Berlin, den 30.07.2013

Alexander Baltz

Unterschrift _____

For R.T. Gihihi

Table of Contents

Abstract

Eukaryotic gene expression is extensively regulated at the post-transcriptional level. Throughout its lifecycle, mRNA is bound by a dynamic repertoire of RNA-binding proteins (RBPs), which interact with mRNA in a sequence and/or structure dependent manner. The combinatorial binding of RBPs and ribonucleoprotein (RNP) complexes regulates many steps of eukaryotic mRNA processing. Although RBPs are crucial for these processes and malfunctions are often causes of disease, the mRNA-bound proteome was not previously systematically characterized.

We developed a photoreactive nucleoside-enhanced *in vivo* UV crosslinking and oligo(dT) purification approach to identify the mRNA-bound proteome using high-resolution quantitative mass spectrometry. When applied to a human embryonic kidney cell line (HEK293) this method led to the reproducible identification of almost 800 proteins. As expected, known RBPs, as well as the majority of proteins containing RNA-binding domains expressed in HEK293 cells were detected, and represent a large percentage of our data. Surprisingly, nearly one-third of the bound proteins were not previously annotated as RNA-binding, and about half of these were not predictable by computational methods to interact with RNA. We validated direct RNA-binding activity of several putative RNA-interacting proteins by Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP). Amongst the identified proteins was IFIT5, a viral response induced protein with enigmatic RNA-binding function. We investigated the RNA-binding activity of IFIT5 using PAR-CLIP, and found that IFIT5 is binding a broad spectrum of RNAs.

I also used the method to characterize the mRNA-bound proteome in the *Drosophila melanogaster* embryo, where many events are driven by post-transcriptional regulation such as mRNA localization and translational regulation. Before maternal-to-zygotic transition, proteins are synthesized from maternal mRNAs, which are already present at fertilization (maternal effect genes). Although we know that some maternal effect genes are bound and regulated by RBPs, we still lack a global picture about the proteins that mediate posttranscriptional regulation. Therefore we set out to characterize the mRNA-bound proteome of *Drosophila melanogaster* embryos in earliest developmental stages, before the onset of zygotic transcription.

We were able to identify about 1000 proteins in oligo(dT) precipitates. In a preliminary analysis we detected a large number of known RBPs and regulators of translation. We

consider the identification of the mRNA-bound proteome of a precisely defined developmental stage as a first step towards understanding the post-transcriptional gene regulatory code during development and differentiation.

Zusammenfassung

Genexpression wird in Eukaryonten weitgehend auf post-transkriptioneller Ebene reguliert. Während ihrer gesamten Lebensdauer wird die Boten-RNA (mRNA) durch ein dynamisches Repertoire von RNA-bindenden Proteinen (RBP) gebunden, die sequenz- und / oder strukturabhängig mit ihr interagieren. Kombinatorische Interaktionen von RBPs und Ribonucleoprotein-Komplexen regulieren nahezu alle Schritte der eukaryotischen mRNA-Prozessierung. Obwohl RBPs entscheidend für diese Prozesse sind und Störungen oft Ursachen von Krankheiten sind, wurde das mRNA-gebundene Proteom bisher nicht systematisch charakterisiert.

Wir entwickelten einen *in vivo*-Ansatz, basierend auf UV-induzierten, kovalenten Bindungen zwischen Proteinen und RNA mithilfe photoreaktiver Nucleoside, um das mRNA-gebundene Proteom zu isolieren und untersuchen. Die Protein-RNA-Komplexe lassen sich mittels magnetischen, oligo(dT)-beschichteten Partikeln aus dem Zelllysat isolieren, anschließend werden die Proteine mittels hochauflösender, quantitativer Massenspektrometrie identifiziert.

Die Anwendung dieser Methode in einer humanen Zelllinie (HEK293) führte zur Identifizierung von nahezu 800 Proteinen. Wie erwartet wurde die Mehrzahl der in HEK293-Zellen vorhandenen RBPs detektiert. Überraschenderweise war fast ein Drittel der Proteine nicht als RNA-bindend annotiert, und etwa die Hälfte davon konnte nicht mit rechnergestützten, theoretischen Methoden vorhergesagt werden. Für eine große Mehrheit einer Stichprobe dieser Proteine ließ sich in Folgeexperimenten die RNA-Bindung bestätigen.

Viele der identifizierten mRNA-bindenden Proteine scheinen Teil von post-transkriptionellen Netzwerken zur Regulation der Genexpression zu sein, darunter IFIT5, ein Protein dass als Antwort auf virale Infektionen vermehrt gebildet wird und dessen Aktivität als RNA-bindendes Protein vor unserer Studie ungewiss war.

Der nächste Schritt war die Entwicklung einer Methode, die die Untersuchung des mRNA-gebundenen Proteoms in einem Modellorganismus mit komplexem Entwicklungsprogramm, genauer dem Embryo der Fruchtfliege *Drosophila melanogaster*, erlaubt.

Waehrend der *Drosophila*-Embryogenese werden viele entscheidende Schritte durch post-transkriptionelle Regulation gesteuert. Vor allem mRNA-Lokalisation und die Regulation der Translation spielen hierbei eine entscheidende Rolle. Im *Drosophila*-Embryo werden Proteine von mRNAs synthetisiert die bereits bei der Befruchtung vorhanden sind. Es gibt zahlreiche gut untersuchte Einzelbeispiele, die die Notwendigeit der translationalen Regulation in den ersten Stunden der Embryogenese untermauern, jedoch ist die Anzahl der so regulierten Gene und die unterliegenden Mechanismen weitgened unbekannt. Ihnen gemein ist aber die Interaktion von Proteinen mit RNA. Jedoch gab es noch keine Studie, die versucht hat, alle mRNA-interagierenden Proteine in einem präzise definierten Entwicklungsstadium, der frühen Embryogenese, zu identifizieren.

Durch eine Anpassung unserer Methode konnten wir etwa 1000 Proteine identifizieren, die vermutlich im *Drosophila*-Embryo mit mRNA interagieren. In einer vorläufigen Analyse entdeckten wir eine große Anzahl von bekannten RBPs, darunter einige Translations-Regulatoren. Darüber hinaus entdeckten wir auch hier Proteine, die nicht als RNA-bindend bekannt waren. Einige dieser Proteine wurden auch in HEK293 Zellen identifiziert, was auf eine evolutionäre Konservierung der RBP-RNA-Interaktion hindeutet.

Die Identifizierung des mRNA-gebundenen Proteoms in einem genau definierten Entwicklungsstadium ist ein erster Schritt zum Verständnis des post-transkriptionellen Codes der Genexpression, und seiner Funktion während der Entwicklung und Differenzierung eines biologischen Organismus.

Abbreviations

4SU	4-thiouridine
4TU	4-thiouracil
6SG	6-thioguanosine
bp	Basepairs
C	Celsius
cCL	conventional Crosslink
CDS	Coding sequence
Da	Dalton
dFCS	Dialyzed fetal calf serum
DGE	Digital Gene Expression
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Desoxyribonucleic acid
dsRNA	Double stranded RNA
FCS	Fetal calf serum
GO	Gene Ontology
h	Hour(s)
iBAQ	Intensity based absolute quantification
k	kilo
LC-MS/MS	Liquid chromatography coupled with tandem mass spectrometry
LiDS	Lithium dodecyl sulfate
LOESS	locally weighted scatterplot smoothing
m/z	Mass-to-charge ratio
min	minute(s)
miRNA	microRNA
mm	Millimeter
mRNA	messenger RNA
MS	Mass spectrometry
MZT	Maternal-to-Zygotic transition

nt	Nucleotides
PAR-CL	Photoactivatable Ribonucleoside Crosslink
PAR-CLIP	PAR-Crosslinking and Immunoprecipitation
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
piRNA	Piwi-interacting RNA
PPI	Protein-protein-interaction
RCF	Relative centrifugal force
RNA	Ribonucleic acid
RPKM	reads per kilobase of exon model per million mapped reads
rpm	Rotations per minute
rRNA	ribosomal RNA
RT	room temperature (25C)
RT-PCR	Reverse transcription polymerase chain reaction
s	Second(s)
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	Stable isotope labeling by amino acid in cell culture
siRNA	Small interfering RNA
TBS	Tris buffered Saline
TCA	Tricarboxylic acid
tRNA	transfer RNA
UTR	untranslated region
V	Volt

1. Introduction

1.1. Post-transcriptional gene regulation in eukaryotes

The central dogma of molecular biology, as first formulated by Francis Crick in 1958 (Crick, 1958), states that DNA is transcribed to mRNA, which is translated into protein. The amount of protein produced from DNA can hereby be regulated at different levels. Gene expression is tightly regulated at the level of transcription, by transcription factors and the chromatin state, which controls how much RNA is being transcribed from DNA (Berger, 2007; Spitz and Furlong, 2012). After transcription, gene regulation is achieved on the one hand through control of the mRNA levels in the cell, i.e. mRNA stability and decay rate, and on the other hand, through control of how much mRNA is actively translated into protein (Moore, 2005; Sonenberg and Hinnebusch, 2009).

The central dogma of molecular biology can be applied to all living organisms, but the extent to which gene expression is regulated, generally increases with organism complexity. Prokaryotes, for example, lack a nuclear compartment, thus transcription of mRNA is directly coupled to translation into protein, which limits the possibilities for post-transcriptional gene regulation. In eukaryotes however, mRNA has to be transcribed and transported out of the nucleus into the cytoplasm, where translation takes place.

Messenger RNA is produced in the nucleus by transcription, followed by RNA-processing, which includes splicing, addition of a poly(A)-tail to the 3'-end and addition of a 7-methylguanylate (m7G)-cap at the 5'-end. One of the interesting features of eukaryotic mRNAs is that a number of processing events, such as alternative splicing and polyadenylation can alter the mRNA. This leads to a wide diversity of transcript isoforms being produced from a single gene, which controls the features of the protein being produced, as well as the fate of the mRNA by regulation of localization and degradation (Moore, 2005). Eukaryotic mRNA is thus subject to extensive post-transcriptional gene regulation before being translated into protein.

Gene expression must be tightly controlled, since correct gene expression is crucial for the survival of the organism. However, although gene expression is tightly controlled, ensuring the production of the right protein at the right time and at the proper location, it is also a dynamic process and must allow a certain degree of flexibility. One reason for

the dynamic behavior is, that living systems need to react to external stimuli and adapt to stress, thus maintaining cellular homeostasis which ultimately allows survival (Orphanides and Reinberg, 2002). This can be achieved either at the transcriptional or at the post-transcriptional/translational level, depending on the nature of the stimuli and the most efficient way to respond.

Transcriptional control is energetically favorable since RNAs are only transcribed upon demand and thus transcriptional control would prevent "uneconomic" mRNA-production.

If a fast response to a stimulus is required, it is beneficial to regulate gene expression at the translation step, since the need for time-consuming mRNA transcription and processing could be bypassed (Holcik and Sonenberg, 2005). This form of gene expression regulation requires mechanisms to prevent translation, for example by inactivating translation initiation factors (eIFs) (Sonenberg and Hinnebusch, 2009). On the other hand, translation can also be specifically upregulated by RNA binding proteins (RBPs) and microRNAs (miRNAs) (Vasudevan, 2012; Vasudevan and Steitz, 2007; Vasudevan et al., 2007).

Changing the transcripts stability and decay rate can also regulate the transcript levels in the cell, especially for short-lived transcripts (de Nadal et al., 2011).

Post-transcriptional and translational gene regulation is the main reason, why the levels of mRNA do not always correlate with the protein levels in the cell. Recent studies observed that this trend increases with organism complexity. The correlation between mRNA and protein level was $R^2 \sim 0.8$ in *E. coli*, (Taniguchi et al., 2010), ~0.6 in *S. cerevisiae* (de Sousa Abreu et al., 2009) and ~0.4 in human (Schwanhausser et al., 2011) which is indicative for extensive post-transcriptional and translational regulation in higher eukaryotes. But what is the major contribution to the difference between mRNA and protein level in the cell?

According to the study from the lab of M. Selbach, where concentration and turnover of thousands of mRNAs and proteins were measured in parallel in a human cell line, about 40 % of the variation in cellular protein concentration can be explained by mRNA concentration and about 41 - 54% by differences in translation rates. The degradation rates of mRNA and protein on the other hand, seem to have a surprisingly small role in regulating protein concentrations (Schwanhausser et al., 2011). Based on this observation, this study proposed that in a human cell line post-transcriptional and

translational regulation are at least as important as transcriptional control in determining the protein levels in the cell.

The main mediators of post-transcriptional gene regulation are RNA-interacting proteins such as RNA-binding proteins (RBPs), RNA-helicases, nucleases and RNA-modifying enzymes and non-coding regulatory RNAs such as micro-RNAs (miRNA). During and immediately after transcription, nascent mRNA is bound by RBPs, forming messenger nucleoprotein (mRNP) complexes, which modulate and regulate virtually all aspects of mRNA metabolism. Throughout their life cycle, mRNP complexes consist of a dynamically changing repertoire of proteins and RNAs that modulate processing, cellular localization, and the decay and translation rate of mRNAs (Figure 1) (Martin and Ephrussi, 2009; Moore and Proudfoot, 2009; Sonenberg and Hinnebusch, 2009).

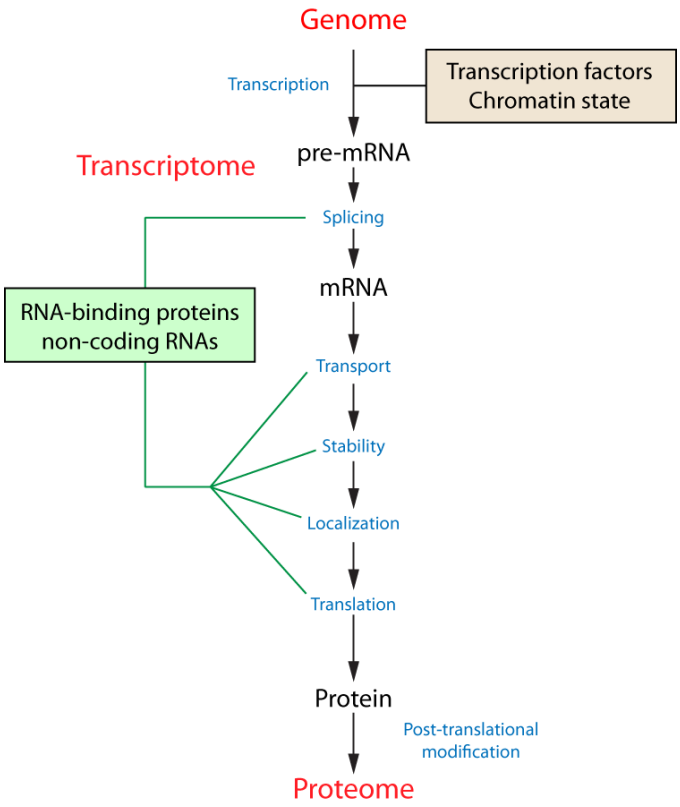


Figure 1: RNA-binding proteins are involved in all aspects of mRNA metabolism and function. In eukaryotic cells, the current model states that RNA-binding proteins bind mRNAs throughout their life cycles. RNA-binding proteins coordinate and regulate the interconnected steps of post-transcriptional regulation.

The individual steps of post-transcriptional regulation (as illustrated in Figure 1) are in many instances interconnected. Regulatory RBPs and RNPs often participate in

numerous mRNA-processing events, which physically couple the processes and allow direct feedback between sequential and non-sequential steps in mRNA metabolism (Komili and Silver, 2008; Moore and Proudfoot, 2009; Muller-McNicoll and Neugebauer, 2013). For example, nuclear RNA processing can influence the subsequent localization of mRNA. It was shown that the correct positioning of *oskar* mRNA to the posterior pole of the *D. melanogaster* oocyte depends on mRNA splicing events in the nucleus (Ghosh et al., 2012; Hachet and Ephrussi, 2004). The crosstalk between transcriptional, post-transcriptional and translational regulation thus adds another level of complexity to eukaryotic gene expression as illustrated in Figure 2.

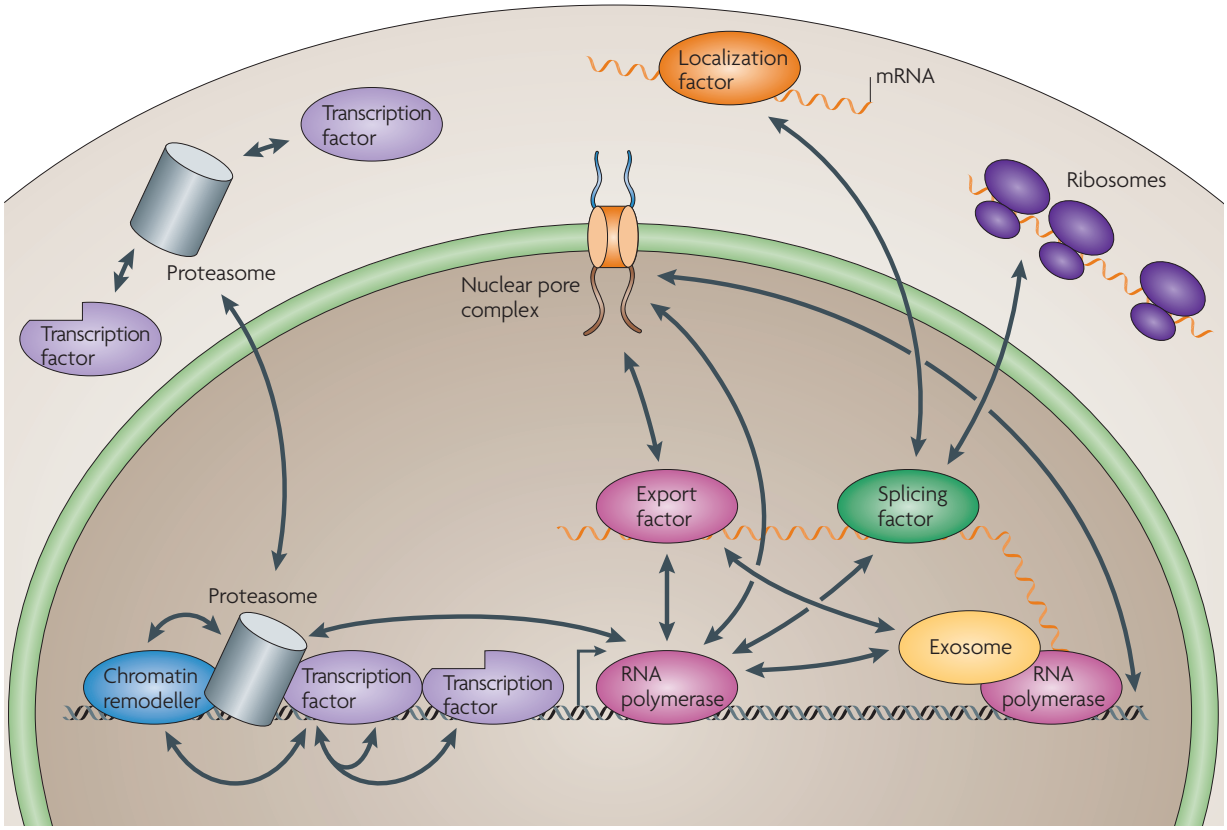


Figure 2: Eukaryotic gene regulatory processes are interconnected. The transcription of a gene is regulated by transcription factors, chromatin remodelers, subnuclear localization and activation by the proteasome. In the cytoplasm, the proteasome can also process certain transcription factors into their 'active' form, thus modulating gene expression. The different stages of mRNA processing are also intricately connected, including coupling between transcription, splicing, export and degradation. These nuclear processing events also affect the mRNA cytoplasmic fate, regulating both its cytoplasmic localization and translation level. The figure also shows ribosomes and the exosome, an enzymatic complex involved in the regulation of mRNA metabolism. Reprinted by permission from Macmillan Publishers Ltd: [Nature Reviews Genetics] (Komili and Silver, 2008), copyright (2008).

1.2. RNA-binding proteins at the heart of post-transcriptional gene regulation

RBPs are involved in virtually every aspect of PTGR.

To our current knowledge, from transcription to degradation, RNA is always present in complex with RNA-binding proteins in the cell. The protein population bound to mRNA changes during the lifetime of the RNA and the dynamic assembly and composition of mRNPs determines partially the fate of the transcript. An average mRNA molecule in *S. cerevisiae* is bound by at least 10 different known RBPs during its life cycle from transcription to degradation (Hogan et al., 2008). It is reasonable to assume that with increased organism complexity, the number of proteins interacting with RNA is also increasing, which can be partly explained by the increasing length of the untranslated regions.

How do RNA-binding proteins bind their targets?

RBPs are involved in many processes of post-transcriptional gene regulation, and thus interact with a multitude of different RNA targets and binding sites. The broad variety in the sequence and structure of RNA molecules would suggest a comparably large structural repertoire of protein domains to allow specific interaction of RBPs with their targets. However, most RBPs are built in a modular structure from a combination of common RNA-binding domains (RBDs) (Lunde et al., 2007). Evolution has generated a wide range of RNA-binding domains in proteins. One of the most frequently occurring RBD in nature is the RNA recognition motif (RRM) which is present in about 0.5 – 1 % of all human genes. RRMs generally bind to specific single stranded ssRNA sequences. The double stranded dsRBD on the other hand interacts preferably with dsRNA, and some domains, like the Zinc fingers and the KH domain (heterogeneous nuclear (hn) RNP K-homology domain) can interact with both ssRNA and ssDNA (Lunde et al., 2007)).

The specificity of an RBP is on the one hand determined by the specificity of the RNA-binding domain itself, based on the peptide sequence and structure fold, and on the other hand by the combination and spatial organization of those. This organization can be seen in analogy to a molecular toolkit, which allows developing RNA-binding proteins handling specific functions from a common set of building blocks. These domains can act together as RNA-recognition units, with biological functions that either cannot be achieved by one RNA-binding domain alone or might combine several biological functions into one functional unit.

Based on the presence of one or more domains that interact with RNA, over 600 proteins in the mammalian genome have been predicted to interact with RNA (de Lima Morais et al., 2011). Since it was known for many RBPs that they interact with mRNA through specific domains and combinations thereof, computational tools were designed to predict RNA-binding proteins based on the knowledge of known RNA-binding proteins and their RNA-binding domains (Han et al., 2004). However, this approach ignores all RBPs with uncharacterized RNA-binding domains and also proteins that interact with RNA, despite the absence of canonical RNA-binding domains

The cytosolic aconitase, also known as iron-regulatory protein 1 (IRP1), is an example of an RBP, which would be missed by this approach, since it binds RNA through an unconventional mechanism without a recognizable canonic RNA-binding domain. Aconitase has both enzymatic (aconitase) and RNA-binding (IRP1) activity, both mediated through overlapping active sites (Walden et al., 2006). Binding to the iron-responsive element (IRE) in mRNA induces an extensive conformational change, which explains the alternate functions of IRP1 as enzyme or mRNA translation regulator, which controls its specific target mRNAs depending on cellular iron levels (Kennedy et al., 1992).

This and other examples of RNA-binding activity of unexpected proteins highlight the need to identify RNA-binding proteins in an unbiased experimental approach.

Where do RNA-binding proteins bind mRNA?

Post-transcriptional and translational regulation of mRNA is mediated by *trans*-acting factors (e.g. RBPs) binding to *cis*-acting regulatory elements (RBP-binding sites on mRNA and pre-mRNA) (Gebauer et al., 2012). These *cis*-acting elements are usually located in the untranslated regions of the mRNA (UTRs), although binding to CDS is also observed. The 3'UTR of an mRNA contains a plethora of cis-regulatory elements, which determine the fate of the mRNA (Gebauer et al., 2012; Kuersten and Goodwin, 2003). The average length of a 3'UTR varies between species and can range from 200 nt (plants and fungi) to 800 nt (humans), but the variation is even more distinct within species (Mignone et al., 2002). Longer 3'UTRs are generally associated with increased potential for post-transcriptional regulation of the mRNA, since they usually contain more *cis*-regulatory elements than their shorter counterparts. 3'UTR length can vary over two

orders of magnitude, ranging from 60 nt to over 4000 nt in human cells (Hesketh, 2005).

RBPs that bind to *cis*-regulatory sequences in the 5'UTR and 3'UTR can regulate the translation and localization of mRNA (Lasko, 2011; Sonenberg and Hinnebusch, 2009). For example, the 3'UTR of a single transcript usually contains several *cis*-acting elements, which can be bound in a combinatorial fashion. The importance of *cis*-regulatory sequences in 3'UTRs is emphasized by several recent studies, linking alternative 3'UTR formation to diverse cellular processes and diseases (Danckwardt et al., 2011; Mayr and Bartel, 2009; Wang et al., 2008). The difference in 3'UTR length is hereby critical for the sensitivity of the mRNA to the combination of RBPs and RNPs in the cellular environment.

The regulation also depends on the concentration and activity of the *trans*-acting factors in the cell. The combination of both allows a context-specific regulation of the transcript by integrating signals from multiple cellular processes (Gebauer et al., 2012). Also, a *trans*-acting factor can regulate the fate of several mRNAs, which contain similar corresponding *cis*-regulatory elements (Kanitz and Gerber, 2010).

MicroRNA-containing ribonucleoprotein complexes (miRNPs) are an example for an RNA-containing trans-acting factor, regulating mRNA stability and translation. Functional miRNA-protein complexes regulate gene expression by translational repression or induced degradation of the target mRNA (Bartel, 2009). A single miRNP can hereby modulate and "fine-tune" the expression of hundreds of proteins (Selbach et al., 2008). In a recent study, several miRNPs (and RBPs) were assayed for RNA targets and binding sites on a transcriptome-wide scale (Hafner et al., 2010). A surprising result was that each of the tested miRNP or RBP bound between 5 - 30% of the more than 20,000 transcripts detectable in HEK293 cells, which suggests that a transcript will on average be bound, and presumably regulated, by a combination of miRNPs and/or RBPs, that ultimately determines the transcript's fate.

It is thus not surprising that defects in miRNA mediated post-transcriptional gene regulation contribute significantly to human diseases (Mendell and Olson, 2012; van Kouwenhove et al., 2011). Notably, RBPs also control the biogenesis of miRNAs, either globally, by associating with key components of the miRNA biogenesis pathway, or for a subset of miRNAs, by recognizing specific miRNA features (Krol et al., 2010; van Kouwenhove et al., 2011).

Identification of RNA binding protein target sites

In order to investigate the protein-mRNA interactome, it is critical to know which targets are bound by RBPs and where the interactions take place. Generally, we can classify methods to identify RBP targets and binding sites as *in vitro* studies and *in vivo* studies. SELEX (Tuerk and Gold, 1990) and RNAcompete (Ray et al., 2009; Ray et al., 2013) are typical examples for *in vitro* selection of RNA fragments that hybridize to the RBP of interest under the chosen conditions. A major drawback of *in vitro* approaches is the artificial environment in which the binding reaction takes place. For example, a RBP might bind a certain sequence motif under the chosen experimental condition, but this interaction cannot take place *in vivo* because either the binding site is not accessible or a high affinity interaction takes place only in combination with other cellular proteins or metabolites. A recent study on transcription factors showed that binding preferences observed in SELEX experiments could be enhanced by binding with a single cofactor (Slattery et al., 2011). It is reasonable to assume that similar mechanisms can influence the results of *in vitro* experiments on RNA-binding proteins as well.

A standard attempt to identify the functional targets of an RBP *in vivo* uses RNA immunoprecipitation, coupled to microarray analysis (RIP-Chip) or high-throughput sequencing (RIP-seq) (Keene et al., 2006). This assay, however, cannot distinguish direct from indirect interaction partners and, depending on the lysis and immunoprecipitation conditions, even artificial protein-RNA interactions might form, leading to false positive interactions (Mili and Steitz, 2004). Another limitation is that RIP-Chip/RIP-seq cannot provide precise information about the binding sites of the RBP.

Current approaches to identify the targets of RBPs and the binding site on RNA use UV-crosslinking to covalently bind proteins to RNA. The combination of *in vivo* UV-crosslinking (*in vivo* here means that the crosslink is introduced in intact cells, and not in cell lysates) and immunoprecipitation (CLIP) allows isolating and analyzing protein-bound RNA fragments, which represent putative protein-RNA interaction sites (Ule et al., 2003). In combination with high-throughput sequencing, the CLIP protocol can identify RBP targets transcriptome-wide (Konig et al., 2010; Licatalosi et al., 2008). However, the method has some drawbacks as well, for example the difficulty in separating specific from non-specific bound RNAs (which are co-precipitated in the IP).

In addition, several proteins are not efficiently crosslinked to RNA with conventional UV_{254nm} irradiation (Ascano et al., 2012; Kishore et al., 2011; Milek et al., 2012)

Those limitations were addressed with the advent of photoactivatable ribonucleoside-enhanced CLIP (PAR-CLIP) (Hafner et al., 2010). The improvement consists in the use of photoreactive thionucleosides, 4-thiouridine (4SU) and 6-thioguanosine (6SG) which are incorporated into newly transcribed RNA. For several tested proteins, photoreactive nucleosides seem to increase the crosslinking efficacy between protein and RNA, but the probably most striking feature are the characteristic and highly specific nucleotide transitions at the exact position of the crosslink (Hafner et al., 2010). If 4SU is used for RNA labeling, T is converted to C (T-to-C transition) and if 6SG is used, G is converted to A (G-to-A transition). Therefore, the nucleotide transitions can be identified in the sequence reads and used to determine the exact crosslink position, which in turn allows binding site identification at nucleotide resolution. In addition the frequency of nucleotide transitions is particularly high at the crosslink positions and can thus be used to separate signal from noise (Hafner et al., 2010).

1.3. RNA-binding proteins and disease

Considering their large number, their role in all aspects of RNA metabolism and their variety of functions, it is not surprising that aberrant RBPs have been implicated in a wide range of human genetic diseases, from neurologic disorders to cancer (Cooper et al., 2009; Darnell, 2010; Lukong et al., 2008).

In several cases, the pathological phenotype of a defective RNA-binding protein is connected to its main biological function. For example, defects in RBPs involved in alternative splicing can give rise to pathologies impairing the functionality of the brain, where alternative splicing is known to play an important role (Lukong et al., 2008). Also, chromosomal rearrangements on genes encoding for RBPs involved in cell growth and proliferation can have a causative role in cancer (Darnell, 2010). As examples of clinical manifestations, Fragile X syndrome is known to derive from the loss of function of the RNA binding protein FMRP (Dictenberg et al., 2008). Myotonic distrophy is associated with the loss of function of MBNL1 and the gain of function of CUG-BP1 (Wheeler and Thornton, 2007), and TLS/FUS translocations are prominent features in sarcomas (Riggi et al., 2007).

In most cases, it is still unclear if the effects are directly or indirectly mediated by the RBP, and if the RNA-binding function is involved in the onset or the progression of the respective disease. A deeper knowledge of the RBP targets and the functional consequences of target binding could help elucidating the disease-causing mechanisms and might provide powerful tools for drug development.

Screening for novel RBPs might assign RNA-binding function to previously known disease related proteins. This could be important from the clinical point of view, especially if we consider that we are approaching the era of personalized medicine based on individual genome/transcriptome profiling (Chen et al., 2012). In this context, we can infer the usefulness of the generation of a comprehensive catalogue of RNA binding proteins (which comprises also RBPs whose RNA binding function could not be predicted), coupled to follow up experiments mapping their interactions with RNA targets.

1.4. Identification and characterization of RNA-binding proteins

Since RBPs are involved in virtually all aspects of PTGR, they have been extensively studied in the last decades; indeed, initial attempts to identify and characterize messenger ribonucleoprotein (mRNP) particles date back to the late 1970s. The first studies to isolate mRNPs and identify their components relied on co-sedimentation of proteins with RNA in sucrose density gradients. (Martin et al., 1978). However, these fractionation-based methods were prone to contamination with non-RNP proteins, which might associate with RNA during the purification procedure, while some genuine RNPs might dissociate and thus be missed in the analysis.

Other methods were therefore developed in the hope to identify pure mRNPs. An elegant approach to identify poly(A)+ RNA binding proteins is the purification of polyadenylated RNA by oligo(dT) chromatography and analysis of co-precipitated proteins (Lindberg and Sundquist, 1974). However, although this approach had some advantages over sedimentation-based methods, the problem of false positives (non-specific proteins in the precipitates) persisted, since proteins might interact with the oligo(dT) chromatography column. The method was therefore constantly refined to identify only proteins, which are in direct contact with mRNA *in vivo*. A breakthrough was hereby the introduction of ultraviolet (UV) light induced photo-crosslinking of

RBPs to RNA *in vivo*. Since UV-crosslinking generates a covalent bond between RBP and RNA, subsequent isolation steps can be performed under stringent protein-denaturing and disulfide-bond reducing conditions, therefore limiting the probability for false positives and contaminants (Dreyfuss et al., 1984a; Dreyfuss et al., 1984b; Mayrand and Pederson, 1981; Mayrand et al., 1981; van Eekelen et al., 1981; Wagenmakers et al., 1980).

The ability of shortwave UV irradiation to crosslink proteins to nucleic acids by covalent bonds was discovered early (Smith, 1976). Greenberg and colleagues performed the first studies using UV-irradiation to facilitate the identification of proteins associated with polyadenylated RNAs *in vitro* (Greenberg, 1979). In this study, polysomal fractions were isolated from cell lysate and *in vitro* UV-irradiated the samples to covalently bind proteins to RNA before isolation by oligo(dT) sepharose chromatography. The conclusion was that most eukaryotic mRNAs in polyribosomes are in close contact with proteins and that RBPs can successfully be crosslinked to RNA and isolated by oligo(dT) precipitation (Greenberg, 1979).

Greenberg and colleagues argued that crosslinking would allow the isolation of mRNA-protein complexes under conditions that minimize non-specific binding of proteins to RNA. They went even further and proposed the ability of a protein to crosslink to RNA as a criterion to identify specific RNA-binding proteins, since the major requirement for photocrosslinking is a close contact between protein and nucleic acid. However, some proteins might be specific RNA-binders, but not efficiently crosslink to RNA, due to steric and chemical properties of the protein-RNA contact site (Meisenheimer and Koch, 1997).

From the earliest studies on RNA-binding proteins described above, followed by landmark studies, such as the characterization of the heterogeneous nuclear ribonucleoparticles (hnRNP) (Choi and Dreyfuss, 1984) and the mRNA polyadenylate-binding protein (Adam et al., 1986), great progress has been made and several studies in the last decade aimed to identify most, if not all, proteins that interact with mRNA and generate a catalog of RNA-binding proteins. Tsvetanova and colleagues used a dual approach to identify RNA-binding proteins in *S. cerevisiae*. The first method was based on oligo(dT) precipitation of RNPs from cell lysates and the second used protein arrays,

which contained about 80% of the *S. cerevisiae* proteome, and were hybridized with an yeast RNA pool.

However, the study had some flaws in the experimental design. For example, oligo(dT)-poly(A)$^+$RNA hybridization and subsequent washing steps were not performed under protein-denaturing conditions, increasing the risk to identify false positive interactors. False positives can also occur in both approaches if protein-RNA interactions are identified, which are not present in the cellular context. The same is true for the protein array, where RNAs might bind to certain proteins, which would be not accessible in the cellular context.

Nevertheless, using this approach confirmed previously annotated RBPs and a common set of 35 novel putative RBPs were identified by either one or both of the applied methods, amongst them several metabolic enzymes (Tsvetanova et al., 2010).

A slightly different approach by Scherrer and colleagues probed high-density protein microarrays with fluorescently labeled total RNA or mRNA and identified around 200 proteins that reproducibly interact with different types of RNA in *S. cerevisiae* (Scherrer et al., 2010). The same pitfalls mentioned above also apply to this study and most other *in vitro* studies.

In a similar approach, Butter and colleagues tried to identify RBPs that interact with a particular RNA *in vitro* by quantitative mass spectrometry (Butter et al., 2009). However, the experimental setup did not allow distinguishing direct from indirect protein-mRNA interactions. In addition, re-association of RNA-binding proteins might occur after cell lysis and during the purification protocol (Mili and Steitz, 2004) and thus bona fide RBPs cannot be distinguished from false-positive interactors. This is especially problematic if proteins, which would have no possibility to interact in the native cellular context, re-associate due to spatial or temporal restrictions, essentially facing the same problems as in the studies mentioned above.

Despite those drawbacks, these studies succeeded in identifying new RNA-binding proteins and future experiments will show how many of the identified proteins present truly novel RBPs. In summary, at the time of our study, the attempts to identify RBPs on a global scale were mainly performed in yeast and without adequate means to distinguish false positives from true interactors and direct from indirect interactions. In addition, yeast is in many respects a relatively simple organism compared to human cells, and thus fails to reflect the whole complexity of post-transcriptional gene

regulation in higher eukaryotes. For example, most genes in human cells are alternatively spliced, which dramatically increases the transcriptome and proteome diversity in the cell (Johnson et al., 2003; Matlin et al., 2005; Modrek et al., 2001). In contrast, in *S. cerevisiae* only 5 % of all genes contain introns and almost all of those genes contain only a single intron (Nash et al., 2007).

In order to cover the entity of post-transcriptional gene regulation of higher eukaryotes we pursued a comprehensive experimental definition of the mRNA-bound proteome of a human cell line.

1.5. Identification of RNA-binding proteins in *D. melanogaster* embryos

In early *D. melanogaster* development most of the embryonic gene regulation seems to be mediated by transcriptional regulation. This was demonstrated by impressive mutagenesis screens performed in the late 1970s by Eric Wieschaus and Christiane Nuesslein-Volhard (Nusslein-Volhard and Wieschaus, 1980). After the genes underlying these mutations were cloned, mostly by the labs of Walter Gehring and Herbert Jackle, it was found that more than 90% encoded for transcription factors. However, many crucial developmental processes are driven by post-transcriptional regulation, like the establishment of embryo polarity by mRNA localization and local translation regulation (Lecuyer et al., 2007; Martin and Ephrussi, 2009).

In *D. melanogaster*, the egg chamber is composed of an oocyte and 15 sister nurse cells surrounded by a somatic follicular epithelium. Nurse cells synthesize maternal RNAs, proteins and organelles that are required for the growth and development of the oocyte during oogenesis (Becalska and Gavis, 2009). The supply of maternal factors by the nurse cells is critical, since the early embryo is essentially transcriptionally quiescent and proteins are produced from maternally deposited mRNAs (maternal effect genes) before the onset of zygotic transcription during maternal-to-zygotic transition (MZT) (Figure 3). The expression of the protein products before MZT is thus mainly regulated at the post-transcriptional level (Richter and Lasko, 2011).

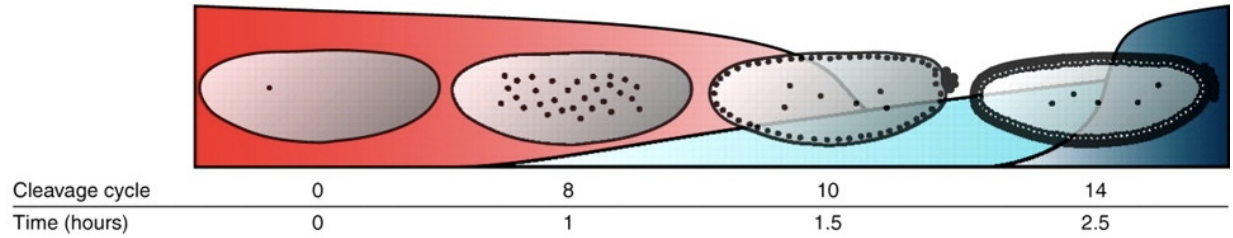


Figure 3: MZT transcription in *D.melanogaster* embryos. The embryonic stages are depicted schematically and the corresponding cleavage cycle and time after fertilization are noted below. Red curves represent the degradation profiles of maternal effect genes. The blue curves illustrate the minor (light blue) and major (dark blue) waves of zygotic genome activation. The last embryonic stage presented here is the developmental time point at which major zygotic transcription is required. Modified from (Tadros and Lipshitz, 2009), with permission.

One surprising example is the demonstration that the transcription factor Bcd represses translation of caudal mRNA *in vitro* by binding to a Bcd response element (BRE) in the caudal mRNA 3'UTR (Rivera-Pomar et al., 1996). Bcd binding to the 3' end of caudal RNA blocks translation initiation at the 5' end, which creates an asymmetric distribution of Caudal (Cad) protein (Niessing et al., 2002), a transcription factor and master regulator of posterior patterning in the early embryo (Olesnicky et al., 2006). In fact, the mechanism of translation inhibition was recently shown to depend on d4EHP, an eIF4E related cap-binding protein, which interacts with the mRNA cap structure and Bcd protein that binds the BRE in the cad-mRNA (Cho et al., 2005). The role of Bcd-protein as a translation regulator might be more widespread than previously believed, meaning that Bcd might interact not only with caudal mRNA but also with (many) other maternal mRNAs and that other proteins bearing similar domains might be involved in both transcriptional and translational regulation as well.

Several other mechanisms of translational regulation were discovered in the *D. melanogaster* embryo, like the regulation of maternal hunchback (hb) mRNA by Nanos (Nos), Pumilio (Pum), and Brain tumor (Brat) proteins (Cho et al., 2006) or the activation of specific germ line mRNAs by Vasa (Vas) (Liu et al., 2009).

In addition to translational regulation, mRNA localization is believed to contribute substantially to embryo patterning. Indeed, a recent study showed that 71% of about 3000 examined mRNAs in the *D. melanogaster* embryo were asymmetrically distributed in a large number of divers patterns (Lecuyer et al., 2007).

However, in those very early steps of embryogenesis (and also for subsequent developmental stages), we still lack global knowledge about which proteins interact

with mRNA, where they bind their targets and how those interactions ultimately mediate post-transcriptional gene regulation.

2. Aims of this Thesis

2.1. The mRNA bound proteome in HEK293 cells

The main focus of this study was to purify stringently the mRNA-bound proteome and to comprehensively identify all proteins that interact with mRNA in a human cell.

RNA-binding proteins mediate and regulate virtually all aspects of post-transcriptional and translational gene regulation. However, we still lack a complete picture of the proteins that interact with RNA in a human cell.Thus, we aimed to define the mRNA-bound proteome of a human cell *in vivo*. Our method freezes mRNA-protein interactions as they occur in the native cellular context by irradiating live cells with UV-light, which introduces a covalent bond between protein and RNA. We like to describe our method in analogy to taking a "snapshot" of the interactions between RBP and RNA. The covalent crosslink between RBP and RNA permits subsequent purification under protein denaturing conditions and should therefore avoid contaminant proteins, which facilitates the detection of true interaction partners.

We developed this oligo(dT) affinity purification approach in human embryonic kidney (HEK293) cells. We chose HEK293 cells since the photoreactive nucleoside labeling and UV-crosslinking protocol was already established for PAR-CLIP experiments (Hafner et al., 2010). Available PAR-CLIP data can therefore be used to define positive controls. In addition, HEK293 cells are fast growing and easy to handle, which is a great advantage for large-scale proteomics studies. The identified proteins will be validated and analyzed for RNA-binding targets, binding sites and binding preferences by PAR-CLIP.

We think that this study will broaden our current knowledge about the plethora of RBPs and how they regulate gene expression at the post-transcriptional level.

2.2. The mRNA bound proteome in *D. melanogaster* embryos

During the first steps of embryogenesis, post-transcriptional regulation is crucial for the patterning of the early embryo as shown by studies on localized regulation of translation. However, only a few examples of these regulatory events have been identified, but we hypothesize that this mechanism is more widespread than commonly believed and that many more proteins could be involved in similar processes.

We thus set out to identify the mRNA-bound proteome in early *D. melanogaster* embryogenesis, based on the same method we developed in human cells. This would allow for the first description of an mRNA-bound proteome of a precisely defined developmental stage, namely the early *D. melanogaster* embryogenesis before maternal-to-zygotic transition.

In addition, comparing the mRNA-bound proteomes of fruit flies and humans might highlight evolutionarily conserved mechanisms and shed light on those that are unique to a particular organism.

To address those questions, we intended to develop an *in vivo* UV-crosslinking protocol, which allows to covalently crosslink proteins to mRNA in *D. melanogaster* embryos. The covalent crosslink between proteins and RNA should allow the identification of the mRNA-bound proteome by oligo(dT)–based precipitation of polyadenylated RNAs. In addition, the effective *in vivo* UV-crosslinking protocol is a first step towards PAR-CLIP in *D. melanogaster* embryos, which would enable to identify RBP targets and binding sites transcriptome-wide and with high resolution during a defined developmental stage.

3. Materials and Methods

3.1. Oligonucleotides

Small RNA cloning adapters

5’ adapter

rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC

3’ barcoded adapters (bar-code is underlined)

NBC1: AppTCTAAAATCGTATGCCGTCTTCTGCTTG-InvdT

NBC2: AppTCTCCCATCGTATGCCGTCTTCTGCTTG-InvdT

NBC3: AppTCTGGGATCGTATGCCGTCTTCTGCTTG-InvdT

NBC4: AppTCTTTTATCGTATGCCGTCTTCTGCTTG-InvdT

NBC5: AppTCTCACGTCGTATGCCGTCTTCTGCTTG-InvdT

NBC6: AppTCTCCATTCGTATGCCGTCTTCTGCTTG-InvdT

NBC7: AppTCTCGTATCGTATGCCGTCTTCTGCTTG-InvdT

NBC8: AppTCTCTGCTCGTATGCCGTCTTCTGCTTG-InvdT

cDNA amplification for recombinant cell lines (restriction sites are underlined)

ALKBH5:

5’-TTCAGTCGACATGGCGGCCGCCAGCGGCTACACGGACCTGCGTGAGAAG;

5’- CTATTGATGCCAACAGCCTTTCCATC,

PGK1:

5’-ATGTCGCTTTCTAACAAGCTGACGCTG;

5‘-ATAAGAATGCGGCCGCCTAAATATTGCTGAGAGCATCCACCCCAG,

qRT-PCR primers

RNU61: 5'-GTGCTCGCTTCGGCAGC; 5'-TGGAACGCTTCACGAATTTGC

GAPDH: 5'-AGCCACATCGCTCAGACAC; 5'-GCCCAATACGACCAAATCC,

DNAJC4: 5’-ACAAACAAAACAAACAAGTGCTG; 5’-GCATCTGCTTCACCTTCCTG

DDX1: 5’-AACACAAGACCTGGTGCTAATAGTC; 5’-TGCTTGATCCATCTTATGTTCC,

EEF2: 5’-5'-GTCGGGAGAGCATATCATCG; 5'-CGACCGGGTCAGATTTCTT,

RIP/RT-PCR primer

C22orf28: 5'-TCAAGACTATCTGAAGGGAATGG; 5'-CAGGGGTTGTGTTGAAGACC
CAPRIN1: 5'-GCTAGAGGCTTGATGAATGGA; 5'-GAAGGGCGGTAACCATCATA
GPI: 5'-CATCAACAGCTTTGACCAGTG; 5'-GCCATCAAGCTCAGGCTCTA
MACF1: 5'-CCGATTGCATCACAACCAT; 5'-TTAGCCCATGTCAGGACCTC
MSH6: 5'-GCTGTGCGCCTAGGACAT; 5'-CCCTTAATGAATTTATAGAGGAACGTA
RPL22: 5'-AAATTGTGCCCTGCGAGTT; 5'-ATGGGAGCCAAGGTAGGACT
XBP1: 5'-AAACAGAGTAGCAGCTCAGACTGC; 5'-TCCTTCTGGGTAGACCTCTGGGAG

3.2. Antibodies

anti-HA.11 (COVANCE, 16B12)
anti-FLAG (SIGMA, F1804)
anti-HNRNPK (EPITOMICS, EP943Y)
anti-mouse immunoglobulins (DAKO, P0447)
anti-rabbit immunoglobulins (DAKO, P0448)
anti-Aubergine (*D. melanogaster*) antibody (source: Rabbit) was kindly provided by the laboratory of Ruth Lehmann (NYU Langome medical Center, NYC, USA).

3.3. HEK293 cell line culture conditions and transfections

Flp-In T-Rex 293 cells (Invitrogen) were grown in D-MEM high glucose with 10% (v/v) fetal bovine serum, 1% (v/v) 2 mM L-glutamine, 1% (v/v) 10,000 U/ml penicillin/10,000 µg/ml streptomycin, 100 µg/ml zeocin and 15 µg/ml blasticidin. Cell lines stably expressing His/FLAG/HA-tagged proteins were generated by co-transfection of pFRT/TO/His/FLAG/HA constructs with pOG44 (Invitrogen) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's recommendations.
Cells were selected by exchanging zeocin with 100 µg/ml hygromycin. Expression of epitope-tagged proteins was induced by addition of 200 – 1000 ng/ml doxycycline 15 to 20 h before crosslinking. The expression of His/FLAG/ was assessed by Western analysis using mouse anti-HA.11 monoclonal antibody (Covance).
For quantitative proteomics, cells were grown in SILAC medium as described in (Ong et al., 2002). Briefly, Dulbecco's Modified Eagle's Medium (DMEM) Glutamax lacking arginine and lysine (a custom preparation from Gibco) supplemented with 10% dialyzed fetal calf serum (dFCS, Gibco) was used. Heavy (H) and light (L) SILAC media

were prepared by adding 84 mg/l $^{13}C_6$ $^{15}N_4$ L-arginine plus 146 mg/l $^{13}C_6$ $^{15}N_2$ L- lysine or the corresponding non-labeled amino acids (Sigma), respectively. Labeled amino acids were purchased from Sigma Isotec.

3.4. Plasmids

pDONR vectors were largely obtained from the ORFeome library. pENTR constructs were generated by PCR amplification of the respective coding sequences (CDS) from Flp-In 293 T-REx (Invitrogen) cDNA, followed by restriction digest and ligation into pENTR4 (Invitrogen). pDONR and pENTR vectors carrying CDS were recombined into pFRT/TO/His/FLAG/HA-DEST destination vector (Invitrogen) using GATEWAY LR recombinase (Invitrogen) according to manufacturer's protocol to allow for doxycycline- inducible expression of stably transfected His/FLAG/HA-tagged protein in Flp-In 293 T- REx (Invitrogen) from the inducible TO/CMV-promoter.

3.5. Isolation of mRNA-interacting proteins in HEK293 cells

HEK293 cells were grown for 16 hr in SILAC medium supplemented with 4SU and 6-thioguanosine at a final concentration of 200 µM each. An additional labeling pulse with 100 µM of each photoreactive nucleoside was applied 2 h prior to UV-irradiation to ensure the labeling of short-lived transcripts. Living cells, grown in light SILAC medium, were irradiated with 365 nm UV light (0.2 J/cm2), while control cells grown in heavy SILAC medium were not UV-crosslinked. In a label swap experiment, cells grown in heavy SILAC medium were crosslinked and cells grown in light SILAC medium were used as non-crosslinked control. Following crosslinking, cells were harvested by scraping them off with a rubber policeman, washed with ice-cold PBS and collected by centrifugation (2000 RCF, 4°C, 10 min). Resulting cell pellets were lysed in 10 pellet volumes of lysis/binding buffer (100 mM Tris HCl, pH 7.5, 500 mM LiCl, 10 mM EDTA pH 8.0, 1% (w/v) lithium-dodecylsulfate, 5 mM DTT, Complete Mini EDTA-free protease inhibitor (Roche)). For a typical proteomics experiment 300 µl oligo(dT) Dynabeads (bed volume) were added to cell extract prepared from thirty 15 cm tissue culture plates and incubated for 1 h at room temperature on a rotating wheel. After incubation supernatants were saved for multiple rounds of oligo(dT)-precipitation. Beads were washed 2 times in 150 ml lysis/binding buffer for 30 min at room temperature, followed by 3 washing steps in 150 ml NP40 washing buffer (50 mM Tris HCl, pH 7.5, 140 mM LiCl, 2 mM EDTA pH 8.0, 0.5 % NP40, 0.5 mM DTT). Protein-mRNA complexes

were heat eluted from beads in low-salt elution buffer (10 mM Tris HCl, pH 7.5) for 2 min at 80°C. For subsequent mass spectrometry analysis, RNA was digested by incubation with RNase I (10 U/ml) and benzonase (125 U/ml) for 3 h at 37° in elution buffer containing 1mM MgCl2. After nuclease treatment, proteins were precipitated with trichloroacetic acid, washed with acetone and denatured in SDS- PAGE loading buffer before separation on a NuPAGE Novex 4 to 12% gradient gel (Invitrogen) followed by in-gel trypsin-digest.

3.6. Mass Spectrometry Analysis of Precipitated Proteins (HEK293)

In-gel digestion of oligo(dT) precipitated proteins

Preparations of oligo(dT)-precipitated protein-RNA complexes for mass spectrometry analysis using in-gel digestion mRNA-bound proteins were separated on a NuPAGE Novex 4 to 12% gradient gel (Invitrogen) under reducing conditions. Proteins were fixed in fixative solution (50% methanol (v/v), 10% acetic acid (w/v)) and stained afterwards with the Colloidal Blue staining Kit (Invitrogen). Gel lanes were cut into 12 gel slices, which were individually subjected to reduction, alkylation and in-gel digestion with sequence grade modified trypsin (Promega) according to a standard protocol (Shevchenko et al., 2006). After in-gel digestion peptides were extracted and desalted using StageTips (Rappsilber et al., 2007) prior to analysis by mass spectrometry.

HPLC and mass spectrometry

Reversed-phase liquid chromatography (rpHPLC) was performed employing a Eksigent NanoLC – 1D Plus system using self-made fritless C18 microcolumns (Ishihama et al., 2002) (75 µm ID packed with ReproSil-Pur C18-AQ 3-µm resin, Dr. Maisch GmbH) connected on-line to the electrospray ion source (Proxeon) of an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher). Peptide samples were loaded onto the column with a flow rate of 250 nl/min followed by sample elution at a flow rate of 200 nl/min with a 10 to 60 % acetonitrile gradient over 6 h in 0.5% acetic acid. The LTQ-Orbitrap Velos instrument was operated in the data dependent mode (DDA) with a full scan in the Orbitrap followed by up to 20 consecutive MS/MS scans in the LTQ. Precursor ion scans (m/z 300–1700) were acquired in the Orbitrap part of the instrument (resolution R = 60,000; target value of 1 x 106), while in parallel the 20 most intense ions were

isolated (target value of 3,000; monoisotopic precursor selection enabled) and fragmented in the LTQ part of the instrument by collision induced dissociation (CID; normalized collision energy 35 %; wideband activation enabled). Ions with an unassigned charge state and singly charged ions were rejected. Former target ions selected for MS/MS were dynamically excluded for 60 s. Total cycle time for one full scan plus up to 20 MS/MS scans was approximately 2 s.

Processing of mass spectrometry data

Identification and quantification of proteins was carried out with the MaxQuant software package (Cox and Mann, 2008). In essence, the Quant.exe module extracts, re-calibrates and quantifies isotope clusters and SILAC doublets in the raw data files (medium labels: Arg6 and Lys4; heavy labels: Arg10 and Lys8; maximum of three labeled amino acids per peptide; polymer detection enabled; top 6 MS/MS peaks per 100 Da). Generated peak lists (msm-files) were submitted to a MASCOT search engine (version 2.2, MatrixScience) and searched against the IPI human database (v. 3.72) supplemented with common contaminants (e.g. trypsin, BSA). The database was modified in-house to obtain a concatenated target-decoy database as described previously (Elias and Gygi, 2007). Full tryptic specificity was required and a maximum of two missed cleavages and a mass tolerance of 0.5 Da for fragment ions applied. Oxidation of methionine and acetylation of the protein N-terminus were used as variable modifications, carbamidomethylation of cysteine as a fixed modification. Filtering of putative MASCOT peptide identifications, assembly of protein groups and re-quantification was performed with Identify.exe. A minimum peptide length of 6 amino acids was required. False discovery rates were estimated based on matches to reversed sequences in the concatenated target-decoy database. A maximum false discovery rate of 1% at both the peptide and the protein level was allowed. Protein ratios were calculated from the median of all normalized peptide ratios using only unique peptides or peptides assigned to respective protein groups with the highest number of peptides ("Occam's razor" peptides). Only protein groups with at least three SILAC counts (peptide ratios) were kept for further analysis.

SILAC proteomics data analysis

Fold changes were computed by MaxQuant (Cox and Mann, 2008) for proteins and protein groups in case of ambiguities. We considered only fold changes that were supported by at least three measured peptide ratios in a single experiment or three measured peptide ratios over all three mass spectrometry experiments (L1, L2 and H1). The quantified protein groups were associated with NCBI Reference Sequence (Refseq) protein IDs by BLASTing the leading protein of the protein group against the human protein database.

Intensity-based absolute quantification (iBAQ) of proteins

The MaxQuant software computes protein intensities as the sum of all identified peptide intensities (maximum detector peak intensities of the peptide elution profile, including all peaks in the isotope cluster). Protein intensities were divided by the number of theoretically observable peptides (calculated by in silico protein digestion with a Perl script, all fully tryptic peptides between 6 and 30 amino acids were counted while missed cleavages were neglected). iBAQ intensities correlate well with absolute protein abundance and can therefore be used for comparison of protein levels within the experiment (Schwanhausser et al., 2011).

3.7. RNA-binding protein validation assays and PAR-CLIP

Cells were grown in medium supplemented with 100µM 4SU for 16h prior to harvest.

UV 365 nm crosslinking

For UV-crosslinking, the growth medium was removed completely while cells were still attached to the plates. Cells were irradiated on ice with 365 nm UV light (0.2 J/cm2) in a Stratalinker 2400 (Stratagene), equipped with light bulbs for the appropriate wavelength. Cells were scraped off with a rubber policeman in 2 ml PBS per plate and collected by centrifugation at 800 × g for 4 min at 4°C.

Cell lysis and first partial RNase T1 digestion

The pellets of cells crosslinked with UV 365 nm were resuspended in 3 cell pellet volumes of NP40 lysis buffer (50 mM Tris HCl, pH 7.5, 140 mM LiCl, 2 mM EDTA, pH 8.0, 0.5% (v/v) NP40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)) and incubated on ice for 10 min. The typical scale of such an experiment was 3 ml of cell

pellet. The cell lysate was cleared by centrifugation at 13,000 × g. RNase T1 (Fermentas) was added to the cleared cell lysates to a final concentration of 1 U/µl and the reaction mixture was incubated in a water bath at 22°C for 10 min and subsequently cooled for 5 min on ice before addition of antibody-conjugated magnetic beads.

Preparation of Dynabeads Protein G magnetic beads

10 µl of Dynabeads Protein G magnetic particles (Invitrogen) per ml cell lysate were washed twice with 1 ml of citrate-phosphate buffer (4.7 g/l citric acid, 9.2 g/l Na2HPO4, pH 5.0) and resuspended in twice the volume of citrate-phosphate buffer relative to the original volume of bead suspension. 0.25 µg of anti-FLAG M2 monoclonal antibody (Sigma) per ml suspension was added and incubated at room temperature for 40 min. Beads were then washed twice with 1 ml of citrate-phosphate buffer to remove unbound antibody and resuspended again in twice the volume of citrate-phosphate buffer relative to the original volume of bead suspension.

Preparation of ANTI-FLAG M2 Magnetic Beads

20 µl of ANTI-FLAG M2 magnetic beads (Sigma-Aldrich) per ml cell lysate were washed twice with 1 ml of citrate-phosphate buffer and resuspended in one original volume of citrate-phosphate buffer.

Immunoprecipitation, second RNase T1 digestion and dephosphorylation

10 µl antibody-conjugated Protein G magnetic beads or 20 µl of ANTI-FLAG M2 magnetic beads were added per ml of partial RNase T1 treated cell lysate. Incubation was performed in 15 ml tubes on a rotating wheel for 1 hr at 4°C. Magnetic beads were collected on a magnetic particle collector (Invitrogen). Manipulations of the following steps were carried out in 1.5 ml microfuge tubes. The supernatant was removed from the bead-bound material. Beads were washed 2 times with 1 ml of IP wash buffer (50 mM HEPES-KOH, pH 7.5, 300 mM KCl, 0.05% (v/v) NP40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)) and resuspended in one volume of IP wash buffer. RNase T1 (Fermentas) was added to obtain a final concentration of 50 U/µl, and the bead suspension was incubated at 22°C for 8 min, and subsequently cooled for 5 min on ice. Beads were washed 3 times with 1 ml of high-salt wash buffer (50 mM HEPES-KOH, pH 7.5, 500 mM KCl, 0.05% (v/v) NP40, 0.5 mM DTT, complete EDTA- free

protease inhibitor cocktail (Roche)) and resuspended in two bead volumes of dephosphorylation buffer (50 mM Tris-HCl, pH 7.9, 100 mM NaCl, 10 mM MgCl2, 1 mM DTT). Calf Intestinal Alkaline Phosphatase (CIP) was added to obtain a final concentration of 0.5 U/µl, and the suspension was incubated for 60 min at 37°C. Beads were washed twice with 1 ml of phosphatase wash buffer (50 mM Tris-HCl, pH 7.5, 20 mM EGTA, 0.5% (v/v) NP40) and twice with 1 ml of Polynucleotide Kinase (PNK) Buffer (50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 10 mM MgCl2, 5 mM DTT). Beads were resuspended in one original bead volume of PNK buffer.

Radiolabeling of RNA segments crosslinked to immunoprecipitated proteins

To the bead suspension described above, γ-32P-ATP was added to a final concentration of 0.25 µCi/µl and T4 Polynucleotide Kinase (PNK) to 1 U/µl in one original bead volume. The suspension was incubated for 30 min at 37°C. Thereafter, nonradioactive ATP was added to obtain a final concentration of 100 µM and the incubation was continued for another 5 min at 37°C. The magnetic beads were then washed 5 times with 800 µl of PNK Buffer and resuspended in 20 µl of SDS-PAGE Loading Buffer (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).

RNase and DNase digestion assay (only performed for EDF1 nucleic acid binding test)

Immunoprecipitation of epitope-tagged protein of interest was performed as described in the above section and radioactively labeled at the 5'end. Following radiolabeling, the beads were washed twice with PNK buffer and resuspended in PNK buffer. The sample was divided in three aliquots and incubated at 37°C for 30 min with either RNase I (0.1 U/µl) or DNase I (0.1 U/µl). A control sample was incubated at 37°C without addition of Nucleases. After incubation, the beads were washed 5 times with 800 µl of PNK Buffer and resuspended in 20 µl of SDS-PAGE Loading Buffer.

SDS-PAGE and Western Blotting

Beads were taken up in SDS-PAGE loading buffer and boiled 5 min at 95°C. The supernatant was loaded on a 4-12 % NuPAGE Bis-Tris gradient gel (Invitrogen) and exposed to a phosphorimager screen for detection of radioactive signal. The gel was subsequently transferred to a nirocellulose membrane and protein-RNA complexes

were immunodetected by incubation with anti-HA.11 antibody. After incubation with HRP- conjugated secondary anti-mouse antibody, the tagged protein was visualized using ECL Western Blot detection reagent (GE-Healthcare).

Electroelution of crosslinked RNA-protein complexes from gel slices

The radioactively labeled RNA-protein complex migrating at the expected molecular weight of the target protein was excised from the gel and electroeluted in 800 µl SDS running buffer using D-Tube Dialyzer Midi (Novagen) according to the instructions of the manufacturer.

Alternatively, the RNA-protein complex was blotted onto a nitrocellulose membrane and the complex of the expected molecular weight was excised from the membrane and protein-bound RNA was eluted by Proteinase K digestion.

Proteinase K digestion

An equal volume of 2x Proteinase K Buffer (100 mM Tris-HCl, pH 7.5, 150 mM NaCl, 12.5 mM EDTA, 2% (w/v) SDS) with respect to the electroeluate was added, followed by the addition of Proteinase K (Roche) to a final concentration of 1.2 mg/ml, and incubation for 30 min at 55°C. The RNA was recovered by acidic phenol/chloroform extraction followed by chloroform extraction and ethanol precipitation. The pellet was dissolved in 6 µl water.

cDNA library preparation and deep sequencing

The recovered RNA was prepared for sequencing using a cDNA library preparation protocol, originally described for cloning of small RNAs (Dolken et al., 2008; Hafner et al., 2008). In contrast to the published cloning procedure the 3'adapter ligation was carried out with a barcoded 3'adapter (see oligonucleotides) in a 20 µl reaction volume using 6 µl of the recovered RNA. PAR-CLIP libraries were sequenced on Illumina Genome Analyzer GAII and HiSeq platform using the 1X50BP single read protocol.

3.8. PAR-CLIP computational analysis

Illumina PAR-CLIP cDNA sequencing reads were aligned to the human genome assembly (hg18), allowing for up to one mismatch, insertion or deletion. Only uniquely mapping reads were retained. We identified clusters of aligned PAR-CLIP reads

continuously covering regions of pre-mRNA sequence based on the condition that a sequence cluster requires sequence coverage from both duplicate libraries PAR-CLIP libraries for each protein, whereas a read with T-to-C transition is only needed from one of the two libraries (consensus assumption) or from both libraries (conservative assumption). The number of T-to-C mismatches served as a crosslink score. We also assigned a quality score based on the number and positions of distinct reads contributing to the cluster.

As the reads should originate from protein-bound transcripts we regarded clusters aligning antisense to the annotated direction of transcription as false positives. We were thus able to select cutoffs on both scores such as to keep the estimated false positive rate below 5%. After filtering by these cutoffs we expect each remaining cluster to harbor at least one binding site (Lebedeva et al., 2011).

7-mer enrichment analysis around crosslink site

A search for enriched 7-mers over the background of all possible 7-mers was performed in a 41nt-window around the IFIT5-crosslink site using an in-house pearl script.

Consensus Motif search

We used the MEME suite (Bailey et al., 2009) to search for enriched motifs in the top 500 clusters. Parameters: minimum motif width of 6 nt and a maximum width of 75 nt.

IFIT5 cluster distribution over transcript regions

The position of the preferred crosslinking sites (as determined by T-to-C transition frequency) was determined within the Refseq transcript compartments (CDS or UTR) from the UCSC genome browser. The relative position of the site was noted using the bedtools software suite. Zero is hereby the most anterior part of the respective compartment and 1 indicates the position at the very end. The density of the resulting positioning scores were calculated with an in house R-script and plotted for the CDS and the 3'UTR.

3.9. RIP and RT-PCR

Cells stably expressing the FLAG epitope-tagged protein of interest were harvested, washed in ice-cold PBS and collected by centrifugation (2000 RCF, 4°C, 10 min).

Resulting cell pellets were resuspended in 3 volumes of NP40 lysis buffer (50 mM HEPES-KOH at pH 7.4, 150 mM KCl, 2 mM EDTA, 0.5% (v/v) NP40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail) and incubated on ice for 10 min. Lysates were cleared by centrifugation (16,000 RCF, 4°C, 15 min). 1/33 of the total volume was mixed with 3 volumes of TRIzol and 0.2 volumes of chloroform to extract total cellular RNA. Phases were separated by centrifugation (16,000 RCF, 4°C, 10 min.) and RNA was ethanol precipitated. The remaining cleared extract was incubated with FLAG-conjugated Protein G Dynabeads (Invitrogen) or Protein G Dynabeads only and incubated 1 h at 4°C. Beads were washed 3 times with IP wash buffer (50 mM HEPES-KOH at pH 7.4, 150 mM KCl, 2 mM EDTA, 0.05% (v/v) NP40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail), resuspended in one original volume of Proteinase K solution (200 mM Tris-HCl at pH 7.5, 300 mM NaCl, 25 mM EDTA, 2% (w/v) SDS, 0.6 mg/ml Proteinase K) and incubated 20 min at 65°C. RNA was phenol chloroform extracted and ethanol precipitated. The resulting pellet was dried at room temperature and resuspended in H_2O.

Single-stranded cDNA was synthesized from total RNA using an 18nt oligo-dT primer and Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. After reverse transcription, the target transcripts of the protein of interest were amplified by PCR, using a set of intron-spanning PCR primers (see Oligonucleotides) that amplified a 100-150 nt sequence, which was subsequently analyzed by agarose gel electrophoresis.

3.10. Quantitative real-time PCR

Single stranded cDNA was synthesized as described above and subjected to quantitative real-time PCR using Power SYBR Green PCR master mix (Applied Biosystem) and the StepOne Real-Time PCR System (Applied Biosystem) following the manufacturer's instructions.

3.11. Identification of mRNA-crosslinked proteins by Western analysis

Cell lines stably expressing the protein of interest were induced with doxycycline and grown in the presence of 4SU and 6SG as described above. Crosslinking, cell lysis and oligo(dT)-precipitations were performed as described above. Input, supernatant after precipitation and the oligo(dT)-purified material were treated with RNase before TCA-

precipitation and the protein was analyzed on a 4-12 % NuPAGE Bis-Tris gradient gel (Invitrogen). After Western analysis, the nitrocellulose membrane was incubated either with anti-HA.11 antibody (for FLAG/HA or HIS/FLAG/HA tagged proteins) or with an antibody against the endogenous protein. HRP-conjugated secondary antibodies were used and the proteins were visualized using the ECL Western blot detection reagent (GE-Healthcare)

3.12. Sequence analysis of oligo(dT)-purified RNA

Standard mRNA purification (mRNA-seq)

HEK293 cells grown in SILAC medium were harvested as described previously and the cell pellet was immediately resuspended in TRIzol reagent (Invitrogen) and homogenized, according to the manufacturers protocol. One volume chloroform was added to five volumes TRIzol solution, vigorously mixed and incubated at room temperature. After centrifugation (13,000 g, 5 min, 4°C) the aqueous phase was transferred to a clean RNase-free tube and 1 volume phenol/chloroform/isoamyl alcohol (25/24/1, v/v) was added. The sample was mixed vigorously, incubated 5 min at room temperature and centrifuged at 13,000 g (5 min, 4°C). The aqueous phase was mixed with 1 volume isopropanol and precipitated on ice for 1 h. After centrifugation (13,000 g, 30 min, 4°C) the pellet was washed with 80 % (v/v) ethanol. The pellet was dried at room temperature and resuspended in nuclease-free water. Poly(A)+ RNA was purified from total RNA by two rounds of precipitation with oligo(dT) beads according to the manufacturer's instructions and resuspended in nuclease-free water.

RNA oligo(dT) precipitation from 4SU and 6SG labeled non-irradiated cells ("4SU +6SG no UV")

We isolated mRNA from HEK293 cells grown in SILAC medium supplemented with 4SU and 6SG by oligo(dT) precipitation as described for the isolation of mRNA-bound proteins, but without UV-irradiation. The isolated mRNA was ethanol precipitated, washed and resuspended in nuclease-free water.

Purification of 4SU and 6SG labeled RNA ("4SU + 6SG purified RNA") by biotinylation

mRNA was isolated by direct oligo(dT)-precipitation from HEK293 cell lysates grown on SILAC medium supplemented with 4SU and 6SG, without UV-crosslinking as

described before. mRNA was ethanol precipitated to remove traces of DTT before biotinylation. Biotinylation and pull-down of labeled RNA using streptavidin-conjugated beads was performed as described previously (Dolken et al., 2008).

RNA oligo(dT) precipitation from 4SU and 6SG labeled UV-irradiated cells ("4SU +6SG UV")

mRNA was isolated from HEK293 cells grown in SILAC medium supplemented with 4SU and 6SG by oligo(dT) precipitation of crosslinked protein-RNA complexes as described for the isolation of mRNA-bound proteins. After elution from oligo(dT) beads, protein-RNA complexes were Proteinase K digested in Proteinase K reaction buffer (800 mM Guanidinium-HCl, 50 mM EDTA, 5% Tween 10, 0.5% Triton-X 100) for 3 h at 55°C with a final Proteinase K concentration of 2 mg/ml. The RNA was recovered by acidic phenol/chloroform extraction, ethanol precipitation and resuspended in nuclease-free water.

cDNA library preparation for transcriptome sequencing

RNA obtained from the four different RNA isolation conditions described above was analyzed by next-generation sequencing. cDNA libraries were prepared from the recovered RNA, following the mRNA sequencing protocol provided by Illumina. Briefly, poly(A)+RNA was fragmented using 5x fragmentation buffer (200mM Tris-acetate, pH 8.1, 500 mM potassium-acetate, 150 mM magnesium-acetate) by heating at 94°C for 3.5 min. After ethanol precipitation, first- and second-strand cDNA synthesis was performed with random hexamer primers. cDNA fragments were end-repaired using T4 polymerase, T4 PNK and Klenow DNA polymerase. A protruding "A" base was added to the 3' end of the DNA fragments to enable ligation of Illumina adapters with "T" overhangs. After adapter ligation, cDNA in the size range of 200 +/- 25 bp was selected for PCR amplification and sequenced on an Illumina GAII or HiSeq platform for 1X36 bp (single-end sequencing protocol) according to the manufacturer's instructions.

3.13. Computational Analysis

For a detailed description of the computational analysis of overrepresented protein domains in the mRNA-bound proteome refer to the Supplemental Experimental Procedures in (Baltz et al., 2012).

A description of the algorithm to predict mRNA-interacting proteins, including a description of the benchmarking procedure, appears in the Supplemental Experimental Procedures. The Supplemental Experimental Procedures also contain a detailed description of the generation of the RNA-binder association network and the gene ontology (GO) terms enrichment analysis.

3.14. *D. melanogaster* culture conditions

For collections, adult flies were transferred to collection cages and fed with the respective brewers yeast medium for >2 days before starting the embryo collections to adapt the adults flies to the environment. The flies were kept at 25°C in 12h day-night cycles. Collections were performed at random times, without any preference to collect during night or day cycles. However, the 0-2h old embryos were always kept in the dark before crosslinking to avoid possible photocrosslinking through environmental light. We optimized the collections by increasing the surface area of the collection cage (with paper "curtains" attached to inner side of the cage). The fly population density hereby did not exceed 6 flies/cm^2.

0 – 2 h old embryos were collected on white grape juice agar plates for subsequent photoreactive nucleoside incorporation analysis and UV-crosslinking experiments.

In total, we collected between 2500 – 6000 mg embryos (0- 2h old) for the large-scale experiments.

3.15. *D. melanogaster* strains

yw fly strain

yw-flies (wild type fly strain with mutation in body color (brown to yellow) and eyes(red to white) were used for all experiments without application of photoreactive nucleosides.

X490 fly strain: GAL4-UAS driven UPRT-expression in nurse cells

All TU-tagging based experiments were performed with a fly strain described here and termed X490. The goal was to express UPRT in the oocyte and early embryo at the highest level possible for TU-tagging based RNA-labeling and UV_{365nm}-crosslinking experiments. We created a fly strain expressing UPRT in nurse cells under the control of nurse cell specific nanos (NOS) promoter, which is active only during oogenesis. Only flies containing NOS-GAL4 and UAS-UPRT will express the UPRT construct in the egg

cell and the early embryo. We used standard genetic crosses to generate homozygous flies that express both GAL4 and UAS-UPRT from two copies of each transgene.

NOS-GAL4 line: (termed X368, generous gift from Gergen Lab)
NGT-40A nos-Gal4 (strong expression, insertion on chromosome II)
described in (Tracey et al., 2000)

UPRT line: Bloomington stock #27603 (insertion on chromosome III)
Stock List Description w[*]; wg[Sp-1]/CyO; P[w[+mC]=UAS-UPRT.M]3.1

The following scheme illustrates how the double homozygous NOS-Gal4/UPRT flies were created.

Cross X68 to double balancer line X188:

$$\text{♀(X368)}\ \frac{GAL4}{GAL4};\frac{+}{+}\ \text{X}\ \frac{CyO}{Sco};\frac{Tb}{Sb}\ \text{♂ (X188, } double\ balancer\ line\text{)}$$

select tubby pupae, pick curly winged females and cross to X368 males:

$$\text{♀}\ \frac{GAL4}{CyO};\frac{+}{Tb}\ \text{X}\ \frac{GAL4}{GAL4};\frac{+}{+}\ \text{♂(X368)}$$

select tubby pupae, pick females (without curly wings) and cross to 27603 males:

$$\text{♀}\ \frac{GAL4}{GAL4};\frac{+}{Tb}\ \text{X}\ \frac{+}{CyO};\frac{UAS}{UAS}\ \text{♂ (27603)}$$

select tubby pupae, pick curly winged females and males, cross them:

$$\text{♀}\ \frac{GAL4}{CyO};\frac{UAS}{Tb}\ \text{X}\ \frac{GAL4}{CyO};\frac{UAS}{Tb}\ \text{♂}$$

The back cross results in a double homozygous fly, no selection necessary since *Cyo* and *Tb* are homozygous lethal:

$$\frac{GAL4}{GAL4};\frac{UAS}{UAS}$$

The crossing yielded a fly homozygous for GAL4 and UAS-UPRT to obtain a high UPRT expression in *D. melanogaster* nurse cells.

3.16. Metabolic labeling of *D. melanogaster* embryo RNA with 4-thiouridine

Preparation of 4-thiouracil containing brewers yeast paste medium

4-thiouracil stock solutions (ranging from 0 - 400mM) were prepared in 50% DMSO. One weight volume of brewers yeast (Acros organics, 368080010) was mixed with one weight volume of 4-thiouracil containing stock solution of the respective concentration. The yeast-paste was heat inactivated at 55°C for 10 min and stored at 4°C. 4-thiouracil containing fly food was fed to the UPRT-expressing X490-fly. 4SU containing yeast-paste was prepared the same way, but using 4-thiouridine instead of 4-thiouracil.

3.17. Photoreactive Nucleoside incorporation assays

Biotinylation of photoreactive nucleoside labeled transcripts and detection by Dotblot

D. melanogaster total RNA was isolated by TRIzol extraction according to the manufacturers instructions. RNA was ethanol precipitated to remove traces of DTT before biotinylation. Thiol-specific biotinylation, dot-blot assays were carried out as described previously (Dolken et al., 2008). For dot-blot assays, typically 1 µg total biotinylated RNA was spotted on the membrane. As a positive control we used a 5'-biotinylated 55-mer random DNA oligo, RNA from flies fed on brewers yeast without photoreactive nucleosides served as negative control.

Quantification of 4SU incorporation by LC-MS

Total RNA was digested and dephosphorylated to single nucleosides for LC-MS analysis using a modified protocol similar as described by Andrus and Kuimelis (Andrus and Kuimelis, 2001). Briefly, 60 µg of total RNA were incubated for 16 hours at 37 °C with 297.5 U/ml of bacterial alkaline phosphatase (Worthington Biochemical) and 112 U/ml

of snake venom phosphodiesterase (Worthington Biochemical) in 60 µl buffer containing 13 mM MgCl2, 1.66 mM Zn Cl2 and 33 mM Tris- HCl at pH 8.5.

To separate the nucleosides from the residual salts and enzymes 8 µl 3M sodium acetate (pH5.5) and 200 µl ethanol were added to the reaction mixture and samples were chilled on dry ice for 15 minutes. After centrifugation (5 min, 20,000rcf, 4°C) the clear supernatant was transferred to a new tube, 600 µl ethanol were added and samples were chilled on dry ice for 10 minutes. After centrifugation the supernatant containing the nucleosides was completely dried under vacuum and the samples were dissolved in 30 µl water prior to LC-MS analysis.

The nucleoside mixture was separated by UPLC (Agilent 1290) on a Zorbax Eclipse Plus C18 (2.1 × 50mm, 1.8 micron particle size) reverse phase column (Agilent). Separation was performed using 5% methanol in water containing 0.1% formic acid with a flow rate of 0.1 ml/min isocratic for 9 minutes. For detection a TSQ Quantum Vantage triple quadrupole (Thermo) was used. The instrument was operated in SRM mode to monitor the transitions 245 -> 113 (positive mode) and 243 -> 110 + 245 -> 200 (negative mode) for uridine and the transitions 261 - > 129 (positive mode) and 259 -> 116 + 259 -> 216 (negative mode) for 4SU. For quantification 11 different concentrations of Uridine and 4SU, ranging from 12.5pg/µl to20 ng/µl, were used for calibration. Each sample was analyzed at 5 different dilutions in triplicate (ranging from 1:10000 to 1:1), to enable the quantification of each analyte in the linear range of the calibration curve.

3.18. UV crosslinking of *D. melanogaster* embryos

UV_{365nm}-Crosslinking

For UV-crosslinking at 365nm, embryos were fed with photoreactive nucleosides and UV_{365nm}-irradiated with 4J/cm2. For the large scale experiment, adult UPRT-expressing X490 flies were fed with 400mM 4-thiouracil yeast paste and 0-2h old embryos were collected for UV-irradiation after the adult flies were fed on 4TU containing yeast paste for at least 3 days, ensuring incorporation of photoreactive nucleosides in *D. melanogaster* embryo RNA. Before UV-irradiation, the yeast paste was removed from the grape juice agar plates containing the embryos and a thin layer of cold water was added to the agar plates. The 0-2h old embryos were carefully brushed from the plate with a soft brush, allowing them to float on a thin layer of water. The embryos on the

agar plate were immediately cooled down on ice and UV_{365nm}-irradiated in a Spectrolinker 1500A with 4 J/cm^2. During UV-irradiation the embryos were constantly chilled on ice to inhibit further development.
Embryos were washed with deionized water and dechoroniated with bleach before freezing and storing at -80°C.

UV_{254nm}-Crosslinking
For the large-scale experiment, adult *yw* flies were fed with standard yeast paste and 0-2h old embryos were collected for UV-irradiation. Embryo preparation and collection were performed as described above for UV_{365nm}-crosslinking, but the embryos were irradiated with UV light at 254nm in a Spectrolinker 1500A with 4 J/cm^2.

3.19. Isolation of mRNA-interacting proteins in *D. melanogaster* embryos

Embryos of *D. melanogaster* females fed with photoreactive nucleoside containing yeast paste, were harvested 2h after the eggs have been laid and UV-irradiated as described above. The frozen embryos were thawed on ice and lysed 10 - 20 ml lysis/binding buffer per gram embryo weight (100 mM Tris HCl, pH 7.5, 500 mM LiCl, 10 mM EDTA pH 8.0, 1% (w/v) lithium-dodecylsulfate, 5 mM DTT, Complete Mini EDTA-free protease inhibitor (Roche)). For a typical proteomics experiment 25 ml oligo(dT) Dynabeads (NEB) (bead volume in original storage buffer) were added to embryo extract prepared from 2500 – 6000 mg of embryos. For each proteomics experiment, the same amount of non-irradiated embryos was processed in parallel. The oligo(dT)-precipitation was performed essentially as for HEK293 cells with the following modifications:
The washing steps were increased (4 times with lysis/binding buffer and 6 times with NP40 washing buffer) and after elution, the RNA was digested by incubation with RNase I at 25 U/ml and benzonase (62.5 U/ml) for 3 h at 37°C in elution buffer containing 1mM MgCl2 and stored at -80C for further analysis.

3.20. Identification of mRNA-crosslinked proteins by Western analysis

Crosslinking, embryo lysis and oligo(dT)-precipitations were performed as described above. Input, supernatant after precipitation and the oligo(dT)-purified material was nuclease treated and the protein was analyzed on a 4-15 % Mini PROTEAN TGX

gradient gel (Biorad). After Western analysis, the nitrocellulose membrane was blocked and incubated with an antibody against the endogenous protein. HRP-conjugated secondary antibodies were used and the proteins were visualized using the ECL select Western blot detection reagent (GE-Healthcare)

3.21. Mass spectrometry analysis of Precipitated Proteins (*D. melanogaster* embryos)

After nuclease treatment proteins were concentrated with an Amicon Ultra-15 centrifugal filter unit (10kDa-cutoff). We analyzed 25% of the samples from UV_{254nm} – irradiated embryos in an initial mass spectrometry test-run.

In-gel digestion of oligo(dT) precipitated proteins

Preparations of oligo(dT)-precipitated protein-RNA complexes for mass spectrometry analysis using in-gel digestion mRNA-bound proteins were isolated as described in experimental procedures and separated on a NuPAGE Novex 4 to 12% gradient gel (Invitrogen) using reducing conditions. Proteins were fixed in fixative solution (50% methanol (v/v), 10% acetic acid (w/v)) and stained afterwards with the Colloidal Blue staining Kit (Invitrogen). Gel lanes were cut into 10 gel slices, which were individually subjected to reduction, alkylation and in-gel digestion with sequence grade modified trypsin (Promega) according to standard protocols (Shevchenko et al., 2006). After in-gel digestion peptides were extracted and desalted using StageTips (Rappsilber et al., 2007) prior to analysis by mass spectrometry.

HPLC and mass spectrometry

Reversed-phase liquid chromatography (rpHPLC) was performed employing a EASY-nLC II (Thermo Fisher) using self-made fritless C18 microcolumns (Ishihama et al., 2002) (75 µm ID packed with ReproSil-Pur C18-AQ 3-µm resin, Dr. Maisch GmbH) connected on-line to the electrospray ion source (Proxeon) of a Q Exactive mass spectrometer (Thermo Fisher). Peptide samples were eluted at a flow rate of 250 nl/min with a 5 to 95 % acetonitrile gradient over 2 h in 0.5% acetic acid. Settings for MS analysis is as follows: one full scan (resolution 70,000; m/z 300–1,700) followed by

top 10 MS/MS scans using higher-energy collisional dissociation (HCD) (min. signal required, 21,000; isolation width, 2; normalized collision energy, 26).The Q Exactive instrument was operated in the data dependent mode (DDA) with a full scan in the Orbitrap followed by up to 10 consecutive MS/MS scans. Ions with an unassigned charge state and singly charged ions were rejected. Former target ions selected for MS/MS were dynamically excluded for 30 s. Total cycle time for one full scan plus up to 10 MS/MS scans was approximately 2 s.

Processing of mass spectrometry data

All raw data were analyzed and processed by MaxQuant (v1.3.0.5). Search parameters included two missed cleavage sites, cysteine carbamidymethyl fixed modification, and variable modifications including methionine oxidation and protein N-terminal acetylation. The peptide mass tolerance was 6 ppm and the MS/MS tolerance was 10 ppm. Database search was performed with Andromeda against uniprot.DROME.2012-06 with common serum contaminants and enzyme sequences. False discovery rate (FDR) was set to 1% at peptide and at protein level.

4. Results

4.1. Identification of mRNA-interacting proteins in HEK293 cells

In order to characterize the mRNA-bound proteome, we aimed to develop a method, which allows "freezing" protein-RNA interactions as they occur *in vivo* by introducing covalent bonds between proteins and RNA. The covalent crosslinking between proteins and RNA is induced by UV-irradiation, hereafter named photocrosslinking.

4.2. Optimization of mRNP oligo(dT) affinity purification

A key feature of our approach is the use of photoreactive nucleoside analogs, 4-thiouridine (4SU) and 6-thioguanosine (6SG) (Favre et al., 1993). The use of photoreactive nucleoside analogs can dramatically increase the crosslinking efficacy of proteins to RNA with UV irradiation at 365 nm (Hafner et al., 2010). Both nucleoside analogs are readily taken up by mammalian cells and are incorporated into newly synthesized RNA (Melvin et al., 1978).

Using photoreactive nucleosides 4SU and 6SG offers several advantages:

The photoactivation occurs at 365nm, which allows selective excitation of the photoreactive thionucleosides, minimizing nucleic acid damage (Meisenheimer and Koch, 1997). The activated intermediate is highly reactive, enabling efficient crosslinking at lower energy doses with good yields ranging from 10% up to 90% (Tanner et al., 1988). Furthermore, the photoreactive nucleoside analogs used for this study introduce only minor structural perturbance by replacing an oxygen atom (1.4 Å) with sulfur (1.85 Å) at the respective position (Figure 4). This ensures that polymerases accept the thionucleoside-triphosphates as substrates without affecting base pairing properties.

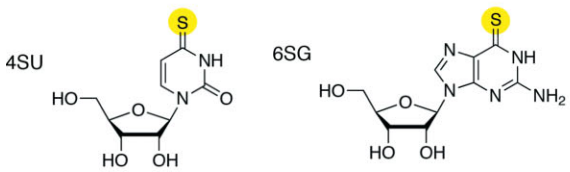


Figure 4: Photoreactive nucleosides used for labeling of HEK293 mRNA.
(4SU = 4-thiouridine, 6SG = 6-thioguanosine)

Another advantage of RNA labeling with photoreactive nucleoside analogs is the absence of detectable incorporation into DNA under the experimental conditions used for this study (Baltz et al., 2012), which reduces the probability of undesirable protein-DNA crosslinks.

By generating covalent bonds between proteins and RNA in living cells, photocrosslinking stabilizes the protein-RNA contacts in mRNP complexes, which in turn facilitates their isolation by oligo(dT) affinity purification (Setyono and Greenberg, 1981; Wagenmakers et al., 1980).

Because oligo(dT) affinity purification is based on nucleic acid hybridization between the oligo(dT) molecules immobilized on magnetic beads and the poly(A)-tail of mRNAs, stringent purification conditions can be used (here: 500mM lithium chloride and 1% (w/v) lithium dodecyl sulfate). Working with protein-denaturing conditions minimizes contaminations and ensures a stringent isolation of proteins in direct contact with mRNA. More stringent purification conditions, such as high concentrations of urea and guanidinum salts, would cause denaturation of the oligo(dT) hybrids and prohibit purification (Priyakumar et al., 2009).

UV-crosslinking creates covalent bonds between proteins and RNA if they are in direct contact within 1 Angstrom contact distance, (Ule et al., 2005), but it does not promote protein-protein cross-linking (Greenberg, 1979), which is a prerequisite for our approach to isolate only proteins that interact directly with mRNA. After stringent washing, the protein content of oligo(dT) affinity-purified mRNP complexes is analyzed by quantitative mass spectrometry (Figure 5).

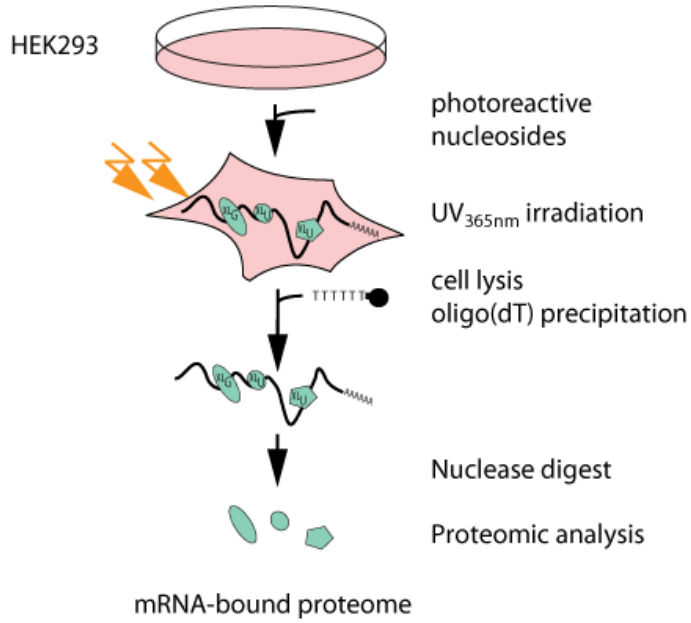


Figure 5: Identification of mRNA-interacting proteins by mass spectrometry. The scheme illustrates the experimental setup to identify proteins binding to polyadenylated RNA. Cellular mRNA is labeled with photoreactive nucleosides and proteins are covalently crosslinked to RNA by *in vivo* UV-crosslinking. The mRNPs are purified with oligo(dT) magnetic beads and nuclease digested before proteomic analysis by mass spectrometry.

I initially tested this approach by purifying protein-mRNA complexes using oligo(dT) beads from extracts of *in vivo* UV-irradiated human embryonic kidney (HEK) 293 cells. HEK293 cells were grown in medium supplemented with 4SU, 6SG or both. As a negative control, we used non-irradiated cells and cells grown on medium without photoreactive nucleosides.

The oligo(dT) purified mRNP complexes were eluted from the beads using a low salt elution buffer and high temperature to disrupt the oligo(dT)-poly(A) hybridization. The eluate was treated with RNase, resolved by SDS-PAGE and visualized by silver staining. The combination of metabolic labeling of RNA with photoreactive nucleosides and irradiation at UV 365 nm allows a high recovery of proteins (Figure 6).

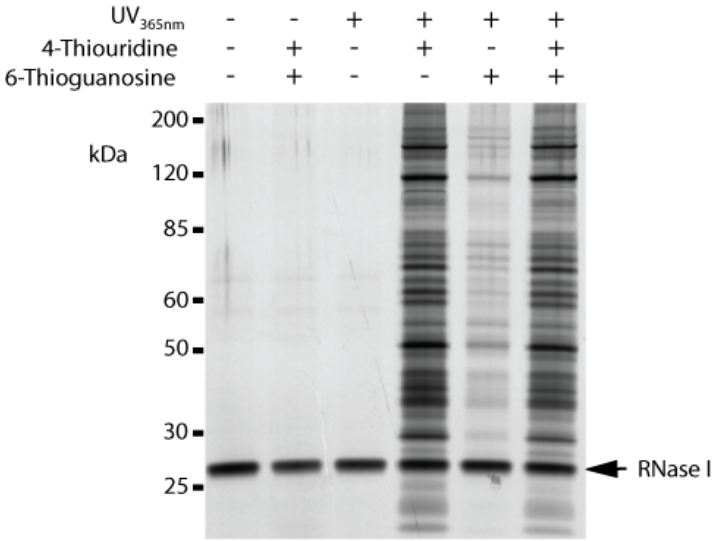


Figure 6: SDS-PAGE analysis of proteins crosslinked to polyadenylated RNA.
HEK293 cells were grown in medium supplemented with photoreactive 4SU and/or 6-thioguanosine and irradiated with UV-light at 365 nm. Non-irradiated cells and cells grown in medium without photoreactive nucleosides, served as negative control. Cells were lysed under denaturing conditions, and polyadenylated mRNA was isolated by oligo(dT)-bead based precipitation. Protein-RNA complexes were eluted from oligo(dT) beads, treated with RNase I and separated on an SDS gradient gel. Proteins were visualized by silver staining.

To assay the amount of mRNA precipitated from extracts of UV-irradiated and non-irradiated cells, we measured depletion of GAPDH-mRNA in the cell lysate supernatant after oligo(dT) precipitation by qRT-PCR (Figure 7). In addition, we monitored the enrichment of GAPDH mRNA in the oligo(dT) precipitate. We chose U6-spliceosomal RNA as endogenous control since it is a short and non-poly-adenylated transcript and thus unlikely to be precipitated by oligo(dT) beads. The qRT-PCR analysis showed that comparable amounts of GAPDH mRNA were precipitated from UV-irradiated and non-irradiated cells, suggesting that labeling of RNA and UV irradiation had only a minor effect on the mRNA pull-down efficacy (Figure 7).

Since GAPDH mRNA is a highly abundant transcript in HEK293 cells, as determined by digital gene expression analysis (DGE), we wanted to find out if we could detect low abundance transcripts as well. In addition to GAPDH, we tested EEF2 mRNA (high abundance transcript in HEK293) and two transcripts with low expression levels in HEK293 cells (DNAJC4 and DDX1), with the same precipitation protocol and UV-irradiated cells. Transcripts of high abundance in HEK293 cells (e.g. GAPDH or EEF2

mRNA) as well as transcripts of low abundance (DNAJC4, DDX1) were enriched in the precipitate (Figure 7).

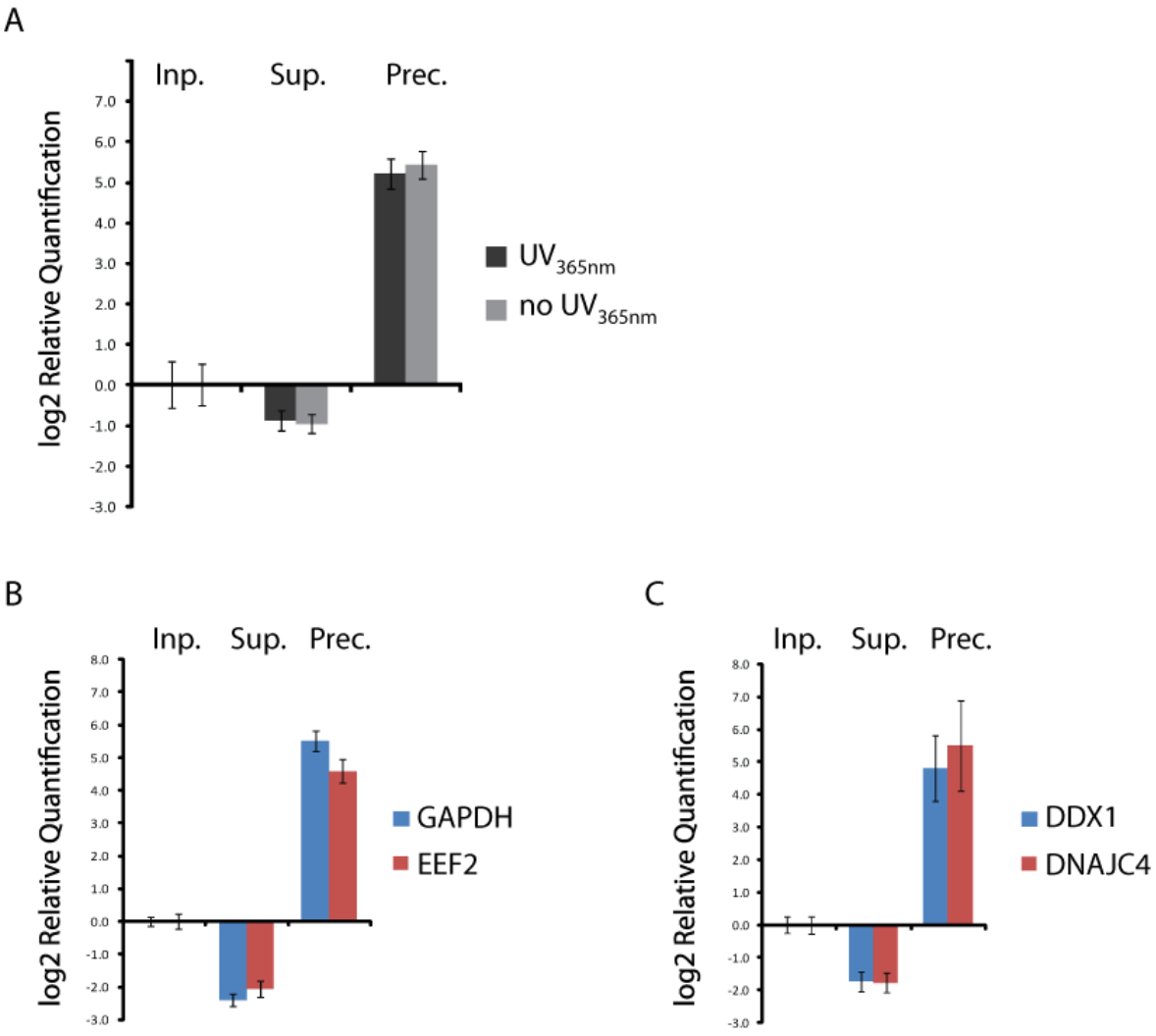


Figure 7: Analysis of oligo(dT) precipitate by qRT-PCR
(A) GAPDH-mRNA precipitation efficacy from extracts of UV-crosslinked and non-crosslinked cells. HEK293 cells were incubated with 4SU + 6SG and either crosslinked by UV_{365nm} irradiation or not crosslinked.
(B)(C) mRNA precipitation efficacy for transcripts with high copy number (GAPDH, EEF2) and low copy number (DDX1, DNAJC4) in HEK293 cells. HEK293 cells were incubated with 4SU + 6SG and crosslinked by UV_{365nm} irradiation.
Lysis was performed under denaturing conditions and polyadenylated RNA was precipitated with oligo(dT) beads. The amount of probed mRNA was examined by qRT-PCR in RNA of input, supernatant and precipitate. Values on y-axis show the log2-fold change of probed mRNAs compared to endogenous control (U6 spliceosomal RNA). The error bars display the standard error of the mean expression level with a 95% confidence interval.

Most of the tested known RBPs were detected by Western analysis of the oligo(dT) precipitate. Some of the tested RBPs showed only a weak signal in the Western analysis, such as Quaking (QKI) and the Insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1) (Figure 8A), indicating insufficient protein precipitation. However, a strong signal could be observed, when re-probing the same precipitate with an antibody against the endogenous heterogeneous nuclear ribonucleoprotein K (HNRNPK)(Figure 8A). The Argonaute protein (AGO2/EIF2C2) and the cytoplasmic poly(A) binding protein 4 (PABC4) were not detectable by the initial oligo(dT) precipitation protocol (Figure 8B).

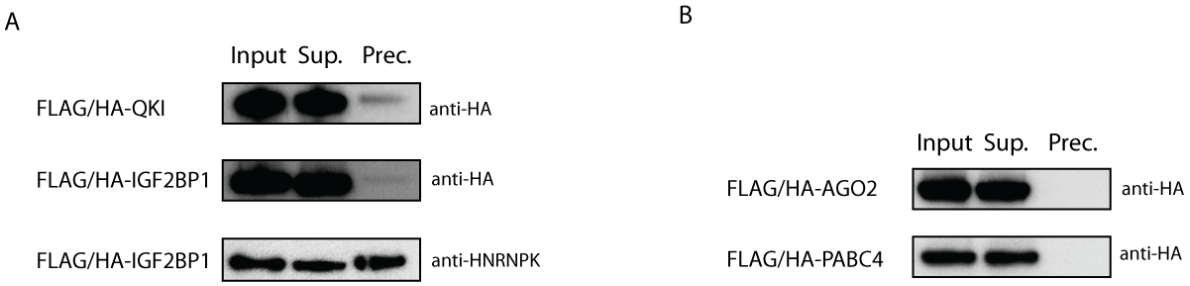


Figure 8: Western analysis of FLAG/HA-tagged RNA-binding proteins QKI, IGF2BP1, AGO2 and PABC4 of UV-crosslinked and non-crosslinked cells. Input extract, supernatant after precipitation (Sup.), and oligo(dT)-purified material (Prec.) were probed with either anti-HA antibody or with anti-HNRNPK antibody where noted.

Since AGO2, QKI and IGF2BP1 were shown to crosslink efficiently to RNA in PAR-CLIP experiments (Hafner et al., 2010), I aimed to improve the detection of those proteins by scaling up the experiment, using more cell lysate on the one hand and increasing the amount of oligo(dT)-beads per precipitation on the other hand.

By increasing the cell mass, a sufficient amount of IGF2BP1, QKI and PABC4 could be precipitated, for detection by Western analyses (Figure 8A). No protein was detected in the non-irradiated negative controls, demonstrating the specificity of our protocol.

I compared different elution methods in order to reduce possible contaminant proteins interacting with the bead surface. For that, an RNase-based elution protocol by RNase A/T1 treatment of the oligo(dT)-beads precipitated mRNP complexes (RNase elution, RE) was compared to the low salt and high temperature elution protocol (standard elution, SE) (Figure 9). We reasoned that an RNase based elution might reduce possible

bead-surface bound contaminants in the precipitate, as previously published (Michlewski and Caceres, 2010).

However, some proteins (such as IGF2BP1 and PABC4) were not efficiently eluted by RNase treatment alone (Figure 9). Therefore, the low salt and high temperature based elution method was preferred to elution by RNase treatment and used for subsequent experiments.

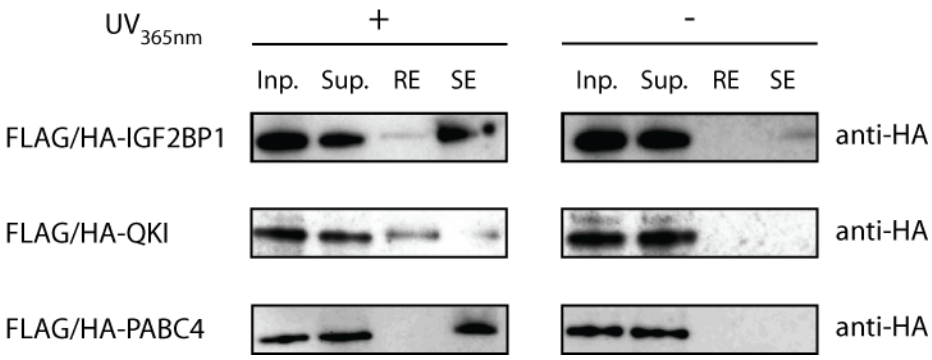


Figure 9: Western analysis of FLAG/HA-tagged RNA-binding proteins IGF2BP1, QKI and PABC4 of UV-crosslinked and non-crosslinked cells. Input extract (Inp.), supernatant after precipitation (Sup.), and oligo(dT)-purified material after RNase-elution (RE) and standard elution (SE)

Another benefit of the RNase-free elution is the possibility to re-use oligo(dT) beads for multiple precipitation rounds. This modification of the protocol can dramatically reduce the cost of the experiment. In addition it helps increasing the protein yield in the precipitates as demonstrated in the following experiment for the Ago2 protein.

Ago2 could not be detected after a single oligo(dT) precipitation (Figure 10B), likely due to the insufficient precipitation of mRNAs and/or incomplete capture of mRNAs with shortened poly(A) tails, such as microRNA/AGO-targeted mRNAs.

I therefore performed an experiment to monitor the depletion of mRNA after several consecutive rounds of oligo(dT) precipitations with the same set of oligo(dT) beads. The depletion of GAPDH from the supernatant was measured after each oligo(dT) precipitation round. The GAPDH transcript is abundant and targeted by AGO proteins (Hafner et al., 2010; Kishore et al., 2011). Figure 10A shows how after one precipitation round, only about 70% of this transcript was depleted in the supernatant when compared to input RNA. Three additional consecutive pull-downs from the same extract reduced the amount of GAPDH mRNA in the supernatant to about 5% (Figure 10A).

Four consecutive precipitation rounds were therefore sufficient to deplete over 95% of a very abundant mRNA in HEK293 cells.

The eluates of four oligo(dT) purifications were pooled and the co-precipitated proteins were analyzed by Western analysis. AGO2 protein was now detected in the precipitate (Figure 10B), indicating that multiple consecutive oligo(dT) precipitations are required to precipitate crosslinked AGO2 protein.

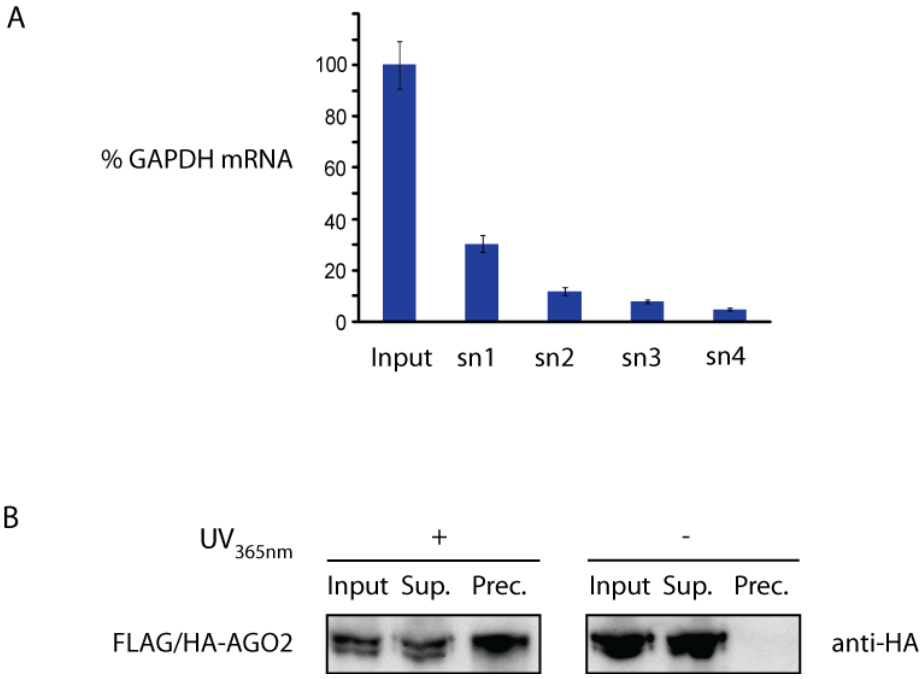


Figure 10: Multiple consecutive rounds of oligo(dT) precipitations increase the precipitation efficiency. (A) *GAPDH* mRNA depletion. qRT-PCR analysis of *GAPDH* mRNA in supernatants (SN1 to SN4) after each round of oligo(dT) bead precipitation (four in total) compared to *GAPDH* mRNA in extract before precipitations (input) are shown as percent of input. The error bars display the calculated maximum and minimum expression levels that represent the standard error of the mean expression level with a 95% confidence interval.
(B) Western analysis of FLAG/HA-tagged RNA-binding protein AGO2 in input extract, supernatant after precipitation (Sup.), and oligo(dT)-purified material (Prec.) of UV-crosslinked and non-crosslinked cells.

4.3. Characterization of the oligo(dT)-purified RNA

In addition to the analysis of individual transcripts by qRT-PCR, we analyzed the RNA in the precipitates of oligo(dT) purifications by next-generation sequencing (NGS). NGS–based digital gene expression analysis (DGE) provides a detailed picture of the RNA

species and their quantities, and thus precisely monitors which RNA classes are enriched in the oligo(dT)-precipitation protocol.

I isolated RNA from the pooled precipitates of four consecutive oligo(dT) purifications, and constructed a cDNA library through random priming. As for the experimental setup to identify mRNA-bound proteins, the RNA was labeled with both 4SU and 6SG. RNA was isolated from both irradiated and non-irradiated cells. Digital gene expression analysis of the cDNA library of non-irradiated cells, labeled with 4SU and 6SG, revealed that about 88% of the sequenced reads mapped to mRNA, and only 8% to ribosomal RNA (rRNA) genes (Figure 11). In RNA precipitates obtained from UV-irradiated cells the rRNA content increased about 4.5-fold (from 8% to 36%), likely reflecting the crosslinking of ribosomes to mRNA transcripts (Figure 11).

For comparison, we used a standard mRNA purification procedure to generate a reference transcriptome, starting from TRIzol-extracted RNA of untreated cells and involving only a single oligo(dT) precipitation, which yielded 96% mRNA and 2% rRNA (Figure 11).

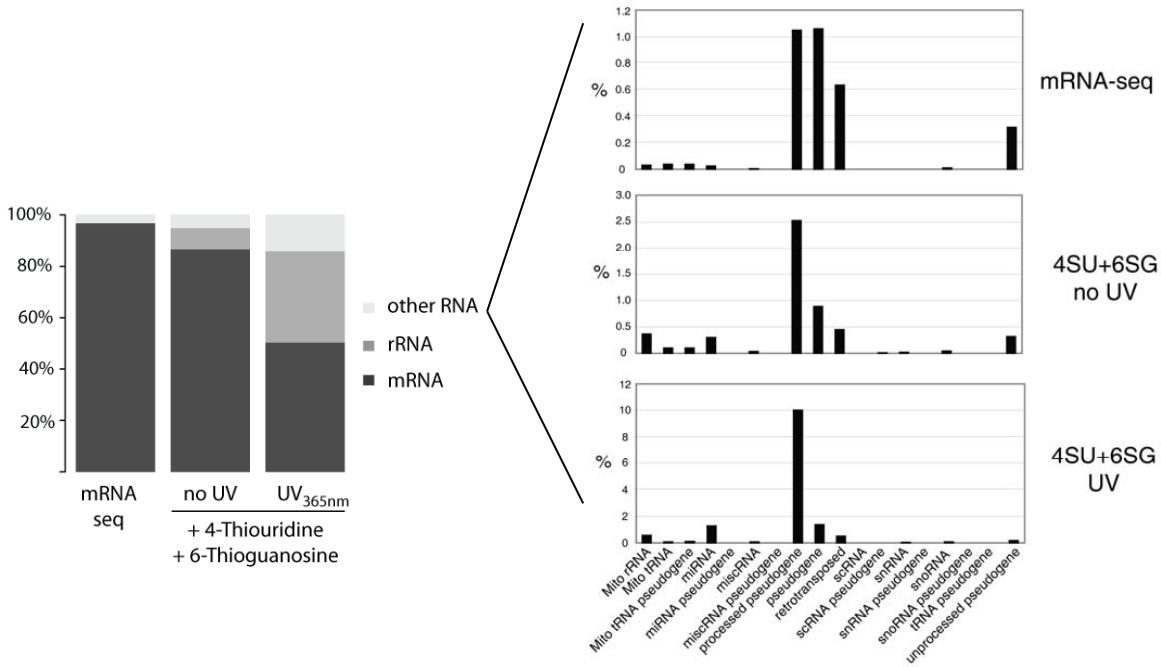


Figure 11: Read count distribution over different RNA types.
mRNA was purified by a single oligo(dT) precipitation from TRIzol-extracted total RNA (mRNA seq) and used as reference transcriptome.
Analogous to the protocol for isolation of mRNA-bound proteins, RNA was isolated by direct precipitation from cellular extract by four consecutive rounds of oligo(dT)

precipitation. RNA was precipitated from cell lysate of both UV-irradiated and non-irradiated cells (4SU+6SG UV and 4SU+6SG no UV, respectively). Prior to RNA analysis, crosslinked proteins were removed by Proteinase K digest. The recovered RNA was analyzed by next-generation sequencing. The read count distribution over different RNA classes (mRNA, rRNA, and other) was inferred by multiplication of the FPKM values with the respective length of the longest transcript of a given gene.
(A) Read count distribution depending on the protocol used for precipitation.
(B) Detailed description of the RNAs summarized as "other".

We compared the different RNA libraries to understand how precipitated transcripts, derived from UV-crosslinked and non-crosslinked cells, correlate with mRNA obtained from untreated cells (reference transcriptome, "mRNA seq").

For this analysis we compared the FPKM values (fragments per kilobase of transcript per million mapped reads) for each transcript, which is an indicator of transcript abundance in the precipitate. A direct comparison of the FPKM values of the different RNA libraries showed that the Pearson-correlation coefficient between the HEK293 reference transcriptome and the transcriptome obtained through our protocol was 0.87 for non-irradiated cells and 0.82 for UV-irradiated cells, indicating that the oligo(dT)-precipitated mRNA closely reflected the cellular mRNA pool (Figure 12).

Ribosomal and other RNAs that are more abundant in the UV-crosslinked sample explain the lower Pearson correlation coefficient between the "mRNA seq" sample and the UV-crosslinked sample. Indeed, when comparing the UV-crosslinked sample to the non-crosslinked sample, a shift in the "cloud" of RNA reads towards moderate expression levels was observed for most transcripts, probably caused by the large amount of reads mapping to rRNAs and to other RNAs (Figure 11).

Also, the LOESS regression line shows a distinct kink towards the UV-crosslinked sample in the range of transcripts with higher FPKM values, which can also be explained by the increased amount of reads that mapped to ribosomal and other RNAs in the UV-irradiated sample (Figure 12, red line).

Since photoreactive nucleosides are required for efficient protein-RNA crosslinking with UV light at 365nm (Hafner et al., 2010), we monitored the incorporation of photoreactive nucleosides into RNA.

For this, the thiol-groups of 4SU and 6SG were *in vitro* biotinylated and the biotinylated RNA was isolated with streptavidin coated magnetic beads (Dolken et al., 2008). I isolated the photoreactive nucleoside containing sub-fraction of the oligo(dT)-

precipitated RNA from 4SU- and 6SG-labeled, non-crosslinked cells by biotinylation and streptavidin purification. Direct comparison of the thionucleotide-containing RNA ("4SU + 6SG purified RNA") with the cellular mRNA ("mRNA seq") revealed a Pearson correlation coefficient of 0.90 and also LOESS regression analysis indicated efficient and unbiased metabolic labeling of cellular transcripts with photoreactive nucleosides (Figure 12).

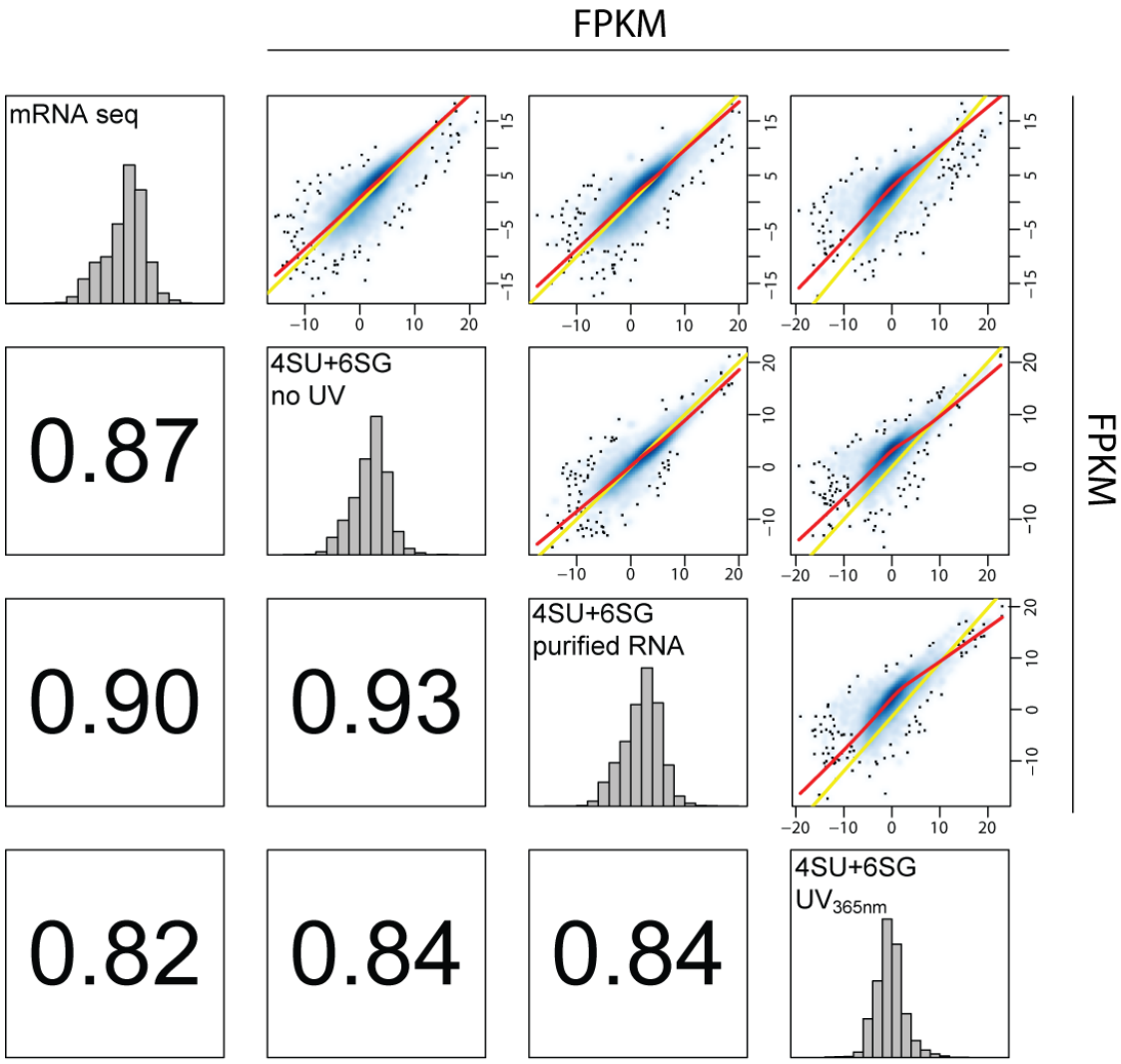


Figure 12: Pair-wise correlation between RNA abundance expressed as log2 FPKM (fragments per kilobase of transcript per million mapped reads) of the RNA described in Figure 11 and 4SU+6SG purified RNA. To monitor the photoreactive nucleoside incorporation into RNA, the 4SU- and 6SG-containing RNA was purified from oligo(dT)-precipitated RNA of non-crosslinked cells by biotinylation and streptavidin pull-down (4SU+6SG purified RNA) and analyzed by next-generation sequencing. The diagonal is shown as yellow line for each pairwise comparison, whereas a LOESS (locally weighted scatterplot smoothing) regression line is shown in red.

In summary, based on the Western analysis of precipitated proteins in Figure 10 and the analysis of precipitated RNA by qRT-PCR and NGS, we concluded that at least four consecutive oligo(dT) purifications are required for efficient purification of cellular mRNA-protein complexes, with a significant and unbiased enrichment of mRNA over other classes of RNA in the resulting precipitates. Multiple consecutive precipitations with oligo(dT)-magnetic beads are thus an experimentally and economically reasonable alternative to increased oligo(dT)-magnetic bead volumes used per experiment.

4.4. Identification of mRNA-bound Proteins by Quantitative Mass Spectrometry

We aimed to design an experimental setup, which allows us to specifically identify "true" RNA-interacting proteins, and distinguish them from contaminant. Since we used stringent, protein-denaturing conditions for cell-lysis and subsequent washing steps during oligo(dT)-precipitations, we assumed that true RNA-interacting proteins would be highly enriched over background in the precipitates, thus displaying the highest peptide counts in mass spectrometry analysis.

Nevertheless, we searched for an additional criterion to distinguish between RBPs, which are covalently crosslinked to RNA and thus co-precipitate in oligo(dT) purifications, and possible contaminant proteins that might be detected in the precipitates due to their high abundance in the cell, formation of stable protein-protein complexes despite denaturing conditions or interaction with the oligo(dT) beads surface, regardless of the conjugated oligo(dT) sequences.

The additional criterion should therefore be able to distinguish between proteins that originate from the UV-irradiated cell population, in which RNA-interacting proteins are covalently crosslinked to RNA, and proteins that originate from a non-irradiated control population. We thus assumed that unspecific background proteins should be detected in both samples, while specific RNA-interacting proteins should be enriched over background in the UV-irradiated sample.

Mass spectrometry allows a relative comparison of ion signals and thus relative quantification of peptides either between two samples that were analyzed in the same manner or, if stable isotopes are used, within the same analysis. Stable isotopes allow distinguishing between chemically equivalent peptides by introducing different mass-to-charge ratios, which can be detected by mass spectrometry. We therefore chose

stable isotope labeling-based quantitative mass spectrometry to analyze the proteomic content of the oligo(dT) precipitates. Stable isotope labeling allows outstanding quantification accuracy, since the signal intensities of the peptide forms are compared within the same mass spectrometry run and differences in sample preparation and ionization efficiencies between runs can be neglected.

We used SILAC (stable isotope labeling by amino acids in cell culture), a metabolic labeling technique, based on supplementing growth media of living cells with heavy stable-isotope amino acids (Ong and Mann, 2006). The labeled amino acids in the "heavy" (labeled) SILAC medium are incorporated into virtually all proteins through cellular growth and protein turnover. A control cell population is grown in "light" (unlabeled) SILAC medium.

Cells were grown in medium supplemented with "light" or "heavy" stable isotope labeled amino acids and either UV- crosslinked or not (control). We performed oligo(dT) purifications, from both cell lysates in parallel, combined the precipitates and compared the protein abundance from crosslinked cells to that of non-crosslinked cells (Figure 13). We expected to see a low heavy-to-light (H/L) ratio for RNA-interacting proteins, if the cells grown in "light" SILAC medium were UV-irradiated. Unspecific background proteins should therefore exhibit an approximately 1:1 ratio, which allows a robust separation of signal from noise.

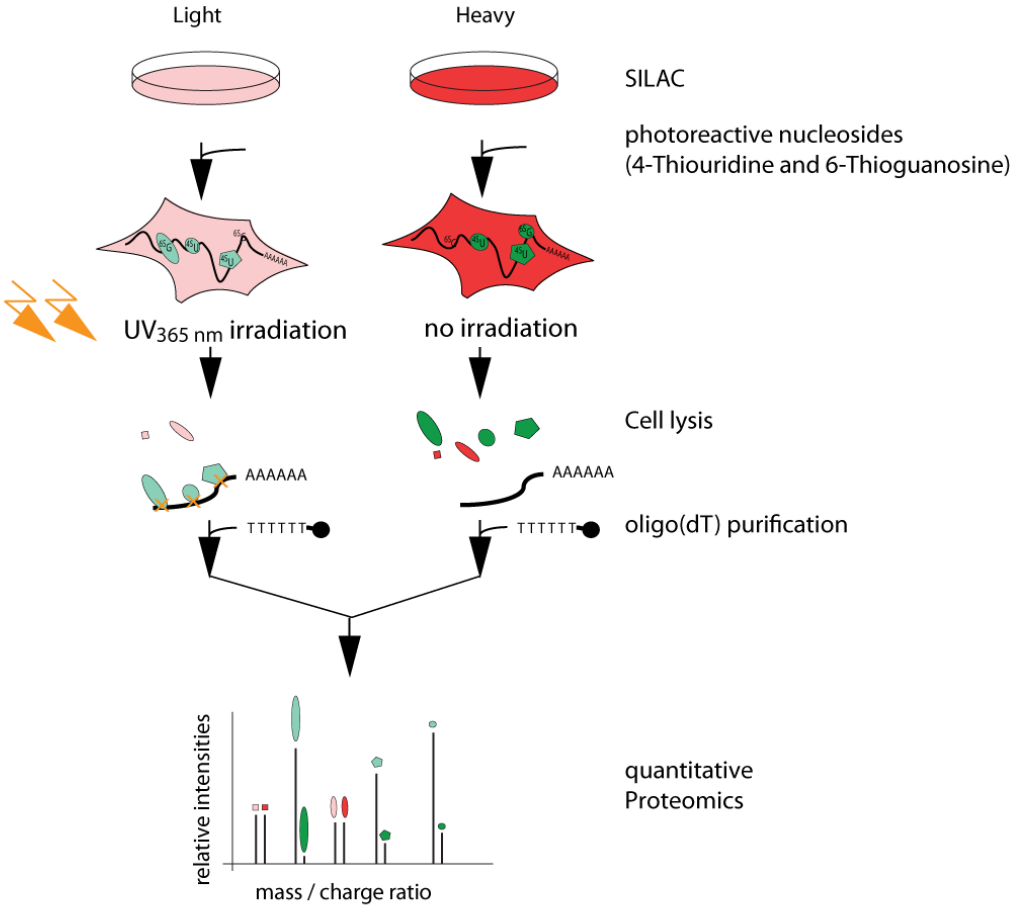


Figure 13: Quantitative Proteomic Analysis of mRNA-bound Proteins using SILAC (Stable Isotope Labeling with Amino acids in Cell culture). HEK293 cells were grown in either light or heavy SILAC-media. Photoreactive 4-thiouridine (4SU) and 6-thioguanosine (6-SG) were supplied to the medium and incorporated in newly synthesized RNA. Crosslinking was performed with UV_{365nm} irradiation to covalently bind RBPs to mRNAs in living cells. Cells were lysed under denaturing conditions and polyadenylated RNA was purified by precipitation with oligo(dT)-beads. After stringent washing under denaturing conditions, the Protein-RNA complexes were eluted from the beads and RNA was removed by nuclease treatment. The nuclease treated samples from crosslinked and non-crosslinked cells were combined and analysed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Specific mRNA-interacting proteins can be distinguished from nonspecific background proteins by their SILAC heavy-to-light ratios. mRNA-bound proteins are detected by low heavy-to-light ratios in the experiments where "light" labeled cells are UV-irradiated and a high heavy-to-light ratio in the crossover experiment, where "heavy" labeled cells are UV-irradiated (label swap experiment). In contrast, nonspecific binders show a heavy-to-light ratio close to 1 in both experimental conditions.

In two biological replicate experiments the "light" labeled cells were UV irradiated and proteins in the oligo(dT) precipitations were compared to the precipitate of non-crosslinked "heavy" labeled cells. We termed those experiments L1 and L2 thereafter,

since the “light” labeled cells were UV-irradiated. In a single ‘‘label swap’’ experiment (H1), we UV-irradiated the ‘‘heavy’’ labeled cells and the recovered proteins were compared to those of ‘‘light’’ labeled non-crosslinked cells. In a label swap control experiment, we expected to observe inverted heavy-to-light SILAC ratios for mRNA-interacting proteins.

It is almost impossible to avoid common contaminants in mass spectrometry (e.g. keratins), which are introduced during sample preparation. Those contaminant proteins should be revealed by the label swap experiment, since the proteins introduced after stable isotope labeling should exhibit low heavy-to-light SILAC ratios.

For each experiment, we pooled the precipitates of four consecutive oligo(dT) purifications, nuclease-treated the protein-RNA complexes to remove residual RNA and combined the precipitated samples from crosslinked and non-crosslinked cells before protein concentration by TCA-precipitation. The concentrated protein solutions were separated by SDS-PAGE followed by in-gel trypsin-digest.

In all three mass spectrometry experiments we acquired a total 1,383,856 MS/MS spectra after 300 h of data acquisition. For data processing, we used the MaxQuant software package (Cox and Mann, 2008) and identified a total of 1,326 proteins with at least one unique peptide. The number of identified proteins in each experiment is shown in Figure 14A (“identified”). In addition, only proteins with at least three observed SILAC-peptide ratios were considered reliably quantified (Figure 14A, “quantified”).

To evaluate the reproducibility of the method, we compared the overlap between the biological replicate experiments. Both the overlap of identified proteins (proteins which were identified with at least one unique peptide in each experiment) and quantified proteins (with at least three observed SILAC-peptide ratios in each experiment) were analyzed.

Seven hundred ninety proteins were identified in all three proteomic analyses, and 561 of those were quantified with at least three observed SILAC-peptide ratios in each experiment (Figure 14B, venn diagram), demonstrating a substantial overlap between experiments.

A

experiment	identified	quantified
L1	999	694
L2	1147	839
H1	1050	754

B

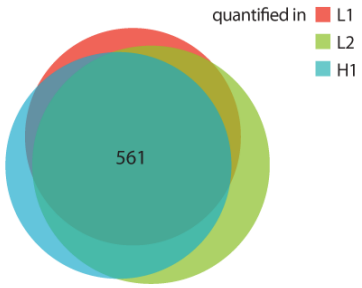


Figure 14: Reproducibility between proteomic experiments (A) Summary of proteomic experiments. In two replicates, the proteomic composition of oligo(dT) precipitates was analyzed for ‘‘light’’ labeled crosslinked cells (experiments L1 and L2) and one experiment for ‘‘heavy’’ labeled crosslinked cells (H1). The table indicates the number of identified (as determined by detection of unique peptides) and quantified proteins (as determined by SILAC ratios of proteins) in each experiment. (B) The overlap of proteins quantified in the replicate experiments is shown in the Venn diagram.

We also examined how well the SILAC peptide ratios correlate between replicate experiments and compared log2 SILAC ratios from biological replicates L1 and L2 (Figure 15A). Out of 800 proteins identified in both experiments, 778 proteins were specifically enriched in the precipitates of UV-crosslinked cells relative to the non-irradiated control cells (SILAC log2 fold changes < 0 in both cases), which indicates that 97% of all identified proteins show UV-crosslinking specific enrichment in both biological replicates.

Next, we wanted to examine if the oligo(dT) precipitation is biased to precipitate proteins with high copy numbers in HEK293 cells. We compared the enrichment of proteins in oligo(dT) precipitations (as determined by the mean SILAC ratios in L1 and L2) to the abundance of the respective proteins in HEK293 cells (as determined by intensity-based absolute quantification, iBAQ). We observed no significant correlation between the enrichment of proteins in oligo(dT) precipitates and the protein abundance rank, suggesting that the degree of enrichment was independent of the number of protein molecules present in the cell (Figure 15B).

Additionally, we plotted the log2 SILAC ratios from the label swap experiment H1 against each of the both biological replicate experiments L1 and L2. As expected, most SILAC ratios were inverted in the label swap experiment, since the cells grown on "heavy" SILAC medium were UV-irradiated (Figure 15C,D). 135 proteins in the label swap experiment showed negative log2 SILAC ratios. A close look at the proteins with low SILAC ratios in both the biological experiments and the label swap experiment (Figure 15C,D, lower-left corners), revealed that the majority of the proteins in this group were known contaminants such as trypsin, LysC and keratins. We therefore excluded proteins with negative SILAC ratios in the label swap experiment. Among the 135 excluded proteins were six proteins, which were annotated to interact with RNA (ELAVL3, PA2G4, PDCD11, RBM16, RBPMS and SNRPE). We therefore expect a false negative rate below 5 %.

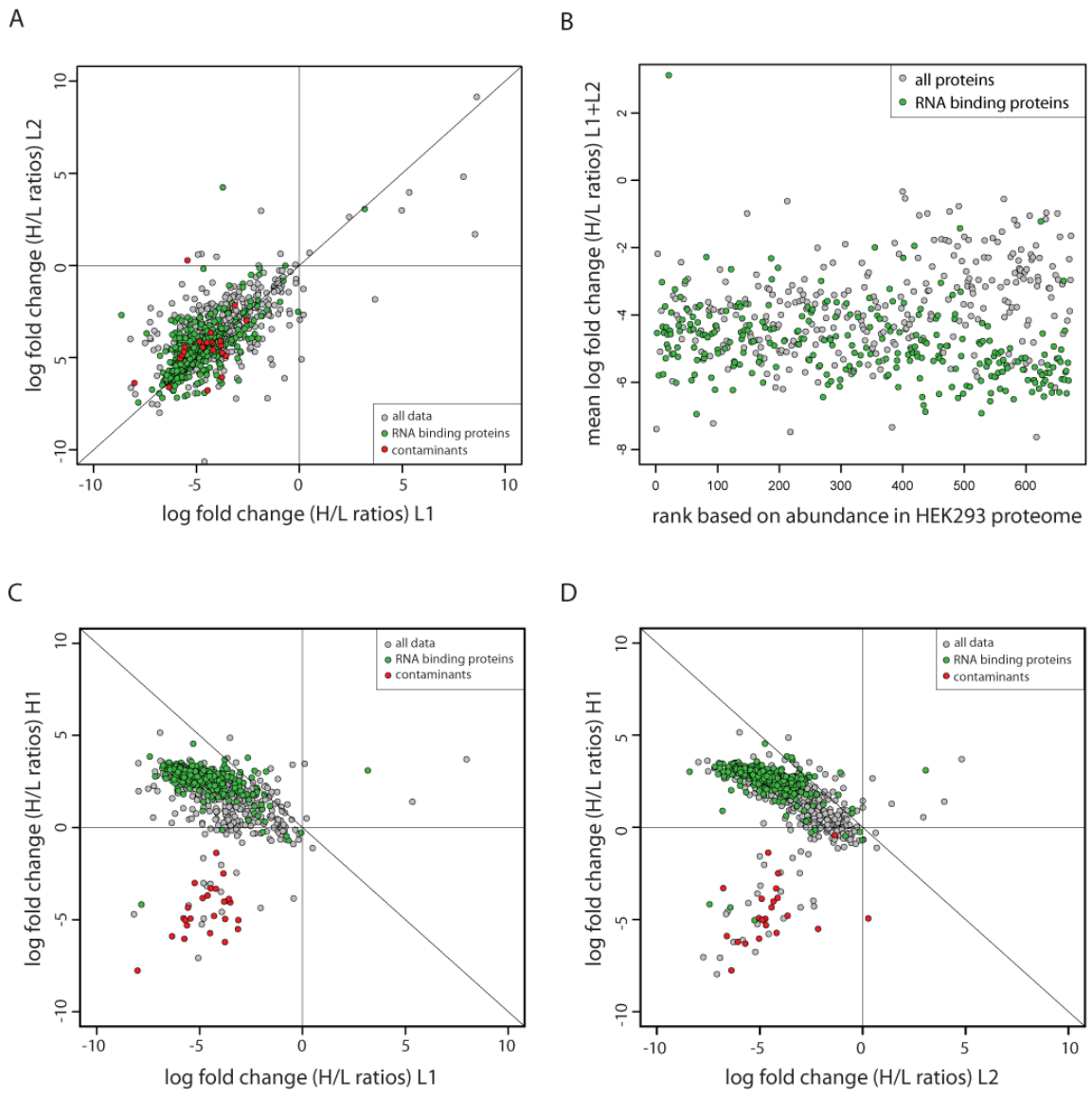


Figure 15: reproducibility of SILAC-ratios between biological replicates.
(A) Comparison of the log2 fold changes (LFC) of ‘‘heavy’’ to ‘‘light’’ SILAC ratios (H/L ratios) of proteins quantified in biological replicates L1 and L2. Previously known RNA-binding proteins are indicated in green, and known contaminants in red.
(B) Correlation of mean log2 fold (LFC) SILAC ratios of proteins identified in L1 and L2 and abundance rank of the respective proteins in HEK293 proteome (based iBAQ data of 3019 identified proteins).
(C)(D) As in (A), proteins were quantified in L1 (C) or L2 (D) plotted against proteins quantified in the label swap experiment H1.

In order to further reduce the number of possible false positives, we required an enrichment of at least 3-fold (determined by SILAC ratios) in at least one of the three analyses. This led to the exclusion of 54 proteins, which did not meet this criterion. We thereby ended up with 889 proteins, which passed all our criteria and, in order to facilitate further analysis, we retained only the longest isoforms of the identified proteins.

The non-redundant list, which was used for all further analyses, comprised 797 proteins, which I refer to as the mRNA-bound proteome of HEK293 cells (Table S1.)

All 797 proteins in Table S1 were ranked in three classes, according to the number of experiments in which they were detected as highly, i.e. more than three-fold enriched. 505 proteins (63% of all proteins in Table S1) met this requirement in all three proteomic analyses and were categorized as class I, which makes up the largest group of proteins in our final set.

190 proteins (24%) showed 3-fold enrichment in two out of three proteomic experiments (class II). The remaining 102 proteins (13%) passed the 3-fold enrichment criterion in one out of three experiments (class III).

4.5. Overview of identified mRNA-interacting proteins

I classified the 797 identified proteins into functional categories based on their gene annotations. As expected, the majority of identified proteins fell into gene annotation categories related to RNA metabolism. The most frequent protein categories were ribosomal proteins, helicases, translation factors and other RNA-binding proteins, making up close to 70% of all identified proteins (Figure 16A). In contrast, we identified only a small number of highly expressed cellular proteins with no apparent RNA-binding capabilities, such as metabolic enzymes, heat-shock proteins or histones.

We compared the cellular abundance of the proteins by category, to exclude systematic bias towards protein categories containing highly abundant cellular proteins. The median relative abundance of RNA-binding proteins, ribosomal proteins and translations factors was slightly higher but comparable to that of other functional groups of proteins (Figure 16B).

In summary, the results presented here indicate that the oligo(dT) precipitations specifically enrich for RNA-interacting proteins and are not biased to co-purify highly abundant cellular proteins.

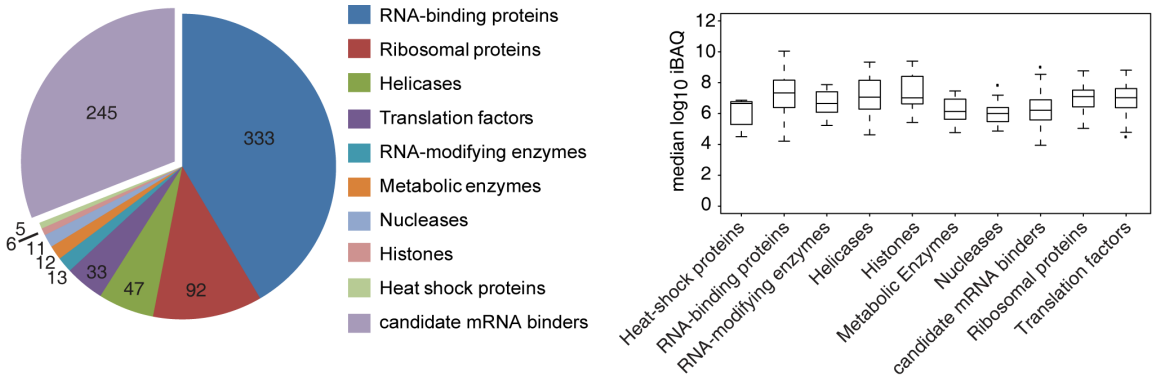


Figure 16: Overview of the identified mRNA-interacting Proteins.
(A) Number of identified proteins classified in functional categories. The category "RNA-binding proteins" covers proteins that were annotated as RNA-binding at the time of the study, but did not fall into any of the categories above. Proteins that were identified in our study, but not annotated as RNA-binding proteins before, were summarized as "candidate mRNA binders".
(B) Median relative number of protein molecules (iBAQ) belonging to functional categories as determined in (A). Protein amounts were calculated as the sum of all peptide peak intensities divided by the number of theoretically observable tryptic peptides (Schwanhausser et al., 2011). The iBAQ values are shown as box-and-whisker plots, where the median is shown as horizontal line, the surrounding box defines the upper and lower quartile and the sample range is defined by the whiskers. The dots indicate potential outliers.

A closer look at the identified proteins revealed that we discovered a large number of proteins we expected to find based on their biochemical function. During their lifetime, RNA-molecules interact with RNA-binding proteins and form dynamic messenger ribonucleoprotein particles (mRNPs) (Moore, 2005; Muller-McNicoll and Neugebauer, 2013). Thus, we expected to detect proteins present in complexes that modulate mRNA metabolism, especially those present in mRNPs that bind to most transcripts in the cell. An example for such a central mRNP complex present on most mRNAs in the cell is the exon junction complex (EJC), which plays a major role in posttranscriptional regulation of mRNA in metazoans (Bono and Gehring, 2011). During splicing of mRNAs the EJC is deposited upstream of the exon-exon junction and transported to the cytoplasm where

it influences the translation and localization of the spliced mRNA, and it is also involved in cytoplasmic mRNA surveillance and nonsense-mediated decay (Chang et al., 2007).
The structure of the core was solved at 2.2 A resolution and reveals how the four core proteins eIF4A3, CASC3/Btz, MAGOH and RBM8A/Y14 interact with mRNA and the bound ATP molecule (Bono et al., 2006). The crystal structure indicates that eIF4A3 encloses an ATP molecule and provides binding sites for six ribonucleotides. CASC3/Btz on the other hand is in contact with the 5' nucleotide of the bound mRNA, whereas MAGOH and RBM8A mostly contact eIF4A3 and stabilize the ATP-bound state of the complex (Figure 17).

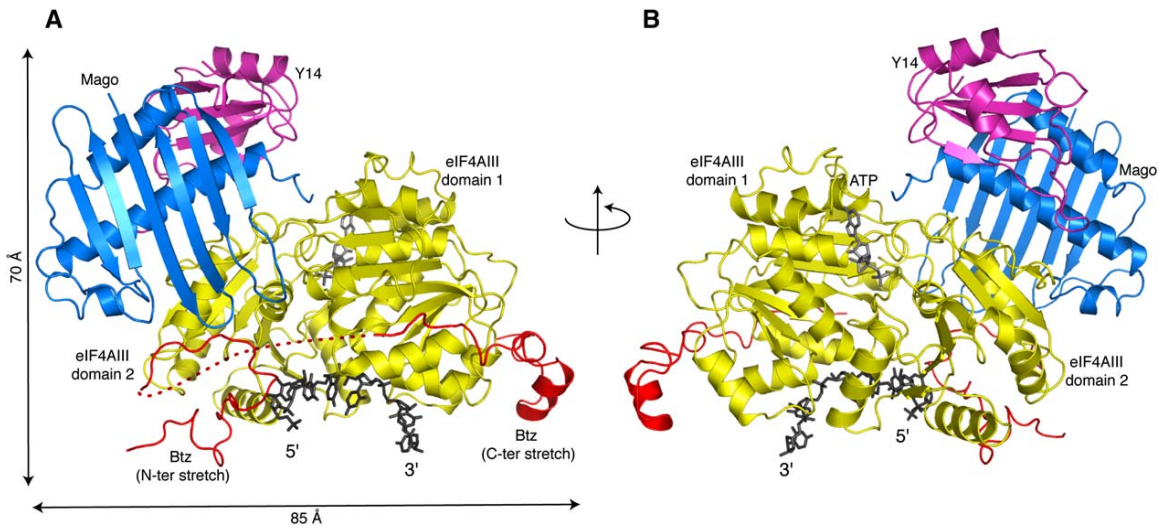


Figure 17 Structure of Exon-Junction-Complex enclosing an ATP molecule (gray) and RNA (black).
The EJC is shown in two orientations by a 180º rotation about a vertical axis. The relative positions of Btz (red), the DEAD-box helicase eIF4AIII (yellow), Mago (blue) and Y14 (magenta) indicate that both eIF4AIII and Btz are in direct contact with RNA (black). The dotted line in red shows the path of a portion of Btz not present in the electron density. Reprinted from (Bono et al., 2006), with permission from Elsevier.

We detected all four core-proteins of the exon junction complex (RBM8A/Y14, MAGOH, EIF4A3, and CASC3/BTZ (Bono et al., 2006). In accordance with the structural model and as predicted from UV-crosslinking experiments (Ballut et al., 2005; Shibuya et al., 2004), we identified an order of magnitude more molecules of the EJC core proteins EIF4A3 and CASC3, which directly interact with RNA, compared to RBM8A and MAGOH, which are not in direct contact with mRNA. However, it is important to consider that RBM8A and MAGOH might interact with mRNA in an alternative EJC-independent way.

In addition to EJC proteins, we observed proteins present in the translation initiation complex (EIF4A1, EIF4B, EIF4E, EIF4G1, and EIF4H (Jackson et al., 2010)). We identified three out of four Argonaute proteins expressed in HEK293 cells. EIF2C4 the AGO protein with the lowest expression in 293 was not detected in the precipitates, which indicates that some proteins are not detectable despite being expressed in HEK293 cells.

The identified RNA-binding proteins also overlapped to a certain extent with the spliceosome and the nucleolus. The spliceosome is a multi-subunit complex consisting of the small nuclear RNAs (snRNA) U1, U2, U4/U6 and U5 and a large number of core and associated proteins (Wahl et al., 2009). The spliceosome removes introns from precursor mRNAs (pre-mRNAs) and ligates the 5' and 3' exons to form mRNA in a process termed splicing and first described in 1977 (Berget et al., 1977).

We identified 99 out of 172 proteins that were detected in the spliceosome (Bessonov et al., 2008). This overlap is not surprising, since the spliceosome is deposited on RNA co-transcriptionally and can remain bound to mRNA after transcription has been completed (Bhatt et al., 2012).

Our lysis protocol disrupts the nuclear structure and accordingly, our set of identified RNA-binding proteins also overlaps with the nuclear protein content and with proteins that were also found in the nucleolus. 243 out of 548 proteins found in the nucleolus proteome (Andersen et al., 2005) were also detected in our analysis (figure 18A), which indicates, that our method allows identifying proteins from various cellular compartments and mRNP complexes.

Most RNA-binding proteins typically contain specific RNA-binding domains, which recognize sequence and/or structure elements and mediate binding to RNA. RBPs often use a modular combination of conserved domains, which determines in part the RNA-binding specificity of the protein (Lunde et al., 2007).

RNA-binding domains are indicative for RBP function and have thus been used to computationally predict RBPs and to estimate the number of RBPs in the cell (de Lima Morais et al., 2011). We used this structural feature of RNA-binding proteins to estimate how many of the RNA-binding domain containing proteins in HEK293 cells we can identify with our approach. For that, we compared how many of the proteins containing at least one specific RNA-binding domain were also identified in the mRNA-bound

proteome. The majority of the RNA-binding domain-containing proteins observed in the HEK293 deep proteome (Geiger et al., 2012) were also identified by our analysis (Figure 18B).

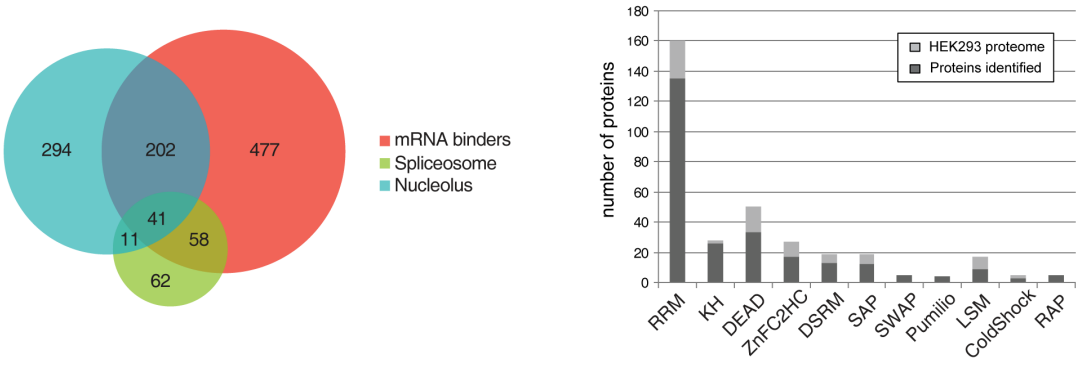


Figure 18
(A) Overlap of the identified proteins (mRNA binders) with proteins present in the nucleolus and in the spliceosome.
(B) The number of identified proteins with at least one specific RNA-binding domain (dark gray) was compared to the respective number of the RNA-binding domain-containing proteins that were identified in the HEK293 proteome (light gray).

A surprising result of the study was the identification of 245 proteins, which were not previously annotated as RNA binding at the time of the study (Figure 16, Table S1, candidate mRNA binders). Eighty percent of these proteins (197 out of 245) were detected in at least two out of three experiments and 47% (115 out of 245) were detected in all three. Less than 20% (48 out of 245) of the candidate mRNA binders were detected in only one proteomic analysis.

In order to assess if those candidate mRNA binding proteins could be predicted by computational analysis, we applied an adaptation of a multiple association network integration algorithm (Mostafavi et al., 2008) to predict proteins with RNA-binding function.

For this analysis we considered gene ontology (GO) data, InterPro and Pfam protein domain data, gene co-expression, protein-protein interaction, and structural similarity data (Drew et al., 2011). The most informative data types for predicting RNA-binding were determined to be GO process, GO localization, and 3D protein structure similarity, followed to a lesser degree by protein-protein interaction data, gene co-expression, and InterPro domains. We applied the algorithm to the candidate mRNA-interacting proteins in a manner that minimizes false negative predictions at the expense of false

positive predictions; yet 112 out of 245 candidate RNA-binding proteins could not be predicted as RNA binders even at a very low precision level R 20% (Table S1). Almost half of the proteins that could not be predicted (48 out of 115) were identified in all three proteomic analyses (class 1 proteins). Those results indicate that the identified proteins interact with RNA by distinct and/or highly divergent protein domains. In addition, the candidate mRNA binders might act largely isolated from known protein association networks connected to RNA-metabolism and –regulation, which might explain the difficulty to classify them as RNA-interacting proteins by the multiple association network algorithm.

Many of the identified candidate mRNA-binders were poorly investigated and annotated at the time of the study, for example the proteins encoded by the open reading frames C1orf35, C16orf80, C11orf31, C9orf114 and C19orf47. On the other hand, we identified proteins that were initially annotated as transcription factors (JUN, NXF1) or putative protein kinases (FASTKD1, FASTKD2, FASTKD5). Their function in RNA metabolism is still enigmatic, but can be explored by PAR-CLIP, also in combination with functional knockdowns, or similar experimental setups.

4.6. Overrepresentation of nucleic acid binding domains

RNA-binding proteins typically interact with their target RNAs through RNA-binding protein domains (RBDs). Most well studied RNA-binding proteins have modular structures, built from a few (and often repetitive) RNA-binding domain blocks. The combination and structural arrangement of RBDs can define the RNA-binding specificity of an RBP (Lunde et al., 2007).

In the following chapter we analyzed the mRNA-bound proteome for enriched protein domains and protein structure folds. We searched for RNA-binding domains, which were annotated at the time of the study, as well as domains, which were not previously shown to mediate RNA-binding, but were enriched in our set mRNA-binders and might thus represent novel RNA-binding protein structures.

To do so, the identified proteins were classified either based on their three-dimensional structure or based on their amino acid sequence. For the structural classification, we first queried the set of mRNA-interacting proteins against the Protein Folding Project database (Drew et al., 2011). The protein folding database provided SCOP (structural

classification of proteins (Murzin et al., 1995)) superfamily classifications derived from sequence similarity (psi-blast), fold recognition, and de novo structure prediction with Rosetta (Rohl et al., 2004) for the identified RNA-interacting proteins. As expected, when we analyzed the mRNA-bound proteome for enriched superfamilies, we detected an overrepresentation of folds associated with single and double-stranded RNA (dsRNA)-binding function (RNA-binding domain, eukaryotic type KH domain, and dsRNA- binding domain-like), helicases (P-loop containing nucleoside triphosphate hydrolases) and nucleases (Pin domain-like) with a corrected p value % 0.05 (Table S2, part A).

Besides the annotated RNA-binding domains, we also found two superfamilies significantly enriched that are associated with DNA binding (“winged helix” DNA-binding domain, present in a number of RNA helicases, and AlbA-like, present in POP7 and in C9orf23, previously described as involved in RNA binding (Aravind et al., 2003)) (Table 1). The significant overrepresentation of these domains in the mRNA-bound proteome strongly suggests that these DNA-binding protein folds could also interact with RNA.

Table 1 Selected SCOP superfamily folds enriched in the mRNA-bound proteome

RNA-binding domains

Domain	*Representative Protein*	*SCOP*	*Corrected p-value*
RBD	PABPC1	d.58.7	9.90E-124
KH (type I)	HNRNPK	d.51.1	1.80E-21
dsRNA	STAU1	d.50.1	2.70E-10
LSM	LSM14A	b.38.1	8.10E-08

Putative RNA-binding domains

Domain	*Representative Protein*	*SCOP*	*Corrected p-value*
HMG box	HMGB1	a.21.1	4.30E-11
“Winged helix” DNA-binding	DDX54	a.4.5	0.0004
AlbA-like	C9orf23	d.68.6	0.0003

The identified proteins were also analysed through Pfam and Inter- Pro domain enrichment analysis, with the aim of obtaining an additional perspective of the mRNA-bound proteome associated structures. Even if most of the significantly enriched domains were the commonly recognized RNA-interaction motifs, we also found an overrepresentation of several domains with putative RNA-binding activity (Table S2, part B), such as the SWAP/SURP domain and the RNA-binding domain abundant in Apicomplexans (RAP) domain, already suggested as RNA-binding, based on sequence comparison (Denhez and Lafyatis, 1994; Lee and Hong, 2004).

We also found a domain of unknown function in our data (DUF_1605). This domain is found towards the C terminus of the DEAD-box helicases, which are involved in unwinding nucleic acids and in other various aspects of RNA metabolism. Further studies employing mutation of the domain and monitoring RNA-binding function might elucidate if the domain is mediating direct RNA-binding.

We repeated the analysis for the over-representation of protein domains for the 112 identified proteins that could not be predicted to bind RNA. Among the 112 identified non-predictable RNA interactors we found the RAP domain significantly enriched (Table S2). In addition, several SCOP superfamily folds were also enriched among the unpredicted proteins.

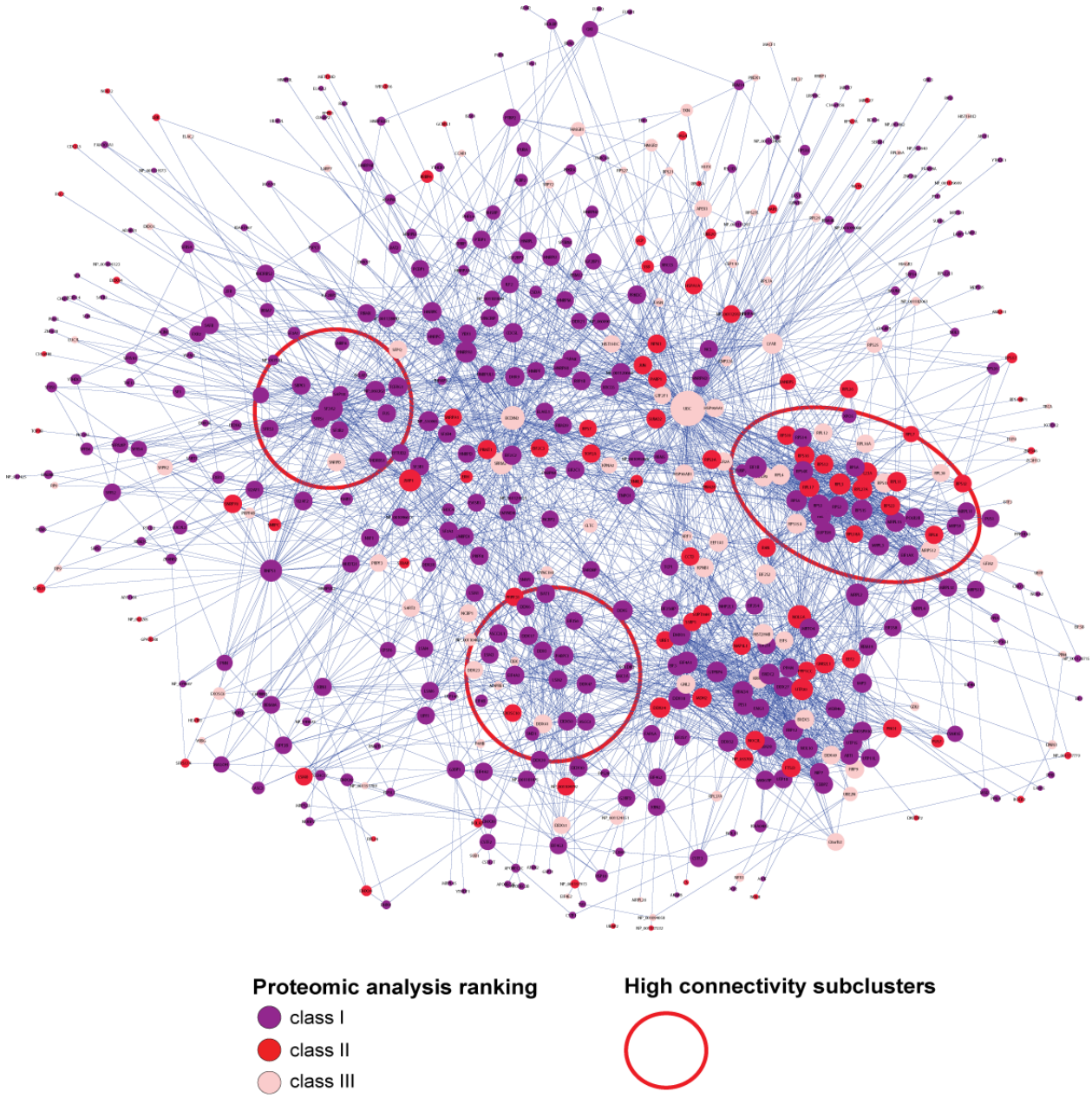
4.7. Connections between posttranscriptional regulation to DNA-related processes

In the cellular environment, proteins often act in complexes to exert their function. In order to systematically examine the connectivity of the identified mRNA binders and their potential connection to biological processes, which are not related to mRNA metabolism and regulation, we generated a network based on protein-protein interactions (PPI) (Figure 19).

The PPI network of the mRNA bound proteome reflects major nodes of highly connected proteins, which are explained by their tight interactions in the cell, such as the proteins of the large ribosomal subunit (RPL) and the small ribosomal subunit (RPS), both accumulated in high density at the right side of the network in Figure 19. Several other PPI clusters of high connectivity are detectable, such as the network of

DEAD box helicases (DDX) and eukaryotic initiation factors (eIF) in the lower part of the network, or the splicing factors and spliceosome proteins (SF/SFRS) on the left.

In general, when comparing the PPI network of mRNA binders to a random network of equal size, we observed a higher average clustering coefficient (Figure 19, lower table). This indicates the presence of highly interconnected protein clusters within the network, which is exemplified by the spliceosome or the ribosome, as described above (Figure 19).



	Average Clustering	Average Degree	Characteristic Path Length
mRNA-bound Proteome	0.166	6.589	2.992
Random Network	0.0466 (0.0126)	1.460 (0.230)	4.695 (0.546)

Figure 19 Protein-Protein Interaction (PPI) Network for mRNA-Interacting Proteins.
Shown is the network of protein-protein interactions between the proteins detected in our assay. Node color denotes number of times the protein was seen (purple = seen in three experiments, red = seen in two experiments, and pink = seen in one experiment). Blue edges indicate previously reported protein-protein interactions. The node size represents the relative connectivity of the respective node in the network. Representative PPI clusters of high connectivity are indicated by red circles.
Table: The average clustering coefficient of a node (number of links observed between the first neighbors of the node divided by the total theoretically observable number of connections between them), average degree (mean number of neighbors per node in the

network) and the characteristic path length (average over the distances between nodes in the network) are shown for the network of RNA-binders compared to a random network of the same size, which was generated by iterative random selection of proteins from a theoretical expressed HEK293 proteome. The standard deviations for the randomly generated networks are shown in parenthesis.

These PPI clusters are indicative of functional modules that mediate and regulate complex biological processes.

We analyzed the set of 797 mRNA binders and their first neighbors in the PPI network for enriched GO terms linked to biological processes (Ashburner et al., 2000). As expected from about 70% annotated RNA interacting proteins in our set, the most significantly overrepresented GO terms were associated with mRNA splicing, localization, processing, and translation (Table 2). Intriguingly, we observed several DNA-related processes, namely ''response to DNA damage,'' ''DNA-dependent transcription,'' and ''DNA duplex unwinding'' being overrepresented in our data.

Table 2: GO-term overrepresentation for the mRNA-bound proteins and their first neighbors (based on the protein-protein-interaction network)

GO ID	*Term*	*Count*	*%*	*p-value*
GO:0008380	RNA splicing	163	7.4	8.2E-64
GO:0006397	mRNA processing	170	7.8	1.2E-59
GO:0006412	translation	142	6.5	2.8E-36
GO:0006974	response to DNA damage	122	5.6	3.6E-36
GO:0006351	transcription	101	4.6	7.4E-18
GO:0050658	RNA transport	44	2.0	1.3E-12
GO:0032508	DNA duplex unwinding	12	0.5	7.0E-06

We analyzed the overrepresented GO-term "response to DNA damage'' in more detail. For that we created a PPI sub-network containing the proteins linked to the term "response to DNA damage". The members of the network and their direct interaction are shown in Figure 20.

XRCC6 (Ku70), XRCC5 (Ku80) and the DNA-activated protein kinase (PRKDC) are central hubs in this network and identified in all three proteomic analyses (Table S1). Ku proteins are known to act as heterodimers and possess the rare capability to act in

different biological processes by two distinct modes of binding. Ku heterodimer the can bind sequence-nonspecific to double-stranded DNA and has an important role in DNA double-strand break repair and recombination. In addition, the Ku proteins have been shown to interact with RNA structures, such as the RNA-stem loop region in yeast telomerase TLC1 and the RNA-component of human telomerase (Ting et al., 2005), which is a prerequisite for nuclear localization of the RNP. The RNA-binding is a conserved function of the heterodimer in yeast, and the mechanism of RNA recognition was recently described (Dalby et al., 2013). In addition, XRCC6 had been suggested to bind internal ribosomal entry site (IRES) elements and is likely involved in the regulation of IRES- mediated mRNA translation (Silvera et al., 2006).

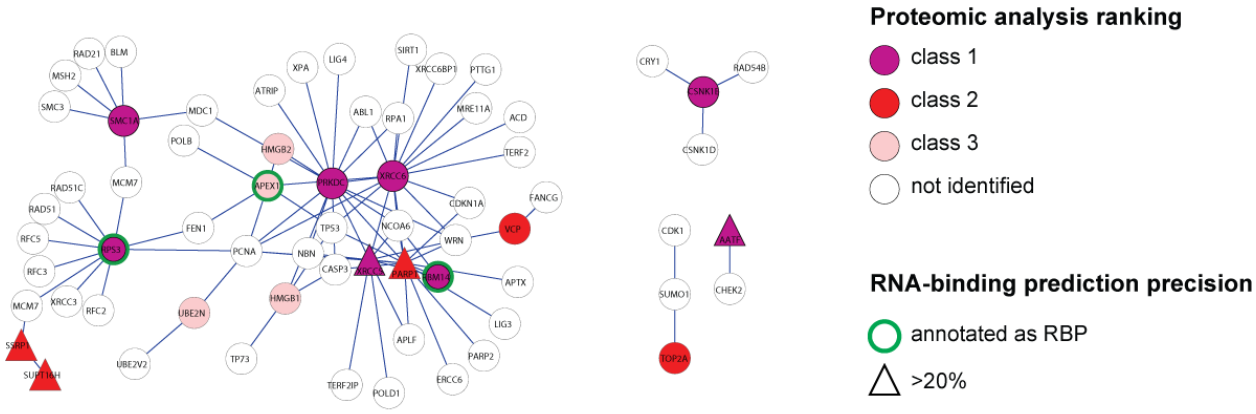


Figure 20: Members of the cluster enriched for "DNA damage response" depicted as nodes in the PPI network. The identified mRNA-bound proteins and their first neighbors based on protein-protein interactions were analyzed for overrepresented Gene Ontology terms. Only proteins in the cluster "DNA damage response" with direct interactions are shown. The node color represents the number of times the proteins were identified in our analysis (purple, detected in all three experiments; red, detected in two experiments; pink, detected in one experiment), and the blue edges indicate direct protein interactions between the cluster members. The node shape indicates the precision of RNA-binding prediction and annotated RNA-binding proteins are marked by green node border color.

When we repeated the analysis for enriched processes and pathways with the 112 identified proteins that could not be predicted to bind RNA, we observed only the GO term "ATP/purine-binding" to be statistically significant enriched.

4.8. Validation of direct RNA-binding function of candidate mRNA interactors

To validate the RNA-binding activity of a subset of the identified proteins, we applied PAR-CLIP. In brief, HEK293 cell lines, stably expressing epitope-tagged proteins, were grown in the presence of 4SU and UV irradiated. Immunopurified and RNase-treated protein-RNA complexes were radiolabeled with T4 polynucleotide kinase, separated by SDS-PAGE, and blotted onto a membrane. The radiolabeled protein-RNA complexes were visualized by phosphoimaging, while protein precipitation was monitored by Western analysis (Figure 21).

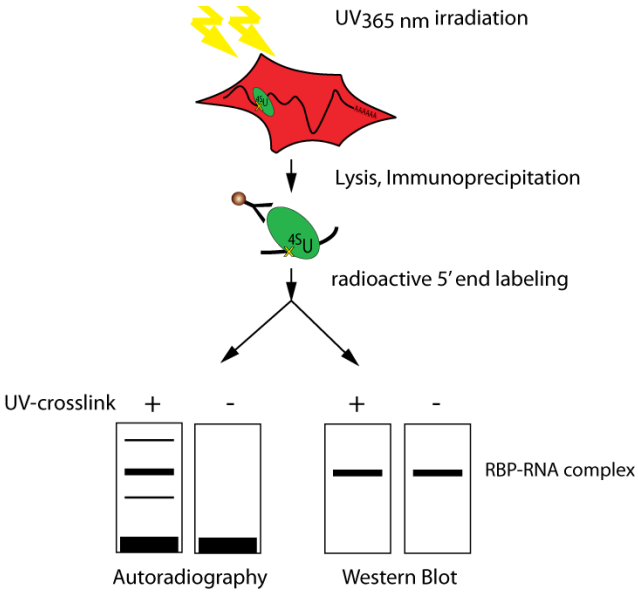


Figure 21: Illustration of the experimental setup to validate mRNA-interacting proteins in HEK293 cells. PAR-CLIP was used to validate direct RNA-binding activity of candidate RBPs. HEK293 cells expressing a FLAG-HA-tagged version of the protein of interest were incubated with the photoreactive nucleoside 4SU and UV-irradiated. Non-irradiated cells served as negative control. Protein-RNA complexes were separated by SDS-PAGE and blotted onto nitrocellulose membrane. Western analysis with an anti-HA antibody confirmed the correct size and equal loading of the immunoprecipitated protein. Phosphorimaging indicated efficient radioactive labeling of covalently bound nucleic acid in the mRNP complex.

We used five known RNA-binding proteins as positive controls: CAPRIN1 (Shiina et al., 2005), HNRNPD (Knapinska et al., 2011), HNRNPR (Hassfeld et al., 1998), HNRPNU (Kiledjian and Dreyfuss, 1992), as well as MYEF2, which is a transcriptional repressor (Haas et al., 1995) with an RNA recognition motif (RRM) domain. As expected, the immunoprecipitated proteins from UV-irradiated cells, efficiently crosslinked to RNA, when compared to proteins that were immunoprecipitated from non-irradiated cells (Figure 22A).

As negative controls, we used two metabolic enzymes that were not identified in our proteomic analyses as potential RNA binders: phosphoglycerate kinase 1 (PGK1) and lactate dehydrogenase A (LDHA). Both proteins are among the top 15% most abundant proteins in HEK293 cells, (based on iBAQ data of 3019 identified proteins). We were not able to detect radiolabeled protein-RNA complexes in immunoprecipitations of both negative controls (Figure 22B), indicating high specificity of our validation protocol.

We generated HEK293 cell lines stably expressing 29 putative mRNA-interacting proteins as epitope-tagged versions. 21 proteins could be immunoprecipitated in the

PAR-CLIP assay (Table S3). We tested the RNA-binding activity of 18 candidate RBPs belonging to class I and three members of class II, BTF3, C16orf80, and PRDX1 (Figure 22). For all proteins, except BZW1 and C16orf80, we observed an increased radioactive signal in immunoprecipitations of irradiated cells when compared to the non-irradiated control population. The increase in radioactivity is most likely caused by bound nucleic acids (Hafner et al., 2010), indicating that these proteins were crosslinked to RNA, and thus are likely in close contact or directly binding to RNA. The relatively strong signal in some of the non-crosslinked samples could on the one hand be explained by background crosslinking due to exposure to daylight and on the other hand by the presence of a kinase in the immunoprecipitated protein/RNA complex that utilizes gamma-ATP to phosphorylate the protein of interest, which can lead to a stronger than expected background signal.

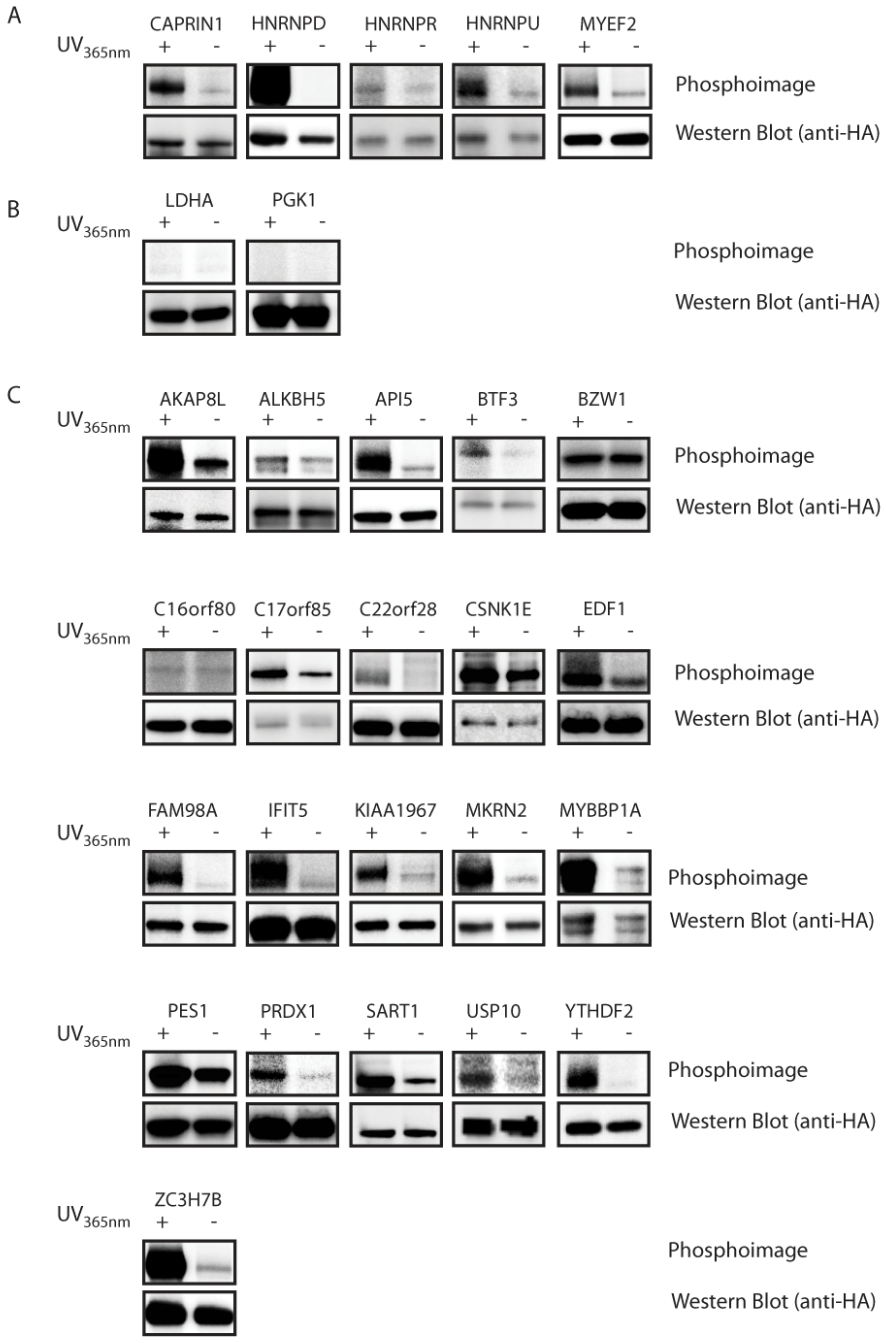


Figure 22: Validation of direct RNA-binding activity of candidate mRNA Binders. RNA-binding activity of candidate mRNA binders was determined by PAR-CLIP as described in Figure 17. The assay was performed at least twice for each protein. Representative results are shown.
(A) RNA-binding proteins CAPRIN1, HNRNPD, HNRNPR, HNRNPU, and MYEF2 were detected in the mRNA-bound proteome and served as positive controls.
(B) Metabolic enzymes LDHA and PGK1, both not detected in oligo(dT) precipitations and highly abundant in HEK293 cells, served as putative negative controls.
(C) Results of PAR-CLIP assay for 21 putative mRNA binders are shown. The radioactive signal in non-crosslinked immunoprecipitates is likely due to the presence of protein kinases and/or background crosslinking. See also Table S3.

The literature related to the identified candidate mRNA-binding proteins assigns diverse functional roles to the candidates. Several proteins possess enzymatic activities: ALKBH5 (2-oxoglutarate oxygenase), C22orf28 (RNA ligase), CSNK1E (kinase), MKRN2 (ubiquitin ligase), PRDX1 (peroxidase), and USP10 (ubiquitin thioesterase). Others have been implicated in transcriptional regulation either by acting as transcription factors (BTF3, MYBBP1A, and EDF1) or by inhibiting histone deacetylases (KIAA1967).

EDF1 (Endothelial differentiation-related factor 1) is a highly conserved protein, which was proposed to regulate endothelial cell differentiation and act as transcriptional co-activator interconnecting regulatory proteins with the basal transcription machinery (Dragoni et al., 1998). EDF1 contains a prokaryotic-type helix-turn-helix motif, suggesting that this protein may function in DNA binding.

To examine the nature of the crosslinked nucleic acid, an immunoprecipitate of UV-irradiated EDF1 protein was divided in three aliquots and each aliquot was incubated with either RNase I, DNase I or no nuclease as negative control. RNase I treatment, but not DNase I, reduced the radioactive signal of the complex, indicating that EDF1 was crosslinked to RNA (Figure 23).

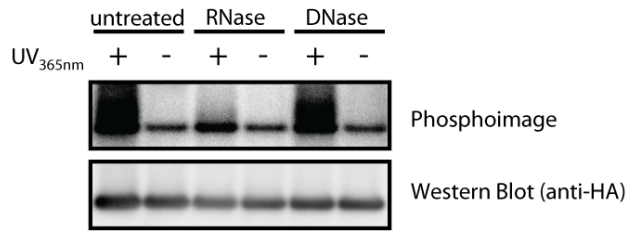


Figure 23: UV_{365nm}-irradiation crosslinks EDF1 protein to RNA. HEK293 cells were stably transfected with HIS/FLAG/HA-EDF1, grown in medium supplemented with 4SU and 6SG and UV-crosslinked. A non-crosslinked control population was processed in parallel. After cell lysis and mild RNase T1 treatment, HIS/FLAG/HA-EDF1 was immunopurified using an anti-FLAG antibody. The precipitate was mildly treated with RNase T1 and washed. Co-precipitated nucleic acid was dephosphorylated using calf alkaline phosphatase (CIP) and labeled with gamma-32P-ATP at the 5'-end, using T4 polynucleotide kinase. The labeled protein-nucleic acid complex was either treated with RNase I, DNase I or incubated without nucleases (untreated) as control. The protein-nucleic acid complexes were resolved by SDS-PAGE and blotted onto nitrocellulose membrane. The radioactive labeling of covalently bound nucleic acid was monitored by phosphorimaging. Equal loading of EDF1 protein was confirmed by Western analysis.

4.9. Identification of RNA-targets and RNA-binding Sites of Candidate mRNA Interacting Protein IFIT5 by PAR-CLIP

To confirm that identified RNA interactors are binding mRNA transcripts, we applied PAR-CLIP in combination with next-generation sequencing (Hafner et al., 2010).

IFIT5 (Interferon-induced protein with tetratricopeptide repeats 5) is up-regulated upon viral infection (Fensterl and Sen, 2011) and interferon treatment (Niikura et al., 1997) and was detected in our screen as candidate RBP. At the time of our study, no mRNA-binding activity for IFIT5 was reported. We therefore set out to identify the RNA-targets and binding sites of IFIT5 by PAR-CLIP.

IFIT5 bind predominantly to 3'UTR and in coding sequences (CDS)

Two biological replicate PAR-CLIP experiments were preformed to identify the RNA targets and binding sites of IFIT5 protein. The PAR-CLIP cDNA sequencing data was analyzed using the PAR-CLIP analysis pipeline, which is described in detail in (Lebedeva et al., 2011). We obtained around 15 million reads per experiment, of which around 1 million mapped uniquely to the genome (Table 3).

Table 3: PAR-CLIP reads obtained from the biological replicate experiments.

PAR-CLIP	3'-adapter	reads	unique reads
IFIT5 4SU 1	NBC6	15.8 M	1.0 M
IFIT5 4SU 2	NBC8	14.7 M	0.9 M

In PAR-CLIP, the RBP binding sites are identified as clusters of sequencing reads, containing the characteristic T-to-C nucleotide transitions. The diagnostic T-to-C transitions were prevalent in the sequencing reads of both biological replicate experiments, confirming efficient labeling and crosslinking (Figure 24).

Uniquely aligning reads were grouped into read clusters with overlapping and adjacent reads and the clusters were scored based on the number of reads and characteristic nucleotide transitions. Clusters mapping antisense to known transcripts were used to estimate the false discovery rate.

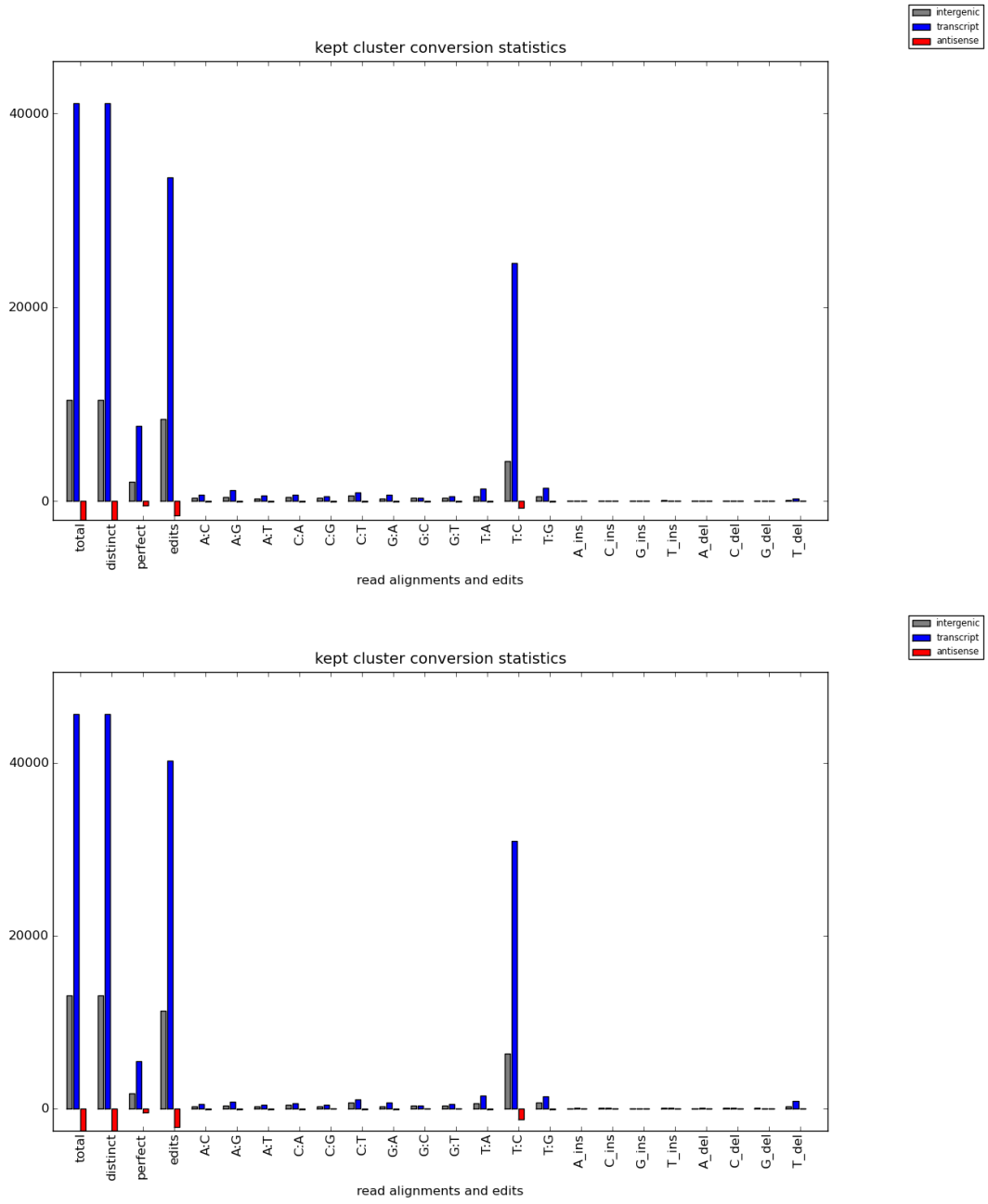


Figure 24: T-C transitions are the prevalent mutation in sequencing reads. Elevated numbers of T-to-C transitions in both biological replicate experiments confirm efficient labeling and crosslinking.

The analysis of the two biological replicates confirmed that IFIT5 protein binds predominantly to the CDS and 3'UTR of target mRNAs (Figure 25). The distribution of binding sites across transcript categories was comparable in both biological replicates (Figure 25A). In the biological replicate experiments, the clusters mapped to 3831 target genes (experiment 1) and 3490 target genes (experiment 2). 2231 target genes

were identified in both, indicating high reproducibility between biological replicates (figure 25B).

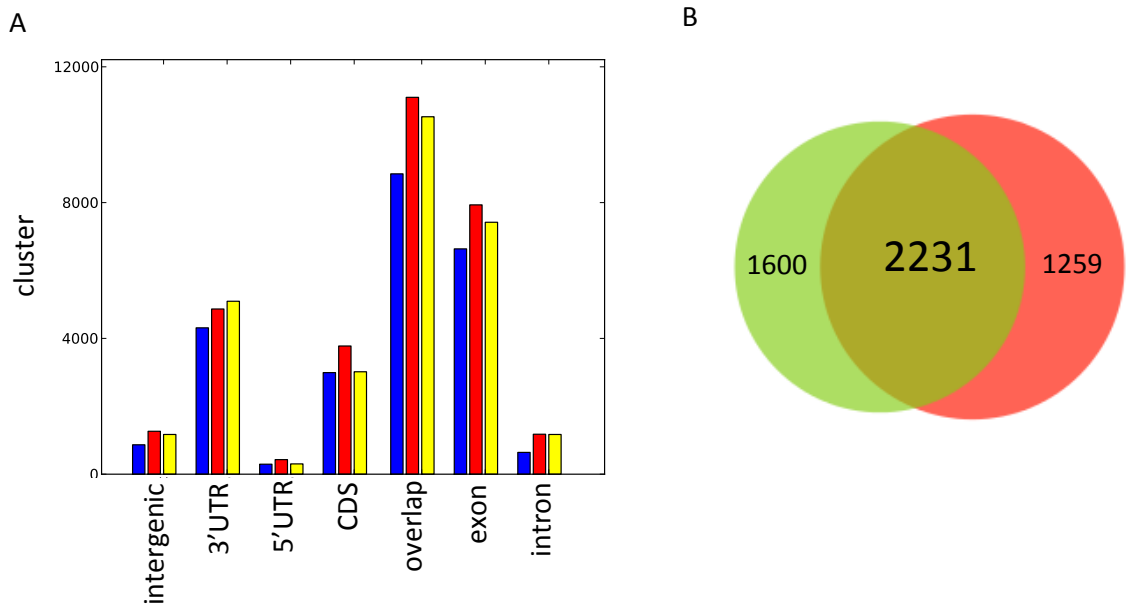


Figure 25: IFIT5 protein binds mRNA predominately in CDS and 3'UTR.
(A) Distribution of IFIT5 binding clusters across transcripts. The clusters derived from IFIT5 PAR-CLIP 1 (red), PAR-CLIP 2 (yellow) and combined from both biological replicate experiments (blue).
(B) Overlap of the target genes identified in the two biological replicate PAR-CLIP experiments (green: PAR-CLIP 1, red: PAR-CLIP 2).

For the following we used a conservative assumption, considering only IFIT5 clusters that contained reads from both biological replicate experiments and at least one read from each library with at least one T-to-C transition. Using these conservative criteria, we identified 9342 IFIT5 clusters (cutoff 5% FDR), of which 7203 could be mapped to 3176 target genes. 2139 could not be annotated to Refseq genes (hg18). A list of the top 200 IFIT5 targets based on the cluster frequency is shown in Table S4.

IFIT5 mRNA target preferences

Out of the 3176 significant mRNA targets we selected the top 200 transcripts and analysed them for overrepresented Gene Ontology (GO) terms using DAVID (Huang da et al., 2009). IFIT5 target genes had significantly enriched annotated functions in mRNA processing (splicing), regulation of translation and intracellular protein transport. In

addition, IFIT5 seems to bind preferentially to mRNAs encoding for metabolic enzymes and proteins involved in regulation of apoptosis.

IFIT5 was also found to bind its own mRNA, indicating a possible feedback loop. The most significantly enriched GO-terms associated with IFIT5-targets are shown in Table S5.

A similar analysis performed for all 3176 significant targets yielded almost identical GO-term enrichment (data not shown).

IFIT5 binds preferentially to AU-rich motifs

We performed a motif-enrichment analysis for the IFIT5 binding sites. For this analysis we searched for enriched 7-mers in a 40-nucleotide window around the crosslink centered region and compared their abundance in order to find indications for their binding preferences. We found U-rich motifs to be predominant among IFIT5-bound sequences and AU-rich 7-mers were highly enriched. Figure 26 shows the highest enriched 7-mers around the crosslink site.

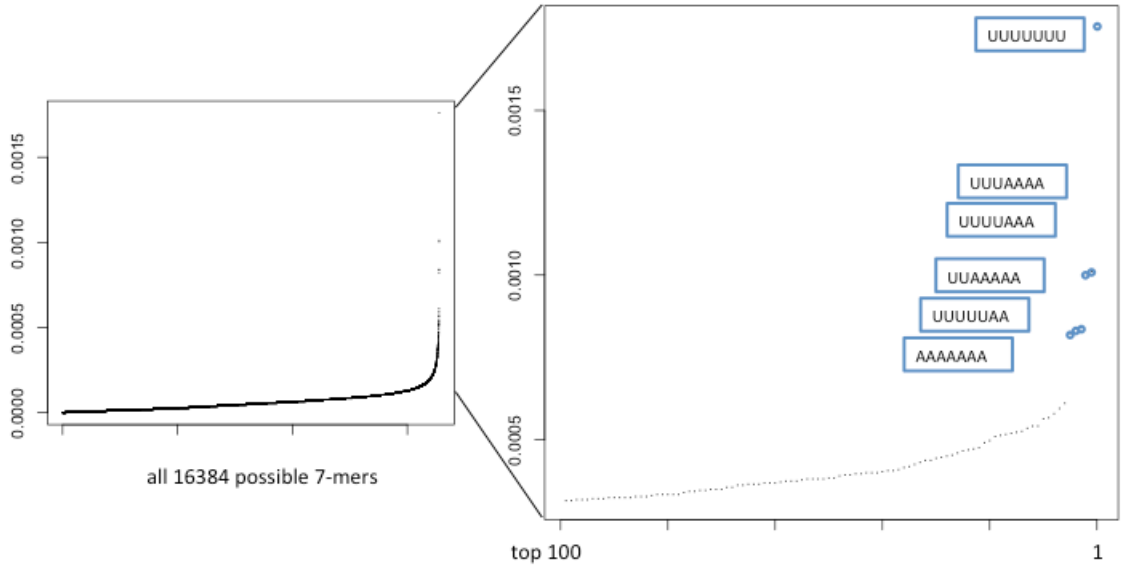


Figure 26 Enriched 7-mers around crosslink site. A 41-nucleotide window around the crosslink site was scanned for enriched 7-mers out of 16384 possible combinations (left: distribution of all 7-mers, right top 100 enriched 7-mers) The top three 7-mers enriched in the PAR-CLIP are UUUUUUU, UUUAAAA, and UUUUAAA.

Putative IFIT5 binding sites showed an enrichment for AU-rich 7-mers. Interestingly, we found an A-rich 7-mer (AAAAAAA) suggesting that IFIT5 might bind to the polyA-

site, or to the to the AU-rich polyadenylation signal (AAUAAA), which is in close proximity to polyA-tail.

We analyzed the binding pattern of IFIT5 in the 3'UTR and the CDS separately (Figure 27) and observed different binding distributions, depending on the transcript compartment. IFIT5 seems to avoid binding to the proximity of the start and the end of the CDS. However, when bound to the 3'UTR, IFIT5 binds preferentially in proximity to the CDS and we could not detect a binding preference close to the terminal end of the 3'UTR, which would indicate binding to the polyadenylation signal.

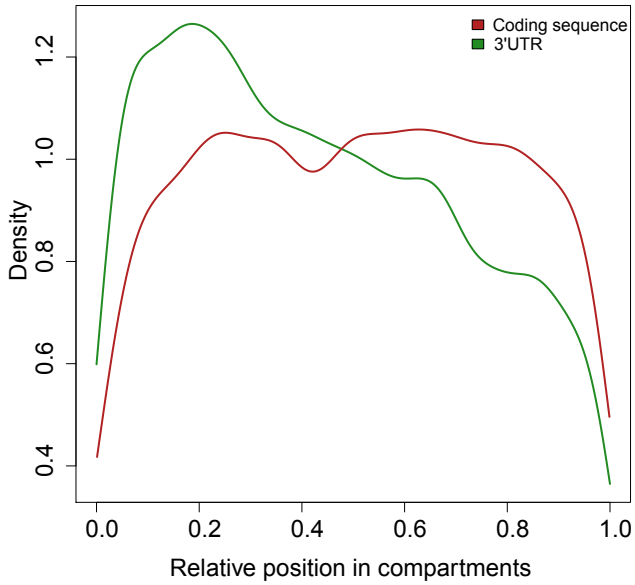


Figure 27: Densitiy of IFIT5 binding sites over transcript compartments. The position of the preferred crosslinking sites (as determined by T-to-C transition frequency) was determined within the CDS or 3'UTR. 0 is hereby the most 5'end of the respective compartment and 1 indicates the position at the very end.

IFIT5 binds to tRNA

Since IFIT5 was reported to interact with tRNA transcripts (Hogg and Collins, 2007), we mapped the clusters to the UCSC-track "tRNA_subgroup" (identified with repeatmasker (Smit et al., 1996-2010))

We identified overlap with 334 out of 1956 tRNA genes (with at least one cluster mapping), indicating substantial binding to tRNA. For example, 9 out of 37 iMet-tRNAs were bound with at least one cluster.

Our results are in agreement with previous publications, which propose a tRNA binding function for IFIT5 (Hogg and Collins, 2007; Katibah et al., 2013). Indeed, one of the most significant clusters identified in the combined PAR-CLIP data mapped to tRNA (tRNA-Ile-ATA).

Consensus motif discovery in top500 clusters

We analyzed the top500 ranked clusters (about 5% of the 9342 clusters) for overrepresented motifs using the MEME suite (Bailey et al., 2009). The analysis was limited to the 500 clusters because of the limited input character space of the MEME suite. The maximum width was chosen based on the average cluster length for the top 500 clusters analyzed (72.5 nt). Two binding motifs were discovered as highly enriched (Figure 28C,D). Interestingly, the motif was mainly found in clusters binding to tRNAs. A close inspection of the motif and the structural features of tRNA secondary structure revealed that the motifs overlapped with two conserved structural components of tRNAs, the TψC-loop and the D-loop (figure 28B). The characteristic sequence in the consensus motif and the respective positions in the tRNA "cloverleaf" shaped secondary structure are highlighted. This enrichment is expected due to the structural conservation of tRNA, requiring a L-shaped tertiary structure for proper function in translation. The crosslinking site, as determined by the T-to-C transition, was hereby not restricted to a specific location on the tRNA molecule, indicating that IFIT5 can contact tRNA on several positions. Interestingly, crosslinking events were also observed outside of the annotated mature tRNA transcript. These additional sequences are cleaved from the pre-tRNA transcript during tRNA maturation, indicating a possible role for IFIT5 in tRNA biogenesis regulation.

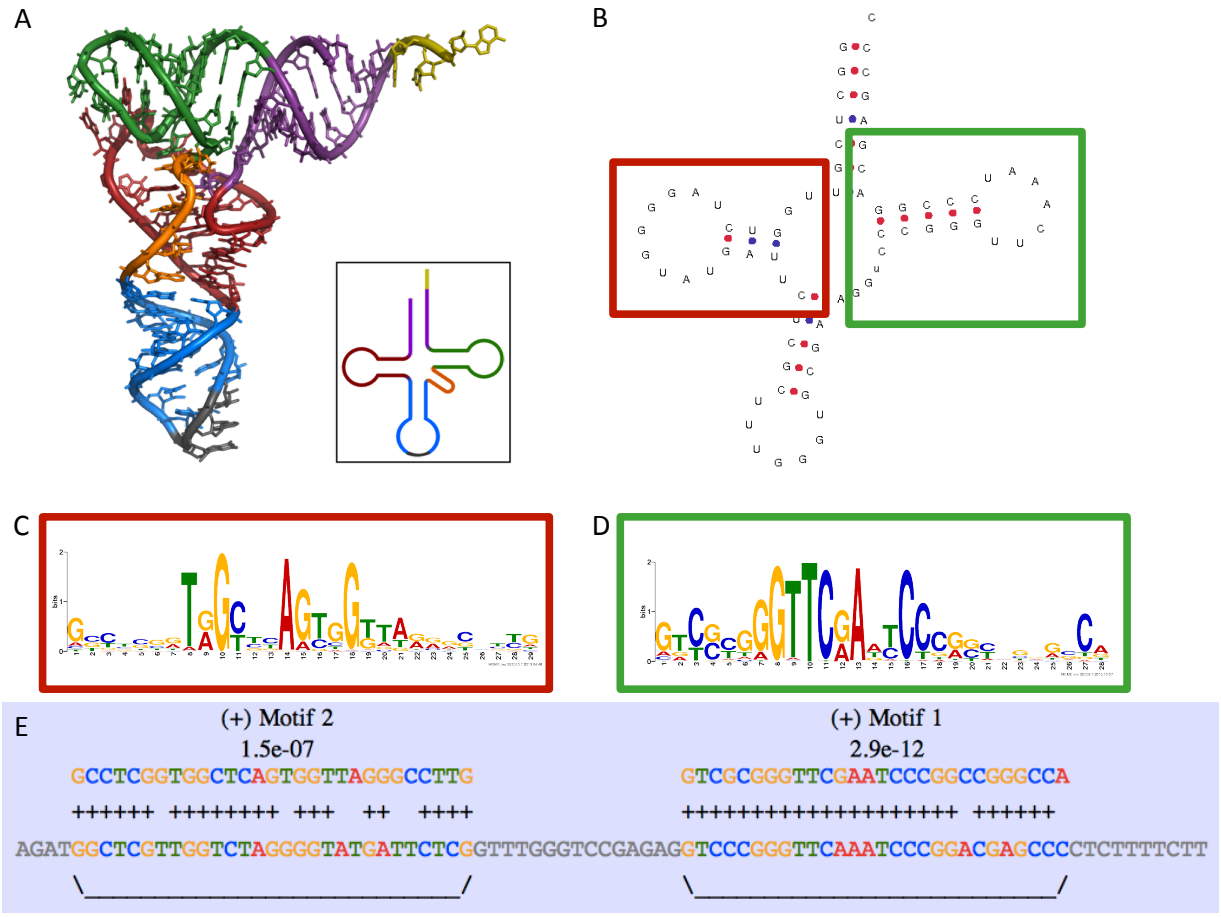


Figure 28: Overrepresented motifs represent structural features of tRNA.
(A) tRNA tertiary structure (tRNA-Phe, from yeast) and schematic secondary structure forming the "cloverleaf". The structural elements are: *CCA tail* in yellow, *Acceptor stem* in purple, *Variable loop* in orange, *D arm* in red, *Anticodon arm* in blue with *Anticodon* in black, *T arm* in green.
(B) secondary structure of the human proline tRNA (tRNA-ProTGG)
(C)(D) highest enriched binding motifs among top 500 IFIT5 binding site clusters.
(C) motif 2 (width 29nt, e-value 1.6e-232, found in 120 sites), represents *D arm*
(D) motif 1 (width 28nt, e-value 5.2e-612, found in 155 sites): represents *T arm.*
(E) mapping of the 2 motifs in cluster (CID-001888) which covers the tRNA12-ProTGG

In addition to tRNA binding, we found other non-coding RNAs bound by IFIT5 and accordingly, the other motifs detected by MEME mapped to rRNA and vaultRNA (Kickhoefer et al., 1993).

In order to identify a consensus motif on mRNAs, which might have escaped the previous analysis due to the overrepresentation of tRNA- and rRNA-related motifs, we repeated the MEME analysis, only considering clusters that mapped to mRNA. No significantly overrepresented motif was detected besides the one obtained from the clusters that mapped to the histone RNA 3'UTR, which is not surprising, due to the high sequence conservation of the histone RNA 3'UTR loop (Davila Lopez and Samuelsson, 2008).

4.10. Identification of RNA-binding sites of several candidate mRNA interactors

In addition to IFIT5, we performed PAR-CLIP experiments for five additional proteins, which were identified in the protein-mRNA interactome study:

ALKBH5, C22orf28, C17orf85, ZC3H7B and the known RNA-binding protein CAPRIN1 (Baltz et al., 2012). As for IFIT5, all PAR- CLIP sequencing data were analyzed with a computational analysis pipeline (Lebedeva et al., 2011) and the diagnostic T-to-C transitions in aligned reads demonstrated efficient RNA-protein crosslinking. The PAR-CLIP experiments confirmed that the five tested proteins bound predominately to mRNA. The mRNA targets for the respective proteins are summarized in (Baltz et al., 2012) and the binding sites are available at http://dorina.mdc-berlin.de/cgi-bin/hgTracks/ (Anders et al., 2012). In addition, we confirmed some of the interactions of these proteins with their top mRNA targets by RNA immunoprecipitation (RIP) coupled to semiquantitative RT-PCR (Baltz et al., 2012), Supplemental Figure S5B).

The combined PAR-CLIP data of IFIT5 and the other tested proteins strongly indicate that the oligo(dT) based precipitation truly captures mRNA-binding proteins. Furthermore, the PAR-CLIPseq data is a rich resource for further characterization of the newly discovered RNA-binding proteins.

4.11. Summary of the mRNA-bound proteome of a human cell line

We developed a photoreactive nucleotide-enhanced UV crosslinking and oligo(dT) purification approach to identify the mRNA-bound proteome using quantitative proteomics. Application to a human embryonic kidney cell line identified close to 800 proteins. Nearly one-third of those were not annotated as RNA binding at the time of the study, and about 15% were not predictable by computational methods to interact with RNA. We confirmed direct RNA-binding of six candidate proteins by PAR-CLIP and identified mRNA-targets and binding sites by PAR-CLIPseq. Our observations indicate the presence of a large number of mRNA binders with diverse molecular functions participating in posttranscriptional gene-expression networks.

4.12. Identification of mRNA-interacting Proteins in *D. melanogaster* Embryos

I aimed to adapt the developed method to study protein-RNA interactions *in vivo* and in a defined developmental stage and chose early stage *D. melanogaster* embryos, before MZT, as a reference point to study the dynamics of the mRNA-bound proteome during development and differentiation of a complex organism.

4.13. Complementary UV-Crosslinking Strategies to study *in vivo* Protein-RNA Interactions

We chose two UV-crosslinking strategies to covalently bind proteins to RNA:

- PAR-crosslinking using UV light at 365nm and photoreactive 4-thiouridine
- Conventional-crosslinking with UV light at 254nm

Both UV crosslinking methods have their advantages and disadvantages. In brief UV_{365nm} crosslinking causes less damage to nucleic acids and shows a higher crosslinking efficacy for a number of proteins (Hafner et al., 2010). In addition when used for PAR-CLIPseq, the T-to-C transition allows precise determination of the crosslink-site and the possibility to distinguish crosslinked RNA from background. The same is true for transcriptome-wide protein occupancy profiling studies as described in (Baltz et al., 2012).

UV_{245nm} crosslinking on the other hand does not require labeling with photoreactive nucleosides prior to crosslinking, which is favorable when photoreactive nucleoside labeling is inefficient or impossible. Lower crosslinking efficacy and increased nucleic acid damage have to be considered for this alternative approach.

Taken into account the possibilities and limitations, we decided to use a dual approach by using both PAR-CL and cCL to define the mRNA-bound proteome in *D. melanogaster* embryos.

4.14. Labeling of *D. melanogaster* mRNA in adults and embryos with photoreactive nucleosides (TU-tagging)

For PAR-CL we first needed to label RNA with photoreactive nucleosides before UV_{365nm}-irradiation. Our approach to label RNA with 4-thiouridine is based on a protocol known as ThioUridine-tagging (TU-tagging), which was developed in *D. melanogaster* (Miller et al., 2009) and recently applied to mouse (Gay et al., 2013). TU-tagging allows tissue specific labeling of RNA with photoreactive 4-thiouridine.

The method is based on the *Toxoplasma gondii* nucleotide salvage enzyme uracil phosphoribosyltransferase (UPRT), which catalyzes the conversion of uracil and 5-phosphoribosyl-1-R-diphosphate to uridine monophosphate (UMP) (Cleary et al., 2005). In TU-tagging, the nontoxic photoreactive uridine precursor 4-thiouracil is supplied with the food and subsequently incorporated in the organism.

UPRT then converts 4-thiouracil to 4-thiouridine monophosphate, which is the active compound incorporated into newly synthesized RNA (Miller et al., 2009). TU-tagging allows tissue-specific conversion of the photoreactive nucleoside precursor 4-thiouracil to 4-thiouridine by spatially restricted uracil phosphoribosyltransferase (UPRT) expression. 4-thiouridine is thus only incorporated into nascent RNA of the UPRT-expressing tissue. We reasoned that the combination of tissue specific labeling of RNA followed by UV_{365nm} crosslinking enables selective tissue-specific crosslinking of RBPs to RNA.

4.15. The UPRT-fly

Since the early stage embryo before MZT is largely transcriptionally silent, we searched for a way to label the maternal RNA already in the developing oocyte.

In early stages of oocyte development, the maternal factors are delivered from the nurse cells to the developing oocyte via ring canals (Becalska and Gavis, 2009). At stage 10 of oogenesis, the nurse cells contract to extrude their contents into the oocyte, accompanied by ooplasmic streaming, mixing the incoming nurse cell cytoplasm with the ooplasm. We reasoned that tissue-specific UPRT expression would allow labelling of mRNAs with 4SU in nurse-cells. Because most transcripts in the oocyte and early embryo are maternal and produced in the nurse-cells (Becalska and Gavis, 2009), we assumed that this approach would allow labelling embryo RNA with photoreactive 4SU which is in turn a prerequisite for efficient protein-RNA crosslinking by UV_{365nm}-irradiation.

I created a fly strain that expresses UPRT in the nurse cells in order to label RNA with 4-thiouridine in the developing oocyte. The 4-thiouridine labeled RNA in the nurse cells will then be transferred into the developing oocyte during oogenesis. Labeling maternal RNA in the oocyte would therefore allow photoreactive nucleoside labeling of embryonic RNA. Our experimental setup to label RNA with 4-thiouridine is described in Figure 29. The GAL4-UAS system was used to express UPRT in nurse cells under the control of the nurse-cell specific nanos promoter (Phelps and Brand, 1998; Tracey et al., 2000).

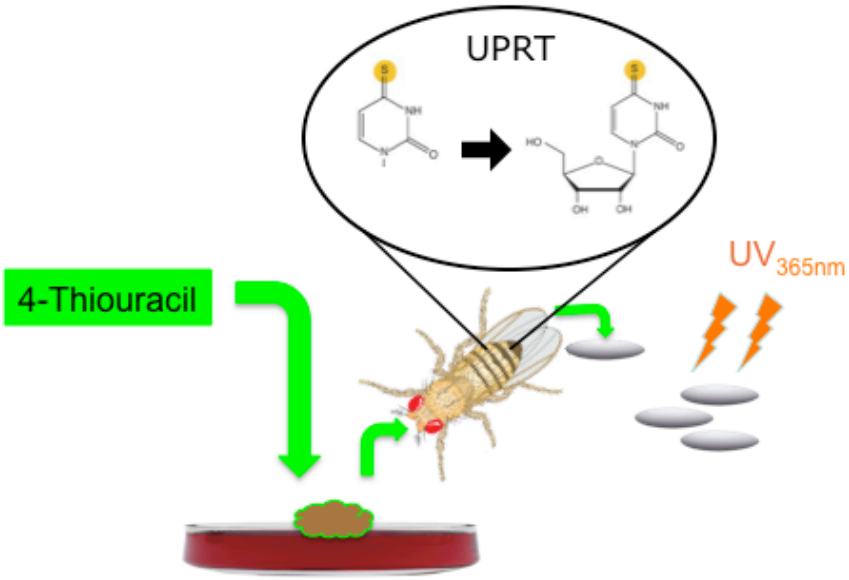


Figure 29: Experimental setup for photoreactive nucleoside based UV-crosslinking in *D. melanogaster* embryos. 4-thiouracil containing yeast paste was fed to adult flies. After several days of feeding, 4-thiouracil is incorporated in the adult fly. Nurse cell specific UPRT

expression converts the precursor 4-thiouracil to 4-thiouridine in nurse cells of adult flies. 4-thiouridine is incorporated in newly synthesized RNA in the nurse cells. Nurse cells transfer RNA to the developing oocyte via ring canals. Since the nurse cells release their content into the oocyte, labeling RNA in nurse cells allows labeling of oocyte RNA and thus maternal embryonic RNA. After egg laying the 0-2h old embryos are collected and UV_{365nm}-irradiated in order to covalently bind proteins to RNA for further analysis.

4.16. *D. melanogaster* embryos can be efficiently labeled with photoreactive nucleosides

We monitored the incorporation of photoreactive nucleosides into adult and embryonic RNA with two independent methods. First, we used an assay that detects thionucleosides, to test the incorporation of 4SU into *D. melanogaster* RNA. RNA that contains 4-thiouridine is biotinylated at the thiol-group, immobilized on a membrane and detected by streptavidin-hybridization (as described in (Dolken et al., 2008)). For our initial labeling experiments we used photoreactive nucleoside concentrations similar to those employed for TU-tagging (1 mM 4-thiouracil)(Miller et al., 2009) and increased the concentration until we detected a signal in embryonic RNA, since the labeling efficiency is much lower than for HEK293 cells (Hafner et al., 2010) or *C. elegans* (Jungkamp et al., 2011).

We probed embryonic RNA from UPRT-flies grown on yeast medium containing 4-thiouracil, as well as embryonic RNA extracted from *yw*-flies grown on yeast medium with increasing amounts of 4-thiouridine (Figure 30).

Both 4-thiouracil feeding to UPRT-expressing flies and 4-thiouridine feeding to *yw*-flies, allowed photoreactive nucleoside incorporation into embryonic RNA (figure 30). We could improve the photoreactive nucleoside incorporation into embryo RNA by increasing the nucleoside concentration in the fly food. Based on the intensities observed in the dotblot, we assume that feeding the UPRT-fly with approximately twice the amount of 4-thiouracil allows a labeling efficacy similar to the RNA from the yw fly fed with 4-thiouridine, probably because 4-thiouracil is not completely converted into 4-thiouridine by UPRT.

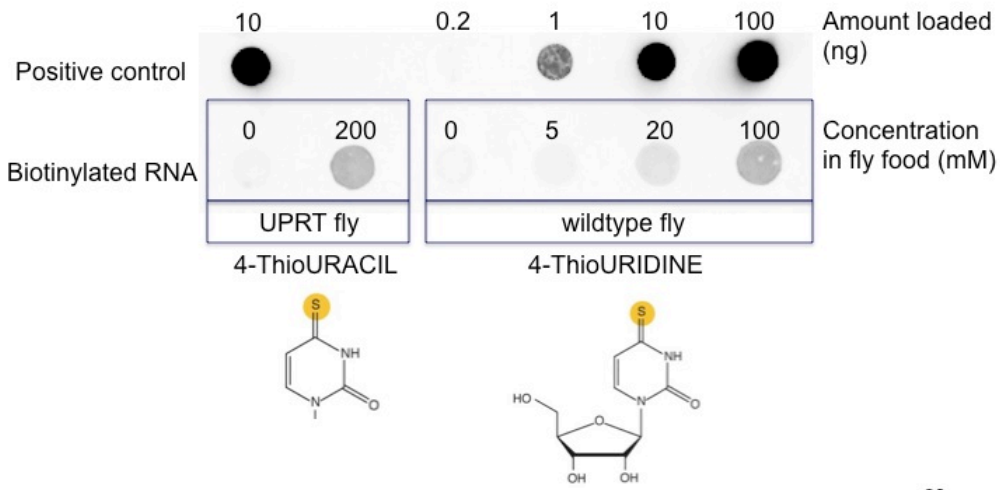


Figure 30. Labeling of nascent RNA with photoreactive nucleosides in *D. melanogaster* embryos. Dotblot assays with thiol-specific biotinylated total RNA extracted from *D. melanogaster* embryos. Signal was compared to a biotinylated DNA-oligo positive control. Adult flies were labeled with different concentrations of 4-thiouridine (yw-fly, "wildtype") or 4-thiouracil (UPRT-fly). Equal amounts of RNA extracted from embryos were spotted on the membrane. Labeling efficiency increases with higher concentrations of photoreactive nucleosides in the fly food.

We analyzed total RNA, extracted from embryos and adult flies by a LC-MS/MS-based method that allows to quantitatively measure uridine and 4-thiouridine and thus to calculate how many uridine bases in the labeled RNA are substituted by 4-thiouridine. With this method, we determined that 0.1% of all uridine bases were replaced by 4SU when we labeled with 100 mM 4-thiouridine.

4.17. PAR-Crosslink (UV_{365nm} irradiation)

After establishing photoreactive nucleoside labeling of embryonic RNA, I searched for suitable conditions for UV-crosslinking of RBPs to labeled transcripts in *D. melanogaster* embryos and noticed, that with increased UV-irradiation energy, more proteins were detected in oligo(dT) precipitates (data not shown). Experiments with increasing doses of UV-irradiation were performed in order to determine the optimal amount of UV-irradiation. For large-scale experiments, we chose 4J/cm², a value that allowed successful *in vivo* UV-crosslinking in *C. elegans* (Jungkamp et al., 2011) as well as *D. melanogaster* embryos. Under the conditions used for labeling (400 mM 4-thiouracil in yeast paste) and UV-irradiation (4J/cm²), almost none of the crosslinked embryos survived and developed into larvae. In this viability assay, we transferred UV-irradiated embryos to vials and counted how many embryos survived and developed into larvae

compared to a non-irradiated control. The survival rate of non-irradiated but 4SU-labeled embryos was comparable to the non-labeled control population. We therefore reasoned, that the 4SU-labeled and UV-irradiated embryos died because of covalent crosslinking of proteins to the 4SU-containing transcripts.

We chose to proceed with the isolation of mRNA-interacting proteins from *D. melanogaster* embryos applying the same protocol as used for HEK293 cells. Minor modifications were introduced to optimize the protocol for *D. melanogaster* embryo lysates, such as homogenization of the embryo lysate with a glass homogenizer and clearance by centrifugation prior to oligo(dT) precipitation (Figure 31).

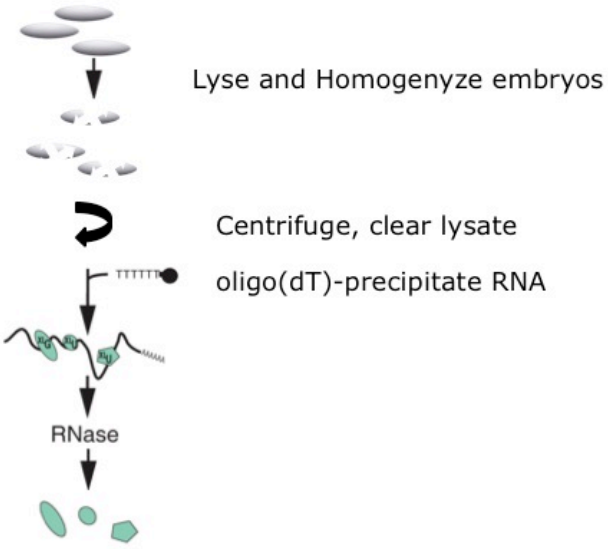


Figure 31: Experimental setup to isolate mRNA-bound proteins from UV-irradiated embryos. Embryos were homogenized and mRNA-bound proteins were isolated from the cleared lysate by oligo(dT)-based precipitation. Protein-RNA complexes were nuclease digested and analyzed by SDS-PAGE or Western analysis.

I performed a large scale oligo(dT) precipitation experiment using ~2600 mg of each UV-crosslinked and non-crosslinked fly embryos. After three consecutive rounds of oligo(dT) precipitations, the precipitates were analyzed by SDS-PAGE and Western blot (Figure 32). A characteristic protein pattern was observed in the precipitates of UV-irradiated embryos, whereas the precipitates from non-irradiated embryos were remarkably clear. The largest amount of protein was precipitated during the first two oligo(dT) precipitation rounds, indicating that three consecutive precipitations were sufficient to precipitate most of the mRNA-bound proteins from the embryo lysate. The protein pattern differed substantially from the input sample (embryo lysate before

oligo(dT) precipitation), indicating that the method is not biased to precipitate highly abundant proteins.

I probed the precipitates for Aubergine, a known RNA-binding protein, which is expressed in the early embryo. Western analysis against the endogenous RBP confirmed its presence in the precipitate. In agreement with the silver staining result, the majority of the mRNA-bound protein was precipitated in the first oligo(dT) pull-down, indicating that three consecutive rounds of oligo(dT) precipitations are sufficient to isolate most of the mRNA bound Aubergine protein from the embryo lysate (Figure 32).

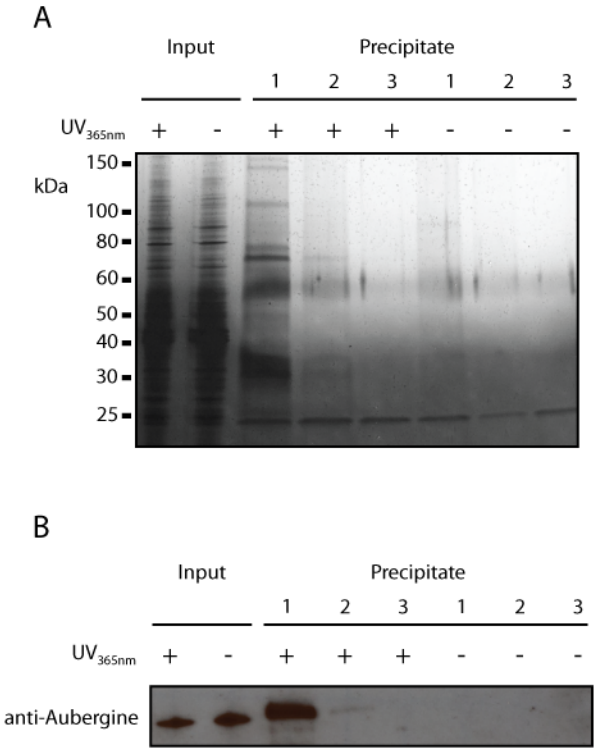


Figure 32 SDS-PAGE analysis of proteins crosslinked to polyadenylated RNA.
D. melanogaster adult flies were fed with brewer's yeast supplemented with photoreactive 4-thiouridine. 0-2h old embryos were collected and irradiated with UV-light at 365 nm. Non-irradiated embryos served as negative control. Cells were lysed under denaturing conditions, and polyadenylated mRNA was isolated by three consecutive oligo(dT)- precipitations. Protein-RNA complexes were eluted from oligo(dT) beads, nuclease treated and separated on a SDS gradient gel.
(A) Proteins were visualized by silver staining.
(B) Western analysis of the endogenous RNA-binding protein Aubergine.

These results indicate that RBPs can be covalently crosslinked to mRNA in *D. melanogaster* embryos and isolated by oligo(dT)-precipitation.

4.18. Conventional Crosslink (UV254nm irradiation)

Despite successful labeling of *D. melanogaster* embryo RNA with photoreactive nucleosides, we were concerned that the labeling efficiency might be the bottleneck for successful protein-RNA crosslinking, since the uridine to 4-thiouridine substitution rate was lower than in HEK293 cells (~2.5% (1 in 40 uridines is substituted by 4-thiouridine) (Hafner et al., 2010)) and *C. elegans* (about 1%, (Jungkamp et al., 2011)). Therefore, we decided to analyze the mRNA-bound proteome with both PAR-CL and "conventional" crosslinking (cCL). The experimental setup resembles the PAR-CL based protocol, but with the following modifications:

- *yw*-flies were used instead of UPRT-expressing flies
- Flies were fed with standard brewers yeast without addition of photoreactive nucleosides.
- Protein-RNA crosslink is performed by UV-irradiation at 254nm (instead of 365nm).

I performed a large-scale experiment using the cCL-protocol and analyzed the precipitates by SDS-PAGE and Western analysis. Again, a characteristic pattern of proteins was detected by SDS-PAGE and the presence of the RNA-binding protein Aubergine could be confirmed in the precipitates of UV-irradiated embryos, but not in the non-irradiated negative control (Figure 33).

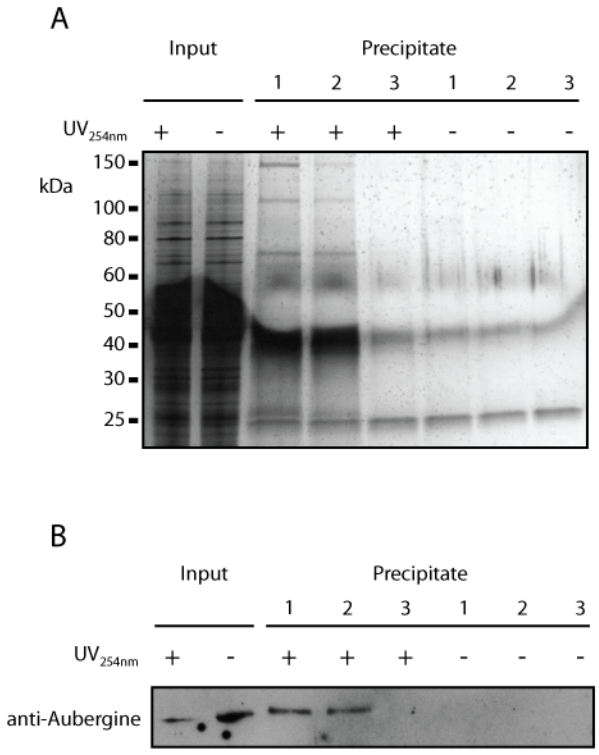


Figure 33: Analysis of proteins crosslinked to polyadenylated RNA.
D. melanogaster adult flies were fed with brewer's yeast medium. 0-2h old embryos were collected and irradiated with UV-light at 254 nm. Non-irradiated embryos served as negative control. Cells were lysed under denaturing conditions, and polyadenylated mRNA was isolated by three consecutive oligo(dT)-bead based precipitations. Protein-RNA complexes were eluted from oligo(dT) beads, nuclease treated and separated on a SDS gradient gel.
(A) Proteins were visualized by silver staining. 0.002% of input and 6% of each precipitate were analyzed.
(B) Western analysis of the endogenous RNA-binding protein Aubergine.

After two rounds of oligo(dT) precipitations, no Aubergine protein could be detected in further precipitations by Western analysis, indicating that the majority of mRNA-bound Aubergine protein is precipitated by three consecutive rounds of oligo(dT) precipitations.

4.19. Comparison of PAR-CL with cCL

The nuclease-digested eluates from three consecutive precipitation rounds were pooled and concentrated by ultrafiltration and analyzed by SDS-PAGE. Silver staining revealed a characteristic protein pattern with remarkable similarity in the precipitates of both protocols (Figure 34). However, some proteins seem to crosslink more efficient,

depending on the UV-crosslinking method used. Almost no protein could be detected in the control samples from non-irradiated embryos.

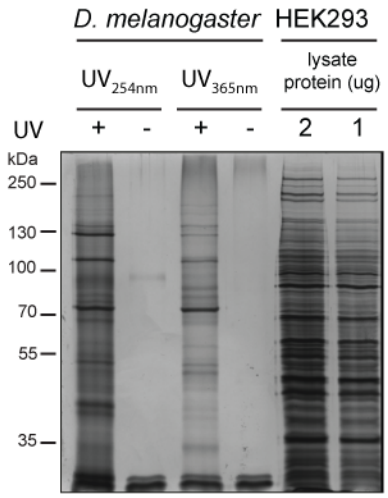


Figure 34: Direct comparison of precipitated proteins from UV_{254nm}-irradiated and UV_{365nm}-irradiated embryos. Embryos from *yw*-flies fed on standard brewers yeast medium were irradiated with UV-light at 254nm, whereas embryos from UPRT-expressing flies, fed with 4-thiouracil containing yeast paste, were irradiated with UV-light at 365nm. Two reference HEK293 protein lysate samples were loaded as control to estimate the amount of protein on the gel. The protein amount loaded on the gel equals 10% of the sample volume used for the mass spectrometry analysis test run.

4.20. Identification of mRNA-bound Proteins in early *D. melanogaster* Embryogenesis by Mass Spectrometry Analysis

We used 25 % of the precipitated protein from each of the cCL samples (*yw*-flies + UV_{254nm} crosslink and *yw*-flies without crosslink) for an initial mass spectrometry test run.

Due to time constraints, we were only able to measure one condition in an initial mass spectrometry test run. We chose to measure the UV_{254nm} crosslinking approach first, because we observed more protein in the precipitate and thus expected to identify a higher number of proteins from the 25 % test sample. Since we cannot support the preliminary results with a biological replicate experiment or the complementary UV_{365nm} crosslinking approach, we have to be careful in interpreting the data and drawing conclusions. Nevertheless, since our and similar protocols proved robust in HEK293 cells (Baltz et al., 2012), HeLa cells (Castello et al., 2012) and yeast (Mitchell et

al., 2013), we give a brief overview of the identified proteins, which might be representative for the mRNA-bound proteome of *D. melanogaster* embryos before MZT.

We identified 1269 proteins with at least two unique peptides and assayed the protein quantity in the samples by intensity based absolute quantification (iBAQ). We compared the iBAQ values for the identified proteins in both experiments and calculated the enrichment in the UV-irradiated sample by dividing over the non-irradiated control (Figure 35). A stringent cutoff was set, requiring at least 10-fold enrichment in the UV-irradiated sample compared to the non-irradiated control. Out of 1269 identified proteins, 371 proteins fulfilled this criterion and 659 proteins were only detected in the UV-irradiated sample but not in the control, meaning that those proteins were also highly enriched, but no ratio could be calculated. Among the remaining 239 proteins, which were either not detected in the UV-irradiated sample at all, or with a low ratio (< 10-fold), were mostly common contaminants (64 proteins, mainly keratins), heat-shock proteins and metabolic enzymes.

We combined the 371 proteins which were 10-fold enriched in the UV-irradiated sample with the 659 proteins which were detected only in the UV-irradiated sample, removed the three common contaminant proteins (keratin, protease, hemoglobin), which resulted in a list of 1027 highly enriched proteins (Table S6).

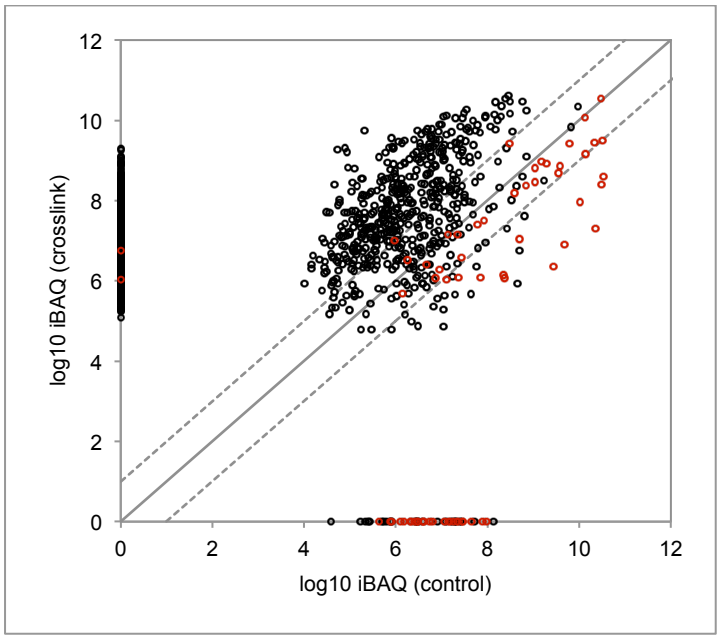


Figure 35: Proteomic analysis of oligo(dT) precipitated proteins from *D. melanogaster* embryos in the conventional crosslink experiment. Scatter plot of protein intensities (iBAQ) comparing the proteins from a UV-irradiated sample (Crosslink) to the non-irradiated sample

(control). Each dot represents one protein. Axes show the log10-fold iBAQ values. The dashed lines represent the threshold for 10-fold enrichment (or depletion). Common contaminant proteins are depicted in red. Proteins, for which no ratio could be calculated, because they were present in only one sample, are shown on the y-axis and x-axis respectively.

We analyzed the preliminary list for overrepresented Gene Ontology (GO) terms using DAVID (Huang da et al., 2009). We hereby considered 720 out of the total set of 1027 proteins for the analysis, which were unambiguous identified by DAVID. The functional annotation analysis revealed that the most overrepresented GO-terms were related to translation, mitotic spindle organization and elongation, post-transcriptional gene regulation, cytoskeleton organisation and gene silencing by RNA (Table S7).

The overrepresented GO terms for mitotic spindle organization can likely be explained by the high number of ribosomal proteins detected in our data, causing the overrepresented GO term. In the first hours of *D. melanogaster* development, the zygotic nucleus undergoes thirteen very rapid mitotic divisions in a single syncytium, ending up with about 6000 nuclei after the thirteenth division cycle. Ribosomal proteins have shown to be involved in mitotic spindle organization, for example ribosomal protein S3, which was also detected in our analysis (Jang et al., 2012).

The overrepresented GO-terms translation (and translation-regulation), post-transcriptional gene regulation and gene silencing by RNA indicate that we mostly precipitated RNA-binding proteins. We observed that most of the high-ranked proteins in the data were known RNA-interacting proteins and RBPs in *D. melanogaster* and identified, we detected Smaug, which is a master regulator of post-transcriptional gene regulation in the early embryo (Benoit et al., 2009). In addition, we detected the nuclear cap binding proteins (CBP80 and CBP20), which suggests that we can precipitate full-length transcripts.

We expected to detect protein factors involved in mRNA degradation, since maternal mRNA is degraded before MZT (Tadros et al., 2007a; Tadros and Lipshitz, 2009). And indeed, we found members of protein complexes involved in 5’ -> 3’ degradation, such as the decapping complex (all known members identified), 5’-exonuclease (XRN1 & XRN2) and the LSM complex. We detected members of protein complexes involved in 3’ -> 5’ degradation as well, such as the CCR4-NOT complex, which is involved in cytoplasmic deadenylation (Temme et al., 2004). Those results indicate, that we can

freeze protein-RNA interactions by crosslinking and thus also detect proteins involved in RNA-degradation.

We analyzed the proteins for enriched domains and observed a strong enrichment common RNA-binding domains, such as the nucleotide binding alpha-beta plaid, the RNA-recognition motif, the K homology domain or the Piwi-Argonaute-Zwille (PAZ) domain found in Argonaute proteins (Table S8).

Taken together, these preliminary results indicate that we precipitate a set of proteins, which is highly enriched for RNA-interacting proteins.

5. Discussion

5.1. A global and unbiased approach to identify mRNA-interacting proteins

Messenger RNA is never alone. During its whole life cycle, from synthesis to decay, mRNA is bound by RNA-interacting proteins such as RBPs, helicases, nucleases and RNA-modifying enzymes, forming messenger-ribonucleoparticles that mediate and regulate most aspects of mRNA metabolism (Martin and Ephrussi, 2009; Moore and Proudfoot, 2009; Muller-McNicoll and Neugebauer, 2013). We therefore set out to identify and characterize the mRNA-bound proteome of a human cell line by *in vivo* UV-crosslinking of proteins to mRNA followed by oligo(dT) affinity purification (Baltz et al., 2012).

The high sensitivity of modern mass spectrometric instruments allows a rapid and accurate detection of proteins from a complex sample. However, dealing with contaminating proteins can be a challenge and it requires careful control experiments to avoid false positives. We used SILAC-based quantitative mass spectrometry to distinguish between direct mRNA-interacting proteins and indirect binders. We identified 797 proteins, which were isolated based on their ability to crosslink to photoreactive-nucleoside-labeled polyadenylated RNA. Sequencing of the oligo(dT) precipitated RNA showed that the majority of transcripts were derived from protein-coding genes.

We assessed the performance of the method through a careful analysis of our results, by evaluating and validating the presence or absence of expected and unexpected RNA-binding proteins. In addition, we assessed the overlap of our dataset with published data obtained through similar methods.

As expected, most of the identified proteins were previously described to interact with RNA basing on their functional annotation as RBPs, helicases, nucleases, ribosomal proteins and RNA-modifying enzymes.

Unexpectedly, among the identified proteins, 245 proteins (31%) were not annotated as RBPs at the time of the study, suggesting that the method would allow us to detect most of the known RBPs expressed in HEK293 and novel RBP candidate proteins as well.

We assayed if we could predict RNA-binding activity for the candidate RNA-binding proteins, using state of the art computational analysis. However, about half of the proteins could not be predicted, not even at very low confidence level. This indicates

that computational prediction attempts, which are based on sequence and/or structure similarity to known RBPs, do not provide an exhaustive catalog of proteins, and emphasizes the need for experimental unbiased and global approach to discover novel RNA-interacting proteins.

Although we believe that our mRNA-bound proteome in HEK293 cells is validated as a complex data set, which contains high probability RBPs, only a careful individual validation of the identified proteins can elucidate if the identified candidate RBPs bind mRNA or other RNA classes. We chose PAR-CLIP as a direct RNA binding assay for the validation of our candidate RBPs. We believe that PAR-CLIP outperforms other validation methods, such as fluorescence-based RNA-RBP co-localization studies, or *in vitro* binding assays which cannot distinguish between direct RNA-binding activity and RNA-binding through protein-protein-interaction, leading to false positives (Riley and Steitz, 2013).

We confirmed that 19 out of 21 tested proteins directly bind RNA. Furthermore, we confirmed mRNA binding and identified targets and binding sites for six of the candidate proteins by PAR-CLIP. A detailed example of the validation by PAR-CLIP is discussed for the candidate RBP IFIT5 in section 5.2. Despite the high validation rate (19 out of 21 tested proteins, 90 %), it remains possible that some of the identified proteins do not directly cross-link to RNA, but interact strongly with other RBPs and thereby elude the stringent wash steps of the oligo(dT) precipitation. Additional validation experiments and comparison with studies in other model organisms are therefore necessary.

During the time of our study, another group also published an mRNA-bound proteome in HeLa-cells (Castello et al., 2012). Castello and colleagues also used an *in vivo* UV-crosslinking and oligo(dT) precipitation approach, however with minor differences in the experimental setup.

Both conventional crosslinking and PAR-crosslinking (however with 4SU only) were used to covalently bind proteins to RNA. Instead of SILAC-based quantitative proteomics, a label-free approach was chosen to assay the enrichment of proteins in oligo(dT) precipitates after UV-irradiation. The differences and advantages of both quantification and irradiation approaches are discussed in 5.4.

With this method, Castello et al. identified 860 proteins, of which 554 are also represented in our dataset (~70 %). The overlap can be increased to ~80% when

considering only the annotated RBPs. The overlap between our data is encouraging, and extensive experimental validation will show if most of the observed differences can be explained by the biological differences between the cell lines (and culture conditions) or by the experimental setup and subsequent data analysis.

A recent study showed that for the identification of the components of a defined protein mixture most of the discrepancy in the results can in most cases be explained by the use of different databases and search engines (Bell et al., 2009). With a standardized analysis method, and an identical experimental approach, we would expect an even higher overlap between our datasets.

Another recent study identified mRNA-binding proteins in yeast (*S. cerevisiae*) by oligo(dT) precipitation and mass spectrometry analysis of co-purified proteins, similar to the studies in human cell lines (Mitchell et al., 2013). The mRNA-bound proteome was studied under glucose-deprivation conditions and a total of 120 proteins were found to bind to mRNA. Of them, 17 proteins were not shown to interact with RNA before, 49 proteins were annotated as RNA binding, but not shown to interact with mRNA, and 92 proteins have human homologues or related protein. Indeed, 56 of those were found in both our study and (Castello et al., 2012). This “core” of conserved RNA-binding proteins contains mostly annotated mRNA-interacting proteins, but also a common set of proteins that have not been described to interact with mRNA before. For example, the pescadillo ribosomal biogenesis factor 1 (PES1), which we validated through a PAR-CLIP based assay, and other proteins involved in rRNA ribosome biogenesis, were identified in both human and yeast studies and possibly represent novel RBPs, whose mRNA-binding function is highly conserved across eukaryotes.

Notably, the number of experimentally identified RBPs in *S. cerevisiae* is much lower than in HEK293 and HeLa cells (Baltz et al., 2012; Castello et al., 2012; Mitchell et al., 2013). Based on GO annotations, the yeast proteome is estimated to contain about 600 RBPs (Cherry et al., 1997). The low number of experimentally detected proteins could be explained by both technical as well as biological reasons. It is reasonable to assume, that in a simpler model organisms like yeast, fewer proteins interact with mRNA and that the number of proteins interacting with mRNA increases with organism complexity. This hypothesis is also supported by the comparison of our study with two previous studies, which both detected approximately 100 - 200 RBPs in yeast (Scherrer et al., 2010; Tsvetanova et al., 2010). Future studies will show if this assumption holds

true and if systems that rely heavily on post-transcriptional regulation, have an increased proportion of proteins interacting with mRNA

Due to the presence of expected RBPs, the validation of unexpected RBPs and the significant overlap with independent studies, we can evaluate our method as an unbiased and reliable approach to define the mRNA-bound proteome. However, proteome means by definition the entire set of proteins.

It is therefore important to note which RNA-binding proteins might have escaped the detection by mass spectrometry and why (false negatives). We classified false negative RBPs as proteins that were not detected in our analysis, although their expression in HEK293 was confirmed by a deep proteomic study (Geiger et al., 2012). Several technical and biological reasons can hereby explain false negatives, such as the human Argonaute protein EIF2C4.

The first and probably most intuitive explanation is that some proteins only interact with a small subset of mRNAs. Although we were able to precipitate low-abundance mRNAs (as validated by qRT-PCR in Figure 7), it might be possible that the amount of co-precipitated protein is below the mass spectrometry detection limit.

Another explanation could be that some RBPs are expressed but do not bind RNA under the experimental conditions. For example, we observed that recombinant FLAG-HA tagged API5 protein is only reproducibly binding RNA when cells are grown in SILAC medium but not in DMEM (as determined by radioactive signal in CLIP experiments, as in 4.8, data not shown) could be due to the different composition of the growth medium. For example, the absence of a particular growth factor in SILAC medium (which is prepared with dialyzed serum) could promote API5 RNA-binding. On the other hand, some proteins only bind RNA under stress conditions. We observed that Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) does not significantly bind to RNA when the cells are grown neither in SILAC or DMEM medium. However, we could detect RNA-binding of GAPDH upon oxidative stress and glucose deprivation (data not shown). These results agree with recent publications stating that GAPDH binds and regulates mRNA of hypoxia-induced genes under oxidative stress, and might thus act as a redox sensor in the cell (Chang et al., 2013; Kondo et al., 2011). Although we observed this phenomenon only for the two proteins tested here, we assume that other RBPs may interact with RNA only under specific conditions, which allows selective post-transcriptional regulation of the target transcripts in response to a stimulus or stress. The architecture

of the protein-RNA contact could prevent in some cases an effective crosslink by UV-irradiation. This problem might be addressed by the application of different UV-irradiation protocols, varying the wavelength and intensity of irradiation or the combinatorial usage of a broader range of photoreactive nucleosides (Meisenheimer and Koch, 1997).

5.2. Analysis of the mRNA bound proteome

RNA binding domains and protein domains without apparent RNA-binding activity

We searched the mRNA-bound proteome and the subset of candidate RBPs for common features such as enriched protein-folds and domains. As expected, we found the common RNA-binding domains significantly enriched but we also detected domains for which no experimental evidence for RNA-binding function was available at the time of the study.

The RAP domain (acronym for RNA-binding domain abundant in Apicomplexans) was previously predicted to be RNA binding based on its sequence (Lee and Hong, 2004). However, no experimental data confirmed the prediction. The domain is conserved down to archea and eubacteria, where it is mostly found alone or in close proximity to the P-loop NTPase fold (IPR027417). In eukaryotes however, the domain is almost exclusively found either alone or in combination with two Fas-activated serine/threonine (FAST) kinase domains, whereby the latter is more frequent with increased organism complexity. In higher eukaryotes like humans, mouse and zebrafish, the domain is found almost exclusively in combination with the (FAST) kinase domain, such as in FASTKD2 and FASTKD5, which were both detected as candidate RBPs in our screen.

The combination of FASTK-domain and RAP-domain might be an example of a beneficial fusion of protein domains, allowing for new protein functions, which might offer an evolutionary advantage. In this case, we assume that the combination of an RNA-binding domain with a protein-kinase domain might connect RNA-binding to protein-kinase function, probably creating a direct link between signal transduction mechanisms and RNA-metabolism.

The DZF (dimerization zinc finger) domain is a domain with so far unknown function and conserved down to *C. elegans*. It is often found in combination with zinc-finger

(ZnF) domains or in combination with the RNA-recognition motif (RRM) but was also found “solo” in interleukin enhancer binding factor 2 (ILF2) protein, which was shown to bind dsRNA (Satoh et al., 1999). Since ILF2 protein is small (less than 400 amino acids), the DZF domain is over 250 amino acids in length and no further annotated domains were found in ILF proteins, we assume that the DZF domain mediates RNA-binding.

Other domains, like the ‘‘winged helix’’ DNA- binding domain (IPR011991) and the high mobility group (HMG) box domain (IPR009071) were so far only described in connection to DNA-binding, for example in transcription factors. The RNA-binding activity of the domains is still elusive and requires further validation.

Taken together, although many RNA-binding protein domains have been described and extensively studied, we believe that unbiased and global screens for RNA-binding proteins will discover more functional RNA-binding domains, which mediate the RNA-binding activity of an RBP, either alone or in combination with known RNA-binding domains.

RBP-screens for evolutionarily diverse species will shed light on the creation, combination and adaptation of RBD “building blocks” which modulate the RBP function and its connection to other cellular processes.

The mRNA-bound proteome might be highly interconnected by protein-protein interactions

To follow up on the connectivity between RNA-metabolism and other regulatory processes in the cell, we searched for protein-protein interactions (PPI) between the members of the mRNA bound proteome and other cellular processes. We used available PPI-data and found that the mRNA-bound proteome is extensively interacting with transcription factors and proteins involved in DNA-repair. Our findings indicate crosstalk between post-transcriptional and transcriptional gene regulation, which is in accordance with the proposed crosstalk between subsequent steps of mRNA-metabolism (Kanitz and Gerber, 2010; Komili and Silver, 2008).

We also observed a higher average cluster coefficient for the RNA-binders in our PPI-network when compared to a random network. This indicates that the proteins of the mRNA-bound proteome are present in highly connected protein clusters, which in turn suggests that RNA-binders tend to be organized in functional modules.

However, it is important to consider how the PPI-data was generated. A recent study in yeast *S. cerevisiae* claimed that many annotated protein-protein interactions might be in fact RNA-dependent, especially between RBPs (Klass et al., 2013). If two or more proteins bind to the same RNA, inadequate RNase treatment of the lysate can provoke a false positive PPI in the screen, especially if proteins interact with RNA in close vicinity, thus sterically hindering RNA cleavage.

A new perspective on diseases through the lens of RNA-binding proteins?

A fraction of the experimentally identified mRNA-binding proteins has been implicated in diseases. Specific genotypic aberrations in 59 out of the 797 observed proteins have been shown to lead to monogenic disorders (Table S1), including neurogenerative diseases, such as Alzheimer's, Fragile X syndrome and amytrophic lateral sclerosis, muscular diseases, such as myotonic dystrophy, and cancers, such as Ewing sarcoma and chondrosarcoma. Thirteen of the 59 disease-associated proteins have, to our knowledge, previously not been shown to interact with RNA, suggesting new roles for RNA metabolism in the respective diseases.

Do RNA-binding proteins moonlight? Extra-work for an evolutionary advantage?

Recent studies identifying RNA-binding proteins in yeast and mammals revealed a large number of cytoplasmic proteins with catalytic activities, many of them acting in metabolism (Castello et al., 2012; Horos, 2012; Mitchell et al., 2013; Scherrer et al., 2010; Tsvetanova et al., 2010). In our study, we identified only a small number of metabolic enzymes (11 out of 797). This discrepancy could on the one hand be explained by differences in the experimental setup, since all of the above mentioned studies used less stringent purification conditions, which could cause contamination of with abundant enzymes. On the other hand, the explanation might be of biological nature, meaning that more metabolic enzymes interact with RNA in HeLa cells or yeast.

However, irrespective of metabolic enzymes, we discovered a number of proteins with non-metabolic catalytic activities, such as ALKBH5. Human ALKBH5 shows similarity to the *Escherichia coli* DNA methylation repair enzyme AlkB, but is only found in vertebrates. Upon hypoxia, ALKBH5 gene transcription is directly induced by hypoxia inducible factor-1 (HIF-1) in a range of cell types (Thalhammer et al., 2011). Thus ALKBH5 was proposed to be participating in the cellular response to hypoxia. Based on

its 2-oxoglutarate oxygenase activity and on our PAR-CLIP data, which confirmed mRNA binding (Baltz et al., 2012), we assumed that ALKBH5 might function in oxidative de-methylation of mRNA, possibly upon hypoxia. Indeed, a recent study showed that ALKBH5 is a mammalian mRNA-demethylase, that oxidatively reverses N(6)-methyladenosine (m(6)A) from mRNA *in vitro* and *in vivo* and that this demethylation activity can affect mRNA metabolism and export (Zheng et al., 2013). m(6)A is the most prevalent internal modification of mRNA in higher eukaryotes and m(6)A sites are enriched near stop codons and in 3' UTRs (Meyer et al., 2012), indicative for regulatory function.

In addition, we found a number of tRNA synthetases in the mRNA bound proteome of both HEK293 cells and *D. melanogaster* embryos and thus tRNA synthetases might indeed interact with mRNA. Non-canonical functions of tRNA synthetases acting on mRNAs and thus modulating stability or translation have been described before (Frugier and Giege, 2003; Nogueira et al., 2001; Sampath et al., 2004). Our study and (Castello et al., 2012) and (Mitchell et al., 2013) indicate that those mechanisms might be more prevalent than previously thought and evolutionary conserved among eukaryotes.

5.3. Characterization of IFIT5, a candidate mRNA binding protein

Amongst many other proteins, the Interferon-induced protein with tetratricopeptide repeats 5 (IFIT5) was detected in our study as a candidate mRNA-binding protein. IFIT5 is a member of the family of tetratricopeptide repeats (TPRs) containing proteins, which comprises four members in humans (IFIT1, 2, 3 and 5). IFIT5 contains unique structural features, which are not found in other family members, such as the “TPR eddy”, a convoluted intra-molecular packing of eight TPRs, which was proposed as scaffold for RNA recognition (Katibah et al., 2013).

Members of the IFIT-family have important antiviral activities, mainly by impeding virus replication. Their expression is stimulated by interferones, and thus IFIT proteins are up regulated in viral response (Zhou et al., 2013). Interestingly, IFIT5 was also found up-regulated in Alzheimer’s disease (Soler-Lopez et al., 2011). Since IFIT5 was amongst our candidate RBPs, the RNA-binding property might add a new perspective to its function in Alzheimer’s disease.

Early studies on IFIT5 protein function showed that IFIT5 immunoprecipitations co-purified a diverse mixture of small RNAs, predominantly smaller than 150 nucleotides in length, indicating that it can interact directly or indirectly with nucleic acids. Interestingly, the results indicated that IFIT5 interacts predominantly with a subset of tRNAs and other non-coding RNAs that are degraded by IFIT5 or an associated nuclease (Hogg and Collins, 2007).

Another study reported that IFIT5 and IFIT1, its family member with the highest sequence homology, are both binding to 5'-triphosphorylated RNA *in vitro* (Pichlmair et al., 2011).

Since we detected IFIT5 in our analysis for mRNA-binding proteins, we assumed that IFIT5 might be a protein with a broad RNA substrate range, covering mRNA, tRNA and 5'PPP-RNAs, and thus we set out to elucidate its enigmatic role in RNA metabolism by PAR-CLIP.

During the time of the PAR-CLIP study and after the publication of (Baltz et al., 2012), two studies published an IFIT5 crystal structure and proposed its mechanism of RNA binding. (Abbas et al., 2013; Katibah et al., 2013). The first group described IFIT5 in complex with RNA bearing a 5'-triphosphate group (PPP-RNA) (Abbas et al., 2013). Binding to 5'-PPP-RNA was shown to be sequence independent and necessary for antiviral activity of IFIT5. The proposed mechanism to distinguish viral RNA from host RNA, was the detection of the triphosphate group at the RNA 5'-end, since cytosolic host ssRNAs do not possess a 5'-triphosphate group but either a 5'-monophosphate, such as rRNA and tRNA or a 5'-cap as in the case of mRNA. The second study proposed that IFIT5 is binding tRNA (Katibah et al., 2013) in approximately the same region as the PPP-ssRNA binding cleft described by Abbas and colleagues (Abbas et al., 2013). It remains to be shown how this apparent contradiction, since Abbas et al. claim that IFIT5 is binding exclusively to PPP-ssRNAs and Katibah et al. report tRNA binding, can be resolved.

Our discovery of IFIT5 in polyA+ribonculeoprotein complexes suggested that IFIT5 might be binding to mRNA sequences. Using PAR-CLIP to determine the cellular binding sites of IFIT5 suggests binding to ribosomal RNAs, vaultRNAs and miRNAs (Table S4). IFIT5 crosslinking was also observed outside of the annotated tRNA transcript, which indicates binding to pre-tRNAs. This mechanism might be used to control the levels of

available tRNAs in the cell, probably in terms of translation regulation, either in response to viral infection or in another functional context.

In addition, we found that IFIT5 binds hundreds of mRNA targets. Binding was mostly observed in the mRNA 3'UTR followed by binding in the CDS. We could not detect any enriched consensus motif for IFIT5 besides the overrepresented motifs that could be explained by the high conservation of the target tRNAs, rRNAs and histone RNA 3'-UTR. IFIT5 might bind to mRNA without consensus motif, but with a preference for AU-rich sequences, since we found them enriched in a 41-nt window around the crosslink-centered-region.

IFIT5 may bind AU-rich elements in a mechanism similar to the one employed by the other ARE-binding IFIT family member IFIT2 (Sen and Fensterl, 2012). AREs are found in the 3'UTRs of many highly unstable mRNAs and represent one of the most common determinants of RNA stability (Chen and Shyu, 1995). IFIT5 might therefore modulate the stability of its targets and this binding activity might complement or mediate the antiviral activities of IFIT5, although this hypothesis still needs experimental validation.

A recent study used a computational approach to identify RNA-interacting proteins based on the chemical properties of the protein-RNA interface (Parisien et al., 2013). Notably, IFIT5 could not be identified as a tRNA-binding protein by this approach, even though the study focused on tRNA-protein interfaces. This result underscores the importance of experimental approaches to define protein-RNA interactions on a genome-wide scale.

Based on our observations, we propose a broader role for IFIT5 in RNA-metabolism than previously assumed. Extensive validation studies, functional analysis by knockdown or PAR-CLIP after viral infection, might reveal connections between the cellular response upon viral infection and adaptations in RNA metabolism and translation regulation.

5.4. The mRNA-bound Proteome in *D. melanogaster* embryos

The choice of the fruit fly *D. melanogaster melanogaster* as a model organism allows us to pose and answer a wider variety of questions than the ones that we can address by using cell culture. Indeed, compared to the early genetic studies in Thomas Hunt Morgan's fly room at Columbia University, where the chromosomal theory of inheritance was confirmed and the very first genetic map created, the fly as a model organism expanded from genetic studies to a wide variety of research that is nowadays performed in *D. melanogaster,* ranging from biophysical examination of morphogen gradient formation in embryos (Kicheva et al., 2012) and its implication for embryo development to unusual questions such as how emotional behavior is encoded in the brain (Dankert et al., 2009).

For our studies we chose *D. melanogaster* embryos, because of both technical and biological reasons. From the technical point of view, our method requires large quantities of starting material and the possibility to perform efficient UV crosslinking *in vivo*. *D. melanogaster* embryos can be collected in large scale without the need for tedious dissection procedures. They are translucent, which allows UV-irradiation based crosslinking. In addition, many established techniques are available for experimental target validation, such as functional knockdowns, in situ hybridization (FISH), and many more. From the biological point of view, the early embryo, before MZT, is a very interesting developmental stage to study post-transcriptional regulatory processes, since the embryo is largely transcriptionally silent and the expression of proteins is mostly regulated at the post-transcriptional level, by mRNA localization and translational regulation (Tadros and Lipshitz, 2005).

We chose two approaches to identify the mRNA-bound proteome in *D. melanogaster* embryos, PhotoActivatable-Ribonucleoside-CrossLinking (PAR-CL) and conventional CrossLinking (cCL). Both protocols have their advantages and limitations. I decided to use both cCL and PAR-CL, because it allows to independently validate RBPs identified with one crosslinking-protocol through the other. We can generate a consensus mRNA-bound proteome and avoid false negatives, i.e. proteins that do not efficiently crosslink in one of both methods. This dual approach was used in HeLa cells (Castello et al., 2012) and showed a significant overlap of proteins identified by both methods. Further experimental validation will show if the majority of proteins selectively identified by only one of the two protocols are indeed direct RNA binding proteins. A recent study

performed in yeast *S. cerevisiae* showed that the efficacy of UV_{254nm} crosslinking varies up to several orders of magnitude, depending on the investigated protein (Klass et al., 2013). So far, no study quantitatively investigated the crosslinking efficacy of a broad range of known RBPs under different and reproducible UV-crosslinking conditions. In summary, these observations suggest that, although the overlap between different methods and the reproducibility within a method can be quite good, a combination of several methods to study RNA-RBP interactions should be preferred when possible.

In vivo labeling of embryo RNA with photoreactive nucleosides

The main difference between cCL and PAR-CL is that the latter depends on photoreactive nucleoside incorporation *in vivo*. For PAR-CL, we chose a combination of tissue specific UPRT expression and the nucleotide precursor 4-thiouracil instead of direct labeling with 4-thiouridine because of several advantages. Firstly, 4-thiouridine is about 40x more costly than 4-thiouracil, which is a factor that needs to be considered for large-scale proteomics and PAR-CLIP experiments.

Another advantage of TU-tagging is that 4-thiouracil is converted to 4-thiouridine only in a restricted cell population by tissue-specific UPRT expression (Miller et al., 2009), limiting the possible effects on the experimental outcome caused by the high concentrations of photoreactive nucleosides in the fly food. The spatial restricted production of photoreactive nucleosides prevents their incorporation into the whole adult fly, which reduces the possibility of detrimental effects on the organism.

We were able to keep fly populations in collection cages over several weeks, fed on 4-thiouracil supplemented yeast paste. Even though we used high concentrations of 4-thiouracil in the yeast paste (up to 400 mM 4-thiouracil in solution), we could observe only minor effects on fly viability when compared to a control population fed on standard yeast paste.

We performed viability assays for the embryos, monitoring if and how many develop into larvae and adult flies. However the results of the viability assay were partly inconclusive, meaning that in some cases we observed a high mortality rate for non-labeled, non-irradiated embryos or, that some labeled embryos developed into larvae despite UV-irradiation. Several repetitions with a larger quantity of embryos and

double-blind studies are necessary to draw final conclusions about the effect of labeling and crosslinking on embryo viability. Transcriptome and proteome profiling and extensive phenotype analysis of labeled embryos and adult flies will also give insights about possible side effects of photoreactive nucleoside labeling.

Nevertheless, despite the high concentration of 4-thiouracil in the fly food, we could still successfully collect embryos from females that were kept up to two weeks in collection cages. *D. melanogaster* is reported to reach an age up to 30 days under standard laboratory conditions. Adult flies lay fewer eggs with increasing age and for this reason we stopped the embryos collections after a maximum of two weeks, and most collections were performed over one week.

For UV-crosslinking experiments, the earliest embryo collections were performed after 3 days feeding on photoreactive nucleoside containing yeast paste, to increase photoreactive nucleoside incorporation into embryo RNA.

We monitored the incorporation of 4-thiouridine into RNA of adult flies and embryos by a biotin-streptavidin based assay and by LC-MS and observed 4SU-incorporation for both adults and embryos. We optimized the labelling procedure to obtain high 4-thiouridine incorporation without introducing a phenotypic effect.

Our results demonstrate that we can achieve incorporation rates in *D. melanogaster* embryos similar to *C. elegans* (0.2 - 1.1 %), which is remarkable if we consider that in our experiments the photoreactive nucleosides need to be transported through the adult fly and get incorporated into the embryo RNA.

UV-irradiation based in vivo RBP-RNA crosslinking in D. melanogaster embryos

Encouraged by the successful labeling protocol, we tested in initial UV-irradiation experiments if we can actually crosslink proteins to RNA in embryos. Our results showed that we can identify co-precipitated proteins in oligo(dT) precipitates upon UV-irradiation and that the non-irradiated control is remarkably clean.

We used the RNA-binding protein Aubergine as positive control and detected it in the oligo(dT) precipitates of UV-irradiated embryos, but not in the control precipitates from non-irradiated embryos. In the UV_{254nm} crosslinking approach without photoreactive nucleoside labeling, we observed the same results. We therefore concluded, that our dual approach successfully crosslinks proteins to RNA in *D. melanogaster* embryos. Therefore, we performed several embryo collections with 3 - 14 days old flies, to obtain

what we considered sufficient starting material for mass spectrometry analysis based on the results of our study in HEK293 cells.

A preliminary analysis of the mRNA-bound proteome in D. melanogaster embryos

For the mass spectrometric analysis of the precipitated proteins, we chose a label free quantification approach, to compare UV-crosslinked and non-crosslinked samples. This approach is generally considered less accurate than stable isotope-labeling based methods but the accuracy is sufficient to identify strong enrichment of a protein against background, which is the situation we expected to face when using our protocol. For the initial attempts to characterize the mRNA-bound proteome in *D. melanogaster* embryos, we aimed to keep the protocol as simple and economic as possible. However, we consider expanding our protocol by using SILAC-flies (Sury et al., 2010), with the benefit of label-swap experiments as performed in HEK293 cells and precise comparison of protein abundances, especially for experiments where protein ratios have to be measured with high accuracy.

Although we prepared samples from both UV_{365nm} and UV_{254nm} irradiated embryos for mass spectrometry analysis, we were only able to measure and include one of the two experiments in this thesis, due to time constraints. We chose UV_{254nm} because of the higher protein concentration in the precipitate. In a preliminary proteomic analysis (using 25% of the protein sample) we were already able to detect a plethora of RNA-binding proteins like translation regulators and post-transcriptional gene regulators such as Smaug (SMG), a major regulator of maternal mRNA decay (Tadros et al., 2007b). We detected key regulators of maternal mRNA translation such as Nanos (Nos), Pumilio (Pum), and Brain tumor (Brat), which regulate hunchback (hb) mRNA translation (Cho et al., 2006).

However, some proteins, which we expected to find, were not detected in our preliminary analysis, such as Bicoid. As described in the introduction, Bicoid is a transcription factor, which can also interact with Caudal mRNA and negatively regulate the translation into Caudal protein (Cho et al., 2005). The reason why we did not detect Bicoid could be, that Bicoid is indeed specifically binding to a single target mRNA and could thus slip detection by mass-spectrometry, because the protein molecules are under the detection limit.

Although this is a preliminary analysis without biological replicates, and demands careful interpretation, we could already observe a clear trend to detect mostly known RNA-binding proteins. Validation experiments, biological replicates, and the complementary approach with UV_{365nm} and photoreactive nucleosides, will help to define a consensus set of RBPs in the *D. melanogaster* embryo, which can then in turn be analysed by *in vivo* PAR-CLIP.

6. Conclusions and Outlook

RNA-binding proteins are important regulators of gene expression and involved in virtually every step of RNA metabolism, yet the universe of RBPs has not been defined and only a small fraction of all known RBPs have been functionally characterized.

Novel approaches to identify RBPs systems-wide are driven by technological advances, such as mass spectrometry based proteomics. Ultrasensitive and fast machines allow us to detect thousands of proteins in parallel from a complex sample and thus to investigate the cellular deep proteome in a reasonable amount of time (Geiger et al., 2012). Technical advances in both mass spectrometry and next generation sequencing, allow studies of protein-mRNA interactions with unprecedented resolution and coverage.

We set out to identify the mRNA-bound proteome of a human cell line and identified about 800 proteins of which one third were not described to bind mRNA before. We were thus able to experimentally define which proteins interact with RNA at a given time.

Since our approach precipitates only poly-A-containing transcripts, we might not identify a number of proteins that only interact with intronic RNA or other non-polyadenylated RNAs. We therefore present a conservative estimate of the total number of RBPs in the respective organism.

The presented work opens many new questions to be investigated in order to understand the post-transcriptional regulatory code.

Firstly, it would be feasible, to characterize novel candidate RBPs of interest, starting from our already published HEK293 mRNA bound proteome. Recent studies showed that RBPs can bind hundreds to thousands of target RNAs (Baltz et al., 2012; Jungkamp et al., 2011; Lebedeva et al., 2011). We assume that this applies to many of the novel RBPs discovered in this study. As shown with our IFIT5 PARCLIP, a closer follow up on novel RBPs can either clarify open questions, or open new doors, especially since many proteins with known “primary” functions can “moonlight” as RNA-interacting proteins.

Since certain intrinsic limitations for functional follow up studies apply to cell culture studies, we adapted our protocol to a complex model system, *Drosophila melanogaster*. This theoretically allows us to study the dynamics of the protein-mRNA interactome during development by simply analyzing and comparing the mRNA-bound proteome of distinct developmental stages. Since labeling and UV-crosslinking proved successful in

D. melanogaster embryos, we have the option to apply *in vivo* PARCLIP to identify RBP targets and binding sites in early embryogenesis. In addition, the photoreactive nucleoside labeling could be used to study the protein occupancy profile on mRNA (Baltz et al., 2012) in the fly embryo. The application of this method during different stages of development and differentiation, could elucidate not just which, but also when regulatory regions on mRNA are bound.

Finally, the UPRT fly bears the possibility to label RNA in a tissue-specific way by ectopic UPRT expression. We could perform tissue-specific crosslinking experiments to study protein-RNA interactions in a defined tissue or cell type.

The comparison of datasets obtained from different organisms (in our case human and *D. melanogaster*) will give valuable insight about evolutionary conservation of protein-RNA interactions. Indeed, by identifying the targets of an RBP in different organisms, we could verify if the examined RBP has conserved, lost or acquired RNA-binding capability, and if the target specificity is evolutionary conserved.

Finally, many post-translational modifications can alter protein function, activity and localization and this holds true for RBPs as well (Thapar and Denmon, 2013). Mass spectrometry analysis allows the parallel identification of protein-modifications (Mertins et al., 2013; Olsen et al., 2006). Investigating post-translational modifications of mRNA-bound proteins might give clues about regulatory mechanisms that allow a certain RBP to bind certain mRNAs only under defined conditions.

We believe that our method can be modified, adapted and widely applied to different biological conditions and systems and that the future applications can greatly contribute to a better understanding of cellular functions of mRNP complexes with the goal to elucidate the posttranscriptional regulatory code that defines growth, differentiation, and disease.

7. Supplementary Information

Table S1: Overview of the identified proteins in HEK293

Table S2: Pfam/Interpro domain and SCOP enrichment analyses

Table S3: Overview on cell lines generated for RNA-binding proteins and CLIP assay results

Table S4: IFIT5 target genes (top 200)

Table S5: GO-term enrichment analysis IFIT5 target genes

Table S6: Overview of the identified proteins in *Drosophila melanogaster* embryos

Table S7: GO-term enrichment analysis (*Drosophila melanogaster* embryos)

Table S8: Protein domain enrichment analysis (*Drosophila melanogaster* embryos)

8. References

Abbas, Y.M., Pichlmair, A., Gorna, M.W., Superti-Furga, G., and Nagar, B. (2013). Structural basis for viral 5'-PPP-RNA recognition by human IFIT proteins. Nature *494*, 60-64.

Adam, S.A., Nakagawa, T., Swanson, M.S., Woodruff, T.K., and Dreyfuss, G. (1986). mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. Molecular and cellular biology *6*, 2932-2943.

Anders, G., Mackowiak, S.D., Jens, M., Maaskola, J., Kuntzagk, A., Rajewsky, N., Landthaler, M., and Dieterich, C. (2012). doRiNA: a database of RNA interactions in post-transcriptional regulation. Nucleic acids research *40*, D180-186.

Andersen, J.S., Lam, Y.W., Leung, A.K., Ong, S.E., Lyon, C.E., Lamond, A.I., and Mann, M. (2005). Nucleolar proteome dynamics. Nature *433*, 77-83.

Andrus, A., and Kuimelis, R.G. (2001). Overview of purification and analysis of synthetic nucleic acids. Current protocols in nucleic acid chemistry / edited by Serge L Beaucage [et al] *Chapter 10*, Unit 10 13.

Aravind, L., Iyer, L.M., and Anantharaman, V. (2003). The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism. Genome biology *4*, R64.

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

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. Nature genetics *25*, 25-29.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic acids research *37*, W202-208.

Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Seraphin, B., and Le Hir, H. (2005). The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. Nature structural & molecular biology *12*, 861-869.

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. Molecular cell *46*, 674-690.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Becalska, A.N., and Gavis, E.R. (2009). Lighting up mRNA localization in Drosophila oogenesis. Development *136*, 2493-2503.

Bell, A.W., Deutsch, E.W., Au, C.E., Kearney, R.E., Beavis, R., Sechi, S., Nilsson, T., and Bergeron, J.J. (2009). A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. Nature methods *6*, 423-430.

Benoit, B., He, C.H., Zhang, F., Votruba, S.M., Tadros, W., Westwood, J.T., Smibert, C.A., Lipshitz, H.D., and Theurkauf, W.E. (2009). An essential role for the RNA-binding protein Smaug during the Drosophila maternal-to-zygotic transition. Development *136*, 923-932.

Berger, S.L. (2007). The complex language of chromatin regulation during transcription. Nature *447*, 407-412.

Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proceedings of the National Academy of Sciences of the United States of America *74*, 3171-3175.
Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Luhrmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. Nature *452*, 846-850.
Bhatt, D.M., Pandya-Jones, A., Tong, A.J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell *150*, 279-290.
Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. Cell *126*, 713-725.
Bono, F., and Gehring, N.H. (2011). Assembly, disassembly and recycling: the dynamics of exon junction complexes. RNA biology *8*, 24-29.
Butter, F., Scheibe, M., Morl, M., and Mann, M. (2009). Unbiased RNA-protein interaction screen by quantitative proteomics. Proceedings of the National Academy of Sciences of the United States of America *106*, 10626-10631.
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.
Chang, C.H., Curtis, J.D., Maggi, L.B., Jr., Faubert, B., Villarino, A.V., O'Sullivan, D., Huang, S.C., van der Windt, G.J., Blagih, J., Qiu, J.*, et al.* (2013). Posttranscriptional control of T cell effector function by aerobic glycolysis. Cell *153*, 1239-1251.
Chang, Y.F., Imam, J.S., and Wilkinson, M.F. (2007). The nonsense-mediated decay RNA surveillance pathway. Annual review of biochemistry *76*, 51-74.
Chen, C.Y., and Shyu, A.B. (1995). AU-rich elements: characterization and importance in mRNA degradation. Trends in biochemical sciences *20*, 465-470.
Chen, R., Mias, G.I., Li-Pook-Than, J., Jiang, L., Lam, H.Y., Chen, R., Miriami, E., Karczewski, K.J., Hariharan, M., Dewey, F.E.*, et al.* (2012). Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell *148*, 1293-1307.
Cherry, J.M., Ball, C., Weng, S., Juvik, G., Schmidt, R., Adler, C., Dunn, B., Dwight, S., Riles, L., Mortimer, R.K.*, et al.* (1997). Genetic and physical maps of Saccharomyces cerevisiae. Nature *387*, 67-73.
Cho, P.F., Gamberi, C., Cho-Park, Y.A., Cho-Park, I.B., Lasko, P., and Sonenberg, N. (2006). Cap-dependent translational inhibition establishes two opposing morphogen gradients in Drosophila embryos. Current biology : CB *16*, 2035-2041.
Cho, P.F., Poulin, F., Cho-Park, Y.A., Cho-Park, I.B., Chicoine, J.D., Lasko, P., and Sonenberg, N. (2005). A new paradigm for translational control: inhibition via 5'-3' mRNA tethering by Bicoid and the eIF4E cognate 4EHP. Cell *121*, 411-423.
Choi, Y.D., and Dreyfuss, G. (1984). Isolation of the heterogeneous nuclear RNA-ribonucleoprotein complex (hnRNP): a unique supramolecular assembly. Proceedings of the National Academy of Sciences of the United States of America *81*, 7471-7475.
Cleary, M.D., Meiering, C.D., Jan, E., Guymon, R., and Boothroyd, J.C. (2005). Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. Nature biotechnology *23*, 232-237.
Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. Cell *136*, 777-793.
Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature biotechnology *26*, 1367-1372.
Dalby, A.B., Goodrich, K.J., Pfingsten, J.S., and Cech, T.R. (2013). RNA recognition by the DNA end-binding Ku heterodimer. RNA *19*, 841-851.

Danckwardt, S., Gantzert, A.S., Macher-Goeppinger, S., Probst, H.C., Gentzel, M., Wilm, M., Grone, H.J., Schirmacher, P., Hentze, M.W., and Kulozik, A.E. (2011). p38 MAPK controls prothrombin expression by regulated RNA 3' end processing. Molecular cell *41*, 298-310.
Dankert, H., Wang, L., Hoopfer, E.D., Anderson, D.J., and Perona, P. (2009). Automated monitoring and analysis of social behavior in Drosophila. Nature methods *6*, 297-303.
Darnell, R.B. (2010). RNA regulation in neurologic disease and cancer. Cancer research and treatment : official journal of Korean Cancer Association *42*, 125-129.
Davila Lopez, M., and Samuelsson, T. (2008). Early evolution of histone mRNA 3' end processing. RNA *14*, 1-10.
de Lima Morais, D.A., Fang, H., Rackham, O.J., Wilson, D., Pethica, R., Chothia, C., and Gough, J. (2011). SUPERFAMILY 1.75 including a domain-centric gene ontology method. Nucleic acids research *39*, D427-434.
de Nadal, E., Ammerer, G., and Posas, F. (2011). Controlling gene expression in response to stress. Nature reviews Genetics *12*, 833-845.
de Sousa Abreu, R., Penalva, L.O., Marcotte, E.M., and Vogel, C. (2009). Global signatures of protein and mRNA expression levels. Molecular bioSystems *5*, 1512-1526.
Denhez, F., and Lafyatis, R. (1994). Conservation of regulated alternative splicing and identification of functional domains in vertebrate homologs to the Drosophila splicing regulator, suppressor-of-white-apricot. The Journal of biological chemistry *269*, 16170-16179.
Dictenberg, J.B., Swanger, S.A., Antar, L.N., Singer, R.H., and Bassell, G.J. (2008). A direct role for FMRP in activity-dependent dendritic mRNA transport links filopodial-spine morphogenesis to fragile X syndrome. Developmental cell *14*, 926-939.
Dolken, L., Ruzsics, Z., Radle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P.*, et al.* (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA *14*, 1959-1972.
Dragoni, I., Mariotti, M., Consalez, G.G., Soria, M.R., and Maier, J.A. (1998). EDF-1, a novel gene product down-regulated in human endothelial cell differentiation. The Journal of biological chemistry *273*, 31119-31124.
Drew, K., Winters, P., Butterfoss, G.L., Berstis, V., Uplinger, K., Armstrong, J., Riffle, M., Schweighofer, E., Bovermann, B., Goodlett, D.R.*, et al.* (2011). The Proteome Folding Project: proteome-scale prediction of structure and function. Genome research *21*, 1981-1994.
Dreyfuss, G., Adam, S.A., and Choi, Y.D. (1984a). Physical change in cytoplasmic messenger ribonucleoproteins in cells treated with inhibitors of mRNA transcription. Molecular and cellular biology *4*, 415-423.
Dreyfuss, G., Choi, Y.D., and Adam, S.A. (1984b). Characterization of heterogeneous nuclear RNA-protein complexes in vivo with monoclonal antibodies. Molecular and cellular biology *4*, 1104-1114.
Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature methods *4*, 207-214.
Favre, A., Moreno, G., Salet, C., and Vinzens, F. (1993). 4-Thiouridine incorporation into the RNA of monkey kidney cells (CV-1) triggers near-UV light long-term inhibition of DNA, RNA and protein synthesis. Photochemistry and photobiology *58*, 689-694.
Fensterl, V., and Sen, G.C. (2011). The ISG56/IFIT1 gene family. Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research *31*, 71-78.

Frugier, M., and Giege, R. (2003). Yeast aspartyl-tRNA synthetase binds specifically its own mRNA. Journal of molecular biology *331*, 375-383.
Gay, L., Miller, M.R., Ventura, P.B., Devasthali, V., Vue, Z., Thompson, H.L., Temple, S., Zong, H., Cleary, M.D., Stankunas, K*., et al.* (2013). Mouse TU tagging: a chemical/genetic intersectional method for purifying cell type-specific nascent RNA. Genes & development *27*, 98-115.
Gebauer, F., Preiss, T., and Hentze, M.W. (2012). From cis-regulatory elements to complex RNPs and back. Cold Spring Harbor perspectives in biology *4*, a012245.
Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012). Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Molecular & cellular proteomics : MCP *11*, M111 014050.
Ghosh, S., Marchand, V., Gaspar, I., and Ephrussi, A. (2012). Control of RNP motility and localization by a splicing-dependent structure in oskar mRNA. Nature structural & molecular biology *19*, 441-449.
Greenberg, J.R. (1979). Ultraviolet light-induced crosslinking of mRNA to proteins. Nucleic acids research *6*, 715-732.
Haas, S., Steplewski, A., Siracusa, L.D., Amini, S., and Khalili, K. (1995). Identification of a sequence-specific single-stranded DNA binding protein that suppresses transcription of the mouse myelin basic protein gene. The Journal of biological chemistry *270*, 12503-12510.
Hachet, O., and Ephrussi, A. (2004). Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. Nature *428*, 959-963.
Hafner, M., Landgraf, P., Ludwig, J., Rice, A., Ojo, T., Lin, C., Holoch, D., Lim, C., and Tuschl, T. (2008). Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. Methods *44*, 3-12.
Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M*., et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129-141.
Han, L.Y., Cai, C.Z., Lo, S.L., Chung, M.C., and Chen, Y.Z. (2004). Prediction of RNA-binding proteins from primary sequence by a support vector machine approach. RNA *10*, 355-368.
Hassfeld, W., Chan, E.K., Mathison, D.A., Portman, D., Dreyfuss, G., Steiner, G., and Tan, E.M. (1998). Molecular definition of heterogeneous nuclear ribonucleoprotein R (hnRNP R) using autoimmune antibody: immunological relationship with hnRNP P. Nucleic acids research *26*, 439-445.
Hesketh, J. (2005). 3′ UTRs and Regulation. eLS.
Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D., and Brown, P.O. (2008). Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS biology *6*, e255.
Hogg, J.R., and Collins, K. (2007). Human Y5 RNA specializes a Ro ribonucleoprotein for 5S ribosomal RNA quality control. Genes & development *21*, 3067-3072.
Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. Nat Rev Mol Cell Biol *6*, 318-327.
Horos, R. (2012). Hepatocytic mRNA interactome. Exp Clin Endocrinol Diabetes *120*, A13.
Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols *4*, 44-57.

Ishihama, Y., Rappsilber, J., Andersen, J.S., and Mann, M. (2002). Microcolumns with self-assembled particle frits for proteomics. Journal of chromatography A *979*, 233-239.
Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol *11*, 113-127.
Jang, C.Y., Kim, H.D., Zhang, X., Chang, J.S., and Kim, J. (2012). Ribosomal protein S3 localizes on the mitotic spindle and functions as a microtubule associated protein in mitosis. Biochemical and biophysical research communications *429*, 57-62.
Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., and Shoemaker, D.D. (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science *302*, 2141-2144.
Jungkamp, A.C., Stoeckius, M., Mecenas, D., Grun, D., Mastrobuoni, G., Kempa, S., and Rajewsky, N. (2011). In vivo and transcriptome-wide identification of RNA binding protein target sites. Molecular cell *44*, 828-840.
Kanitz, A., and Gerber, A.P. (2010). Circuitry of mRNA regulation. Wiley interdisciplinary reviews Systems biology and medicine *2*, 245-251.
Katibah, G.E., Lee, H.J., Huizar, J.P., Vogan, J.M., Alber, T., and Collins, K. (2013). tRNA binding, structure, and localization of the human interferon-induced protein IFIT5. Molecular cell *49*, 743-750.
Keene, J.D., Komisarow, J.M., and Friedersdorf, M.B. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. Nature protocols *1*, 302-307.
Kennedy, M.C., Mende-Mueller, L., Blondin, G.A., and Beinert, H. (1992). Purification and characterization of cytosolic aconitase from beef liver and its relationship to the iron-responsive element binding protein. Proceedings of the National Academy of Sciences of the United States of America *89*, 11730-11734.
Kicheva, A., Cohen, M., and Briscoe, J. (2012). Developmental Pattern Formation: Insights from Physics and Biology. Science *338*, 210-212.
Kickhoefer, V.A., Searles, R.P., Kedersha, N.L., Garber, M.E., Johnson, D.L., and Rome, L.H. (1993). Vault ribonucleoprotein particles from rat and bullfrog contain a related small RNA that is transcribed by RNA polymerase III. The Journal of biological chemistry *268*, 7868-7873.
Kiledjian, M., and Dreyfuss, G. (1992). Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. The EMBO journal *11*, 2655-2664.
Kishore, S., Jaskiewicz, L., Burger, L., Hausser, J., Khorshid, M., and Zavolan, M. (2011). A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nature methods *8*, 559-564.
Klass, D.M., Scheibe, M., Butter, F., Hogan, G.J., Mann, M., and Brown, P.O. (2013). Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in Saccharomyces cerevisiae. Genome research *23*, 1028-1038.
Knapinska, A.M., Gratacos, F.M., Krause, C.D., Hernandez, K., Jensen, A.G., Bradley, J.J., Wu, X., Pestka, S., and Brewer, G. (2011). Chaperone Hsp27 modulates AUF1 proteolysis and AU-rich element-mediated mRNA degradation. Molecular and cellular biology *31*, 1419-1431.
Komili, S., and Silver, P.A. (2008). Coupling and coordination in gene expression processes: a systems biology view. Nature reviews Genetics *9*, 38-48.
Kondo, S., Kubota, S., Mukudai, Y., Nishida, T., Yoshihama, Y., Shirota, T., Shintani, S., and Takigawa, M. (2011). Binding of glyceraldehyde-3-phosphate dehydrogenase to the cis-

acting element of structure-anchored repression in ccn2 mRNA. Biochemical and biophysical research communications *405*, 382-387.
Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nature structural & molecular biology *17*, 909-915.
Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. Nature reviews Genetics *11*, 597-610.
Kuersten, S., and Goodwin, E.B. (2003). The power of the 3' UTR: translational control and development. Nature reviews Genetics *4*, 626-637.
Lasko, P. (2011). Posttranscriptional regulation in Drosophila oocytes and early embryos. Wiley Interdiscip Rev RNA *2*, 408-416.
Lebedeva, S., Jens, M., Theil, K., Schwanhausser, B., Selbach, M., Landthaler, M., and Rajewsky, N. (2011). Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. Molecular cell *43*, 340-352.
Lecuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell *131*, 174-187.
Lee, I., and Hong, W. (2004). RAP--a putative RNA-binding domain. Trends in biochemical sciences *29*, 567-570.
Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X.*, et al.* (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464-469.
Lindberg, U., and Sundquist, B. (1974). Isolation of messenger ribonucleoproteins from mammalian cells. Journal of molecular biology *86*, 451-468.
Liu, N., Han, H., and Lasko, P. (2009). Vasa promotes Drosophila germline stem cell differentiation by activating mei-P26 translation by directly interacting with a (U)-rich motif in its 3' UTR. Genes & development *23*, 2742-2752.
Lukong, K.E., Chang, K.W., Khandjian, E.W., and Richard, S. (2008). RNA-binding proteins in human genetic disease. Trends in genetics : TIG *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.
Martin, K.C., and Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. Cell *136*, 719-730.
Martin, T., Billings, P., Pullman, J., Stevens, B., and Kinniburgh, A. (1978). Substructure of nuclear ribonucleoprotein complexes. Cold Spring Harbor symposia on quantitative biology *42 Pt 2*, 899-909.
Matlin, A.J., Clark, F., and Smith, C.W. (2005). Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol *6*, 386-398.
Mayr, C., and Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell *138*, 673-684.
Mayrand, S., and Pederson, T. (1981). Nuclear ribonucleoprotein particles probed in living cells. Proceedings of the National Academy of Sciences of the United States of America *78*, 2208-2212.
Mayrand, S., Setyono, B., Greenberg, J.R., and Pederson, T. (1981). Structure of nuclear ribonucleoprotein: identification of proteins in contact with poly(A)+ heterogeneous nuclear RNA in living HeLa cells. The Journal of cell biology *90*, 380-384.
Meisenheimer, K.M., and Koch, T.H. (1997). Photocross-linking of nucleic acids to associated proteins. Critical reviews in biochemistry and molecular biology *32*, 101-140.

Melvin, W.T., Milne, H.B., Slater, A.A., Allen, H.J., and Keir, H.M. (1978). Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. European journal of biochemistry / FEBS *92*, 373-379.
Mendell, J.T., and Olson, E.N. (2012). MicroRNAs in stress signaling and human disease. Cell *148*, 1172-1187.
Mertins, P., Qiao, J.W., Patel, J., Udeshi, N.D., Clauser, K.R., Mani, D.R., Burgess, M.W., Gillette, M.A., Jaffe, J.D., and Carr, S.A. (2013). Integrated proteomic analysis of post-translational modifications by serial enrichment. Nature methods *10*, 634-637.
Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell *149*, 1635-1646.
Michlewski, G., and Caceres, J.F. (2010). RNase-assisted RNA chromatography. RNA *16*, 1673-1678.
Mignone, F., Gissi, C., Liuni, S., and Pesole, G. (2002). Untranslated regions of mRNAs. Genome biology *3*, REVIEWS0004.
Milek, M., Wyler, E., and Landthaler, M. (2012). Transcriptome-wide analysis of protein-RNA interactions using high-throughput sequencing. Seminars in cell & developmental biology *23*, 206-212.
Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. RNA *10*, 1692-1694.
Miller, M.R., Robinson, K.J., Cleary, M.D., and Doe, C.Q. (2009). TU-tagging: cell type-specific RNA isolation from intact complex tissues. Nature methods *6*, 439-441.
Mitchell, S.F., Jain, S., She, M., and Parker, R. (2013). Global analysis of yeast mRNPs. Nature structural & molecular biology *20*, 127-133.
Modrek, B., Resch, A., Grasso, C., and Lee, C. (2001). Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic acids research *29*, 2850-2859.
Moore, M.J. (2005). From birth to death: the complex lives of eukaryotic mRNAs. Science *309*, 1514-1518.
Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688-700.
Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C., and Morris, Q. (2008). GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome biology *9 Suppl 1*, S4.
Muller-McNicoll, M., and Neugebauer, K.M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. Nature reviews Genetics *14*, 275-287.
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. Journal of molecular biology *247*, 536-540.
Nash, R., Weng, S., Hitz, B., Balakrishnan, R., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hirschman, J.E.*, et al.* (2007). Expanded protein information at SGD: new pages and proteome browser. Nucleic acids research *35*, D468-471.
Niessing, D., Blanke, S., and Jackle, H. (2002). Bicoid associates with the 5'-cap-bound complex of caudal mRNA and represses translation. Genes & development *16*, 2576-2582.

Niikura, T., Hirata, R., and Weil, S.C. (1997). A novel interferon-inducible gene expressed during myeloid differentiation. Blood cells, molecules & diseases *23*, 337-349.
Nogueira, T., de Smit, M., Graffe, M., and Springer, M. (2001). The relationship between translational control and mRNA degradation for the Escherichia coli threonyl-tRNA synthetase gene. Journal of molecular biology *310*, 709-722.
Nusslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. Nature *287*, 795-801.
Olesnicky, E.C., Brent, A.E., Tonnes, L., Walker, M., Pultz, M.A., Leaf, D., and Desplan, C. (2006). A caudal mRNA gradient controls posterior development in the wasp Nasonia. Development *133*, 3973-3982.
Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell *127*, 635-648.
Ong, S.E., and Mann, M. (2006). Identifying and quantifying sites of protein methylation by heavy methyl SILAC. Current protocols in protein science / editorial board, John E Coligan [et al] *Chapter 14*, Unit 14 19.
Orphanides, G., and Reinberg, D. (2002). A unified theory of gene expression. Cell *108*, 439-451.
Parisien, M., Wang, X., Perdrizet, G., 2nd, Lamphear, C., Fierke, C.A., Maheshwari, K.C., Wilde, M.J., Sosnick, T.R., and Pan, T. (2013). Discovering RNA-protein interactome by using chemical context profiling of the RNA-protein interface. Cell reports *3*, 1703-1713.
Phelps, C.B., and Brand, A.H. (1998). Ectopic gene expression in Drosophila using GAL4 system. Methods *14*, 367-379.
Pichlmair, A., Lassnig, C., Eberle, C.A., Gorna, M.W., Baumann, C.L., Burkard, T.R., Burckstummer, T., Stefanovic, A., Krieger, S., Bennett, K.L.*, et al.* (2011). IFIT1 is an antiviral protein that recognizes 5'-triphosphate RNA. Nature immunology *12*, 624-630.
Priyakumar, U.D., Hyeon, C., Thirumalai, D., and Mackerell, A.D., Jr. (2009). Urea destabilizes RNA by forming stacking interactions and multiple hydrogen bonds with nucleic acid bases. Journal of the American Chemical Society *131*, 17759-17761.
Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nature protocols *2*, 1896-1906.
Ray, D., Kazan, H., Chan, E.T., Pena Castillo, L., Chaudhry, S., Talukder, S., Blencowe, B.J., Morris, Q., and Hughes, T.R. (2009). Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. Nature biotechnology *27*, 667-670.
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.
Richter, J.D., and Lasko, P. (2011). Translational control in oocyte development. Cold Spring Harbor perspectives in biology *3*, a002758.
Riggi, N., Cironi, L., Suva, M.L., and Stamenkovic, I. (2007). Sarcomas: genetics, signalling, and cellular origins. Part 1: The fellowship of TET. The Journal of pathology *213*, 4-20.
Riley, K.J., and Steitz, J.A. (2013). The "Observer Effect" in genome-wide surveys of protein-RNA interactions. Molecular cell *49*, 601-604.
Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W.J., and Jackle, H. (1996). RNA binding and translational suppression by bicoid. Nature *379*, 746-749.
Rohl, C.A., Strauss, C.E., Misura, K.M., and Baker, D. (2004). Protein structure prediction using Rosetta. Methods in enzymology *383*, 66-93.

Sampath, P., Mazumder, B., Seshadri, V., Gerber, C.A., Chavatte, L., Kinter, M., Ting, S.M., Dignam, J.D., Kim, S., Driscoll, D.M., *et al.* (2004). Noncanonical function of glutamyl-prolyl-tRNA synthetase: gene-specific silencing of translation. Cell *119*, 195-208.
Satoh, M., Shaheen, V.M., Kao, P.N., Okano, T., Shaw, M., Yoshida, H., Richards, H.B., and Reeves, W.H. (1999). Autoantibodies define a family of proteins with conserved double-stranded RNA-binding domains as well as DNA binding activity. The Journal of biological chemistry *274*, 34598-34604.
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.
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.
Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58-63.
Sen, G.C., and Fensterl, V. (2012). Crystal structure of IFIT2 (ISG54) predicts functional properties of IFITs. Cell research *22*, 1407-1409.
Setyono, B., and Greenberg, J.R. (1981). Proteins associated with poly(A) and other regions of mRNA and hnRNA molecules as investigated by crosslinking. Cell *24*, 775-783.
Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nature protocols *1*, 2856-2860.
Shibuya, T., Tange, T.O., Sonenberg, N., and Moore, M.J. (2004). eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. Nature structural & molecular biology *11*, 346-351.
Shiina, N., Shinkura, K., and Tokunaga, M. (2005). A novel RNA-binding protein in neuronal RNA granules: regulatory machinery for local translation. The Journal of neuroscience : the official journal of the Society for Neuroscience *25*, 4420-4434.
Silvera, D., Koloteva-Levine, N., Burma, S., and Elroy-Stein, O. (2006). Effect of Ku proteins on IRES-mediated translation. Biology of the cell / under the auspices of the European Cell Biology Organization *98*, 353-361.
Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., *et al.* (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. Cell *147*, 1270-1282.
Smit, A., Hubley, R., and Green, P. (1996-2010). RepeatMasker Open-3.0.
Smith, K. (1976). Photochemistry and Photobiology of Nucleic Acids. Wang, SY, Ed, 187.
Soler-Lopez, M., Zanzoni, A., Lluis, R., Stelzl, U., and Aloy, P. (2011). Interactome mapping suggests new mechanistic details underlying Alzheimer's disease. Genome research *21*, 364-376.
Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell *136*, 731-745.
Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nature reviews Genetics *13*, 613-626.
Sury, M.D., Chen, J.X., and Selbach, M. (2010). The SILAC fly allows for accurate protein quantification in vivo. Molecular & cellular proteomics : MCP *9*, 2173-2183.
Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007a). SMAUG is a major regulator of

maternal mRNA destabilization in Drosophila and its translation is activated by the PAN GU kinase. Developmental cell *12*, 143-155.
Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007b). SMAUG Is a Major Regulator of Maternal mRNA Destabilization in Drosophila and Its Translation Is Activated by the PAN GU Kinase. Developmental cell *12*, 143-155.
Tadros, W., and Lipshitz, H.D. (2005). Setting the stage for development: mRNA translation and stability during oocyte maturation and egg activation in Drosophila. Developmental dynamics : an official publication of the American Association of Anatomists *232*, 593-608.
Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. Development *136*, 3033-3042.
Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533-538.
Tanner, N.K., Hanna, M.M., and Abelson, J. (1988). Binding interactions between yeast tRNA ligase and a precursor transfer ribonucleic acid containing two photoreactive uridine analogues. Biochemistry *27*, 8852-8861.
Temme, C., Zaessinger, S., Meyer, S., Simonelig, M., and Wahle, E. (2004). A complex containing the CCR4 and CAF1 proteins is involved in mRNA deadenylation in Drosophila. The EMBO journal *23*, 2862-2871.
Thalhammer, A., Bencokova, Z., Poole, R., Loenarz, C., Adam, J., O'Flaherty, L., Schodel, J., Mole, D., Giaslakiotis, K., Schofield, C.J.*, et al.* (2011). Human AlkB homologue 5 is a nuclear 2-oxoglutarate dependent oxygenase and a direct target of hypoxia-inducible factor 1alpha (HIF-1alpha). PloS one *6*, e16210.
Thapar, R., and Denmon, A.P. (2013). Signaling pathways that control mRNA turnover. Cellular signalling *25*, 1699-1710.
Ting, N.S., Yu, Y., Pohorelic, B., Lees-Miller, S.P., and Beattie, T.L. (2005). Human Ku70/80 interacts directly with hTR, the RNA component of human telomerase. Nucleic acids research *33*, 2090-2098.
Tracey, W.D., Jr., Ning, X., Klingler, M., Kramer, S.G., and Gergen, J.P. (2000). Quantitative analysis of gene function in the Drosophila embryo. Genetics *154*, 273-284.
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*.
Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505-510.
Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005). CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods *37*, 376-386.
Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212-1215.
van Eekelen, C.A., Riemen, T., and van Venrooij, W.J. (1981). Specificity in the interaction of hnRNA and mRNA with proteins as revealed by in vivo cross linking. FEBS letters *130*, 223-226.
van Kouwenhove, M., Kedde, M., and Agami, R. (2011). MicroRNA regulation by RNA-binding proteins and its implications for cancer. Nature reviews Cancer *11*, 644-656.
Vasudevan, S. (2012). Posttranscriptional upregulation by microRNAs. Wiley Interdiscip Rev RNA *3*, 311-330.
Vasudevan, S., and Steitz, J.A. (2007). AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. Cell *128*, 1105-1118.

Vasudevan, S., Tong, Y., and Steitz, J.A. (2007). Switching from repression to activation: microRNAs can up-regulate translation. Science *318*, 1931-1934.
Wagenmakers, A.J., Reinders, R.J., and van Venrooij, W.J. (1980). Cross-linking of mRNA to proteins by irradiation of intact cells with ultraviolet light. European journal of biochemistry / FEBS *112*, 323-330.
Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.
Walden, W.E., Selezneva, A.I., Dupuy, J., Volbeda, A., Fontecilla-Camps, J.C., Theil, E.C., and Volz, K. (2006). Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. Science *314*, 1903-1908.
Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470-476.
Wheeler, T.M., and Thornton, C.A. (2007). Myotonic dystrophy: RNA-mediated muscle disease. Current opinion in neurology *20*, 572-576.
Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.M., Li, C.J., Vagbo, C.B., Shi, Y., Wang, W.L., Song, S.H., *et al.* (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Molecular cell *49*, 18-29.
Zhou, X., Michal, J.J., Zhang, L., Ding, B., Lunney, J.K., Liu, B., and Jiang, Z. (2013). Interferon induced IFIT family genes in host antiviral defense. International journal of biological sciences *9*, 200-208.

Supplementary table S1: Summary of identified proteins

Class	Offical gene symbol	log2FC experiment L1	log2FC experiment H1 (label swap)	log2FC experiment L2	mean log2FC (experiment 1-3)	log10 iBAQ	Functional Group	RNA binding	spliceosome	nucleolus	Phenotype MIM number	RNA-binding prediction (% precision)	mRNA-binding prediction (% precis	ssRNA-binding prediction (%
I	EDF1	-7.97	3.49	-6.98	6.15	6.96	candidate mRNA binder			x		20		
I	SYNE1	-6.90	3.61	-7.55	6.02	7.15	candidate mRNA binder				612998, 610743	0		
I	PPIB	-4.62	2.49	-10.64	5.92	6.33	candidate mRNA binder				259440	0		
I	TCP1	-7.16	3.03	-7.51	5.90	7.75	candidate mRNA binder					0		
I	IGF2BP1	-6.66	3.28	-7.18	5.71	9.24	RNAbinding	x						
I	IGF2BP3	-6.81	3.22	-6.94	5.66	8.89	RNAbinding	x						
I	SYNCRIP	-6.61	3.43	-6.40	5.48	9.29	RNAbinding	x	x	x				
I	PABPC1	-6.13	3.46	-6.80	5.46	8.39	RNAbinding	x	x					
I	HSPA8	-6.46	3.50	-6.23	5.40	6.78	candidate mRNA binder					0		
I	LSM6	-6.34	3.25	-6.55	5.38	6.50	RNAbinding	x	x					
I	SND1	-6.43	3.43	-6.27	5.38	9.16	RNAbinding	x						
I	RBMX	-6.48	3.53	-6.04	5.35	9.19	RNAbinding	x	x	x				20
I	LSM3	-6.64	2.66	-6.73	5.34	6.64	RNAbinding	x	x					
I	HNRNPA0	-6.59	3.17	-6.26	5.34	9.57	RNAbinding	x		x			80	20
I	HNRNPCL1	-6.66	2.86	-6.49	5.34	6.96	RNAbinding	x						
I	LRPPRC	-6.51	3.26	-6.17	5.31	9.32	RNAbinding	x			220111			
I	HNRNPL	-6.15	3.28	-6.51	5.31	9.72	RNAbinding	x		x			50	20
I	SFPQ	-6.23	3.32	-6.39	5.31	9.31	RNAbinding	x		x			20	
I	FUBP3	-6.11	3.03	-6.77	5.30	9.00	RNAbinding	x						
I	HNRNPR	-6.37	3.22	-6.28	5.29	9.02	RNAbinding	x	x	x			50	20
I	HNRNPK	-6.44	3.24	-6.17	5.28	9.42	RNAbinding	x	x	x			20	
I	ELAVL1	-6.05	3.01	-6.54	5.20	8.98	RNAbinding	x	x	x				
I	DHX9	-6.26	3.17	-6.13	5.19	9.34	helicase	x					20	
I	HNRNPD	-5.90	3.51	-6.13	5.18	9.06	RNAbinding	x	x				50	20
I	HNRNPAB	-6.12	3.34	-6.08	5.18	9.42	RNAbinding	x	x	x				
I	DDX17	-6.44	3.14	-5.95	5.18	8.97	helicase	x	x	x			20	20
I	PHF5A	-6.35	3.07	-6.06	5.16	6.51	RNAbinding		x			50		
I	RALY	-6.17	3.22	-6.04	5.14	8.54	RNAbinding	x	x				50	20
I	TIAL1	-6.10	3.13	-6.19	5.14	9.00	RNAbinding							
I	PUM1	-5.44	2.88	-7.02	5.12	8.13	RNAbinding	x						
I	MATR3	-5.95	3.27	-6.11	5.11	9.14	RNAbinding	x			606070			
I	CCBL2	-4.95	3.44	-6.93	5.11	7.04	RNAbinding	x		x		0		
I	KHSRP	-6.00	3.03	-6.27	5.10	9.00	RNAbinding	x					50	
I	LSM1	-6.63	2.81	-5.82	5.09	6.90	RNAbinding	x					20	
I	FUS	-6.18	3.32	-5.73	5.08	9.08	RNAbinding	x		x	608030		20	
I	RPS28	-6.78	2.92	-5.53	5.08	7.90	ribosomal					80		
I	PURA	-6.01	3.04	-6.14	5.06	8.92	RNAbinding	x				20	20	
I	SERBP1	-5.92	3.17	-6.02	5.04	8.53	RNAbinding	x						
I	EIF3A	-5.78	3.38	-5.93	5.03	8.42	translationfactor							
I	PURB	-6.33	3.01	-5.70	5.02	8.69	RNAbinding	x						
I	ILF3	-5.92	3.16	-5.94	5.01	8.81	translationfactor	x						
I	ILF2	-6.12	3.14	-5.76	5.01	8.79	RNAbinding	x	x	x				
I	NONO	-5.74	3.15	-6.10	5.00	9.31	RNAbinding	x		x				
I	TAF15	-5.97	3.17	-5.80	4.98	8.00	RNAbinding	x			612237			
I	UPF1	-5.75	2.90	-6.27	4.97	8.33	helicase	x						
I	HNRNPH3	-6.16	3.05	-5.71	4.97	9.03	RNAbinding	x	x	x			50	20
I	MRPS11	-6.11	2.59	-6.20	4.96	7.31	ribosomal							
I	HNRNPA3	-5.99	3.08	-5.82	4.96	9.10	RNAbinding	x	x					20
I	DDX3X	-5.53	3.24	-6.11	4.96	8.84	helicase	x	x	x				
I	PABPC4L	-5.94	3.20	-5.73	4.96	7.00	RNAbinding	x				50	20	20
I	HNRNPA2B1	-5.91	2.93	-5.99	4.94	9.95	RNAbinding	x	x				50	20
I	LSM2	-6.33	2.52	-5.96	4.94	6.88	RNAbinding	x	x				20	
I	API5	-5.71	3.10	-5.96	4.92	4.98	candidate mRNA binder					20		
I	PABPC4	-5.35	3.34	-6.05	4.92	8.65	RNAbinding	x					20	
I	EIF4B	-5.70	3.27	-5.76	4.91	8.38	translationfactor	x					20	
I	PTBP2	-6.12	2.86	-5.73	4.90	8.59	RNAbinding	x						20
I	FUBP1	-5.73	3.14	-5.84	4.90	8.34	helicase	x						
I	CSDE1	-5.47	3.14	-6.08	4.90	8.41	RNAbinding	x				50	20	
I	SMC1A	-5.39	2.68	-6.62	4.90	5.84	candidate mRNA binder			x	300590	0		
I	SF3A3	-5.73	3.05	-5.90	4.89	7.40	RNAbinding		x	x			20	
I	TARDBP	-6.11	3.21	-5.33	4.88	8.54	RNAbinding	x			612069			
I	MOV10	-5.69	2.92	-6.04	4.88	8.37	helicase	x						
I	CNBP	-4.58	3.04	-7.03	4.88	7.92	RNAbinding	x			602668			
I	FAM120A	-5.75	3.05	-5.85	4.88	8.47	RNAbinding	x						
I	KHDRBS1	-5.63	3.25	-5.75	4.87	8.86	RNAbinding	x		x				
I	HNRNPH2	-5.63	2.96	-6.03	4.87	8.30	RNAbinding	x	x				50	
I	THOC4	-5.99	3.18	-5.43	4.87	9.25	RNAbinding	x		x			50	20
I	RBM3	-5.94	3.13	-5.51	4.86	8.61	RNAbinding	x		x			20	
I	YBX1	-5.91	3.36	-5.29	4.85	9.40	RNAbinding	x	x					
I	HNRNPM	-5.63	2.96	-5.94	4.84	9.19	RNAbinding	x	x	x			20	
I	MFAP1	-5.03	3.02	-6.48	4.84	7.91	candidate mRNA binder		x			0		
I	RPS5	-7.08	2.81	-4.59	4.83	7.87	ribosomal			x				
I	HNRNPU	-5.70	3.09	-5.69	4.82	9.27	RNAbinding	x	x				20	
I	CHTOP	-5.64	3.23	-5.60	4.82	8.43	RNAbinding							
I	IGF2BP2	-5.63	2.90	-5.91	4.81	8.54	RNAbinding	x			125853			
I	EIF3C	-5.83	2.98	-5.58	4.80	8.16	translationfactor	x						
I	HNRNPC	-6.01	2.98	-5.39	4.79	9.53	RNAbinding	x						
I	CSDA	-6.00	3.31	-5.04	4.78	9.44	RNAbinding	x						
I	EIF4H	-5.76	3.10	-5.47	4.78	7.87	translationfactor	x			194050			
I	HNRNPA1	-5.66	3.00	-5.67	4.78	10.04	RNAbinding	x	x	x			50	
I	MRPS23	-6.24	2.80	-5.28	4.77	7.50	ribosomal					50		
I	RPS10	-5.68	2.96	-5.66	4.76	8.17	ribosomal			x	613308	80		
I	SF3B1	-5.53	3.05	-5.68	4.76	8.13	RNAbinding		x					
I	MYO18A	-6.47	3.36	-4.42	4.75	5.11	candidate mRNA binder					0		
I	SAFB2	-5.78	3.30	-5.17	4.75	7.78	RNAbinding	x						

I	RBM38	-6.28	2.68	-5.25	4.74	8.09	RNAbinding	x							
I	DDX5	-5.31	3.26	-5.59	4.72	8.84	helicase	x	x	x					20
I	EIF4A1	-5.65	2.91	-5.60	4.72	8.46	translationfactor	x		x					
I	RPS3	-5.40	2.89	-5.86	4.72	8.76	ribosomal								
I	TNRC6A	-5.40	3.58	-5.15	4.71	6.26	RNAbinding	x						20	
I	HNRNPF	-5.62	3.25	-5.22	4.70	9.17	RNAbinding	x	x	x				80	
I	HNRNPH1	-5.54	3.07	-5.43	4.68	9.49	RNAbinding	x	x	x				80	
I	XRCC5	-5.70	3.14	-5.18	4.67	8.02	helicase			x			20		
I	DDX39B	-5.47	2.90	-5.64	4.67	8.63	helicase		x					20	20
I	HNRNPUL1	-5.62	3.08	-5.29	4.66	8.59	RNAbinding	x						20	
I	POLDIP3	-5.77	2.84	-5.36	4.66	7.83	RNAbinding	x						20	
I	G3BP1	-5.39	3.19	-5.37	4.65	8.60	helicase	x							
I	RAVER1	-5.81	2.86	-5.28	4.65	8.14	RNAbinding	x							
I	MAGOH2	-6.50	2.79	-4.66	4.65	6.93	RNAbinding	x							
I	HNRPDL	-5.54	3.20	-5.19	4.64	8.27	RNAbinding	x		x					
I	SF3B4	-4.86	3.20	-5.86	4.64	8.23	RNAbinding	x	x					50	
I	ROD1	-5.59	2.89	-5.42	4.63	8.16	RNAbinding	x							
I	SAFB	-5.60	3.18	-5.12	4.63	7.70	RNAbinding	x							
I	SRSF3	-5.58	3.10	-5.21	4.63	8.61	RNAbinding	x	x	x				50	
I	HDLBP	-5.18	2.98	-5.72	4.63	8.17	RNAbinding								
I	PTBP1	-5.96	3.02	-4.88	4.62	9.94	RNAbinding	x		x					
I	UBAP2L	-5.00	3.20	-5.65	4.62	7.73	candidate mRNA binder						0		
I	RPS15	-6.33	3.01	-4.50	4.61	7.60	ribosomal								
I	XRCC6	-5.27	3.13	-5.45	4.61	8.39	helicase						0		
I	MBNL1	-5.94	3.14	-4.75	4.61	7.77	RNAbinding	x						80	
I	TBRG4	-5.79	2.58	-5.47	4.61	8.04	candidate mRNA binder						0		
I	GSPT2	-6.27	1.88	-5.67	4.60	6.15	candidate mRNA binder								
I	UNK	-7.50	2.10	-4.20	4.60	5.88	candidate mRNA binder						20		
I	R3HDM1	-5.60	2.46	-5.74	4.60	5.25	candidate mRNA binder						20		
I	EIF4A3	-5.77	2.80	-5.21	4.59	7.92	translationfactor	x	x	x					
I	TIA1	-5.30	3.07	-5.38	4.59	7.80	RNAbinding	x							
I	SFRS7	-5.61	2.95	-5.19	4.58	8.79	RNAbinding	x	x	x				50	
I	PCBP1	-5.78	3.01	-4.95	4.58	8.35	RNAbinding	x		x				50	
I	PSPC1	-5.33	3.26	-5.15	4.58	8.40	RNAbinding	x		x				20	
I	MPHOSPH10	-6.42	2.73	-4.58	4.58	7.09	candidate mRNA binder			x			80		
I	EIF3G	-6.21	2.95	-4.56	4.57	7.62	translationfactor	x							
I	RC3H1	-5.86	2.83	-5.02	4.57	6.65	RNAbinding	x							
I	ESRP2	-5.30	2.96	-5.45	4.57	7.55	RNAbinding	x							
I	EWSR1	-5.38	3.28	-5.02	4.56	8.12	RNAbinding	x		x	612219		20		
I	ATXN2L	-5.53	2.84	-5.29	4.55	8.05	RNAbinding	x					0		
I	GRSF1	-5.45	2.85	-5.34	4.54	8.19	RNAbinding	x							
I	KIAA1967	-5.32	2.98	-5.32	4.54	7.95	RNAbinding						20		
I	MSI2	-5.20	3.04	-5.35	4.53	8.34	RNAbinding	x							
I	SRSF1	-5.42	2.80	-5.36	4.53	8.99	RNAbinding	x	x	x				50	
I	RPS26	-5.83	2.54	-5.20	4.52	8.43	ribosomal				613309				
I	FXR2	-5.25	2.97	-5.32	4.51	8.26	RNAbinding	x							
I	MRPL2	-5.54	2.45	-5.53	4.51	6.51	ribosomal						50		
I	BZW1	-7.22	2.25	-4.00	4.49	5.67	candidate mRNA binder						50		
I	TRA2B	-5.39	2.79	-5.28	4.49	8.72	RNAbinding	x	x						
I	DDX21	-5.20	2.91	-5.35	4.49	8.76	helicase	x		x					20
I	PRRC2A	-5.63	2.78	-5.03	4.48	8.06	candidate mRNA binder						20		
I	RPUSD4	-4.78	3.50	-5.14	4.48	6.64	RNAmod	x							
I	DAZAP1	-5.31	2.84	-5.26	4.47	8.97	RNAbinding	x							
I	LSM14A	-4.67	3.05	-5.67	4.46	7.98	RNAbinding	x					80	20	
I	HNRNPUL2	-5.44	3.03	-4.92	4.46	8.57	RNAbinding	x					0		
I	FMR1	-5.31	3.09	-4.98	4.46	5.98	RNAbinding	x			300624, 300623, 311360				
I	LARP4B	-5.46	3.06	-4.84	4.45	7.98	RNAbinding	x						20	
I	U2AF2	-5.15	2.99	-5.22	4.45	8.42	RNAbinding	x	x	x				50	
I	PRKDC	-4.84	3.11	-5.40	4.45	7.61	candidate mRNA binder			x			0		
I	AKAP8	-4.93	3.16	-5.26	4.45	7.80	candidate mRNA binder						20		
I	CPSF7	-4.80	3.10	-5.45	4.45	7.38	RNAbinding	x						20	
I	KHDRBS3	-4.96	3.02	-5.36	4.45	7.44	RNAbinding	x							
I	BCLAF1	-5.44	2.62	-5.28	4.45	7.85	candidate mRNA binder		x				20		
I	LARP4	-5.78	2.76	-4.80	4.44	7.15	RNAbinding	x							
I	RSL1D1	-4.51	2.92	-5.90	4.44	8.27	ribosomal			x					
I	MRPL40	-5.91	2.31	-5.06	4.42	7.11	ribosomal						50		
I	G3BP2	-4.97	3.15	-5.15	4.42	7.75	RNAbinding	x							
I	MAGOH	-5.67	2.75	-4.84	4.42	6.03	RNAbinding	x	x					50	
I	CSTF2T	-4.61	1.72	-6.92	4.42	6.06	RNAbinding	x						20	
I	MRPS31	-5.75	2.95	-4.52	4.41	7.26	ribosomal						50		
I	ZFR	-5.47	2.95	-4.80	4.40	8.14	RNAbinding	x							
I	CPEB4	-5.18	3.05	-4.95	4.39	8.28	RNAbinding	x							
I	C2orf15	-4.24	2.72	-6.21	4.39	5.16	ribosomal								
I	SFRS9	-5.33	2.89	-4.95	4.39	8.92	RNAbinding	x	x					50	
I	MRPL13	-6.07	1.87	-5.23	4.39	5.63	ribosomal						50		
I	PRR3	-6.03	2.61	-4.51	4.39	6.62	RNAbinding	x					20		
I	ZCCHC24	-5.78	2.95	-4.43	4.39	6.92	candidate mRNA binder						0		
I	MYEF2	-5.31	2.76	-5.08	4.38	7.70	RNAbinding	x							
I	SLIRP	-5.80	2.86	-4.49	4.38	8.16	RNAbinding	x							
I	FXR1	-5.21	2.90	-5.03	4.38	7.82	RNAbinding	x							
I	PCBP2	-6.36	2.91	-3.85	4.38	9.35	RNAbinding	x	x	x				20	
I	RBM14	-5.09	3.08	-4.95	4.37	8.63	RNAbinding	x							
I	PATL1	-5.11	2.77	-5.22	4.36	7.18	RNAbinding	x							
I	PEG10	-5.60	2.32	-5.15	4.36	7.80	candidate mRNA binder						20		
I	PTCD3	-5.22	2.85	-4.98	4.35	8.15	candidate mRNA binder						20		
I	SSBP1	-5.36	2.43	-5.24	4.34	7.97	RNAbinding	x					20		
I	PRRC2C	-5.31	2.83	-4.88	4.34	7.40	candidate mRNA binder			x			0		
I	SNRPA	-5.82	2.84	-4.35	4.34	7.48	RNAbinding	x	x	x				50	
I	DDX39	-5.75	2.72	-4.52	4.33	7.01	helicase	x					80	50	20
I	ZNFX1	-5.31	2.52	-5.16	4.33	6.24	candidate mRNA binder						0		
I	SF3B14	-5.93	2.54	-4.52	4.33	6.97	RNAbinding	x	x	x				50	
I	SFRS6	-5.59	2.63	-4.74	4.32	8.62	RNAbinding	x	x	x				80	
I	CPSF6	-5.10	3.01	-4.85	4.32	8.00	RNAbinding	x		x					
I	UTP15	-5.13	2.98	-4.84	4.32	7.41	candidate mRNA binder			x			50		
I	MRPS28	-5.69	2.97	-4.25	4.31	6.34	ribosomal						50		
I	CPSF4	-5.60	2.95	-4.33	4.29	7.73	RNAbinding	x							

I	PNN	-5.17	2.80	-4.90	4.29	7.98	candidate mRNA binder		x			80	20		
I	UTP11L	-5.38	3.10	-4.39	4.29	7.15	RNAbinding			x					
I	NCL	-4.88	2.97	-4.99	4.28	9.22	RNAbinding	x		x					
I	RBMS1	-4.64	3.12	-5.03	4.27	7.21	RNAbinding	x					20		
I	ZNF326	-4.85	3.08	-4.86	4.26	8.31	RNAbinding			x		20			
I	PABPN1	-4.35	2.90	-5.54	4.26	8.25	RNAbinding	x	x		164300		50		
I	CIRBP	-5.17	3.06	-4.55	4.26	8.23	RNAbinding	x						20	
I	U2AF1	-5.24	2.77	-4.74	4.25	8.13	RNAbinding	x	x	x			20		
I	DIMT1L	-4.51	2.46	-5.75	4.24	6.67	RNAmod	x		x					
I	AKAP8L	-4.91	2.76	-5.05	4.24	7.73	helicase					20			
I	SF1	-5.16	2.87	-4.68	4.24	7.76	RNAbinding	x					20		
I	RBM47	-5.16	2.55	-4.99	4.23	7.99	RNAbinding	x					20		
I	RBM8A	-6.06	2.11	-4.52	4.23	7.05	RNAbinding	x	x	x					
I	APOBEC3B	-5.11	2.28	-5.29	4.23	6.90	RNAbinding	x							
I	CSTF2	-4.94	2.83	-4.87	4.21	7.45	RNAbinding	x					50		
I	SRP14	-6.04	1.87	-4.72	4.21	8.01	RNAbinding	x		x					
I	EIF3D	-4.85	2.97	-4.81	4.21	7.72	translationfactor	x							
I	MEX3C	-5.52	2.56	-4.53	4.20	7.61	RNAbinding	x							
I	PUF60	-5.49	2.73	-4.38	4.20	7.47	RNAbinding	x					20		
I	TSNAX	-5.49	2.11	-5.00	4.20	6.88	RNAbinding	x				20			
I	DDX6	-5.25	2.80	-4.55	4.20	7.32	helicase	x						20	
I	C11orf68	-4.06	2.75	-5.78	4.20	6.14	candidate mRNA binder					50			
I	FBLL1	-5.21	2.37	-4.99	4.19	7.19	RNAbinding	x				80	80	20	
I	RBM15	-4.81	2.97	-4.79	4.19	7.50	RNAbinding	x			606077				
I	NGRN	-5.03	2.58	-4.96	4.19	5.56	candidate mRNA binder					0			
I	RNMTL1	-7.05	1.98	-3.54	4.19	5.30	RNAbinding	x							
I	AKAP1	-4.94	3.02	-4.56	4.17	6.57	RNAbinding	x							
I	NUFIP2	-4.59	2.72	-5.20	4.17	7.46	RNAbinding	x					20		
I	ZCCHC3	-5.17	2.79	-4.54	4.17	7.72	candidate mRNA binder					20			
I	PUM2	-4.86	2.65	-4.99	4.17	7.65	RNAbinding	x					20		
I	C17orf85	-4.52	2.96	-5.01	4.17	7.19	candidate mRNA binder		x			20			
I	YTHDF1	-6.03	2.49	-3.96	4.16	7.09	RNAbinding	x				20			
I	RBM22	-4.84	2.86	-4.78	4.16	7.99	RNAbinding	x	x				20		
I	LARP1B	-6.02	1.99	-4.44	4.15	5.91	RNAbinding	x							
I	SRSF10	-5.23	2.82	-4.41	4.15	8.65	RNAbinding	x	x				50		
I	MRPS35	-5.52	2.16	-4.77	4.15	7.09	ribosomal					50			
I	CELF1	-5.01	2.77	-4.67	4.15	7.75	RNAbinding	x							
I	LSM14B	-5.50	2.66	-4.29	4.15	6.09	RNAbinding	x				50	20		
I	MSI1	-4.66	2.62	-5.17	4.15	8.39	RNAbinding	x							
I	MCAT	-5.20	2.08	-5.16	4.14	6.89	metabolicEnz					0			
I	THRAP3	-5.46	2.66	-4.30	4.14	8.20	RNAbinding					20			
I	SYF2	-5.69	2.33	-4.40	4.14	5.91	candidate mRNA binder		x			80	50		
I	LSM4	-5.79	1.81	-4.82	4.14	6.79	RNAbinding	x	x						
I	ZC3H7B	-5.20	2.55	-4.64	4.13	7.45	candidate mRNA binder					20			
I	ZC3H11A	-4.64	3.01	-4.75	4.13	7.80	candidate mRNA binder					20			
I	EIF4G1	-4.54	2.58	-5.20	4.11	7.22	translationfactor	x			614251	50			
I	CASC3	-5.12	2.75	-4.46	4.11	7.43	RNAbinding	x					50		
I	PPIL4	-6.17	2.39	-3.76	4.11	6.12	RNAbinding	x							
I	STAU1	-4.63	2.87	-4.79	4.10	7.78	RNAbinding	x							
I	CPEB2	-5.81	2.75	-3.71	4.09	6.59	RNAbinding	x					20		
I	XRN2	-5.00	2.71	-4.56	4.09	7.83	nuclease	x		x		50			
I	FAM98B	-5.44	2.74	-4.08	4.08	7.66	candidate mRNA binder			x		0			
I	MRPL41	-6.07	1.99	-4.17	4.08	6.24	ribosomal					50			
I	NUPL2	-4.02	2.26	-5.96	4.08	5.78	RNAbinding	x				20			
I	RBM27	-4.44	3.17	-4.62	4.08	7.29	RNAbinding	x		x			20		
I	MRPL4	-4.74	2.60	-4.89	4.08	7.11	ribosomal					50			
I	RBM45	-5.82	2.46	-3.94	4.08	7.02	RNAbinding	x							
I	NUDT21	-5.79	2.38	-4.05	4.07	7.26	RNAbinding	x							
I	MRPL1	-5.19	2.70	-4.33	4.07	7.27	ribosomal								
I	NCBP2	-3.86	3.02	-5.34	4.07	5.90	RNAbinding	x	x						
I	MRPS5	-5.56	2.44	-4.18	4.06	7.54	ribosomal								
I	LARP1	-4.76	2.74	-4.67	4.06	8.21	RNAbinding	x							
I	ADARB1	-3.77	2.01	-6.39	4.06	6.02	RNAmod	x							
I	FASTKD2	-4.70	2.63	-4.83	4.05	7.78	candidate mRNA binder				220110	0			
I	SFRS2	-5.02	2.62	-4.52	4.05	8.70	RNAbinding	x	x				50		
I	ZC3H4	-4.58	1.90	-5.66	4.05	7.00	RNAbinding	x				20			
I	RBM4	-4.52	2.86	-4.75	4.04	8.01	RNAbinding	x		x					
I	CELF2	-4.68	2.63	-4.80	4.04	7.56	RNAbinding	x					20		
I	SRRT	-4.48	2.70	-4.92	4.03	7.39	RNAbinding	x				50			
I	MRPS9	-5.26	2.68	-4.15	4.03	7.17	ribosomal					50			
I	APOBEC3F	-5.31	2.65	-4.12	4.03	7.83	RNAmod	x							
I	FBL	-4.90	2.50	-4.67	4.02	8.02	RNAbinding	x		x					
I	ELAVL2	-4.77	2.74	-4.57	4.02	7.92	RNAbinding	x							
I	FAM98A	-5.00	2.94	-4.11	4.02	7.62	candidate mRNA binder					20			
I	DDX1	-5.33	2.47	-4.26	4.02	7.06	helicase	x		x					
I	RBM39	-4.80	2.60	-4.63	4.01	7.95	RNAbinding	x	x	x					
I	YTHDC1	-4.83	3.00	-4.20	4.01	7.49	RNAbinding	x				80	20		
I	MRPL9	-4.40	2.54	-5.08	4.01	6.98	ribosomal					50			
I	FAM120C	-4.69	2.71	-4.62	4.01	7.63	candidate mRNA binder					20			
I	NOVA2	-4.81	3.08	-4.12	4.00	7.62	RNAbinding	x							
I	DDX50	-4.82	2.49	-4.69	4.00	7.71	helicase	x		x					
I	RBM7	-3.86	2.62	-5.52	4.00	6.42	RNAbinding	x	x						
I	FASTKD1	-3.52	4.86	-3.61	4.00	6.35	candidate mRNA binder					0			
I	MYBBP1A	-3.78	2.69	-5.51	3.99	7.87	candidate mRNA binder			x		20			
I	ANKHD1	-4.63	2.91	-4.44	3.99	5.95	RNAbinding	x							
I	YTHDF2	-4.27	2.98	-4.72	3.99	7.93	RNAbinding	x				0			
I	RBM6	-4.60	2.94	-4.43	3.99	6.86	RNAbinding	x							
I	YTHDF3	-4.51	2.78	-4.66	3.98	7.43	RNAbinding	x				20			
I	MEX3D	-5.53	2.45	-3.95	3.98	7.57	RNAbinding	x					20		
I	USP10	-4.72	2.48	-4.72	3.97	6.89	candidate mRNA binder					0			
I	EMG1	-5.80	2.65	-3.46	3.97	6.53	RNAbinding	x		x	211180				
I	PTCD1	-4.61	2.52	-4.77	3.97	7.02	candidate mRNA binder					0			
I	RPUSD3	-4.49	2.68	-4.72	3.96	6.11	RNAmod	x							
I	RBM34	-5.14	2.29	-4.44	3.96	6.93	RNAbinding	x		x					
I	CPEB3	-4.66	2.75	-4.46	3.96	7.01	RNAbinding	x					20		
I	SF3B2	-4.74	2.46	-4.64	3.95	7.38	RNAbinding		x	x		80	20		

I	EIF1	-5.64	2.28	-3.92	3.95	7.08	translationfactor	x							
I	NXF1	-4.79	2.62	-4.42	3.95	7.41	RNAbinding	x							
I	SFRS5	-5.03	2.42	-4.39	3.95	8.02	RNAbinding	x	x	x				50	
I	UTP3	-3.67	2.44	-5.72	3.94	7.09	candidate mRNA binder			x			20		
I	DDX28	-5.11	2.66	-4.04	3.94	7.38	helicase	x							
I	IFIT5	-5.18	2.46	-4.16	3.93	6.11	candidate mRNA binder						0		
I	SF3A2	-3.91	2.61	-5.27	3.93	7.43	RNAbinding	x	x					20	
I	RPS14	-5.68	2.14	-3.96	3.93	8.00	ribosomal			x	153550				
I	TRUB2	-5.43	2.57	-3.77	3.93	7.50	RNAmod								
I	TRA2A	-4.31	2.74	-4.72	3.92	8.28	RNAbinding	x	x					50	
I	LIN28B	-5.19	2.53	-4.00	3.91	7.79	RNAbinding	x						20	
I	STRBP	-5.25	2.62	-3.83	3.90	6.97	RNAbinding	x							
I	IMP3	-4.46	1.75	-5.49	3.90	5.79	ribosomal	x		x					
I	GIGYF2	-5.03	2.78	-3.88	3.90	6.21	candidate mRNA binder				607688		0		
I	FYTTD1	-4.43	3.05	-4.21	3.90	7.72	RNAbinding	x							
I	STAU2	-4.68	2.74	-4.25	3.89	7.23	RNAbinding	x							
I	RBFOX2	-3.98	2.75	-4.93	3.89	8.13	RNAbinding	x						50	
I	YTHDC2	-4.72	2.54	-4.39	3.88	7.19	helicase	x					20		
I	EIF1AX	-4.78	2.31	-4.53	3.87	6.77	translationfactor	x							
I	RBM10	-4.50	2.76	-4.36	3.87	7.55	RNAbinding	x			311900		50		
I	SRSF12	-5.10	2.76	-3.72	3.86	6.56	RNAbinding	x						80	
I	AQR	-4.16	2.55	-4.87	3.86	7.47	RNAbinding	x	x					50	
I	SARNP	-5.50	2.98	-3.10	3.86	7.13	RNAbinding	x		x					
I	MEX3A	-4.59	2.42	-4.53	3.85	7.73	RNAbinding	x							
I	EIF4A2	-5.73	2.81	-2.99	3.84	6.84	translationfactor	x							
I	MTHFSD	-5.46	2.58	-3.48	3.84	6.96	RNAbinding	x							
I	NTPCR	-5.05	1.71	-4.77	3.84	6.20	candidate mRNA binder						0		
I	RBMS2	-4.34	2.79	-4.39	3.84	7.20	RNAbinding	x						20	
I	ADAR	-4.50	2.79	-4.22	3.84	7.01	RNAmod	x		x	127400				
I	C22orf28	-5.37	2.49	-3.66	3.84	6.92	candidate mRNA binder						20		
I	YLPM1	-3.76	2.66	-5.08	3.83	6.78	candidate mRNA binder						50		
I	CEBPZ	-4.29	2.61	-4.60	3.83	7.04	candidate mRNA binder			x			0		
I	SAP18	-5.60	2.65	-3.24	3.83	7.13	candidate mRNA binder			x			0		
I	TRIM56	-5.08	2.30	-4.02	3.80	7.39	candidate mRNA binder						20		
I	ZNF207	-4.40	2.42	-4.59	3.80	6.89	candidate mRNA binder			x			0		
I	ATXN2	-3.86	2.82	-4.70	3.79	7.38	RNAbinding	x			183090			20	
I	HELZ	-4.45	2.12	-4.81	3.79	6.63	helicase	x					20		
I	FAM103A1	-3.64	3.16	-4.57	3.79	6.20	RNAbinding						20		
I	EIF2S1	-5.10	2.71	-3.55	3.79	7.05	translationfactor	x		x					
I	SF3A1	-4.45	2.16	-4.75	3.79	6.23	RNAbinding	x	x					20	
I	MBNL2	-4.94	2.39	-4.03	3.79	5.84	RNAbinding	x						20	
I	RBM12	-4.86	2.45	-4.04	3.78	6.24	RNAbinding	x							
I	RBM33	-4.97	2.53	-3.84	3.78	6.13	RNAbinding	x							
I	YARS2	-3.79	2.44	-5.07	3.77	6.48	RNAbinding				613561				
I	SCAF4	-5.88	2.54	-2.87	3.76	5.70	RNAbinding	x						50	
I	SF4	-5.28	2.31	-3.69	3.76	7.06	RNAbinding	x						20	
I	DHX38	-4.82	2.42	-4.03	3.75	5.91	helicase	x	x				80	50	
I	POP7	-4.64	2.64	-3.98	3.75	6.44	nuclease	x		x			50		
I	EIF4G3	-4.05	2.55	-4.65	3.75	7.03	translationfactor	x							
I	CCDC9	-4.14	2.84	-4.28	3.75	6.72	candidate mRNA binder						0		
I	CAPRIN1	-3.71	3.20	-4.34	3.75	7.84	RNAbinding	x						20	
I	NKRF	-4.50	2.61	-4.08	3.73	7.46	RNAbinding	x		x					
I	DHX8	-5.31	2.52	-3.34	3.72	6.73	helicase	x	x	x				50	
I	EIF2C2	-5.25	2.42	-3.46	3.71	7.20	translationfactor	x						50	20
I	PPP1R10	-5.54	2.54	-3.04	3.70	6.39	RNAbinding	x							
I	MRPS7	-5.02	2.23	-3.85	3.70	7.19	ribosomal								
I	RPL22	-3.59	2.49	-5.01	3.70	8.07	ribosomal			x				20	
I	RBM5	-4.52	2.73	-3.84	3.70	6.57	RNAbinding	x							
I	MKRN2	-4.88	2.35	-3.84	3.69	6.39	RNAbinding	x					0		
I	SNW1	-4.48	2.29	-4.25	3.68	6.53	candidate mRNA binder		x	x			80	20	
I	WDR33	-4.27	2.78	-3.97	3.67	7.01	candidate mRNA binder						0		
I	PRKRA	-4.12	2.11	-4.79	3.67	5.73	RNAbinding	x			612067				
I	UTP18	-3.67	2.73	-4.59	3.66	7.20	candidate mRNA binder			x			50		
I	ABT1	-4.01	2.24	-4.73	3.66	6.33	RNAbinding	x		x					
I	MAZ	-4.15	1.97	-4.86	3.66	6.20	RNAbinding	x							
I	SFRS4	-5.01	2.00	-3.96	3.66	7.98	RNAbinding	x	x					50	
I	HNRPLL	-4.46	2.74	-3.75	3.65	7.84	RNAbinding	x		x					20
I	PES1	-3.83	2.20	-4.90	3.64	6.75	candidate mRNA binder			x			50		
I	RBM12B	-4.29	2.79	-3.82	3.63	7.55	RNAbinding	x						20	
I	RPS2	-4.23	2.83	-3.84	3.63	8.68	ribosomal			x					
I	SUGP2	-4.45	2.61	-3.83	3.63	7.26	RNAbinding	x						20	
I	DCAF13	-4.15	2.25	-4.49	3.63	5.90	candidate mRNA binder			x			50		
I	DHX30	-4.15	2.06	-4.67	3.63	8.28	helicase	x							
I	ASCC3	-3.78	3.68	-3.42	3.62	6.33	helicase	x					20		
I	SARS2	-6.07	2.10	-2.70	3.62	5.83	RNAbinding	x			613845		50		
I	EBNA1BP2	-4.17	2.12	-4.54	3.61	7.17	candidate mRNA binder						50		
I	IBA57	-5.35	1.70	-3.77	3.61	5.84	candidate mRNA binder						0		
I	QKI	-4.18	2.27	-4.37	3.61	6.82	RNAbinding	x						20	
I	DHX57	-4.22	2.73	-3.85	3.60	7.10	helicase	x					20		
I	CDC5L	-4.95	1.98	-3.84	3.59	5.97	RNAbinding	x	x					20	
I	XPO5	-3.94	2.67	-4.14	3.58	6.81	RNAbinding	x							
I	ZC3HAV1	-4.15	2.29	-4.29	3.58	7.02	RNAbinding	x							
I	NOP2	-3.67	2.43	-4.59	3.56	7.48	RNAbinding	x		x					
I	MBNL3	-3.81	2.68	-4.20	3.56	6.47	RNAbinding	x						20	
I	EIF4G2	-4.27	2.27	-4.15	3.56	7.10	translationfactor	x							
I	TCERG1	-4.09	2.58	-4.00	3.56	7.08	candidate mRNA binder			x			20		
I	DDX42	-4.34	2.05	-4.26	3.55	6.68	helicase	x	x					20	20
I	MEX3B	-4.77	2.68	-3.19	3.54	6.54	RNAbinding	x							
I	EIF1B	-5.12	2.13	-3.36	3.54	7.02	translationfactor	x							
I	MRPL15	-4.50	2.78	-3.33	3.54	6.71	ribosomal						50		
I	CSTF1	-4.22	2.60	-3.77	3.53	7.35	RNAbinding	x						50	
I	PRPF8	-4.23	2.37	-3.99	3.53	8.25	RNAbinding	x	x	x	600059			20	
I	WDR46	-3.57	2.25	-4.77	3.53	6.77	other			x			0		
I	GAR1	-5.31	2.34	-2.92	3.52	7.34	RNAbinding	x		x					
I	RG9MTD1	-4.71	2.52	-3.32	3.52	7.19	nuclease						50		
I	NGDN	-3.65	2.08	-4.81	3.51	6.98	candidate mRNA binder			x			20		

I	POLR2B	-4.51	2.17	-3.84	3.51	6.09	RNAbinding								
I	BRIX1	-3.96	2.17	-4.39	3.51	7.28	candidate mRNA binder			x					
I	LLPH	-4.36	1.95	-4.21	3.51	6.74	candidate mRNA binder			x			0		
I	DHX15	-4.58	2.52	-3.38	3.49	7.79	helicase	x	x				80	20	
I	KIAA1429	-4.19	2.47	-3.81	3.49	7.38	other						80	50	
I	RBM4B	-5.07	2.32	-3.06	3.48	6.45	RNAbinding	x						20	
I	RNPS1	-4.31	2.45	-3.68	3.48	7.59	RNAbinding	x	x						
I	U2SURP	-4.79	2.28	-3.35	3.47	6.67	RNAbinding	x	x					20	
I	NOVA1	-3.02	3.00	-4.41	3.47	6.58	RNAbinding	x						20	
I	HARS2	-4.38	2.37	-3.64	3.46	6.48	candidate mRNA binder						50		
I	AATF	-3.30	2.37	-4.68	3.45	6.87	candidate mRNA binder			x			20		
I	RBM19	-3.84	1.72	-4.78	3.45	6.33	RNAbinding	x		x					
I	RPL7L1	-4.12	2.25	-3.97	3.45	6.62	ribosomal								
I	EIF3L	-4.07	2.19	-4.08	3.45	6.76	translationfactor	x							
I	ZC3H14	-4.48	2.27	-3.56	3.44	6.26	RNAbinding	x							
I	RRP12	-3.67	2.26	-4.36	3.43	6.99	candidate mRNA binder			x			20		
I	ZNF768	-4.18	2.54	-3.57	3.43	6.20	candidate mRNA binder						20		
I	WBP11	-3.96	2.15	-4.18	3.43	6.04	candidate mRNA binder		x	x			80	20	
I	ACIN1	-4.61	2.25	-3.36	3.41	6.35	RNAbinding	x	x				20		
I	MRPL11	-4.15	2.61	-3.45	3.40	7.58	ribosomal						50		
I	C9orf114	-3.70	2.08	-4.43	3.40	6.34	candidate mRNA binder						0		
I	PPIA	-5.06	1.83	-3.29	3.39	7.12	candidate mRNA binder						20		
I	FASTKD5	-3.69	2.38	-4.10	3.39	6.09	candidate mRNA binder						0		
I	SPTBN1	-5.76	1.61	-2.79	3.39	4.74	candidate mRNA binder						0		
I	ZC3H7A	-3.33	2.64	-4.17	3.38	6.17	RNAbinding	x					20		
I	SUPT6H	-3.90	2.36	-3.87	3.38	6.30	RNAbinding	x							
I	MRPL45	-3.94	2.21	-3.97	3.37	5.96	ribosomal						50		
I	URB1	-3.78	2.30	-4.03	3.37	6.02	candidate mRNA binder			x			20		
I	KCTD12	-4.24	1.81	-4.01	3.35	6.37	candidate mRNA binder						0		
I	SLTM	-3.68	2.63	-3.73	3.35	7.62	RNAbinding	x		x					
I	PEF1	-4.16	2.47	-3.40	3.34	6.16	candidate mRNA binder						0		
I	DHX16	-4.22	2.32	-3.47	3.34	7.00	helicase						80	50	
I	SNRNP200	-3.01	1.98	-5.03	3.34	4.95	helicase	x	x	x			80	20	20
I	RBM15B	-3.85	2.41	-3.74	3.33	6.63	RNAbinding	x						80	
I	RRP8	-3.57	2.28	-4.14	3.33	6.49	candidate mRNA binder			x			50		
I	FIP1L1	-2.86	2.43	-4.67	3.32	6.56	RNAbinding	x							
I	LAS1L	-3.82	2.76	-3.37	3.32	6.55	candidate mRNA binder			x			20		
I	NUDT16L1	-4.46	2.33	-3.14	3.31	6.70	candidate mRNA binder			x			20		
I	ZNF638	-4.16	2.07	-3.69	3.31	5.79	RNAbinding	x						20	
I	TDRD3	-4.33	2.49	-3.09	3.30	6.80	candidate mRNA binder						20		
I	MRTO4	-4.43	1.66	-3.82	3.30	6.68	RNAbinding	x		x					
I	MRPL22	-4.01	2.26	-3.64	3.30	5.91	ribosomal						50		
I	LUC7L3	-3.82	2.26	-3.83	3.30	7.14	RNAbinding	x							
I	XRN1	-2.65	3.79	-3.44	3.29	5.65	nuclease	x							
I	SURF6	-3.26	2.38	-4.23	3.29	7.07	RNAbinding	x		x					
I	PRRC2B	-3.66	2.29	-3.91	3.29	6.35	candidate mRNA binder						0		
I	CSTF3	-4.46	2.63	-2.73	3.27	6.53	RNAbinding	x							
I	CDK11B	-3.49	2.73	-3.58	3.26	6.62	candidate mRNA binder						0		
I	DHX36	-4.31	2.84	-2.63	3.26	6.87	helicase	x					50		
I	RAVER2	-3.55	2.76	-3.42	3.24	6.44	RNAbinding	x							
I	MRPS26	-3.28	2.50	-3.94	3.24	6.43	ribosomal						20		
I	NAF1	-4.83	1.92	-2.94	3.23	6.64	RNAbinding	x							
I	ALG13	-3.76	2.05	-3.85	3.22	6.91	candidate mRNA binder						0		
I	NOP16	-3.78	2.73	-3.14	3.22	6.57	candidate mRNA binder			x			50		
I	DDX27	-3.29	2.35	-4.00	3.21	6.83	helicase	x		x					
I	RRP36	-3.33	2.40	-3.91	3.21	6.87	candidate mRNA binder			x			50		
I	GNL3	-3.92	1.68	-4.03	3.21	6.72	candidate mRNA binder			x			20		
I	RBM25	-4.41	2.40	-2.79	3.20	7.11	RNAbinding	x							
I	SUPT5H	-4.13	2.16	-3.27	3.18	6.70	RNAbinding	x					50		
I	RBM28	-3.08	2.53	-3.91	3.17	7.18	RNAbinding	x		x	612079			50	
I	MRPL3	-4.29	1.72	-3.50	3.17	5.57	ribosomal								
I	ALKBH5	-3.51	1.99	-4.00	3.17	7.25	candidate mRNA binder						0		
I	NOL10	-3.64	2.02	-3.83	3.17	5.74	candidate mRNA binder			x			50		
I	RPS4X	-4.77	1.65	-3.08	3.17	7.32	ribosomal			x				20	
I	REXO4	-3.95	1.93	-3.57	3.15	6.34	nuclease	x		x			50		
I	SPEN	-3.76	1.76	-3.93	3.15	6.72	RNAbinding	x							
I	NOC2L	-2.85	1.87	-4.72	3.14	6.03	candidate mRNA binder						50		
I	EFTUD2	-4.00	2.05	-3.37	3.14	5.95	RNAbinding	x		x				20	
I	SAMD4B	-3.04	1.76	-4.61	3.14	6.46	candidate mRNA binder						0		
I	CNOT1	-3.41	1.85	-4.12	3.13	6.27	candidate mRNA binder						50		
I	RRP1B	-3.51	2.18	-3.69	3.13	6.85	candidate mRNA binder			x			50		
I	NCOA5	-3.91	2.30	-3.16	3.12	6.05	candidate mRNA binder						0		
I	SCAF11	-3.22	3.53	-2.60	3.12	6.73	candidate mRNA binder						80	20	
I	EIF2C1	-3.86	2.42	-3.02	3.10	6.38	translationfactor	x							
I	CHERP	-4.29	2.02	-2.98	3.09	6.72	RNAbinding	x	x						
I	C14orf21	-3.38	1.94	-3.96	3.09	7.02	RNAbinding	x		x					
I	SART1	-4.89	1.89	-2.46	3.08	6.08	candidate mRNA binder		x				80	20	
I	CSNK1E	-3.72	2.37	-3.09	3.06	5.55	candidate mRNA binder						0		
I	PURG	-2.47	2.71	-3.98	3.05	6.93	RNAbinding	x							
I	PUS1	-3.39	1.79	-3.96	3.05	6.05	RNAmod	x			600462				
I	PRPF40A	-3.49	2.10	-3.55	3.04	6.52	candidate mRNA binder						80	20	
I	NOLC1	-4.06	1.68	-3.39	3.04	6.21	RNAbinding	x		x					
I	RPSA	-3.40	2.08	-3.56	3.01	7.13	ribosomal						80		
I	PPAN	-3.11	2.25	-3.68	3.01	6.54	candidate mRNA binder						50		
I	MKI67IP	-3.61	1.88	-3.53	3.01	6.87	RNAbinding	x		x					
I	SRPK1	-3.96	1.89	-3.17	3.01	6.59	candidate mRNA binder			x			50		
I	UTP14A	-2.41	2.44	-4.15	3.00	6.39	candidate mRNA binder			x			50		
I	NSUN2	-3.84	1.82	-3.33	3.00	7.07	RNAmod	x							
I	TNPO1	-2.87	2.14	-3.95	2.98	6.21	candidate mRNA binder						20		
I	DDX52	-3.84	1.70	-3.41	2.98	6.94	helicase	x		x					
I	GRWD1	-3.27	2.51	-3.02	2.93	6.14	candidate mRNA binder			x			50		
I	APOBEC3C	-3.66	2.76	-2.32	2.91	6.73	RNAbinding	x					50		
I	GTPBP4	-3.61	2.15	-2.97	2.91	6.91	candidate mRNA binder						20		
I	DDX18	-2.99	1.83	-3.82	2.88	7.42	helicase	x		x					
I	RRP1	-2.34	1.75	-4.44	2.85	6.47	candidate mRNA binder						50		
I	OTUD4	-3.36	2.19	-2.94	2.83	5.96	candidate mRNA binder						0		

I	SON	-2.79	2.01	-3.68	2.83	7.04	RNAbinding	x					20			
I	ZCCHC6	-2.90	1.74	-3.76	2.80	6.10	RNAbinding	x				50				
I	DDX47	-2.99	1.90	-3.47	2.79	6.20	helicase	x		x						
I	FARSA	-3.28	1.68	-3.32	2.76	5.67	RNAbinding	x								
I	DDX10	-2.70	1.80	-3.78	2.76	6.30	helicase	x		x						20
I	NHP2L1	-4.09	1.65	-2.53	2.76	6.16	RNAbinding	x	x	x						
I	LUC7L2	-3.07	2.22	-2.91	2.73	7.00	RNAbinding	x		x		20				
I	UPF3B	-3.10	1.94	-3.06	2.70	6.80	RNAbinding	x		x	300676					
I	EIF3E	-2.46	2.11	-3.43	2.66	6.61	translationfactor	x								
I	ZCCHC17	-3.08	1.68	-3.23	2.66	5.75	RNAbinding	x								
I	RRS1	-2.65	1.91	-3.35	2.64	7.61	candidate mRNA binder			x		50				
I	ABCF1	-3.41	1.61	-2.83	2.62	5.77	RNAbinding	x		x						
I	SNRPD1	-3.43	1.82	-2.55	2.60	8.01	RNAbinding	x	x				20			
I	NIP7	-3.39	1.59	-2.68	2.55	5.88	ribosomal	x		x						
I	TCOF1	-2.89	1.91	-2.72	2.51	7.06	candidate mRNA binder			x	154500	20				
I	ZRANB2	-2.63	1.98	-2.60	2.40	6.67	RNAbinding	x					20			
I	ZCRB1	-1.77	3.03	-5.33	3.38	6.34	RNAbinding	x		x			20			
I	SRFBP1	-2.16	2.95	-4.69	3.27	6.67	candidate mRNA binder			x		20				
I	MRPS21	-5.09	2.87	-1.82	3.26	7.10	ribosomal					50				
I	HSD17B10	-4.33	3.17	-2.09	3.20	7.46	metabolicEnz				300438, 300705, 300220	20				
I	TOP2A	-1.69	2.14	-5.73	3.19	5.31	candidate mRNA binder					0				
I	PPHLN1	-4.27	2.95	-1.87	3.03	5.83	candidate mRNA binder					0				
I	MRPL54	-4.46	2.17	-2.31	2.98	6.25	ribosomal					50				
I	BMS1	-2.19	2.07	-4.17	2.81	6.29	candidate mRNA binder			x		50				
I	DHX34	-1.98	2.34	-3.96	2.76	4.62	helicase	x								
I	SNRNP70	-2.22	2.15	-3.88	2.75	8.54	RNAbinding	x		x			50			
I	RPL35A	-3.49	1.77	-2.20	2.49	7.25	ribosomal			x	612528					
I	SUMO2	-2.05	2.54	-2.78	2.46	8.03	candidate mRNA binder					0				
I	FTSJ3	-1.84	2.22	-3.28	2.45	7.87	RNAmod	x		x		50				
I	SRRM1	-2.21	1.91	-2.99	2.37	7.82	RNAbinding	x	x				50			
I	NOC3L	-1.78	2.25	-2.60	2.21	7.24	candidate mRNA binder			x		50				
I	SSRP1	-2.35	1.84	-2.25	2.14	6.54	candidate mRNA binder			x		20				
I	DDX46	-2.40	1.79	-2.23	2.14	7.20	helicase	x		x			20	20		
I	SFRS11	-1.86	1.78	-2.75	2.13	7.62	RNAbinding	x					50			
I	HIST1H1E	-2.52	1.69	-2.17	2.13	7.13	histone			x		0				
I	RRBP1	-2.26	1.75	-2.10	2.04	6.99	candidate mRNA binder					20				
II	ZC3H8	NA	3.01	-8.41	5.71	6.37	RNAbinding	x								
II	ZC3H18	-4.96	NA	-6.41	5.68	5.25	RNAbinding	x				20				
II	SNRPA1	-8.64	NA	-2.69	5.67	5.59	RNAbinding	x	x	x						
II	RBM26	-5.11	NA	-5.92	5.52	6.08	RNAbinding	x					20			
II	DROSHA	-5.47	NA	-5.25	5.36	5.31	nuclease	x								
II	MTO1	-6.91	NA	-3.69	5.30	7.00	translationfactor					50				
II	PCDH20	-4.97	NA	-5.56	5.27	6.21	candidate mRNA binder					0				
II	ZNF598	-6.10	NA	-4.19	5.15	5.20	candidate mRNA binder					0				
II	ERI3	-6.03	NA	-3.65	4.84	6.24	nuclease	x				0				
II	C9orf23	-6.68	2.75	NA	4.71	6.00	RNAbinding	x				20				
II	IPO5	-4.16	NA	-5.08	4.62	4.86	candidate mRNA binder					20				
II	CNOT4	-2.95	NA	-6.24	4.60	5.71	RNAbinding	x					20			
II	CDK13	-5.13	NA	-3.91	4.52	5.16	candidate mRNA binder					50				
II	MYO5A	NA	3.15	-5.66	4.41	5.13	candidate mRNA binder				214450	0				
II	C19orf29	-4.74	NA	-3.75	4.25	6.05	candidate mRNA binder		x			80	20			
II	ROCK2	NA	3.94	-4.55	4.25	5.26	candidate mRNA binder					0				
II	TSFM	-3.72	NA	-4.71	4.22	5.74	RNAbinding	x			610505					
II	ZNF579	-2.81	NA	-5.48	4.14	6.81	candidate mRNA binder					0				
II	EIF2C3	-5.43	NA	-2.81	4.12	6.04	translationfactor	x								
II	ZAK	-3.88	NA	-4.31	4.10	5.36	candidate mRNA binder					0				
II	SRSF8	-4.17	NA	-3.89	4.03	7.04	RNAbinding	x					50			
II	PCDHGA9	-1.55	3.20	-7.21	3.98	6.96	candidate mRNA binder					0				
II	C19orf47	NA	2.36	-5.58	3.97	5.27	candidate mRNA binder					0				
II	CCT3	-6.33	1.58	NA	3.96	5.58	candidate mRNA binder					0				
II	PRMT1	-4.84	NA	-3.06	3.95	5.73	candidate mRNA binder					0				
II	UBE2O	-4.20	NA	-3.64	3.92	6.29	candidate mRNA binder					0				
II	CHCHD1	-5.62	2.12	NA	3.87	5.74	candidate mRNA binder					20				
II	HOXB6	-3.65	NA	-4.00	3.83	5.85	candidate mRNA binder					0				
II	FBRSL1	-3.94	NA	-3.66	3.80	6.16	candidate mRNA binder					80	80	20		
II	MTERF	-5.36	2.06	NA	3.71	6.14	candidate mRNA binder					20				
II	GPATCH8	-3.65	3.76	NA	3.71	4.30	candidate mRNA binder					0				
II	ERH	NA	2.72	-4.64	3.68	6.53	candidate mRNA binder			x		20				
II	C3orf26	NA	2.18	-5.15	3.67	5.77	RNAbinding	x		x		20		20		
II	RPS12	-4.84	NA	-2.46	3.65	7.36	ribosomal			x						
II	NAP1L1	-7.25	0.54	-3.08	3.62	6.43	candidate mRNA binder					0				
II	LRRC47	-4.76	NA	-2.42	3.59	5.26	RNAbinding	x								
II	KIF1C	-4.35	NA	-2.76	3.55	5.30	candidate mRNA binder					0				
II	EPPK1	NA	3.12	-3.95	3.53	4.96	candidate mRNA binder					0				
II	PNO1	-4.53	2.53	NA	3.53	6.13	RNAbinding	x		x						
II	EIF4ENIF1	-2.40	NA	-4.53	3.47	5.59	translationfactor	x				20				
II	EXOSC10	-3.95	NA	-2.94	3.44	5.56	RNAbinding	x		x						
II	MDH2	-5.79	1.04	-3.34	3.39	6.91	metabolicEnz			x		0				
II	NAA38	-3.86	2.88	NA	3.37	5.87	RNAbinding	x	x				50			
II	EIF5A	-4.81	1.39	-3.85	3.35	7.00	translationfactor	x								
II	NOL8	NA	1.61	-5.05	3.33	4.20	RNAbinding	x		x						
II	CHD2	-3.46	NA	-3.18	3.32	5.38	helicase					0				
II	TRNAU1AP	NA	2.70	-3.94	3.32	6.32	RNAbinding	x								
II	VCP	-3.51	NA	-3.11	3.31	7.32	candidate mRNA binder				613954, 167320	0				
II	BAG4	NA	2.23	-4.35	3.29	5.55	candidate mRNA binder					0				
II	WDR75	NA	1.99	-4.59	3.29	4.92	candidate mRNA binder			x						
II	RPL18A	-4.93	1.51	-3.34	3.26	6.95	ribosomal			x			20			
II	STRAP	-3.47	1.53	-4.72	3.24	7.45	candidate mRNA binder									
II	PTCD2	-4.19	2.28	NA	3.24	5.85	candidate mRNA binder					0				
II	JUN	NA	2.42	-4.04	3.23	5.85	candidate mRNA binder					0				
II	METTL16	-5.39	0.48	-3.77	3.21	6.56	candidate mRNA binder					20				
II	NSUN5	-4.76	1.66	NA	3.21	5.48	candidate mRNA binder			x		20				
II	RPS16	-4.56	1.47	-3.57	3.20	7.40	ribosomal			x			20			
II	MRPS24	-5.70	2.88	-1.02	3.20	6.44	ribosomal					50				
II	SMG1	-4.07	1.52	-4.00	3.20	5.68	candidate mRNA binder					50				
II	RPS17	-5.13	1.49	-2.93	3.19	7.48	ribosomal			x	612527	80				

II	RBBP6	-3.76	1.39	-4.38	3.18	6.42	candidate mRNA binder		x			0			
II	ANXA11	NA	1.74	-4.49	3.12	4.65	candidate mRNA binder								
II	PFN1	-5.35	0.99	-2.97	3.10	6.58	candidate mRNA binder					0			
II	PRPF31	-3.57	1.45	-4.29	3.10	5.80	RNAbinding	x			600138				
II	DUS3L	-2.93	NA	-3.26	3.10	6.01	candidate mRNA binder					50			
II	NPM1	-4.29	1.48	-3.49	3.08	8.44	RNAbinding	x			601626				
II	SSB	-4.38	1.52	-3.26	3.05	7.41	RNAbinding	x		x					
II	C1orf131	NA	1.92	-4.13	3.03	4.75	candidate mRNA binder								
II	C16orf80	-4.06	1.96	NA	3.01	5.70	candidate mRNA binder					0			
II	CXorf57	NA	2.00	-3.97	2.99	4.88	candidate mRNA binder					20			
II	MRPL44	-3.00	NA	-2.94	2.97	5.03	ribosomal								
II	PWP2	NA	2.12	-3.79	2.95	5.00	RNAbinding								
II	POLRMT	-3.76	1.16	-3.93	2.95	7.22	candidate mRNA binder					20			
II	RPS19BP1	-3.83	2.04	NA	2.93	5.93	candidate mRNA binder			x		50			
II	ESF1	NA	2.62	-3.22	2.92	5.24	RNAbinding			x		50			
II	SECISBP2L	-2.56	NA	-3.27	2.92	5.37	RNAbinding	x				20			
II	DDX24	-3.20	1.40	-4.11	2.90	6.05	helicase	x		x					20
II	TRIM28	NA	2.35	-3.43	2.89	5.30	candidate mRNA binder					20			
II	RPL11	-4.82	1.06	-2.78	2.89	6.88	ribosomal				612562		20		
II	HEATR6	-2.94	NA	-2.81	2.88	5.51	candidate mRNA binder					0			
II	DDX54	-3.56	1.32	-3.71	2.86	7.82	helicase	x		x					
II	RPS13	-4.19	0.86	-3.38	2.81	7.63	ribosomal			x					
II	CS	NA	2.23	-3.38	2.80	6.26	metabolicEnz					20			
II	RCC2	-4.16	1.51	-2.66	2.78	6.21	candidate mRNA binder					0			
II	DNTTIP2	-2.99	NA	-2.56	2.78	6.20	candidate mRNA binder			x		0			
II	RAN	-5.16	0.75	-2.37	2.76	7.06	candidate mRNA binder					20			
II	C1orf35	-3.52	2.00	NA	2.76	5.99	candidate mRNA binder					0			
II	NOL12	NA	2.25	-3.25	2.75	5.51	RNAbinding	x		x					
II	MTDH	-3.26	1.45	-3.53	2.74	6.36	candidate mRNA binder					0			
II	TFB2M	-3.78	0.93	-3.46	2.72	6.43	RNAmod	x				20			
II	NPM3	-4.19	1.04	-2.89	2.71	7.06	candidate mRNA binder			x		50			
II	C11orf31	NA	1.87	-3.54	2.71	5.18	candidate mRNA binder					0			
II	GLTSCR2	NA	2.08	-3.33	2.70	4.88	candidate mRNA binder					20			
II	DAP3	-3.18	1.51	-3.37	2.69	6.63	ribosomal					20			
II	UBA1	NA	2.58	-2.77	2.68	5.76	candidate mRNA binder					0			
II	ANXA7	NA	2.06	-3.28	2.67	4.86	candidate mRNA binder								
II	UTP20	NA	1.88	-3.46	2.67	5.79	candidate mRNA binder			x		50			
II	RPL26	-3.32	1.46	-3.16	2.64	7.45	ribosomal			x			20		
II	GCN1L1	NA	1.99	-3.29	2.64	4.48	translationfactor	x							
II	RPS23	-3.02	1.49	-3.33	2.61	7.07	ribosomal			x		80	20		
II	EIF2AK2	NA	1.71	-3.50	2.60	5.09	translationfactor	x		x					
II	EEF2	-2.89	1.44	-3.45	2.59	6.34	RNAbinding	x		x					
II	SNRPC	-3.20	1.96	NA	2.58	6.02	RNAbinding	x		x					
II	AKAP17A	-2.68	NA	-2.47	2.57	5.70	RNAbinding	x					20		
II	MRPL14	-3.87	1.06	-2.77	2.57	5.18	ribosomal					50			
II	RPS20	-4.18	0.48	-2.96	2.54	6.76	ribosomal								
II	MAP4	NA	1.60	-3.48	2.54	4.37	candidate mRNA binder					0			
II	RPL8	-3.44	1.27	-2.88	2.53	6.88	ribosomal			x			20		
II	ZNF385A	-3.41	2.97	-1.08	2.49	5.97	RNAbinding	x				50			
II	PDAP1	-4.17	0.83	-2.44	2.48	6.57	candidate mRNA binder					0			
II	NUCKS1	NA	2.13	-2.74	2.44	5.80	candidate mRNA binder					0			
II	MRPL27	-3.52	0.90	-2.89	2.44	5.97	ribosomal					50			
II	NSA2	-4.22	2.13	-0.93	2.43	6.41	candidate mRNA binder			x		80			
II	RPS27A	-2.62	1.34	-3.31	2.42	8.60	ribosomal			x		50			
II	PARP1	-3.73	0.86	-2.67	2.42	6.87	candidate mRNA binder					20			
II	NAT10	-2.92	1.11	-3.22	2.41	5.63	metabolicEnz			x		20			
II	RPL7	-3.27	1.20	-2.69	2.39	7.70	ribosomal			x					
II	HEATR1	-0.48	3.41	-3.24	2.38	5.77	candidate mRNA binder			x		80			
II	RPL3	-3.27	1.14	-2.70	2.37	7.06	ribosomal								
II	RPS3A	-3.51	1.16	-2.43	2.37	7.14	ribosomal								
II	RPL27A	-3.03	1.00	-3.02	2.35	7.68	ribosomal								
II	GNB2L1	-3.22	1.11	-2.70	2.34	5.47	candidate mRNA binder		x			20			
II	PPP1CC	-3.72	0.35	-2.81	2.29	5.72	candidate mRNA binder			x					
II	TOP3B	-2.94	0.99	-2.95	2.29	6.85	candidate mRNA binder					0			
II	WBSCR16	-3.29	0.59	-2.92	2.27	5.74	candidate mRNA binder					0			
II	NOP56	-2.58	1.36	-2.87	2.27	7.85	RNAbinding	x							
II	RPS19	-3.74	0.69	-2.32	2.25	6.80	ribosomal			x	105650				
II	ZCCHC11	-2.34	1.15	-3.19	2.23	6.94	RNAbinding	x				50			
II	PNISR	-2.17	NA	-2.28	2.23	5.53	candidate mRNA binder					50			
II	RPS7	-3.37	0.76	-2.53	2.22	7.13	ribosomal			x	612563		20		
II	ZNF346	-2.55	0.58	-3.48	2.20	5.20	RNAbinding	x							
II	ARL6IP4	-2.49	1.14	-2.93	2.19	5.59	candidate mRNA binder			x		80	20		
II	RPL17	-3.07	0.99	-2.44	2.17	6.85	ribosomal			x		80	20		
II	RPL24	-2.78	0.13	-3.58	2.16	7.79	ribosomal								
II	HSPA1A	-3.82	0.27	-2.36	2.15	6.65	HeatShock					20			
II	DDX31	-2.05	NA	-2.20	2.12	5.87	helicase	x		x					
II	PRIC285	-2.32	1.07	-2.91	2.10	5.74	helicase	x							
II	SUPT16H	NA	1.70	-2.44	2.07	5.37	candidate mRNA binder			x		20			
II	PUS7	-4.02	2.66	0.47	2.07	6.08	RNAmod	x							
II	TFAM	-3.02	0.35	-2.81	2.06	6.08	candidate mRNA binder					20			
II	RPS8	-2.72	0.88	-2.52	2.04	7.73	ribosomal			x		80	20		
II	CRNKL1	-2.51	0.71	-2.79	2.00	5.82	RNAbinding	x	x				20		
II	RPL23A	-2.66	0.68	-2.63	1.99	7.76	ribosomal			x					
II	EZR	-2.40	0.90	-2.50	1.93	5.40	candidate mRNA binder					0			
II	DUSP11	-2.74	1.76	-1.18	1.89	6.48	RNAbinding	x		x					
II	MTPAP	-2.62	0.39	-2.67	1.89	7.19	RNAbinding	x			613672		20		
II	NOP58	-2.10	1.25	-2.30	1.88	7.85	RNAbinding	x		x	611523				
II	LBR	-2.39	0.15	-3.03	1.86	5.13	candidate mRNA binder					20			
II	TOP1	-1.81	1.50	-2.11	1.81	5.88	candidate mRNA binder					20			
II	KIAA0020	-1.99	1.23	-1.91	1.71	6.92	RNAbinding	x							
II	RPL14	-2.25	0.90	-1.87	1.67	7.11	ribosomal					0			
II	RPS24	-2.15	0.29	-2.20	1.54	8.16	ribosomal			x	610629	20			
II	RPL15	-1.99	0.97	-1.58	1.52	8.10	ribosomal					0			
II	NOP14	-7.23	1.47	-2.31	3.67	5.98	RNAbinding	x		x					
II	APEX1	-6.19	0.75	-2.06	3.00	6.00	nuclease	x		x					

II	POLR2A	-3.49	NA	-2.29	2.89	4.96	candidate mRNA binder						50			
II	NKAP	-1.80	NA	-3.98	2.89	5.89	candidate mRNA binder						0			
II	SUPV3L1	-3.26	NA	-2.07	2.66	5.74	helicase	x								
II	SRPK2	-3.41	NA	-1.79	2.60	5.96	candidate mRNA binder						50			
II	RPL12	-5.51	0.44	-1.75	2.57	5.97	ribosomal			x						
II	RPS15A	-4.51	0.89	-2.21	2.54	6.33	ribosomal			x				20		
II	PIN4	-4.87	0.24	-2.17	2.43	6.21	candidate mRNA binder						20			
II	RPS27	-3.56	1.12	-2.10	2.26	7.32	ribosomal							20		
II	RPL30	-4.22	0.44	-2.06	2.24	6.59	ribosomal			x						
II	BTF3	-2.16	1.22	-3.32	2.23	5.98	candidate mRNA binder						0			
II	SRRM2	-2.17	0.62	-3.71	2.17	7.36	RNAbinding	x	x					20		
II	DDX23	-2.37	NA	-1.96	2.16	6.27	helicase	x	x	x						
II	RPLP0	-4.09	0.11	-2.20	2.13	6.30	ribosomal							20		
II	GTF2F1	-3.21	1.51	-1.66	2.13	5.57	RNAbinding	x		x			50			
II	RSRC2	-2.45	NA	-1.74	2.09	5.59	candidate mRNA binder						20			
II	LARP7	-2.27	1.86	NA	2.06	5.26	RNAbinding	x						20		
II	RPL27	-3.57	0.31	-2.18	2.02	7.25	ribosomal			x			80			
II	RPS25	-2.73	1.32	-1.99	2.01	7.40	ribosomal			x						
II	MRPL20	-3.71	0.03	-2.14	1.96	7.01	ribosomal									
II	RPL4	-2.34	1.29	-2.24	1.96	7.93	ribosomal			x				20		
II	RPL10A	-3.77	0.01	-2.09	1.96	6.93	ribosomal							20		
II	RPS18	-3.38	0.11	-2.30	1.93	6.82	ribosomal			x				20		
II	EEF1A1	-3.49	0.10	-2.15	1.92	7.38	translationfactor	x		x						
II	HMGB1	-3.61	0.41	-1.72	1.91	7.20	RNAbinding	x					0			
II	UBE2N	-2.65	0.47	-2.26	1.79	5.63	candidate mRNA binder						0			
II	FAM133B	-2.78	0.73	-1.82	1.78	6.48	candidate mRNA binder									
II	HIST1H1C	-1.50	1.94	-1.85	1.76	8.41	histone			x			0			
II	DEK	-1.56	1.75	-1.99	1.76	6.65	candidate mRNA binder			x	125264		0			
II	SUB1	-1.64	1.61	-1.99	1.74	7.30	candidate mRNA binder			x			20			
II	PRDX1	-3.13	0.04	-1.98	1.72	6.14	metabolicEnz			x			0			
II	HMGB2	-3.18	0.36	-1.60	1.71	6.67	candidate mRNA binder						0			
II	DDX51	-1.81	0.32	-2.75	1.63	6.47	helicase	x		x						20
II	HSP90AB1	-2.80	0.38	-1.66	1.62	6.86	HeatShock						0			
II	RPL37A	-2.90	0.14	-1.71	1.58	6.10	ribosomal						80	20		
II	RPL21	-2.52	0.09	-2.06	1.55	6.93	ribosomal							20		
III	MRPL21	-6.04	NA	NA	6.04	5.49	ribosomal									
III	DUSP14	-5.72	NA	NA	5.72	5.76	candidate mRNA binder						0			
III	NFX1	NA	NA	-5.46	5.46	4.49	candidate mRNA binder						0			
III	C1orf204	-5.24	NA	NA	5.24	6.20	candidate mRNA binder						0			
III	ELAC2	NA	NA	-4.91	4.91	4.96	nuclease				176807		50			
III	EIF4E	NA	NA	-4.81	4.81	5.37	translationfactor	x								
III	DDX41	-4.73	NA	NA	4.73	5.75	helicase	x	x	x				20	20	
III	CCDC137	NA	NA	-4.63	4.63	5.51	candidate mRNA binder									
III	CLK3	-4.30	NA	NA	4.30	5.87	candidate mRNA binder						20			
III	NOL11	NA	NA	-4.25	4.25	4.53	candidate mRNA binder			x			50			
III	P4HB	NA	NA	-4.07	4.07	5.62	metabolicEnz						0			
III	MEPCE	-3.92	NA	NA	3.92	6.12	candidate mRNA binder						20			
III	WIBG	NA	NA	-3.91	3.91	6.34	RNAbinding	x						20		
III	ZC3H13	-3.86	NA	NA	3.86	5.59	RNAbinding	x					20			
III	SRP72	NA	0.89	-6.81	3.85	5.72	RNAbinding	x		x				20		
III	FASN	NA	NA	-3.78	3.78	4.77	metabolicEnz						0			
III	EXOSC6	NA	NA	-3.61	3.61	4.86	nuclease	x								
III	MYH9	-0.72	NA	-6.12	3.42	4.86	candidate mRNA binder			x	603622, 153650, 153640,		0			
III	NBPF10	NA	3.37	NA	3.37	5.04	RNAbinding	x					0			
III	DIDO1	-1.31	NA	-5.39	3.35	5.41	candidate mRNA binder						20			
III	ARF1	-6.35	0.25	NA	3.30	5.44	candidate mRNA binder						0			
III	EIF4E2	NA	NA	-3.30	3.30	6.31	translationfactor	x								
III	PRPF38A	-3.29	NA	NA	3.29	6.35	candidate mRNA binder		x				80	20		
III	FLNB	-1.40	NA	-5.07	3.23	4.65	candidate mRNA binder				108721, 108720, 112310,		0			
III	SPATS2L	NA	NA	-3.18	3.18	4.50	candidate mRNA binder			x			50			
III	KPNA2	NA	NA	-3.16	3.16	5.37	candidate mRNA binder			x			0			
III	MRPS14	-4.81	1.43	NA	3.12	5.65	ribosomal						50			
III	RC3H2	NA	2.91	NA	2.91	6.00	RNAbinding	x					0			
III	KPNB1	NA	NA	-2.73	2.73	5.18	candidate mRNA binder						20			
III	CLTC	NA	NA	-2.62	2.62	5.62	candidate mRNA binder						0			
III	RPF1	NA	1.32	-3.73	2.53	4.56	RNAbinding	x		x						
III	GDI2	NA	NA	-2.53	2.53	4.65	candidate mRNA binder						0			
III	MRPS12	-4.13	0.92	NA	2.52	5.91	ribosomal						50			
III	TXN	-4.86	0.17	NA	2.52	5.71	metabolicEnz						0			
III	ANP32A	NA	0.21	-4.82	2.51	5.14	candidate mRNA binder						50			
III	TEFM	-2.51	NA	NA	2.51	6.50	candidate mRNA binder						20			
III	DHX29	-4.66	NA	-0.18	2.42	4.65	helicase	x								
III	DNM1	NA	NA	-2.38	2.38	3.94	candidate mRNA binder						0			
III	CDKN2A	-3.31	1.39	NA	2.35	6.40	candidate mRNA binder				151623, 155755, 155601,		0			
III	RP9	-3.09	1.52	NA	2.30	5.90	candidate mRNA binder				180104		80	20		
III	RARS2	NA	NA	-2.29	2.29	5.64	RNAbinding	x								
III	ZC3H15	NA	0.92	-3.58	2.25	5.91	candidate mRNA binder						20			
III	RPS27L	-3.42	1.02	NA	2.22	5.70	ribosomal						50			
III	HSPA9	NA	NA	-2.22	2.22	4.50	HeatShock			x			20			
III	GSPT1	NA	NA	-2.18	2.18	4.21	candidate mRNA binder									
III	ERAL1	-4.93	NA	0.60	2.17	6.12	RNAbinding	x								
III	LYAR	NA	1.34	-2.97	2.16	5.61	candidate mRNA binder			x			50			
III	FAM50A	NA	1.54	-2.74	2.14	5.14	candidate mRNA binder		x				20			
III	PAPD5	NA	NA	-2.11	2.11	4.33	RNAbinding	x		x						
III	MRPL10	-2.98	1.24	NA	2.11	7.00	ribosomal						20			
III	MRM1	NA	NA	-2.09	2.09	5.23	RNAmod						0			
III	WBSCR22	-4.76	NA	0.62	2.07	5.88	candidate mRNA binder			x			0			
III	SART3	-2.60	1.50	NA	2.05	5.96	RNAbinding	x		x	175900			20		
III	GNL2	NA	1.51	-2.58	2.05	5.30	candidate mRNA binder			x			50			
III	SLC25A11	-3.83	0.25	NA	2.04	6.02	candidate mRNA binder						0			
III	PRPF3	-2.88	1.20	NA	2.04	5.44	RNAbinding	x	x	x	601414		80	20		
III	ALDOA	NA	NA	-2.03	2.03	5.48	metabolicEnz				611881		0			
III	RPL29	-1.35	1.27	-3.46	2.02	7.45	ribosomal									
III	PNPT1	-2.85	NA	-1.16	2.01	5.82	nuclease	x								
III	TBCA	NA	2.67	-1.33	2.00	5.68	candidate mRNA binder						0			

III	EIF5	NA	1.09	-2.89	1.99	5.53	translationfactor	x						
III	NDUFV3	NA	0.99	-2.95	1.97	5.58	metabolicEnz						0	
III	UBTF	NA	1.23	-2.70	1.97	4.81	candidate mRNA binder			x			20	
III	EIF5B	-2.45	1.47	NA	1.96	4.78	translationfactor	x		x				
III	MACF1	-1.19	2.63	NA	1.91	4.71	candidate mRNA binder						0	
III	NCBP1	-2.74	NA	-0.99	1.86	5.72	RNAbinding	x	x					
III	RRP9	NA	1.29	-2.36	1.82	5.38	RNAbinding	x		x				
III	HDAC2	NA	1.42	-2.15	1.79	6.54	candidate mRNA binder			x			0	
III	KRR1	-1.58	1.16	-2.41	1.72	7.13	RNAbinding	x						
III	NOM1	NA	NA	-1.70	1.70	5.49	RNAbinding	x		x				
III	EIF2S2	-1.49	0.92	-2.53	1.65	7.43	translationfactor	x		x				
III	HIST1H1D	-1.05	1.36	-2.49	1.64	5.42	histone			x			0	
III	HSP90AA1	-3.34	0.26	-1.21	1.60	6.77	HeatShock						0	
III	UBC	-1.49	0.63	-2.69	1.60	6.01	candidate mRNA binder						0	
III	H1FX	-1.02	1.12	-2.64	1.59	6.91	histone			x			0	
III	PSIP1	-1.21	1.40	-2.15	1.59	6.60	candidate mRNA binder						20	
III	HMGN5	NA	0.97	-2.17	1.57	5.33	candidate mRNA binder						20	
III	DYNC1H1	-0.46	NA	-2.62	1.54	4.92	candidate mRNA binder						0	
III	H1F0	-1.23	1.93	-1.44	1.53	6.63	histone						0	
III	HMGN2	NA	0.81	-2.18	1.50	6.76	candidate mRNA binder						20	
III	RPL36A	-2.46	0.63	-1.34	1.48	6.05	ribosomal						80	20
III	CCAR1	NA	0.56	-2.39	1.48	4.26	candidate mRNA binder						50	20
III	FLNA	-0.49	NA	-2.46	1.47	5.02	candidate mRNA binder			x	314400, 300321, 305620,		0	
III	MAPRE1	NA	0.38	-2.56	1.47	5.35	candidate mRNA binder						0	
III	PEBP1	NA	0.28	-2.66	1.47	5.56	candidate mRNA binder							20
III	RPL7A	-1.52	0.51	-2.35	1.46	7.63	ribosomal			x				20
III	SREK1	-0.93	NA	-1.97	1.45	6.32	RNAbinding	x						
III	FAU	-1.57	0.60	-1.97	1.38	7.11	RNAbinding			x				
III	RPL10	-2.22	0.66	-1.22	1.36	6.57	ribosomal				300847		80	
III	HSPA1B	NA	0.63	-2.09	1.36	5.29	HeatShock							
III	RPF2	-2.00	0.58	-1.50	1.36	6.80	candidate mRNA binder						0	
III	CANX	NA	0.48	-2.11	1.30	5.12	candidate mRNA binder							
III	PRPF4B	0.20	NA	-2.68	1.24	8.54	RNAbinding		x					20
III	GOT2	NA	0.35	-2.09	1.22	5.72	metabolicEnz			x			0	
III	SNRPB	-2.35	0.76	-0.51	1.21	7.27	RNAbinding	x	x					20
III	HIST2H4B	-0.92	0.16	-2.38	1.15	9.40	histone						0	
III	SNRPD3	-1.94	1.00	-0.50	1.15	7.90	RNAbinding	x	x	x				
III	CCDC86	NA	0.65	-1.63	1.14	5.42	candidate mRNA binder			x			20	
III	PPIG	-1.97	1.43	0.00	1.13	6.55	candidate mRNA binder		x				50	
III	RPL5	-1.87	0.20	-1.19	1.09	6.82	ribosomal			x	612561		0	
III	CKAP4	NA	0.56	-1.59	1.08	6.31	candidate mRNA binder						0	
III	HMGB3	-3.62	0.55	2.95	0.41	5.65	candidate mRNA binder						20	

Supplementary Table S2: Pfam/Interpro domain and SCOP enrichment analyses

Work sheets contain information on summary and detailed description of Pfam, InterPro, SCOP enrichment analyses results significant over-representation at corrected p-value ≤ 0.05

Table S2A: Selected enriched SCOP superfamily folds

RNA-binding domains

Domain	*Representativ*	*SCOP*	*Corrected p-value*
RBD	PABPC1	d.58.7	9.90E-124
KH (type I)	HNRNPK	d.51.1	1.80E-21
dsRNA	STAU1	d.50.1	2.70E-10
LSM	LSM14A	b.38.1	8.10E-08

Putative RNA-binding

HMG box	HMGB1	a.21.1	4.30E-11
"Winged helix" DNA-binding	DDX54	a.4.5	0.0004
AlbA-like	C9orf23	d.68.6	0.0003

Table S2B: Selected enriched Pfam and InterPro domains

RNA-binding domains

Domain	*Representativ*	*Pfam*	*InterPro*	*Corrected P-value*
RRM	PABPC1	PF0007	**IPR000504**	9.80E-112
KH (type I and II)	HNRNPK	PF0001	**IPR004087**	6.30E-21
ZnF-CCCH	U2AF1	PF0064	**IPR000571**	9.30E-19
ZnF-CCHC	LIN28B	PF0009	**IPR001878**	6.40E-11
dsRNA	STAU1	PF0003	**IPR014720**	8.50E-10
OB_NTP_bind	DHX9	**PF0771**	IPR012340	9.30E-09
SAP	HNRNPU	**PF0203**	IPR003034	9.40E-08
LSM	LSM14A	PF0142	**IPR006649**	2.00E-06
YTH	YTH	PF0414	**IPR007275**	5.70E-05
S1	DHX8	PF0057	**IPR022967**	0.0002
PWI	SRRM1	PF0148	**IPR002483**	0.0012
MIF4G	EIF4G1	PF0285	**IPR016021**	0.0036
La	SSB	**PF0538**	IPR006630	0.0052
PurA	PURA	**PF0484**	IPR006628	0.025
Pumilio	Pum1	PF0080	**IPR001313**	0.0634

Putative RNA-binding

SWAP/SURP	SF3A1	PF0180	**IPR000061**	0.0072
DZF	ILF3	**PF0752**	IPR006561	0.0092
RAP	FASTKD1	**PF0837**	IPR013584	0.0609
HAT helix	SART3	PF0218	**IPR003107**	0.1667

Supplementary Table S3: Overview on cell lines generated for RNA-binding proteins and CLIP assay results

Protein	Plasmid used for stable transfection	Expression	Immunoprecipitation	CLIP assay
Controls				
CAPRIN1	pFRT/TO/HIS/FLAG/HA-CAPRIN 1	positive	positive	positive
HNRNPD	pFRT/TO/HIS/FLAG/HA-HNRNPD	positive	positive	positive
HNRNPR	pFRT/TO/HIS/FLAG/HA-HNRNPR	positive	positive	positive
HNRNPU	pFRT/TO/HIS/FLAG/HA-HNRNPU	positive	positive	positive
MYEF2	pFRT/TO/HIS/FLAG/HA-MYEF2	positive	positive	positive
LDHA	pFRT/TO/FLAG/HA-LDHA	positive	positive	negative
PGK1	pFRT/TO/HIS/FLAG/HA-PGK1	positive	positive	negative
novel mRNA binders				
AKAP8L	pFRT/TO/HIS/FLAG/HA-AKAP8L	positive	positive	positive
ALKBH5	pFRT/TO/HIS/FLAG/HA-ALKBH5	positive	positive	positive
API5	pFRT/TO/HIS/FLAG/HA-API5	positive	positive	positive
BTF3	pFRT/TO/HIS/FLAG/HA-BTF3	positive	positive	positive
C17orf85	pFRT/TO/HIS/FLAG/HA-C17orf85	positive	positive	positive
C22orf28	pFRT/TO/HIS/FLAG/HA-C22orf28	positive	positive	positive
CSNK1E	pFRT/TO/HIS/FLAG/HA-CSNK1E	positive	positive	positive
EDF1	pFRT/TO/HIS/FLAG/HA-EDF1	positive	positive	positive
FAM98A	pFRT/TO/HIS/FLAG/HA-FAM98A	positive	positive	positive
IFIT5	pFRT/TO/HIS/FLAG/HA-IFIT5	positive	positive	positive
KIAA1967	pFRT/TO/HIS/FLAG/HA-KIAA1967	positive	positive	positive
MKRN2	pFRT/TO/HIS/FLAG/HA-MKRN2	positive	positive	positive
MYBBP1A	pFRT/TO/HIS/FLAG/HA-MYBBP1A	positive	positive	positive
PES1	pFRT/TO/HIS/FLAG/HA-PES1	positive	positive	positive
PRDX1	pFRT/TO/HIS/FLAG/HA-PRDX1	positive	positive	positive
SART1	pFRT/TO/HIS/FLAG/HA-SART1	positive	positive	positive
USP10	pFRT/TO/FLAG/HA-USP10	positive	positive	positive
YTHDF2	pFRT/TO/HIS/FLAG/HA-YTHDF2	positive	positive	positive
ZC3H7B	pFRT/TO/HIS/FLAG/HA-ZC3H7B	positive	positive	positive
BZW1	pFRT/TO/HIS/FLAG/HA-BZW1	positive	positive	negative
C16orf80	pFRT/TO/HIS/FLAG/HA-C16orf80	positive	positive	negative
AKAP1	pFRT/TO/HIS/FLAG/HA-AKAP1	positive	negative	
CDK13	pFRT/TO/HIS/FLAG/HA-CDK13	positive	negative	
DUSP11	pFRT/TO/HIS/FLAG/HA-DUSP11	positive	negative	
MDH2	pFRT/TO/HIS/FLAG/HA-MDH2	positive	negative	
NKRF	pFRT/TO/FLAG/HA-NKRF	positive	negative	
THRAP3	pFRT/TO/HIS/FLAG/HA-THRAP3	positive	negative	
YARS2	pFRT/TO/HIS/FLAG/HA-YARS2	positive	negative	
ZC3H18	pFRT/TO/HIS/FLAG/HA-ZC3H18	positive	negative	

Supplementary table S4: IFIT5 target genes

Gene	5'UTR	CDS	3'UTR	intron	exon	total
XIST	0	0	0	3	30	33
ERC1	0	0	0	33	0	33
TMPO	0	0	0	0.47	29.53	30
KLHL15	0	0	0	2	28	30
MAP1B	0	0	0	0	29	29
CBX5	0	0	0	0	26.001	26.001
NUCKS1	0	0	0	1	25	26
CEBPG	0	0	0	0	25	25
EEF2	0	0	0	2	19	21
NUFIP2	0	0	0	0	20	20
HNRNPA2B1	0	0	0	5	15	20
GRPEL2	0	0	0	0	19	19
SRSF1	0	0	0	0.045	17.956	18.001
TMBIM6	0	0	0	0	18	18
SRRM2	0	0	0	0	18	18
RLIM	0	0	0	0	18	18
SMAD5	0	0	0	0	17.001	17.001
SON	0	0	0	3	14	17
HNRNPF	0	0	0	0	16.998	16.998
EIF4G2	0	0	0	0	15.001	15.001
ZNF295	0	0	0	0.005	14.995	15
CCND2	0	0	0	1	14	15
ZNF146	0	0	0	0.999	12.999	14.001
SP1	0	0	0	0	14.001	14.001
STC2	0	0	0	1	13	14
PURB	0	0	0	0	14	14
NRAS	0	0	0	0	14	14
NCL	0	0	0	0	14	14
HN1L	0	0	0	0	14	14
C16orf72	0	0	0	12	2	14
PEG10	0	0	0	0	13.002	13.002
ZIC5	0	0	0	0	13	13
SRSF2	0	0	0	0.02	12.98	13
SERBP1	0	0	0	0	13	13
IFIT5	0	0	0	0	13	13
HNRNPA0	0	0	0	0	13	13
ENO1	0	0	0	1	12	13
CDK1	0	0	0	2.02	10.98	13
RBM12	0	0	0	0	12.596	12.596
ZNF711	0	0	0	0	12	12
ZNF529	0	0	0	1.059	10.942	12
GLO1	0	0	0	0	12	12
ARPP19	0	0	0	0	12	12

AMD1	0	0	0	2.01	9.99	12
TMED5	0	0	0	0	11.001	11.001
SLC38A2	0	0	0	0	11	11
SCML1	0	0	0	0	11	11
PPP1CC	0	0	0	0.059	10.941	11
NUPL1	0	0	0	6	5	11
MAZ	0	0	0	1.01	9.99	11
LCOR	0	0	0	0.015	10.985	11
FAM168B	0	0	0	0	11	11
EIF5	0	0	0	1	10	11
CD164	0	0	0	1.014	9.986	11
C5orf51	0	0	0	0	11	11
LUC7L2	0	0	0	7.999	3	10.999
SYNCRIP	0	0	0	0.025	10.977	10.997
CSDE1	0	0	0	0	10.002	10.002
CNBP	0	0	0	1.002	9	10.002
HNRPDL	0	0	0	1.004	8.995	10.001
SMC1A	0	0	0	1	9	10
SET	0	0	0	0.02	9.98	10
RPS23	0	0	0	0	10	10
PTP4A1	0	0	0	0	10	10
OGT	0	0	0	0	10	10
NOLC1	0	0	0	0	10	10
MKI67	0	0	0	0	10	10
MARCKS	0	0	0	0	10	10
KHSRP	0	0	0	3	7	10
GNG12	0	0	0	0	10	10
CCNT1	0	0	0	0	10	10
BUB3	0	0	0	1.059	8.941	10
ARF1	0	0	0	0	10	10
ADNP	0	0	0	0	10	10
TROVE2	0	0	0	5.014	4.983	9.999
BCLAF1	0	0	0	0.999	9	9.999
PAFAH1B2	0	0	0	0.234	8.767	9.001
CGGBP1	0	0	0	1	8.001	9.001
ZNF703	0	0	0	0	9	9
ZNF460	0	0	0	0	9	9
ZNF207	0	0	0	0	9	9
USP1	0	0	0	0	9	9
U2AF2	0	0	0	0	9	9
TUG1	0	0	0	1	8	9
TSPYL1	0	0	0	0	9	9
TRIM28	0	0	0	0	9	9
SOX4	0	0	0	0	9	9
SNHG1	0	0	0	0	9	9
SLFN11	0	0	0	0	9	9

RAB11A	0	0	0	0	9	9
PRPF8	0	0	0	1	8	9
PPP2CA	0	0	0	1	8	9
PPP1R15B	0	0	0	0	9	9
PMAIP1	0	0	0	0	9	9
MATR3	0	0	0	3.01	5.99	9
HNRNPAB	0	0	0	0.01	8.99	9
GNB2L1	0	0	0	2	7	9
EEF1A1	0	0	0	0	9	9
CERS2	0	0	0	0	9	9
CANX	0	0	0	0	9	9
UHMK1	0	0	0	0	8.001	8.001
CELF1	0	0	0	0.055	7.947	8.001
ZNF678	0	0	0	0	8	8
YWHAG	0	0	0	0	8	8
YOD1	0	0	0	0	8	8
UCK2	0	0	0	0	8	8
UBQLN2	0	0	0	0	8	8
TRA2B	0	0	0	0	8	8
TMEM30A	0	0	0	0	8	8
TMED2	0	0	0	0	8	8
TBP	0	0	0	8	0	8
TAB2	0	0	0	0	8	8
SPEN	0	0	0	1	7	8
SLC7A1	0	0	0	0	8	8
SLC38A1	0	0	0	1	7	8
SLC16A1	0	0	0	1	7	8
SF3B3	0	0	0	0	8	8
SCD	0	0	0	0	8	8
PNN	0	0	0	0	8	8
NAA50	0	0	0	0	8	8
MZT1	0	0	0	0	8	8
HNRNPUL1	0	0	0	0	8	8
H3F3B	0	0	0	0	8	8
H2AFX	0	0	0	0	8	8
CCNG1	0	0	0	1	7	8
C6orf228	0	0	0	0	8	8
ATP5B	0	0	0	0	8	8
ARL6IP1	0	0	0	0	8	8
ARL5B	0	0	0	0	8	8
ACTG1	0	0	0	0.005	7.995	8
SHMT2	0	0	0	1.069	6.93	7.999
CREBZF	0	0	0	0.087	7.912	7.999
ZFX	0	0	0	0	7	7
TOMM20	0	0	0	0	7	7
TMEM123	0	0	0	0	7	7

TMED10	0	0	0	0	7	7
STARD7	0	0	0	0	7	7
SRSF7	0	0	0	0	7	7
SREK1IP1	0	0	0	0	7	7
SLC7A5	0	0	0	0	7	7
SERINC1	0	0	0	4	3	7
2-Sep	0	0	0	0	7	7
SARS	0	0	0	1.04	5.96	7
RMND5A	0	0	0	0	7	7
RBM39	0	0	0	3.997	3.003	7
RBM3	0	0	0	0	7	7
PSAT1	0	0	0	0	7	7
PDCL	0	0	0	0	7	7
PCNA	0	0	0	0	7	7
PCBP1	0	0	0	0	7	7
PARP1	0	0	0	0	7	7
LIN28B	0	0	0	0	7	7
KLHDC5	0	0	0	0	7	7
ICMT	0	0	0	0	7	7
HSP90AA1	0	0	0	1	6	7
HNRNPL	0	0	0	2	5	7
GINS1	0	0	0	0	7	7
ELAVL1	0	0	0	0	7	7
EIF1	0	0	0	0	7	7
DEK	0	0	0	1	6	7
DCAF7	0	0	0	0	7	7
CPSF6	0	0	0	0	7	7
CLTC	0	0	0	0	7	7
CKAP2	0	0	0	1	6	7
CENPF	0	0	0	0	7	7
CCNB1	0	0	0	0	7	7
CCDC117	0	0	0	0	7	7
CBX3	0	0	0	1	6	7
C5orf24	0	0	0	0	7	7
BCOR	0	0	0	0	7	7
ABHD13	0	0	0	0	7	7
PKM2	0	0	0	0.03	6.973	6.999
MCL1	0	0	0	0	6.999	6.999
ATP1A1	0	0	0	0	6.999	6.999
ZC3H11A	0	0	0	5.05	1	6.05
H2AFV	0	0	0	4.005	1.996	6.001
ZNF850	0	0	0	0	6	6
ZFP30	0	0	0	0	6	6
YWHAE	0	0	0	1	5	6
YTHDF3	0	0	0	0	6	6
XBP1	0	0	0	0.01	5.99	6

VIM	0	0	0	0	6	6
USP22	0	0	0	0	6	6
TXNIP	0	0	0	0	6	6
TUBA1B	0	0	0	0	6	6
TSPAN3	0	0	0	1.004	4.995	6
TOR1AIP2	0	0	0	0.045	5.956	6
TMEM48	0	0	0	0	6	6
TMED9	0	0	0	0	6	6
SURF4	0	0	0	1	5	6
STMN1	0	0	0	1.017	4.983	6
SP3	0	0	0	0	6	6
SNRPD1	0	0	0	0	6	6
SMNDC1	0	0	0	0	6	6
SFPQ	0	0	0	0	6	6
SF3A1	0	0	0	0	6	6

Supplementary table S5: enriched GO terms among IFIT5 target genes

GO ID	Term	Count	%	PValue
GO:0008380	RNA splicing	20	9.90	5.7E-09
GO:0000377	RNA splicing, via transesterification reactions with bulged a	15	7.43	1.3E-08
GO:0000398	nuclear mRNA splicing, via spliceosome	15	7.43	1.3E-08
GO:0000375	RNA splicing, via transesterification reactions	15	7.43	1.3E-08
GO:0016071	mRNA metabolic process	22	10.89	1.6E-08
GO:0006397	mRNA processing	19	9.41	2.2E-07
GO:0006396	RNA processing	24	11.88	7.4E-07
GO:0010608	posttranscriptional regulation of gene expression	13	6.44	2.2E-05
GO:0006412	translation	15	7.43	1.2E-04
GO:0006414	translational elongation	8	3.96	3.6E-04
GO:0006605	protein targeting	11	5.45	5.4E-04
GO:0046907	intracellular transport	20	9.90	9.9E-04
GO:0006886	intracellular protein transport	14	6.93	1.3E-03
GO:0010605	negative regulation of macromolecule metabolic process	21	10.40	1.5E-03
GO:0006446	regulation of translational initiation	5	2.48	2.0E-03
GO:0034613	cellular protein localization	14	6.93	2.9E-03
GO:0070727	cellular macromolecule localization	14	6.93	3.1E-03
GO:0006096	glycolysis	5	2.48	3.3E-03
GO:0032268	regulation of cellular protein metabolic process	15	7.43	3.8E-03
GO:0006006	glucose metabolic process	8	3.96	4.0E-03
GO:0042981	regulation of apoptosis	21	10.40	4.3E-03
GO:0043067	regulation of programmed cell death	21	10.40	4.8E-03
GO:0010941	regulation of cell death	21	10.40	4.9E-03
GO:0006333	chromatin assembly or disassembly	7	3.47	6.5E-03
GO:0010629	negative regulation of gene expression	15	7.43	6.5E-03
GO:0043066	negative regulation of apoptosis	12	5.94	6.9E-03
GO:0006007	glucose catabolic process	5	2.48	6.9E-03
GO:0034621	cellular macromolecular complex subunit organization	12	5.94	7.4E-03
GO:0016481	negative regulation of transcription	14	6.93	7.4E-03
GO:0043069	negative regulation of programmed cell death	12	5.94	7.7E-03
GO:0060548	negative regulation of cell death	12	5.94	7.8E-03
GO:0034622	cellular macromolecular complex assembly	11	5.45	9.1E-03
GO:0006417	regulation of translation	7	3.47	9.2E-03
GO:0070585	protein localization in mitochondrion	4	1.98	9.8E-03
GO:0006626	protein targeting to mitochondrion	4	1.98	9.8E-03
GO:0015031	protein transport	19	9.41	1.1E-02
GO:0045184	establishment of protein localization	19	9.41	1.2E-02
GO:0019320	hexose catabolic process	5	2.48	1.3E-02
GO:0006839	mitochondrial transport	5	2.48	1.3E-02
GO:0010558	negative regulation of macromolecule biosynthetic proces	15	7.43	1.3E-02
GO:0019318	hexose metabolic process	8	3.96	1.3E-02
GO:0046365	monosaccharide catabolic process	5	2.48	1.4E-02
GO:0031327	negative regulation of cellular biosynthetic process	15	7.43	1.6E-02
GO:0045449	regulation of transcription	47	23.27	1.6E-02
GO:0045934	negative regulation of nucleobase, nucleoside, nucleotide a	14	6.93	1.7E-02
GO:0009890	negative regulation of biosynthetic process	15	7.43	1.8E-02
GO:0045892	negative regulation of transcription, DNA-dependent	11	5.45	1.9E-02
GO:0051172	negative regulation of nitrogen compound metabolic proce	14	6.93	1.9E-02
GO:0051253	negative regulation of RNA metabolic process	11	5.45	2.1E-02

GO:0046164	alcohol catabolic process	5	2.48	2.2E-02
GO:0006334	nucleosome assembly	5	2.48	2.4E-02
GO:0044275	cellular carbohydrate catabolic process	5	2.48	2.5E-02
GO:0007049	cell cycle	18	8.91	2.6E-02
GO:0051276	chromosome organization	13	6.44	2.6E-02
GO:0005996	monosaccharide metabolic process	8	3.96	2.7E-02
GO:0031497	chromatin assembly	5	2.48	2.7E-02
GO:0006457	protein folding	7	3.47	2.9E-02
GO:0065004	protein-DNA complex assembly	5	2.48	3.1E-02
GO:0022411	cellular component disassembly	4	1.98	3.2E-02
GO:0001889	liver development	4	1.98	3.2E-02
GO:0006282	regulation of DNA repair	3	1.49	3.3E-02
GO:0043488	regulation of mRNA stability	3	1.49	3.3E-02
GO:0034728	nucleosome organization	5	2.48	3.4E-02
GO:0007005	mitochondrion organization	6	2.97	3.5E-02
GO:0043487	regulation of RNA stability	3	1.49	3.9E-02
GO:0008104	protein localization	19	9.41	4.0E-02
GO:0033365	protein localization in organelle	6	2.97	4.3E-02
GO:0022403	cell cycle phase	11	5.45	4.6E-02
GO:0006310	DNA recombination	5	2.48	4.9E-02

Table S6: Overview of the identified proteins in D.melanogaster embryos

Protein ID	Gene name	log10 iBAQ YW_ctrl	log10 iBAQ YW_XL	iBAQ Ratio (XL/ctrl)
Q9VJ38	mRpL13	N/A	9.29	XL>ctrl
Q9VYI0	fne	N/A	9.27	XL>ctrl
Q9VXB5	mRpL22	N/A	9.10	XL>ctrl
Q9W253	mRpS29	N/A	9.08	XL>ctrl
Q9VZD5	mRpS6	N/A	9.04	XL>ctrl
Q8SYP9	mRpS5	N/A	9.01	XL>ctrl
Q9V9Z1	mRpL32	N/A	8.95	XL>ctrl
Q9V3D0	mRpL4	N/A	8.94	XL>ctrl
Q9W1V3	Fib	N/A	8.92	XL>ctrl
Q9VQJ0	Rbp9	N/A	8.91	XL>ctrl
Q8IH14	ssx	N/A	8.88	XL>ctrl
Q9W4I3	mRpL30	N/A	8.87	XL>ctrl
Q960Z4	aret	N/A	8.80	XL>ctrl
Q9VHT5	mRpL1	N/A	8.77	XL>ctrl
Q9VFL5	Aats-met	N/A	8.70	XL>ctrl
Q8MRC7	DIP1	N/A	8.67	XL>ctrl
Q9W254	qkr58E-2	N/A	8.67	XL>ctrl
Q9V3E3	mRpL50	N/A	8.66	XL>ctrl
Q7KVQ0	CG4038	N/A	8.64	XL>ctrl
Q9XZ03	mbf1	N/A	8.60	XL>ctrl
Q0KI96	ps	N/A	8.58	XL>ctrl
Q9VKX4	mRpS7	N/A	8.57	XL>ctrl
Q9W334	RpS28b	N/A	8.52	XL>ctrl
A1Z9J6	mRpL53	N/A	8.52	XL>ctrl
P19339-5		N/A	8.52	XL>ctrl
Q9VY48	mRpL38	N/A	8.50	XL>ctrl
Q9VPF6	mRpL15	N/A	8.48	XL>ctrl
Q9VW14	CG14100	N/A	8.46	XL>ctrl
Q9VGW9	mRpL37	N/A	8.46	XL>ctrl
Q9VHN6	mRpL19	N/A	8.44	XL>ctrl
Q9VUX1	mRpS31	N/A	8.43	XL>ctrl
Q9VL89	und	N/A	8.41	XL>ctrl
P10735	tko	N/A	8.40	XL>ctrl
Q27294	caz	N/A	8.40	XL>ctrl
Q7JMZ7	sm	N/A	8.39	XL>ctrl
P50887	RpL22	N/A	8.35	XL>ctrl
Q9W551	CG14817-RA	N/A	8.33	XL>ctrl
Q9I7M2	CG10628	N/A	8.33	XL>ctrl
Q8IP62	mRpS23	N/A	8.33	XL>ctrl
Q7JUX9	trmt10c	N/A	8.31	XL>ctrl
Q8IR99	Imp	N/A	8.31	XL>ctrl
P25158-2	osk	N/A	8.29	XL>ctrl
Q9W362	Larp7	N/A	8.27	XL>ctrl
Q24491	Rsf1	N/A	8.27	XL>ctrl
Q9VAX3		N/A	8.27	XL>ctrl
Q7KY08	AGO1	N/A	8.25	XL>ctrl
A4IJ59	mub	N/A	8.25	XL>ctrl
Q9VIF1	CG9247	N/A	8.24	XL>ctrl
Q95RR6	Pur-alpha	N/A	8.22	XL>ctrl
Q9VHP3	CG8021-RA	N/A	8.22	XL>ctrl
Q24562	U2af50	N/A	8.20	XL>ctrl
E1JH76	fus	N/A	8.18	XL>ctrl
Q8T6B9	pUf68	N/A	8.17	XL>ctrl
Q8MSX1-3		N/A	8.16	XL>ctrl
Q7KNS5	qkr54B	N/A	8.15	XL>ctrl
Q9VSW4	CG4447-RA	N/A	8.14	XL>ctrl
O97428	cib	N/A	8.13	XL>ctrl
Q9V426	vig	N/A	8.13	XL>ctrl
Q9W1V7		N/A	8.10	XL>ctrl
Q9W021	mRpL23	N/A	8.10	XL>ctrl
Q9VLN0		N/A	8.08	XL>ctrl
Q8SXF0	mRpS9	N/A	8.08	XL>ctrl
Q8MLS7	l(2)k09913	N/A	8.07	XL>ctrl
Q7JVK6	trsn	N/A	8.06	XL>ctrl
P41092	RpL27A	N/A	8.06	XL>ctrl

Q7K0A0	mRpL42	N/A	8.03	XL>ctrl
Q9U9Q2	Neos	N/A	8.02	XL>ctrl
Q9VSH4	CG7185	N/A	8.01	XL>ctrl
Q95RB1		N/A	8.01	XL>ctrl
A1Z743	CG1383	N/A	8.01	XL>ctrl
A8JRA4	CG6422-RB	N/A	8.00	XL>ctrl
O77477		N/A	8.00	XL>ctrl
Q86B79	unk	N/A	8.00	XL>ctrl
Q9V431	Aac11	N/A	8.00	XL>ctrl
P49630	RpL36	N/A	7.99	XL>ctrl
Q9W4R7		N/A	7.98	XL>ctrl
Q03334	RpS13	N/A	7.98	XL>ctrl
Q9VU91	bru-3	N/A	7.97	XL>ctrl
Q9VRN5	lin-28	N/A	7.97	XL>ctrl
Q24113	nonA-l	N/A	7.97	XL>ctrl
Q9VUJ0	mRpL39	N/A	7.97	XL>ctrl
Q9VRY5		N/A	7.96	XL>ctrl
Q9W0L7	CG13903	N/A	7.95	XL>ctrl
Q9VUB8	endos	N/A	7.94	XL>ctrl
P43332	snf	N/A	7.94	XL>ctrl
Q9VAQ0	Cul-5	N/A	7.94	XL>ctrl
Q9VBQ7	CG5116-RA	N/A	7.93	XL>ctrl
Q8T3L6	beg	N/A	7.93	XL>ctrl
Q8MSW9	U3-55K	N/A	7.92	XL>ctrl
Q9W413		N/A	7.91	XL>ctrl
Q9W255	qkr58E-1	N/A	7.90	XL>ctrl
Q9VI56		N/A	7.89	XL>ctrl
P16568	BicD	N/A	7.89	XL>ctrl
Q9VPD5	mTerf3	N/A	7.87	XL>ctrl
Q9VPL3	mRpL10	N/A	7.87	XL>ctrl
Q9VNX8	CG7414	N/A	7.84	XL>ctrl
Q9VP13		N/A	7.84	XL>ctrl
P41375	eIF-2beta	N/A	7.84	XL>ctrl
Q7JWR9	CG8635	N/A	7.83	XL>ctrl
Q9I7K0-6	Jupiter	N/A	7.83	XL>ctrl
O44081	Nop60B	N/A	7.82	XL>ctrl
Q9VG07		N/A	7.82	XL>ctrl
Q9VHN4		N/A	7.80	XL>ctrl
O44434	qkr58E-3	N/A	7.80	XL>ctrl
Q9VP20	Mkrn1	N/A	7.79	XL>ctrl
Q86BM8	mRpL27	N/A	7.79	XL>ctrl
Q9W547	mRpL16	N/A	7.78	XL>ctrl
Q9W5W8		N/A	7.78	XL>ctrl
Q8MMC4-2		N/A	7.77	XL>ctrl
Q8INE1	Trax	N/A	7.77	XL>ctrl
Q01617-1	cpo	N/A	7.76	XL>ctrl
Q8SWR8-2	Atx2	N/A	7.75	XL>ctrl
Q9VB04	btz	N/A	7.75	XL>ctrl
Q9W1K4	egl	N/A	7.75	XL>ctrl
P07186	Cp19	N/A	7.74	XL>ctrl
O46036-2	CtBP	N/A	7.73	XL>ctrl
Q9W149	CG11414-RA	N/A	7.73	XL>ctrl
Q9VXR5		N/A	7.72	XL>ctrl
Q86BR2	CG11097	N/A	7.72	XL>ctrl
Q9W2M8	mRpL54	N/A	7.71	XL>ctrl
O18373	SelD	N/A	7.71	XL>ctrl
Q9VPT8	Clp	N/A	7.70	XL>ctrl
Q9VEN9	Patr-1	N/A	7.69	XL>ctrl
Q29QN4		N/A	7.67	XL>ctrl
Q9VXA8	CG9125	N/A	7.67	XL>ctrl
Q9VEN6	mRpS11	N/A	7.67	XL>ctrl
Q9VU36	mRpL20	N/A	7.67	XL>ctrl
Q9VBR0		N/A	7.66	XL>ctrl
O61613	Nmt	N/A	7.66	XL>ctrl
Q9VX63	CG8915-RA	N/A	7.66	XL>ctrl
Q7KTG0		N/A	7.66	XL>ctrl
Q9W0D6		N/A	7.66	XL>ctrl

Q9V9W2	RpL6	N/A	7.65	XL>ctrl
Q9VRL0		N/A	7.65	XL>ctrl
A1ZBF1	CG18190-RB	N/A	7.63	XL>ctrl
Q8WTC1	bonsai	N/A	7.63	XL>ctrl
Q9W424	Spx	N/A	7.63	XL>ctrl
Q9VAY7	CG12259	N/A	7.62	XL>ctrl
Q9VNC1	mRpL44	N/A	7.62	XL>ctrl
Q7K1U0	Arc1	N/A	7.62	XL>ctrl
O76324	dco	N/A	7.61	XL>ctrl
Q9VE52	CstF-64	N/A	7.60	XL>ctrl
Q9I7Q9	mRpL36	N/A	7.58	XL>ctrl
Q9VX15		N/A	7.58	XL>ctrl
Q9VDQ4	trem	N/A	7.58	XL>ctrl
Q9XZ57	Eb1	N/A	7.57	XL>ctrl
Q8IRD2	YT521-B	N/A	7.57	XL>ctrl
A8JPX0		N/A	7.57	XL>ctrl
Q5LK13	Dbp80	N/A	7.57	XL>ctrl
A8JNQ5	A2bp1	N/A	7.55	XL>ctrl
Q9W1H5	Dcp1	N/A	7.54	XL>ctrl
Q8SWT2		N/A	7.53	XL>ctrl
Q8IPB9	Cnot4	N/A	7.52	XL>ctrl
E1JIL5	mRpS10	N/A	7.50	XL>ctrl
Q9VUQ5	AGO2	N/A	7.50	XL>ctrl
Q24537-3	Dsp1	N/A	7.49	XL>ctrl
P17133	snRNP-U1-70K	N/A	7.49	XL>ctrl
Q9VJY9	loqs	N/A	7.49	XL>ctrl
Q9VCC3	mRpS24	N/A	7.48	XL>ctrl
Q7JW66		N/A	7.48	XL>ctrl
Q9VMX0	mRpL28	N/A	7.48	XL>ctrl
Q8I0I5	Pop2	N/A	7.47	XL>ctrl
Q9I7P8		N/A	7.47	XL>ctrl
A8JRH9	CG11337-RD	N/A	7.47	XL>ctrl
Q8MLS0		N/A	7.46	XL>ctrl
Q9VCB6	twin	N/A	7.46	XL>ctrl
Q9VY28	mRpS25	N/A	7.46	XL>ctrl
Q4V619		N/A	7.45	XL>ctrl
Q9VBK9		N/A	7.45	XL>ctrl
Q9VH95	CG16817	N/A	7.45	XL>ctrl
Q8SZ15		N/A	7.43	XL>ctrl
Q9VSF3	CG7375	N/A	7.43	XL>ctrl
Q9VH64		N/A	7.43	XL>ctrl
Q9VNV3	Ddx1	N/A	7.42	XL>ctrl
Q9VLS0		N/A	7.42	XL>ctrl
Q9VL60		N/A	7.42	XL>ctrl
Q7KVU4	CBP	N/A	7.41	XL>ctrl
Q9VAX9	CG4884-RA	N/A	7.40	XL>ctrl
P32234	128up	N/A	7.39	XL>ctrl
Q9VYV4	Amun	N/A	7.39	XL>ctrl
Q9VIW7	CG9987	N/A	7.38	XL>ctrl
Q9VAN6	Slbp	N/A	7.38	XL>ctrl
Q9VVI1		N/A	7.37	XL>ctrl
A1ZBA5	mRpS28	N/A	7.36	XL>ctrl
A1Z830		N/A	7.36	XL>ctrl
P82804	wibg	N/A	7.36	XL>ctrl
Q9W114		N/A	7.36	XL>ctrl
Q24524	sn	N/A	7.35	XL>ctrl
Q9VG62	trus	N/A	7.35	XL>ctrl
P35600	Gnf1	N/A	7.34	XL>ctrl
Q9VQM4	gkt	N/A	7.34	XL>ctrl
Q9VBN5	RpL27	N/A	7.33	XL>ctrl
Q8SXG7		N/A	7.33	XL>ctrl
Q9VVX4	mRpL21	N/A	7.33	XL>ctrl
A1Z7K9		N/A	7.33	XL>ctrl
Q9VVA0	CG9705	N/A	7.33	XL>ctrl
Q9VND7		N/A	7.32	XL>ctrl
Q9VT19		N/A	7.32	XL>ctrl
Q8MS27	mRpL35	N/A	7.31	XL>ctrl

Q9VDE3	slmb	N/A	7.31	XL>ctrl
Q7K4H4	CG2982	N/A	7.31	XL>ctrl
P35875	Parp	N/A	7.31	XL>ctrl
Q9VAJ2	Bub3	N/A	7.30	XL>ctrl
Q9VMH3	stai	N/A	7.30	XL>ctrl
Q9VVL8	CG7441-RA	N/A	7.30	XL>ctrl
Q9VTM5	mtTFB1	N/A	7.30	XL>ctrl
Q8IPZ5		N/A	7.28	XL>ctrl
Q9VFC8-2	CG6904	N/A	7.27	XL>ctrl
Q95RI7		N/A	7.26	XL>ctrl
Q9VGF9	aur	N/A	7.26	XL>ctrl
O01382	Ice	N/A	7.25	XL>ctrl
Q9VUM0	Msh6	N/A	7.25	XL>ctrl
Q9VF02	Hel89B	N/A	7.25	XL>ctrl
O18353	Chi	N/A	7.24	XL>ctrl
Q9VWQ6	CG18259-RA	N/A	7.24	XL>ctrl
Q9VRH6	waw	N/A	7.24	XL>ctrl
Q94522	Scsalpha	N/A	7.23	XL>ctrl
P84029	Cyt-c-p	N/A	7.22	XL>ctrl
Q9VFJ2	mRpL11	N/A	7.22	XL>ctrl
Q9W4D2	Rnp4F	N/A	7.22	XL>ctrl
Q9VDK6		N/A	7.21	XL>ctrl
Q9V9E3	CG3107	N/A	7.21	XL>ctrl
A8JNW5		N/A	7.21	XL>ctrl
Q9VNH1	CG1239	N/A	7.21	XL>ctrl
Q9VCU5	CG17141	N/A	7.20	XL>ctrl
Q9VKM7	CG6089	N/A	7.20	XL>ctrl
P00408	mt:CoII	N/A	7.19	XL>ctrl
Q24439	Oscp	N/A	7.19	XL>ctrl
Q9VZY0		N/A	7.18	XL>ctrl
Q9VSD7	ldbr	N/A	7.18	XL>ctrl
Q9W123	Pof	N/A	7.18	XL>ctrl
Q9VHJ8	skap	N/A	7.18	XL>ctrl
Q9VEB9	Ssdp	N/A	7.17	XL>ctrl
Q8SXK0		N/A	7.17	XL>ctrl
Q9VNR9		N/A	7.14	XL>ctrl
Q7K2G1	CG13349	N/A	7.14	XL>ctrl
Q9VJQ4	l(2)35Bd	N/A	7.14	XL>ctrl
A1Z9L3	pea	N/A	7.14	XL>ctrl
P07764	Ald	N/A	7.14	XL>ctrl
Q8MSB1	CG13638	N/A	7.14	XL>ctrl
Q9VC90		N/A	7.13	XL>ctrl
Q9VE86	WRNexo	N/A	7.12	XL>ctrl
Q8I0J3	mRpS21	N/A	7.12	XL>ctrl
Q9VZP5	eIF5B	N/A	7.12	XL>ctrl
Q9VKN7	ial	N/A	7.09	XL>ctrl
Q9VT38		N/A	7.09	XL>ctrl
P54397	FK506-bp1	N/A	7.09	XL>ctrl
Q0KHZ6		N/A	7.09	XL>ctrl
Q9VIN9	mRpS18B	N/A	7.09	XL>ctrl
P05205	Su(var)205	N/A	7.08	XL>ctrl
Q9V9K7	Ars2	N/A	7.07	XL>ctrl
Q9VK20		N/A	7.06	XL>ctrl
Q9VGP7	mRpL40	N/A	7.06	XL>ctrl
Q9W0Y1	Tina-1	N/A	7.06	XL>ctrl
Q9VXK6	eIF5	N/A	7.06	XL>ctrl
O77277	torp4a	N/A	7.06	XL>ctrl
Q9VWI2	Nat1	N/A	7.05	XL>ctrl
Q7KUB0	Idh	N/A	7.05	XL>ctrl
Q9W0F9	312	N/A	7.04	XL>ctrl
Q7K550	Adam	N/A	7.04	XL>ctrl
Q9V3G3	cyp33	N/A	7.04	XL>ctrl
Q8IPF5	Akap200	N/A	7.03	XL>ctrl
Q9VG06	CG17327-RA	N/A	7.03	XL>ctrl
Q27889	Pp2B-14D;CanA-14F	N/A	7.02	XL>ctrl
Q9W0C3		N/A	7.02	XL>ctrl
Q9VPK7	Pus10	N/A	7.02	XL>ctrl

Q9VZS5	RpL28	N/A	7.02	XL>ctrl
Q9VM33	ico	N/A	7.02	XL>ctrl
Q9V3L6	Cbp20	N/A	7.02	XL>ctrl
P55035	Pros54	N/A	7.02	XL>ctrl
Q9U1H9	sbr	N/A	7.01	XL>ctrl
Q7K7A9	Fen1	N/A	7.01	XL>ctrl
Q9VUU7	CG15715-RA	N/A	7.01	XL>ctrl
Q9VLR5	Ssb-c31a	N/A	7.01	XL>ctrl
Q9VLC5	Aldh	N/A	7.00	XL>ctrl
Q9VPW4	mtRNApol	N/A	7.00	XL>ctrl
A0AQH0	Prosbeta1	N/A	6.99	XL>ctrl
Q9VYW4		N/A	6.99	XL>ctrl
O02195	Trip1	N/A	6.99	XL>ctrl
A1Z6E2	gus	N/A	6.99	XL>ctrl
Q9VSR3	orb2	N/A	6.98	XL>ctrl
Q9V5C6	Prosalpha7	N/A	6.97	XL>ctrl
O16102	Chd3	N/A	6.97	XL>ctrl
Q9VAC9		N/A	6.97	XL>ctrl
Q9VAC1		N/A	6.96	XL>ctrl
Q8MSC4		N/A	6.95	XL>ctrl
Q8IMZ2	orb	N/A	6.95	XL>ctrl
Q9VV39	mRpS34	N/A	6.95	XL>ctrl
A1Z9I6		N/A	6.95	XL>ctrl
Q9XZ06		N/A	6.95	XL>ctrl
Q7K485	cathD	N/A	6.95	XL>ctrl
Q960W0		N/A	6.95	XL>ctrl
Q9VRP3	Txl	N/A	6.95	XL>ctrl
Q24208	Su(var)3-9	N/A	6.95	XL>ctrl
Q9VNS0	mael	N/A	6.95	XL>ctrl
Q9VZD8	CG15014	N/A	6.95	XL>ctrl
Q24276	Cdc37	N/A	6.94	XL>ctrl
P12982	Pp1-87B;Pp1-13C	N/A	6.94	XL>ctrl
Q8IMX8	LSm3	N/A	6.94	XL>ctrl
Q9VW13	CG9330-RA	N/A	6.94	XL>ctrl
Q8SWU7	CG1354	N/A	6.93	XL>ctrl
P48610-3	Argk	N/A	6.93	XL>ctrl
Q9VF26	spn-E	N/A	6.92	XL>ctrl
Q7K1Q7	RpLP0-like	N/A	6.92	XL>ctrl
Q9VVN2	mRpS26	N/A	6.92	XL>ctrl
Q8SX83-4	spen	N/A	6.91	XL>ctrl
Q9VDZ4		N/A	6.91	XL>ctrl
Q9VP77		N/A	6.91	XL>ctrl
Q8IML6		N/A	6.90	XL>ctrl
Q9VNF8	sec23	N/A	6.90	XL>ctrl
Q8SXM8	Aats-lys	N/A	6.90	XL>ctrl
O97067	CG1307	N/A	6.90	XL>ctrl
O62621	betaCop	N/A	6.89	XL>ctrl
A1Z768	rnh1	N/A	6.89	XL>ctrl
P40301	Pros25	N/A	6.89	XL>ctrl
P40945	Arf102F	N/A	6.89	XL>ctrl
Q9VRA1		N/A	6.88	XL>ctrl
P23226-3	Map205	N/A	6.88	XL>ctrl
Q9W3L1	Upf2	N/A	6.87	XL>ctrl
Q9VEJ1	SF1	N/A	6.87	XL>ctrl
Q9VRV8	Trn	N/A	6.87	XL>ctrl
P29052	TfIIB	N/A	6.86	XL>ctrl
P08181	CkIIalpha	N/A	6.86	XL>ctrl
Q9VHS0	CG9601-RA	N/A	6.85	XL>ctrl
Q9V6K1-2	TppII	N/A	6.85	XL>ctrl
Q9W029	yellow-g	N/A	6.85	XL>ctrl
Q9VDV2		N/A	6.85	XL>ctrl
P22769	Pros28.1	N/A	6.85	XL>ctrl
P49762-3	Doa	N/A	6.85	XL>ctrl
P25167	RpIII128	N/A	6.84	XL>ctrl
A1ZAK9	Psi	N/A	6.83	XL>ctrl
Q9VYB0	BthD	N/A	6.82	XL>ctrl
Q7JUY7	ced-6	N/A	6.82	XL>ctrl

Q9VUJ1	Prosbeta2	N/A	6.81	XL>ctrl
Q9VT05		N/A	6.81	XL>ctrl
A1Z9X0	aPKC	N/A	6.81	XL>ctrl
Q9VAV2		N/A	6.81	XL>ctrl
P17210	Khc	N/A	6.81	XL>ctrl
Q9VJJ8	CG4668;CG42389-RF	N/A	6.80	XL>ctrl
Q9U3Z7	hoip	N/A	6.80	XL>ctrl
Q9W2D9	CG10306	N/A	6.80	XL>ctrl
P41043	GstS1	N/A	6.80	XL>ctrl
Q9VDK7	Srp72	N/A	6.79	XL>ctrl
Q7KLW9	Prp19	N/A	6.79	XL>ctrl
Q8IHB0	gish	N/A	6.79	XL>ctrl
Q9VD52		N/A	6.78	XL>ctrl
Q9VKZ8		N/A	6.77	XL>ctrl
Q9VJZ1		N/A	6.77	XL>ctrl
Q9VGZ3	Irp-1B	N/A	6.77	XL>ctrl
Q9I7M3	CG10237-RC	N/A	6.76	XL>ctrl
Q9V6Y3	mRpS16	N/A	6.76	XL>ctrl
Q9VNI3	CG1218	N/A	6.76	XL>ctrl
Q9VIW3	RanGap	N/A	6.76	XL>ctrl
Q9W590	ns3	N/A	6.76	XL>ctrl
P27864	Rrp1	N/A	6.76	XL>ctrl
Q9VGR7	CG6693	N/A	6.75	XL>ctrl
Q9W5D3	EG:34F3.4	N/A	6.75	XL>ctrl
Q8MLU3	Snp	N/A	6.75	XL>ctrl
P18929	mt:ND1	N/A	6.74	XL>ctrl
Q7KVX5	Vap-33-1	N/A	6.74	XL>ctrl
Q9NG60	ligatin	N/A	6.74	XL>ctrl
Q9VU68-2	flr	N/A	6.73	XL>ctrl
Q9VGW7	Art1	N/A	6.73	XL>ctrl
Q9VIV2	swm	N/A	6.72	XL>ctrl
Q8STG9	Sec61alpha	N/A	6.72	XL>ctrl
Q9V9Q4		N/A	6.72	XL>ctrl
P36179	Pp2A-29B	N/A	6.72	XL>ctrl
Q9V4D9	Pur-alpha	N/A	6.71	XL>ctrl
Q7KN97		N/A	6.71	XL>ctrl
Q0E931	Dgp-1	N/A	6.70	XL>ctrl
Q9VH48	Art4	N/A	6.70	XL>ctrl
Q9VX25	CG8188	N/A	6.69	XL>ctrl
Q9VI75	lap	N/A	6.69	XL>ctrl
Q9VKK1	Ge-1	N/A	6.69	XL>ctrl
P20240	Ote	N/A	6.69	XL>ctrl
Q9V3J4	sec13	N/A	6.68	XL>ctrl
Q9V7N5	Vha44	N/A	6.67	XL>ctrl
E1JGM9	par-1	N/A	6.67	XL>ctrl
Q9VJ62	CG10341	N/A	6.67	XL>ctrl
Q9V3J8	wds	N/A	6.67	XL>ctrl
Q02870	hay	N/A	6.67	XL>ctrl
P91926	alpha-Adaptin	N/A	6.66	XL>ctrl
A4GUJ7	AGO3	N/A	6.66	XL>ctrl
Q9W3R9	CG1677-RA	N/A	6.66	XL>ctrl
Q7K3W2	CG8728-RA	N/A	6.66	XL>ctrl
Q9VL78	FKBP59	N/A	6.65	XL>ctrl
Q7JVP2	Rcd-1	N/A	6.65	XL>ctrl
Q9VFV8		N/A	6.64	XL>ctrl
Q9W445		N/A	6.63	XL>ctrl
Q9V3J1-3	VhaSFD	N/A	6.62	XL>ctrl
Q9XYU0	Mcm7	N/A	6.62	XL>ctrl
Q9VGG9		N/A	6.61	XL>ctrl
Q7JZM8	mRpL41	N/A	6.61	XL>ctrl
Q9VN50	eIF3-S5-1	N/A	6.61	XL>ctrl
Q7JVY0	coro	N/A	6.61	XL>ctrl
Q9W3Y3		N/A	6.61	XL>ctrl
A1Z813	sec24	N/A	6.61	XL>ctrl
O02649	Hsp60	N/A	6.60	XL>ctrl
Q9VR56	trmt10a	N/A	6.60	XL>ctrl
Q9W1N4	wmd	N/A	6.60	XL>ctrl

Q9VYQ8	Usp7	N/A	6.59	XL>ctrl
Q24595	mus210	N/A	6.59	XL>ctrl
Q9VIK6	Cen	N/A	6.59	XL>ctrl
Q9VPM7	CG17078-RA	N/A	6.59	XL>ctrl
Q9VL25		N/A	6.58	XL>ctrl
Q9I7M8	Ku80	N/A	6.58	XL>ctrl
P00334	Adh	N/A	6.58	XL>ctrl
P23572	cdc2	N/A	6.57	XL>ctrl
Q9I7Z8		N/A	6.56	XL>ctrl
Q9V6U9	CG16935	N/A	6.56	XL>ctrl
Q9VIH1	RPA2	N/A	6.56	XL>ctrl
Q9VHI8		N/A	6.56	XL>ctrl
Q7KQM6-2	CG11148	N/A	6.56	XL>ctrl
Q9VD07		N/A	6.56	XL>ctrl
Q86BS6-3	metl	N/A	6.55	XL>ctrl
P07909-2		N/A	6.55	XL>ctrl
O76928	cm	N/A	6.55	XL>ctrl
Q9VPP5	CG13690	N/A	6.54	XL>ctrl
P08266	RpII140	N/A	6.54	XL>ctrl
Q7PLI7-2	CG17528	N/A	6.54	XL>ctrl
Q9VV82		N/A	6.54	XL>ctrl
Q9VL18-2	eEF1delta	N/A	6.54	XL>ctrl
Q9V463	Nup154	N/A	6.54	XL>ctrl
Q9VKU3		N/A	6.54	XL>ctrl
Q9VYE9	CG1622-RA	N/A	6.53	XL>ctrl
Q9VJ39		N/A	6.53	XL>ctrl
Q9VB10		N/A	6.53	XL>ctrl
Q9W3N9		N/A	6.53	XL>ctrl
Q9VPU0		N/A	6.52	XL>ctrl
Q9VBX1	Clbn	N/A	6.52	XL>ctrl
Q7K1H0	CG6491	N/A	6.51	XL>ctrl
Q9I7R0		N/A	6.51	XL>ctrl
Q8MRG6	Prp3	N/A	6.51	XL>ctrl
Q9VMX6	Marcal1	N/A	6.51	XL>ctrl
Q9V3F3	mTTF	N/A	6.51	XL>ctrl
P91622	Pdk	N/A	6.51	XL>ctrl
Q9VAQ5	CG11837	N/A	6.51	XL>ctrl
Q9VCH8	tst	N/A	6.50	XL>ctrl
Q8IMV6	Saf-B	N/A	6.49	XL>ctrl
Q9VN03	CG9791	N/A	6.48	XL>ctrl
Q9VNB9	RpL35A	N/A	6.48	XL>ctrl
Q9VKY2	Cand1	N/A	6.48	XL>ctrl
Q5LJP9	CG17486	N/A	6.48	XL>ctrl
Q9XYM5	Su(var)2-10	N/A	6.48	XL>ctrl
Q9VPX7	capt	N/A	6.48	XL>ctrl
Q9W2E7	Rae1	N/A	6.47	XL>ctrl
Q9VUN9	CG7857-RA	N/A	6.47	XL>ctrl
Q7K4N3	Cbp80	N/A	6.47	XL>ctrl
Q961R8	Aats-gly	N/A	6.47	XL>ctrl
Q9VJC7	CG6412	N/A	6.47	XL>ctrl
Q7KQZ4-3	lola	N/A	6.46	XL>ctrl
E1JIC9	TfIIEbeta	N/A	6.46	XL>ctrl
Q9XYP8	l(1)dd4	N/A	6.45	XL>ctrl
Q9W2N5		N/A	6.45	XL>ctrl
Q9VWN9		N/A	6.44	XL>ctrl
Q8INB9-2	Akt1	N/A	6.43	XL>ctrl
Q9V877	sub	N/A	6.43	XL>ctrl
O16810	Orc1	N/A	6.42	XL>ctrl
E1JIL4	Caf1	N/A	6.41	XL>ctrl
O76912	EG:95B7.9	N/A	6.41	XL>ctrl
P52034-3	Pfk	N/A	6.41	XL>ctrl
Q9VMI3		N/A	6.40	XL>ctrl
Q9VVL7		N/A	6.40	XL>ctrl
Q9VEK8	sds22	N/A	6.40	XL>ctrl
Q9U5L1	Gtp-bp	N/A	6.40	XL>ctrl
O16043	Df31	N/A	6.40	XL>ctrl
Q9VLM5	CG13393	N/A	6.40	XL>ctrl

Q0E8J0	MEP-1	N/A	6.39	XL>ctrl
Q9VL68		N/A	6.39	XL>ctrl
Q7K0D3		N/A	6.39	XL>ctrl
Q9VD58		N/A	6.38	XL>ctrl
Q9VRP5-6	scny	N/A	6.38	XL>ctrl
Q9VD13	lqfR	N/A	6.37	XL>ctrl
Q8IRG6	dre4	N/A	6.36	XL>ctrl
Q9XYX1	HDAC6	N/A	6.36	XL>ctrl
Q9VL29		N/A	6.36	XL>ctrl
Q9VV60	Aats-tyr	N/A	6.36	XL>ctrl
P25439-2	brm	N/A	6.36	XL>ctrl
Q6NR30	BEST:LD13441	N/A	6.35	XL>ctrl
A1Z7K8		N/A	6.35	XL>ctrl
O46043	Parg	N/A	6.34	XL>ctrl
Q94547-2	Rga	N/A	6.34	XL>ctrl
Q9VKD3	CG12264	N/A	6.34	XL>ctrl
Q7K126	l(2)NC136	N/A	6.34	XL>ctrl
Q7KRY6	ball	N/A	6.33	XL>ctrl
Q9Y134	l(2)35Df	N/A	6.33	XL>ctrl
Q9TVM2	emb	N/A	6.33	XL>ctrl
A1ZB00	aft	N/A	6.32	XL>ctrl
Q95SY7	sar1	N/A	6.32	XL>ctrl
Q9XYP7-2	Grip84	N/A	6.31	XL>ctrl
Q9VEB4	mTerf5	N/A	6.31	XL>ctrl
Q9V3W0	UK114	N/A	6.30	XL>ctrl
Q9VBL3		N/A	6.30	XL>ctrl
A5XCL5	UGP	N/A	6.30	XL>ctrl
O46106	noi	N/A	6.30	XL>ctrl
Q9VVC5	Nc73EF	N/A	6.30	XL>ctrl
Q9VMJ7	lid	N/A	6.30	XL>ctrl
Q9VKU8	RluA-2	N/A	6.29	XL>ctrl
Q9VB05	ALiX	N/A	6.29	XL>ctrl
P26019	DNApol-alpha180	N/A	6.29	XL>ctrl
Q9W3X8		N/A	6.29	XL>ctrl
Q9VUQ8	CG7427	N/A	6.29	XL>ctrl
Q9VTY6	vih	N/A	6.29	XL>ctrl
Q7KUN5	Dcp2	N/A	6.29	XL>ctrl
Q7K5K3		N/A	6.28	XL>ctrl
Q8MT58		N/A	6.27	XL>ctrl
Q9VJD4	Sgt	N/A	6.27	XL>ctrl
A1Z8J0	CG7741	N/A	6.27	XL>ctrl
Q9VTE6		N/A	6.26	XL>ctrl
Q9VUL1	CTPsyn	N/A	6.26	XL>ctrl
Q9VC18		N/A	6.25	XL>ctrl
Q5BIC3	az2	N/A	6.25	XL>ctrl
Q7K3B9	CG7544	N/A	6.25	XL>ctrl
Q9VZQ8	PHGPx	N/A	6.24	XL>ctrl
Q24044	fzy	N/A	6.23	XL>ctrl
Q8MKK1	CG30185;Gr59f	N/A	6.22	XL>ctrl
Q94523	SdhA	N/A	6.22	XL>ctrl
Q70PP2	Smg1	N/A	6.21	XL>ctrl
Q8IMG2	CstF-50	N/A	6.20	XL>ctrl
Q94517	Rpd3	N/A	6.20	XL>ctrl
Q9W388	AP-1gamma	N/A	6.19	XL>ctrl
O61345	pen	N/A	6.19	XL>ctrl
P20232	TfIIS	N/A	6.19	XL>ctrl
Q9VMH1	WDR79	N/A	6.19	XL>ctrl
Q9VRD4		N/A	6.18	XL>ctrl
Q9VUY4		N/A	6.18	XL>ctrl
P54385-2	Gdh	N/A	6.17	XL>ctrl
Q9VIF6	CG9253-RA	N/A	6.17	XL>ctrl
Q9VUM1	Prp31	N/A	6.17	XL>ctrl
Q8IQ05	cutlet	N/A	6.17	XL>ctrl
Q8SYQ8		N/A	6.16	XL>ctrl
P26802	Dbp73D	N/A	6.16	XL>ctrl
Q03043-4	for	N/A	6.16	XL>ctrl
Q7JX95		N/A	6.15	XL>ctrl

P20028	Rpl135	N/A	6.15	XL>ctrl
Q9V434	Aats-asn	N/A	6.15	XL>ctrl
O97477	Inos	N/A	6.15	XL>ctrl
Q7KUA4	Uba2	N/A	6.15	XL>ctrl
Q9VY54	l(1)G0007	N/A	6.14	XL>ctrl
Q9V3A7	glu	N/A	6.14	XL>ctrl
Q8IH00	Mat89Ba	N/A	6.13	XL>ctrl
Q9XZU1	Cas	N/A	6.13	XL>ctrl
Q9V4F3	CG11360-RA	N/A	6.12	XL>ctrl
Q9VWB8	Nup205	N/A	6.12	XL>ctrl
P48423	Gap1	N/A	6.12	XL>ctrl
Q53ZT0	DnaJ-1	N/A	6.12	XL>ctrl
Q9VYW5		N/A	6.11	XL>ctrl
Q9W107	CG16912	N/A	6.11	XL>ctrl
E1JIX6	Npl4;CG4673	N/A	6.10	XL>ctrl
Q9XZ32		N/A	6.10	XL>ctrl
Q9V475	Cul-3	N/A	6.09	XL>ctrl
Q9VJZ5	Tap42	N/A	6.09	XL>ctrl
Q9VPU4	CG4291-RA	N/A	6.08	XL>ctrl
Q9VQK7		N/A	6.08	XL>ctrl
Q9VEG8	CG7477	N/A	6.08	XL>ctrl
Q9VCV4	Irp-1A	N/A	6.08	XL>ctrl
Q960M4	Prx5	N/A	6.07	XL>ctrl
Q8SY33	gw	N/A	6.06	XL>ctrl
Q9VZI1	Chd64	N/A	6.06	XL>ctrl
Q9VTC1		N/A	6.06	XL>ctrl
Q9VQE0	Drp1	N/A	6.06	XL>ctrl
Q9VZF5	pav	N/A	6.05	XL>ctrl
Q9W5E6	Dredd	N/A	6.04	XL>ctrl
Q9W2Y5	ZAP3	N/A	6.04	XL>ctrl
Q9W5W4	Sdic1;Sdic2;Sdic3;sw	N/A	6.03	XL>ctrl
P14318	Mp20	N/A	6.03	XL>ctrl
Q9W420	Spt6	N/A	6.03	XL>ctrl
Q8SXJ8		N/A	6.03	XL>ctrl
Q8T9L6	CG13900-RA	N/A	6.03	XL>ctrl
Q9NHD5	san	N/A	6.03	XL>ctrl
Q9W246		N/A	6.02	XL>ctrl
Q9VY44	mRNA-cap	N/A	6.01	XL>ctrl
Q8T0G4	CG40286-RA	N/A	6.01	XL>ctrl
Q9VGQ1		N/A	6.01	XL>ctrl
Q9VCW3	Nup133	N/A	6.01	XL>ctrl
Q9VZ82	Aats-leu	N/A	6.00	XL>ctrl
A1Z8P9	Mtor	N/A	5.99	XL>ctrl
Q9VH81	PpD3	N/A	5.99	XL>ctrl
B7Z0X1		N/A	5.98	XL>ctrl
Q9Y0I1	MRG15	N/A	5.97	XL>ctrl
Q9VQR8		N/A	5.96	XL>ctrl
Q9VIG1	CG12050-RB	N/A	5.96	XL>ctrl
Q9VX44	l(1)G0222	N/A	5.95	XL>ctrl
Q9VDP9		N/A	5.94	XL>ctrl
Q9VYQ9	Tango4	N/A	5.94	XL>ctrl
A1Z8X3	ana3	N/A	5.94	XL>ctrl
Q9VWE7	Ranbp21	N/A	5.94	XL>ctrl
P42570	plu	N/A	5.93	XL>ctrl
Q5U0X8		N/A	5.93	XL>ctrl
P25171	Bj1	N/A	5.90	XL>ctrl
Q9VZR6		N/A	5.89	XL>ctrl
P27619-2	shi	N/A	5.87	XL>ctrl
A1Z7S0	Rme-8	N/A	5.87	XL>ctrl
Q9VV48	roq	N/A	5.87	XL>ctrl
Q59E29	fax	N/A	5.86	XL>ctrl
P91875	Rpl1	N/A	5.86	XL>ctrl
Q9V461	Mcm6	N/A	5.86	XL>ctrl
A1A708-2	CG4951	N/A	5.86	XL>ctrl
Q9VP12	Smc5	N/A	5.86	XL>ctrl
Q7K3V6		N/A	5.84	XL>ctrl
Q24169	Orc5	N/A	5.84	XL>ctrl

Q8INM3	sle;CG12592	N/A	5.84	XL>ctrl
Q9VPY2	Plap	N/A	5.82	XL>ctrl
Q9VPC0	Pitslre	N/A	5.82	XL>ctrl
P13607-3	Atpalpha	N/A	5.81	XL>ctrl
Q9VHC8		N/A	5.81	XL>ctrl
A1Z7P0	Pmm45A	N/A	5.80	XL>ctrl
Q9W074	HBS1	N/A	5.78	XL>ctrl
Q9VYA4		N/A	5.76	XL>ctrl
Q7K8Y3	Spn42Da	N/A	5.76	XL>ctrl
Q7K2L1	lat	N/A	5.75	XL>ctrl
Q9W401	kdn	N/A	5.75	XL>ctrl
Q9VMC8	CG9548	N/A	5.75	XL>ctrl
P29845	Hsc70-5	N/A	5.73	XL>ctrl
Q9VIZ3	l(2)37Cb	N/A	5.73	XL>ctrl
Q9VXW9		N/A	5.72	XL>ctrl
Q7K3Z3	p47	N/A	5.72	XL>ctrl
Q9W0E4	Psa	N/A	5.70	XL>ctrl
Q9VM62	SA	N/A	5.70	XL>ctrl
Q9W1B0	gek	N/A	5.70	XL>ctrl
Q32KD2	egg	N/A	5.68	XL>ctrl
Q9VSL8	rhea	N/A	5.68	XL>ctrl
Q9W378	mei-P26	N/A	5.68	XL>ctrl
Q9VSD6	msk	N/A	5.66	XL>ctrl
Q24574	Ubp64E	N/A	5.66	XL>ctrl
Q8MSU4	Sym	N/A	5.64	XL>ctrl
Q9VVA6	nudC	N/A	5.64	XL>ctrl
Q7K4I5	CG8892-RB	N/A	5.63	XL>ctrl
Q9VKH9	cana	N/A	5.63	XL>ctrl
Q9W484	CG4078	N/A	5.63	XL>ctrl
P45437	betaCop	N/A	5.62	XL>ctrl
Q9W3C1	l(1)G0020	N/A	5.62	XL>ctrl
Q9VAY2	Gp93	N/A	5.61	XL>ctrl
Q09332	Ugt	N/A	5.58	XL>ctrl
A1Z6L9	Trap1	N/A	5.57	XL>ctrl
A1Z992	AGBE	N/A	5.56	XL>ctrl
Q9VMD0	DLP	N/A	5.56	XL>ctrl
Q9VIF0	CG9246	N/A	5.55	XL>ctrl
Q9W0P8-2	MED14	N/A	5.55	XL>ctrl
Q9VYV3	prtp	N/A	5.55	XL>ctrl
Q9W252	rad50	N/A	5.51	XL>ctrl
Q9VFE7		N/A	5.51	XL>ctrl
Q8MYL1	Fancd2	N/A	5.51	XL>ctrl
A1Z9P1	tej	N/A	5.48	XL>ctrl
Q9VWC7	CG12702-RA	N/A	5.46	XL>ctrl
Q9XZ11		N/A	5.44	XL>ctrl
O46037	Vinc	N/A	5.42	XL>ctrl
A1Z7J7	sec31	N/A	5.41	XL>ctrl
Q9W2Z4	CG2990-RB	N/A	5.37	XL>ctrl
Q9W1R5	vir	N/A	5.34	XL>ctrl
Q9VPL9	kis	N/A	5.34	XL>ctrl
Q6NNB2		N/A	5.34	XL>ctrl
Q4V4R8	CG32101;CG10638-RB	N/A	5.30	XL>ctrl
P49735	Mcm2	N/A	5.27	XL>ctrl
Q9VQ78		N/A	5.26	XL>ctrl
Q9VZQ3	kst	N/A	5.24	XL>ctrl
Q9W5E4	l(1)1Bi	N/A	5.08	XL>ctrl
Q9VXP3	mRpS30	4.72	9.27	3.6E+04
Q9VCX3	mRpL45	5.32	9.75	2.7E+04
Q9VXQ0	mRpL3	4.90	9.30	2.6E+04
Q9VPT3	KH1	4.93	9.22	1.9E+04
A1Z9A8	CG4679	4.96	8.83	7.4E+03
Q9W592	EG:BACN32G11.3	4.66	8.40	5.5E+03
Q9VRJ7		5.10	8.80	5.1E+03
P41094	RpS18	5.91	9.56	4.5E+03
Q9W1I0	TBPH	5.86	9.32	2.9E+03
Q59DT6	Rb97D	5.63	9.09	2.9E+03
Q8WR53	heph	5.77	9.20	2.7E+03

P41073-3	Pep	5.93	9.34	2.6E+03
P16914-2	elav	6.06	9.46	2.5E+03
P41093	RpL18A	5.06	8.45	2.5E+03
Q960D3	vig2;fdy	5.95	9.30	2.3E+03
B7Z0E2	Unr	5.93	9.29	2.3E+03
Q9VSK9		5.25	8.56	2.0E+03
P13469	mod	6.19	9.50	2.0E+03
Q9VWG3	RpS10b	6.14	9.43	2.0E+03
Q24492	RpA-70	5.27	8.56	1.9E+03
Q9W229	RpS24	6.47	9.74	1.8E+03
Q9VM71	Rat1	4.77	8.03	1.8E+03
Q9W5Y4	CG17018-RC;CG17018-RI	4.49	7.71	1.7E+03
Q9VM69	nop5	5.16	8.37	1.6E+03
O77410	eIF3-S6	5.03	8.23	1.6E+03
Q9VZX6	mRpS35	4.77	7.94	1.5E+03
Q7PLL3	eIF-4B	5.82	8.96	1.4E+03
Q9V4M2	wech	4.56	7.70	1.4E+03
Q9VHR5		4.74	7.88	1.4E+03
P25822	pum	6.40	9.52	1.3E+03
Q9VGH5	glo	6.56	9.65	1.2E+03
P40796	La	6.69	9.79	1.2E+03
Q7KMJ6	nito	6.66	9.72	1.2E+03
Q9W0S7	Tudor-SN	7.04	10.09	1.1E+03
Q0E9B6	RpS11	7.11	10.12	1.0E+03
Q9VHC7	rump	6.79	9.79	1.0E+03
Q9VMY1	mRpL24	5.96	8.91	9.0E+02
Q9VF89	mRpL9	6.96	9.90	8.8E+02
Q9VHQ0	mRpS18A	5.69	8.61	8.4E+02
A1ZAX1	eIF3-S8	6.53	9.44	8.2E+02
A1ZAK7	Psi	6.99	9.90	8.0E+02
Q04047-2	nonA	5.89	8.78	7.8E+02
Q9VK59		5.26	8.14	7.6E+02
Q9W158		6.75	9.59	6.9E+02
Q0KHQ1		6.78	9.62	6.9E+02
Q9VMI5		5.30	8.13	6.7E+02
Q9VK85	Elf	4.88	7.70	6.6E+02
P41374	eIF-2alpha	5.66	8.48	6.6E+02
P41042	RpS4	6.53	9.35	6.5E+02
Q94901	lark	6.93	9.73	6.2E+02
Q9VVI2	Edc3	5.25	8.04	6.1E+02
Q24186	RpS5a	6.77	9.55	6.0E+02
Q9VTU4	CG5642	5.73	8.51	6.0E+02
Q9VKM1	piwi	6.10	8.88	6.0E+02
Q9VYD8	Rbp1-like	6.17	8.94	5.9E+02
P17704	RpS17	6.39	9.15	5.9E+02
P32100	RpL7	5.36	8.13	5.9E+02
Q8T3V6	mRpL47	7.30	10.06	5.7E+02
Q9VVV7		5.98	8.74	5.7E+02
P09052	vas	6.45	9.21	5.7E+02
Q9I7T8	CG11486-RB;CG11486-RI	4.94	7.70	5.7E+02
Q9VUH8		5.47	8.22	5.6E+02
O76922	aub	5.81	8.56	5.6E+02
Q9VD14		6.30	9.04	5.5E+02
Q9VWI1	pcm	5.44	8.17	5.4E+02
Q9W4L1	mRpL33	6.20	8.93	5.3E+02
Q9VJ86	bsf	7.55	10.27	5.3E+02
Q9VCE3	Rox8	7.48	10.19	5.1E+02
Q9VNE9	RpL13A	5.96	8.66	5.0E+02
P39018	RpS19a	6.62	9.32	5.0E+02
Q9VDL2	TFAM	5.05	7.74	4.9E+02
Q9VBQ1	msi	6.74	9.43	4.9E+02
E1JIT0	how	5.58	8.26	4.8E+02
Q9U9Q4	eIF-3p40	5.90	8.57	4.7E+02
Q9VPR5		4.40	7.06	4.6E+02
Q9VN25	eIF3-S10	6.88	9.52	4.4E+02
Q0KHU2	Imp	5.30	7.92	4.2E+02
Q9VNG2		5.12	7.75	4.2E+02

P29327	RpS6	5.69	8.31	4.2E+02
Q7K4L8		6.44	9.05	4.1E+02
Q9W5R8	RpL5	6.71	9.32	4.0E+02
Q9V3E7	Ref1	5.99	8.59	4.0E+02
Q7KUX7	Rbp2	6.72	9.30	3.8E+02
P48810	Hrb87F	7.47	10.05	3.8E+02
Q9VCK0	eIF-3p66	6.76	9.34	3.8E+02
P40798-2	stc	4.81	7.37	3.6E+02
P91632	mub	6.95	9.51	3.6E+02
A8E6M1		6.59	9.15	3.6E+02
A4V1Q1	bol	6.87	9.42	3.5E+02
Q7K3J0		4.88	7.43	3.5E+02
Q9VTF8	mRpL2	7.11	9.65	3.4E+02
Q9VG73	CG5641	6.82	9.35	3.3E+02
Q9VCA8	mask	4.44	6.96	3.3E+02
Q9W4X7	eIF3-S4-1	6.50	9.02	3.3E+02
Q9VLJ9	mRpL51	6.59	9.10	3.2E+02
Q7KN75	Dp1	7.20	9.71	3.2E+02
Q9VGI8	mus309	5.04	7.54	3.2E+02
Q9VJH2	mdy	6.76	9.23	3.0E+02
A1ZAP7		5.94	8.42	3.0E+02
Q960Z0	Klp10A	5.20	7.66	2.9E+02
Q6J5K9-2	armi	4.19	6.62	2.7E+02
Q9VF56		6.12	8.56	2.7E+02
P14130	RpS14a	7.42	9.85	2.6E+02
Q9I7T7-2	CG11505	5.15	7.56	2.6E+02
E1JJD7	tyf	4.84	7.25	2.6E+02
Q9W4C4	spoon	5.92	8.33	2.6E+02
Q9VFE4	RpS5b	6.53	8.93	2.5E+02
Q0E8G6		6.04	8.43	2.4E+02
Q0E9A5	muskelin	5.32	7.70	2.4E+02
Q9VSA3	CG12262	5.19	7.57	2.4E+02
Q9V3P6	Rpn2	4.76	7.13	2.3E+02
Q9VZ55		5.70	8.06	2.3E+02
Q95WY3	Nop56	6.04	8.39	2.2E+02
P55830	RpS3A	7.05	9.38	2.1E+02
Q9VJY6	RpL24	6.76	9.09	2.1E+02
P52304	polo	5.54	7.86	2.1E+02
P13008	RpS26	7.68	9.99	2.0E+02
Q9VDY1	Ino80	4.58	6.88	2.0E+02
Q7KNF1	drosha	4.67	6.97	2.0E+02
Q9V3T8	SC35	6.46	8.76	2.0E+02
P48588	RpS25	5.97	8.26	2.0E+02
Q0E940	eIF3-S9	5.92	8.21	2.0E+02
Q9VNE2	exba	6.12	8.42	2.0E+02
Q9V3V0	x16	6.78	9.07	2.0E+02
Q8MQJ9	brat	6.31	8.60	2.0E+02
Q7K581		7.02	9.30	1.9E+02
P21187	pAbp	8.17	10.45	1.9E+02
P48598	eIF-4E	6.94	9.21	1.9E+02
A1Z9E3	EfTuM	5.07	7.34	1.9E+02
O62602	lic	4.82	7.08	1.8E+02
Q9V3Z4	Rpn5	5.32	7.57	1.8E+02
Q9VEH0	alt	5.19	7.43	1.8E+02
Q94535	U2af38	6.02	8.26	1.8E+02
Q9VHP0	bel	7.58	9.82	1.7E+02
P05990	r	4.17	6.41	1.7E+02
P52029	Pgi	4.87	7.10	1.7E+02
P26686-3	B52	7.54	9.78	1.7E+02
Q9NCC3	SH3PX1	4.99	7.22	1.7E+02
Q9VK69		5.78	8.01	1.7E+02
P34739	lds	5.22	7.43	1.6E+02
Q7JZW2	RpS15	6.98	9.19	1.6E+02
Q9VW68		5.31	7.52	1.6E+02
Q9VEA1	eIF-1A	6.11	8.31	1.6E+02
O18404	scu	6.17	8.37	1.6E+02
O18640	Rack1	6.69	8.89	1.6E+02

Q9VTK9		5.47	7.67	1.6E+02
Q9W3M7		7.03	9.22	1.6E+02
Q9VZE4		6.41	8.60	1.6E+02
Q9VT61		5.51	7.71	1.6E+02
P04052	RpII215	4.34	6.53	1.6E+02
Q2PDU7	Etl1	4.93	7.12	1.5E+02
Q9VRP2		5.85	8.02	1.5E+02
Q9NFU0-2		7.19	9.36	1.5E+02
P48592	RnrS	5.00	7.16	1.5E+02
A4V464		5.82	7.98	1.5E+02
Q08473-2	sqd	8.46	10.62	1.5E+02
Q7KNF2	Pabp2	7.29	9.45	1.4E+02
P55935	RpS9	7.96	10.10	1.4E+02
E2QCF1	ATPCL	4.17	6.30	1.4E+02
P31009	RpS2	8.40	10.53	1.4E+02
Q7JVK1	mRpL18	6.06	8.19	1.3E+02
P09180	RpL4	6.03	8.16	1.3E+02
Q24009	BicC	7.41	9.53	1.3E+02
P19889	RpLP0	6.15	8.27	1.3E+02
Q06559	RpS3	7.95	10.07	1.3E+02
P38979-2	sta	7.32	9.44	1.3E+02
P07909-3	Hrb98DE	7.61	9.73	1.3E+02
P46222	RpL11	5.69	7.78	1.2E+02
Q8T8R1	CG3800	7.78	9.87	1.2E+02
Q26454	dpa	4.58	6.67	1.2E+02
Q9VV75		5.24	7.33	1.2E+02
Q9W4J4		5.83	7.92	1.2E+02
Q7JR58		6.21	8.28	1.2E+02
Q8IPE8	Mtpalpha	6.11	8.18	1.2E+02
Q9W237	RpS16	7.03	9.10	1.2E+02
Q8MSV2-2		6.84	8.89	1.1E+02
Q8I0G5-3	gammaCop	4.83	6.88	1.1E+02
P23128	me31B	8.31	10.35	1.1E+02
Q7KK96	SMC2	4.45	6.49	1.1E+02
Q9V3W7	SF2	7.39	9.42	1.1E+02
Q9W2K2	LSm1	5.78	7.81	1.1E+02
P19109	Rm62	7.74	9.76	1.1E+02
P18431-4	sgg	5.77	7.79	1.0E+02
Q9W0T7	CG16940-RC	5.89	7.91	1.0E+02
Q7KM15	bic	6.27	8.28	1.0E+02
Q86S05-3	lig	6.79	8.79	1.0E+02
Q9VYS4	wisp	6.21	8.20	9.8E+01
P15348	Top2	6.33	8.32	9.8E+01
Q9VD66	fit	6.86	8.86	9.8E+01
Q9VGP4	Ranbp9	4.61	6.60	9.7E+01
Q9VA91	RpS7	7.05	9.03	9.6E+01
Q9VXQ5	Tcp-1zeta	5.80	7.79	9.6E+01
Q95RE4	yps	8.16	10.14	9.6E+01
Q9VKW5		4.63	6.61	9.5E+01
P48809	Hrb27C	8.50	10.47	9.5E+01
P91938-4	Trxr-1	5.35	7.33	9.4E+01
Q8MSV2	shep	6.69	8.67	9.4E+01
Q9V9V7	pasha	5.73	7.70	9.3E+01
Q8MLY8	RpS8	6.66	8.63	9.3E+01
Q9I7D3	Capr	7.04	9.01	9.3E+01
Q9VLM8	Aats-ala	4.89	6.85	9.2E+01
P06606	alphaTub67C	6.10	8.06	9.1E+01
Q8MZI3		7.32	9.28	9.1E+01
P55841	RpL14	7.52	9.47	9.0E+01
A1Z8U4	Cct5	5.69	7.64	8.8E+01
P48159	RpL23	5.84	7.78	8.7E+01
P24785	mle	6.75	8.68	8.6E+01
Q9VY91	Pdcd4	5.72	7.65	8.6E+01
Q9VB55	woc	4.02	5.95	8.5E+01
Q9VC94		5.32	7.25	8.4E+01
Q9NJH0	Ef1gamma	6.24	8.16	8.3E+01
Q9VBU7	Nup358	5.41	7.33	8.2E+01

Q9VW54	Rpn1	5.26	7.17	8.1E+01
Q9W1H4	DNA-ligI	5.44	7.35	8.1E+01
Q9VCN1	DNApol-epsilon	4.45	6.34	7.8E+01
Q9VKX2	Mdh1	5.75	7.63	7.5E+01
O16797-4	RpL3	5.42	7.29	7.5E+01
Q9VIE8	Acon	5.30	7.17	7.5E+01
Q9VDH8	RpS30	6.56	8.43	7.3E+01
Q24311	lin19	4.74	6.60	7.2E+01
Q9VFT4	rin	7.83	9.69	7.2E+01
O18388	Fs(2)Ket	5.49	7.34	7.0E+01
Q9VEG6	Pxt	5.84	7.68	6.9E+01
Q7K3M5		5.71	7.55	6.9E+01
P25159-2	stau	5.62	7.45	6.8E+01
P48605	Cctgamma	6.12	7.94	6.7E+01
Q9VTP4	RpL10Ab	5.82	7.64	6.7E+01
Q9VWH4-2	l(1)G0156	5.26	7.08	6.6E+01
P20480	ncd	5.10	6.90	6.3E+01
Q9XYU1	Mcm3	5.17	6.97	6.3E+01
E1JJM0	Imp	8.21	10.00	6.2E+01
P02844	Yp2	8.06	9.86	6.2E+01
Q9VAW5	larp	7.46	9.25	6.2E+01
Q7JVI3	Tango7	4.87	6.66	6.2E+01
Q9VNH2		4.70	6.49	6.1E+01
P52295	Pen	6.62	8.41	6.1E+01
P80455	RpS12	8.30	10.08	6.1E+01
P50882	RpL9	6.04	7.82	6.0E+01
Q9W2U7	nocte	5.68	7.44	5.8E+01
P47980	Tis11	5.73	7.49	5.8E+01
Q4Z8K6-3	RanBPM	6.03	7.79	5.8E+01
Q9W245		5.05	6.80	5.7E+01
Q9VWE6	CG14215	4.45	6.20	5.7E+01
Q9VVA4		4.48	6.23	5.6E+01
Q86BI3	Zn72D	7.37	9.12	5.6E+01
Q9VUV9	CG5931	5.08	6.82	5.5E+01
Q9VLB7	Gdi	6.03	7.77	5.5E+01
Q02748	eIF-4a	7.45	9.19	5.5E+01
Q9VN21	lost	8.75	10.48	5.3E+01
Q9VQ94	gho	5.45	7.17	5.3E+01
P28750	exu	5.56	7.28	5.2E+01
Q9VZU7		4.74	6.45	5.2E+01
P12613	T-cp1	6.30	8.01	5.1E+01
Q9XYN7	Prosbeta3	5.34	7.04	5.0E+01
Q9W1H8	Thiolase	6.26	7.94	4.9E+01
Q9VMA3	cup	7.70	9.38	4.8E+01
P02843	Yp1	8.49	10.16	4.7E+01
Q9W227		5.78	7.44	4.6E+01
Q95U58	mkg-p	5.08	6.73	4.5E+01
Q9W2T3		5.17	6.82	4.4E+01
Q9VQL1		4.92	6.57	4.4E+01
Q8IPB1	dUTPase	5.95	7.60	4.4E+01
Q7JRH5		7.16	8.80	4.4E+01
Q5KTT4	NAT1	6.57	8.20	4.3E+01
P12881	Pros35	5.44	7.07	4.2E+01
O17445	RpL15	6.87	8.48	4.1E+01
Q0E993	Aats-val	5.12	6.73	4.1E+01
P46223	RpL7A	6.80	8.41	4.1E+01
Q9VL70	yip2	5.68	7.28	4.1E+01
Q9VUC1	Hsc70Cb	5.20	6.81	4.0E+01
Q8T0L3	Uba1	5.92	7.51	4.0E+01
Q9VBU9	RpS27	6.82	8.41	3.9E+01
P06607	Yp3	8.17	9.77	3.9E+01
Q0E965	SRPK	6.35	7.94	3.9E+01
Q9VHN7		6.36	7.94	3.8E+01
Q9VUA0	CG8833	4.84	6.42	3.8E+01
Q9W1G7	Nap1	6.09	7.67	3.8E+01
Q05856	SmB	6.92	8.49	3.7E+01
Q9VFF0		5.84	7.39	3.5E+01

P36241	RpL19	5.57	7.10	3.4E+01
Q9VCH1	eIF4G2	5.14	6.67	3.4E+01
P08928	Lam	5.55	7.08	3.4E+01
Q9V496	Rfabg	5.97	7.49	3.3E+01
P13060	Ef2b	7.47	8.99	3.3E+01
Q9VN44	Karybeta3	4.75	6.26	3.2E+01
Q9W0S6	mRpL17	6.98	8.49	3.2E+01
P54358	DNApol-delta	4.97	6.48	3.2E+01
Q0E8E8	Mpcp	5.51	7.00	3.1E+01
Q9VVU2	RpL26	5.68	7.17	3.1E+01
P46415	Fdh	5.34	6.82	3.1E+01
P04359	RpL32	6.12	7.60	3.0E+01
P55828	RpS20	7.62	9.09	3.0E+01
Q9W414	Rpt4	5.58	7.04	2.9E+01
Q9V414	Smg5	5.55	7.01	2.9E+01
Q9VNA5	Prosbeta7	5.28	6.74	2.9E+01
Q9VHL2	Tcp-1eta	6.75	8.20	2.8E+01
Q7K1Q6	Jabba	6.25	7.69	2.7E+01
Q0KIB3		5.09	6.52	2.7E+01
P54367	CkIalpha	5.75	7.17	2.7E+01
Q00637	Sod2	5.80	7.22	2.6E+01
Q9VCX7		6.80	8.21	2.6E+01
Q9VI58		6.94	8.35	2.6E+01
Q7KUL0		5.86	7.27	2.5E+01
Q8T3U2	RpS23	6.87	8.27	2.5E+01
Q9VPX3		7.25	8.64	2.5E+01
A1ZAB5	clu	6.18	7.57	2.5E+01
Q9VBH8	RpL34a;RpL34b	6.60	8.00	2.5E+01
Q9VTZ0	tral	8.86	10.25	2.5E+01
Q7JV23	ACC	5.05	6.44	2.5E+01
Q7PLL6		6.00	7.38	2.4E+01
Q9V3V3	Pez	4.39	5.77	2.4E+01
Q7KUT2-2	CG8798	4.64	6.01	2.3E+01
Q9XTL9	GlyP	5.93	7.29	2.3E+01
Q9W1B9	RpL12	7.04	8.39	2.3E+01
P48591	RnrL	5.67	7.02	2.2E+01
Q9VJ19	RpL30	7.79	9.14	2.2E+01
Q9V9T4	Acf1	4.65	5.99	2.2E+01
Q05344	Ssrp	4.62	5.96	2.2E+01
P17917	mus209	5.89	7.22	2.2E+01
A8DY44	Nipped-A	4.71	6.05	2.2E+01
P25867	eff	6.27	7.60	2.1E+01
P20432	GstD1	6.38	7.71	2.1E+01
Q9VZS3	CG17737	6.32	7.65	2.1E+01
Q23972	smg	7.62	8.95	2.1E+01
O62619-2	PyK	6.97	8.29	2.1E+01
B7YZQ7	Nurf-38	6.10	7.42	2.1E+01
Q86BY9	rig	7.37	8.69	2.1E+01
Q9V597	RpL31	5.94	7.25	2.0E+01
P22700-2	Ca-P60A	5.57	6.87	2.0E+01
P45594	tsr	6.77	8.08	2.0E+01
A1Z9K0		6.88	8.18	2.0E+01
P25823	tud	6.46	7.75	2.0E+01
O01666	ATPsyn-gamma	5.92	7.20	1.9E+01
Q7KV89	msn	4.60	5.88	1.9E+01
Q9VEB1	Mdh2	6.82	8.11	1.9E+01
Q9W4M9	CG6133	6.62	7.90	1.9E+01
O61380	eIF4G	7.25	8.52	1.9E+01
Q9VQ76	tho2	5.02	6.30	1.9E+01
Q8IP94	Aats-thr	5.54	6.79	1.8E+01
Q8T0M9	Got2	5.31	6.56	1.7E+01
P15007-2	Eno	6.01	7.25	1.7E+01
A8DYI1	CG5544	5.18	6.39	1.6E+01
A1ZAW0	Dcr-2	4.74	5.95	1.6E+01
Q9V3G1	RpL8	7.32	8.51	1.6E+01
A1ZBW1	bl	8.54	9.73	1.5E+01
Q9VWS1		5.47	6.65	1.5E+01

Q9VGW6	Mcm5	5.56	6.74	1.5E+01
Q9VIQ8	CoIV	5.81	6.99	1.5E+01
Q8IP51	wb	7.21	8.39	1.5E+01
P02518	Hsp27	7.06	8.23	1.5E+01
P25161	Rpn3	5.40	6.57	1.5E+01
Q9W0A8	RpL23A	6.09	7.25	1.4E+01
Q86DS1	HIP-R;HIP	5.08	6.24	1.4E+01
P06605	alphaTub84D;alphaTub8	7.16	8.31	1.4E+01
Q9V3U6	26-29-p	6.27	7.41	1.4E+01
Q27268	Hel25E	6.94	8.07	1.3E+01
A1ZBJ2		5.87	7.00	1.3E+01
O96827	Ef1beta	6.03	7.13	1.2E+01
Q7KSN0	MBD-R2	5.33	6.41	1.2E+01
Q9VLV5	SmE	7.20	8.27	1.2E+01
E1JIJ7		5.45	6.52	1.2E+01
P35381	blw	6.92	7.97	1.1E+01
Q7KU24	Chd1	4.67	5.72	1.1E+01
Q94518	Nacalpha	6.81	7.86	1.1E+01
Q9VT75	vnc	5.29	6.32	1.1E+01
Q9VIS4		6.55	7.59	1.1E+01
P54362	g	5.38	6.40	1.1E+01
P46150-2	Moe	7.20	8.22	1.1E+01
Q9GU68	eIF-5A	6.85	7.86	1.0E+01
Q9VCU9	Dcr-1	6.61	7.62	1.0E+01
P98163-2	yl	5.85	6.85	9.9E+00
Q9VHC0	RnpS1	6.04	7.02	9.6E+00
P25007	Cyp1	7.33	8.29	9.2E+00
Q9VSE6	RecQ4	5.38	6.34	9.2E+00
Q95RL2	row	4.63	5.58	8.9E+00
Q7KMP8	Rpn9	5.45	6.39	8.8E+00
O46102	CG11418-RA	6.46	7.40	8.7E+00
Q7JXC4	P32	5.59	6.52	8.5E+00
O97102	smt3	7.45	8.36	8.3E+00
P54622	mtSSB	8.40	9.31	8.1E+00
P48375	FK506-bp2	6.94	7.83	7.6E+00
Q9VZ23	ran	7.28	8.15	7.4E+00
Q9VU35		5.76	6.62	7.2E+00
P55824-3	faf	5.98	6.83	7.2E+00
Q9VU02	SmD1	7.29	8.14	7.1E+00
Q9VWV6	Tsf1	5.05	5.89	6.9E+00
Q26365-2	sesB	7.39	8.22	6.8E+00
Q9VS34	RpL18	6.75	7.57	6.8E+00
Q9VPH7	eRF1	6.50	7.32	6.6E+00
Q7KTV5	Nopp140	5.59	6.39	6.4E+00
Q9VKI8		6.22	7.02	6.3E+00
Q9VN58	ctrip	6.06	6.85	6.1E+00
Q3YMU0	ERp60	6.13	6.91	6.0E+00
A8DY81	Not1	6.73	7.50	5.8E+00
P02517	Hsp26	7.85	8.60	5.6E+00
Q9VYS3	Upf1	8.23	8.97	5.5E+00
P02828	Hsp83	7.73	8.47	5.5E+00
A1ZAN7	RhoGEF2	4.61	5.34	5.4E+00
P18091-2	Actn	4.91	5.64	5.3E+00
Q8MSW0	Aats-ile	5.54	6.27	5.3E+00
O97159	Mi-2	6.11	6.81	5.0E+00
Q9VEZ3	msps	6.36	7.03	4.7E+00
Q9XZJ4	Prosalpha1	7.05	7.72	4.7E+00
Q9VBR8		4.68	5.34	4.6E+00
Q05825	ATPsyn-beta	7.50	8.16	4.6E+00
O61231	RpL10	7.25	7.90	4.5E+00
P31409	Vha55	6.73	7.37	4.4E+00
Q8INH7	timeout	4.84	5.47	4.3E+00
Q9V677	CG8858	4.55	5.17	4.2E+00
P05661-26	Mhc	5.13	5.74	4.1E+00
Q24253	Bap	5.82	6.42	4.0E+00
Q9W3W6		5.08	5.67	3.9E+00
Q9VXE0	SmG	7.50	8.08	3.9E+00

Q7KN62	TER94	7.34	7.92	3.8E+00
P29613	Tpi	6.52	7.08	3.6E+00
Q9VX32	RhoGAPp190	5.09	5.63	3.4E+00
P47938	dhd	6.94	7.46	3.3E+00
P36872-2	tws	6.60	7.12	3.3E+00
Q8SYG3		6.36	6.87	3.3E+00
Q9VKW8	LSm-4-RB	7.15	7.66	3.2E+00
Q9VRL1	Uev1A	6.06	6.57	3.2E+00
Q9W2Z6		6.01	6.50	3.1E+00
O76927-2	RpS21	6.63	7.12	3.1E+00
Q9VA83	Fer2LCH	7.33	7.81	3.0E+00
P41126	RpL13	6.16	6.62	2.9E+00
Q9W3W8	RpL17	6.21	6.67	2.9E+00
P48149	RpS15Aa	7.51	7.96	2.9E+00
Q9V9W7	rod	5.22	5.66	2.8E+00
Q9Y105	Aats-gln	6.04	6.49	2.8E+00
P29844	Hsc70-3	7.18	7.57	2.4E+00
P15357	RpS27A;RpL40;Ubi-p5E;U	9.97	10.35	2.4E+00
P28668	Aats-glupro	6.00	6.37	2.4E+00
O44437	SmD3	7.61	7.98	2.4E+00
P08879	awd	6.84	7.21	2.3E+00
Q9VY45		6.14	6.50	2.3E+00
Q24583	Vha14-1	5.93	6.28	2.3E+00
P07487	Gapdh2	6.40	6.72	2.1E+00
A1Z8U0	Prp8	6.93	7.25	2.1E+00
Q8IMW2		4.98	5.27	2.0E+00
Q95029-2	Cp1	6.54	6.83	1.9E+00
A8JV36	Trf4-1	6.49	6.77	1.9E+00
Q9NG98	Top3alpha	7.69	7.96	1.8E+00
P08736	Ef1alpha48D;Ef1alpha10(	8.85	9.10	1.8E+00
Q9VGM2	fabp	7.32	7.56	1.8E+00
Q9W0B8	alphaCop	5.62	5.82	1.6E+00
Q9V429-2	Trx-2	6.82	6.99	1.5E+00
O18413	Pros45	6.23	6.41	1.5E+00
Q9VK60		6.56	6.71	1.4E+00
Q7KRU8	Fer1HCH	7.17	7.31	1.4E+00
Q9VGS2	Tctp	7.00	7.13	1.4E+00
P29742	Chc	6.77	6.90	1.3E+00
Q9V3H2	Rpn11	6.81	6.94	1.3E+00
Q24297	SmF	6.35	6.45	1.3E+00
P08570	RpLP1	7.27	7.38	1.3E+00
P23696	mts	6.38	6.49	1.3E+00
P20193	Su(var)3-7	6.73	6.81	1.2E+00
Q9VC45	asp	6.56	6.62	1.2E+00
P11147	Hsc70-4	8.25	8.32	1.2E+00
P25843	chic	7.38	7.43	1.1E+00
Q9NBD7	chb	5.74	5.77	1.1E+00
O96651	Top3beta	9.81	9.84	1.1E+00
Q9V3R3	EG:BACR7A4.17	6.31	6.31	1.0E+00
Q9V3Z6	ck	6.95	6.95	1.0E+00
Q27580	Ahcy13	6.97	6.96	9.8E-01
Q9W3U9		5.45	5.44	9.7E-01
P61851	Sod	6.38	6.35	9.3E-01
A1Z713	Vps13	7.69	7.64	9.0E-01
Q9VL71	Srp54	6.90	6.84	8.7E-01
Q9VJ87	ncm	6.02	5.94	8.2E-01
Q95083	Prosalpha5	6.78	6.67	7.9E-01
Q59E58	zip	6.40	6.28	7.7E-01
Q9VU43	SRm160	7.24	7.09	7.2E-01
Q9VI10	SmD2	7.08	6.94	7.1E-01
Q24560	betaTub56D	8.76	8.60	7.0E-01
Q9I7U4	sls	5.35	5.17	6.7E-01
Q01604	Pgk	6.19	5.97	6.1E-01
A4V364	Syp	6.56	6.32	5.7E-01
P07486	Gapdh1	8.04	7.79	5.7E-01
P08985	His2Av	8.64	8.37	5.5E-01
Q9VSS2	Srp68	7.42	7.15	5.4E-01

O97125	Hsp68;Hsp70Bbb;Hsp70E	7.60	7.33	5.4E-01
Q9V3P0	Jafrac1	8.13	7.85	5.3E-01
O18333	Rab2	6.71	6.41	5.0E-01
P37276-2	Dhc64C	6.48	6.13	4.5E-01
Q27415-2	Nlp	6.75	6.41	4.5E-01
Q8IR12	Cap	6.24	5.90	4.5E-01
Q7KMQ0	Rpt1	5.91	5.55	4.4E-01
Q5U0Y0	Sin3A	6.46	6.09	4.3E-01
Q86BA1	Mical	6.02	5.63	4.1E-01
Q9VHR8-2	DppIII	5.24	4.79	3.6E-01
P35122	Uch	6.64	6.13	3.1E-01
P02283	His2B	8.53	8.01	3.0E-01
Q9VHS8	eIF4AIII	7.19	6.65	2.9E-01
Q9XYZ6		6.27	5.69	2.6E-01
Q9W4M3	Femcoat	6.42	5.83	2.6E-01
Q9VB74	CG6066	6.79	6.16	2.4E-01
Q8I0N3	TwdlL;TwdlB;TwdlM;Twd	6.58	5.95	2.3E-01
P05389	RpLP2	6.58	5.93	2.2E-01
P48601	Pros26.4	6.59	5.93	2.2E-01
Q9VF87	Sra-1	5.46	4.79	2.1E-01
P84040	His4	9.23	8.49	1.8E-01
Q9Y171	BcDNA.GH02220	7.12	6.35	1.7E-01
Q9VIJ9	ns4	6.18	5.41	1.7E-01
Q27331	Vha68-2;Vha68-1	6.82	6.02	1.6E-01
Q9V3V6	Tbp-1	6.67	5.84	1.5E-01
Q9VAX8		7.35	6.43	1.2E-01
P61209	Arf79F	7.79	6.82	1.1E-01
P61857	betaTub85D	7.96	6.96	9.9E-02
Q9VYH1	Fer3HCH	6.21	5.15	8.8E-02
Q9VLT5	poe	5.91	4.79	7.6E-02
Q9V3I2	Rab5	7.23	6.11	7.5E-02
P92177-4	14-3-3epsilon	8.42	7.29	7.3E-02
P02299	His3;His3.3A	8.81	7.61	6.3E-02
Q4ABH0	Msp-300	6.92	5.56	4.3E-02
Q9V405	Rpt3;Rpt3R	6.35	4.98	4.3E-02
Q24117	ctp;Cdlc2	7.77	6.35	3.8E-02
P09491-2	Tm2	7.06	5.63	3.8E-02
Q9VWE1	et	6.82	5.36	3.5E-02
Q0E924	l(2)03709	6.48	4.88	2.5E-02
O18332	Rab1	7.29	5.66	2.3E-02
Q9VQI7	Prx6005	7.04	5.28	1.8E-02
P29310	14-3-3zeta	8.71	6.76	1.1E-02
P29843	Hsc70-1	7.64	5.67	1.1E-02
Q9VH01	Bruce	7.04	4.86	6.6E-03
Q7JVH6		8.67	5.92	1.8E-03
Q9W060		4.58	N/A	0.0E+00
Q9VPN5	Hop	5.23	N/A	0.0E+00
Q9VJY7		5.34	N/A	0.0E+00
P08970	su(Hw)	5.40	N/A	0.0E+00
Q9VJZ4		5.42	N/A	0.0E+00
Q9W1Y1	CG3501	5.70	N/A	0.0E+00
Q9VDG0	Dhc93AB	5.73	N/A	0.0E+00
P54399	Pdi	5.80	N/A	0.0E+00
P24156	l(2)37Cc	5.89	N/A	0.0E+00
Q07152	ras	5.91	N/A	0.0E+00
Q7YZ99		6.43	N/A	0.0E+00
P40304	Pros26	6.48	N/A	0.0E+00
P40946	Arf51F	6.92	N/A	0.0E+00
P32392	Arp66B	7.08	N/A	0.0E+00
P18824-2	arm	7.22	N/A	0.0E+00
P48554	Rac2;Rac1	7.29	N/A	0.0E+00
P45888	Arp14D;Arp2	7.31	N/A	0.0E+00
Q9VGB6	Pglym87	7.41	N/A	0.0E+00
O18335	Rab11	7.71	N/A	0.0E+00
P62152	Cam	8.13	N/A	0.0E+00

Supplementary table S7:GO-term enrichment analysis (Drosophila melanogaster embryos)

GO ID	*Term*	*Count*	*%*	*p-value*
GO:0006412	translation	159	22.1	1.1E-55
GO:0007052	mitotic spindle organization	79	11.0	1.7E-36
GO:0007051	spindle organization	82	11.4	8.7E-34
GO:0000022	mitotic spindle elongation	48	6.7	2.5E-32
GO:0000226	microtubule cytoskeleton organization	93	12.9	2.7E-32
GO:0010608	posttranscriptional regulation of gene expression	60	8.3	3.8E-32
GO:0051231	spindle elongation	48	6.7	5.7E-32
GO:0000278	mitotic cell cycle	95	13.2	6.1E-27
GO:0007017	microtubule-based process	100	13.9	1.2E-25
GO:0007010	cytoskeleton organization	105	14.6	9.7E-24
GO:0000279	M phase	106	14.7	2.6E-23
GO:0022403	cell cycle phase	106	14.7	5.4E-22
GO:0006417	regulation of translation	37	5.1	3.1E-21
GO:0016071	mRNA metabolic process	64	8.9	1.9E-20
GO:0031047	gene silencing by RNA	29	4.0	3.0E-20
GO:0022402	cell cycle process	109	15.1	3.8E-20
GO:0043484	regulation of RNA splicing	34	4.7	1.5E-19
GO:0007049	cell cycle	114	15.8	1.7E-18
GO:0035194	posttranscriptional gene silencing by RNA	26	3.6	2.8E-18
GO:0016441	posttranscriptional gene silencing	26	3.6	2.8E-18
GO:0006396	RNA processing	73	10.1	2.3E-17
GO:0048024	regulation of nuclear mRNA splicing, via spliceosome	31	4.3	3.9E-17
GO:0050684	regulation of mRNA processing	31	4.3	3.9E-17
GO:0034621	cellular macromolecular complex subunit organization	55	7.6	5.8E-16
GO:0000381	regulation of alternative nuclear mRNA splicing, via spliceosome	28	3.9	2.0E-15
GO:0006397	mRNA processing	50	6.9	1.3E-13
GO:0016246	RNA interference	19	2.6	4.8E-13
GO:0032268	regulation of cellular protein metabolic process	39	5.4	6.0E-13
GO:0043933	macromolecular complex subunit organization	61	8.5	6.6E-13
GO:0006413	translational initiation	25	3.5	7.0E-13
GO:0022618	ribonucleoprotein complex assembly	20	2.8	1.3E-12
GO:0034622	cellular macromolecular complex assembly	44	6.1	1.4E-12
GO:0022613	ribonucleoprotein complex biogenesis	31	4.3	1.5E-11
GO:0006259	DNA metabolic process	49	6.8	8.1E-11
GO:0016458	gene silencing	39	5.4	2.0E-10
GO:0040029	regulation of gene expression, epigenetic	41	5.7	2.1E-10
GO:0006402	mRNA catabolic process	16	2.2	4.1E-10
GO:0045727	positive regulation of translation	10	1.4	1.2E-09

GO:0065003	macromolecular complex assembly	50	6.9	1.2E-09
GO:0006401	RNA catabolic process	16	2.2	1.5E-09
GO:0035195	gene silencing by miRNA	13	1.8	1.8E-09
GO:0046011	regulation of oskar mRNA translation	11	1.5	2.3E-09
GO:0010605	negative regulation of macromolecule metabolic process	58	8.1	3.0E-09
GO:0006260	DNA replication	28	3.9	3.7E-09
GO:0000956	nuclear-transcribed mRNA catabolic process	14	1.9	1.0E-08
GO:0006261	DNA-dependent DNA replication	20	2.8	3.8E-08
GO:0051276	chromosome organization	51	7.1	4.8E-08
GO:0019953	sexual reproduction	100	13.9	6.2E-08
GO:0007281	germ cell development	45	6.3	1.7E-07
GO:0043331	response to dsRNA	9	1.3	2.1E-07
GO:0031050	dsRNA fragmentation	9	1.3	2.1E-07
GO:0007276	gamete generation	96	13.3	2.3E-07
GO:0007292	female gamete generation	78	10.8	3.5E-07
GO:0048610	reproductive cellular process	73	10.1	3.9E-07
GO:0007028	cytoplasm organization	20	2.8	4.0E-07
GO:0060341	regulation of cellular localization	20	2.8	8.6E-07
GO:0048477	oogenesis	76	10.6	8.6E-07
GO:0044265	cellular macromolecule catabolic process	41	5.7	1.0E-06
GO:0070727	cellular macromolecule localization	43	6.0	1.3E-06
GO:0006403	RNA localization	28	3.9	1.4E-06
GO:0008380	RNA splicing	31	4.3	1.5E-06
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	10	1.4	1.7E-06
GO:0048609	reproductive process in a multicellular organism	99	13.8	1.8E-06
GO:0032504	multicellular organism reproduction	99	13.8	1.8E-06
GO:0007314	oocyte anterior/posterior axis specification	21	2.9	2.0E-06
GO:0032270	positive regulation of cellular protein metabolic process	10	1.4	3.2E-06
GO:0051247	positive regulation of protein metabolic process	10	1.4	3.2E-06
GO:0008358	maternal determination of anterior/posterior axis, embryo	21	2.9	3.7E-06
GO:0010629	negative regulation of gene expression	43	6.0	4.8E-06
GO:0003006	reproductive developmental process	70	9.7	5.1E-06
GO:0000375	RNA splicing, via transesterification reactions	27	3.8	6.1E-06
GO:0006270	DNA replication initiation	9	1.3	6.2E-06
GO:0035196	gene silencing by miRNA, production of miRNAs	7	1.0	6.2E-06
GO:0007315	pole plasm assembly	17	2.4	7.6E-06
GO:0030261	chromosome condensation	12	1.7	1.0E-05
GO:0007317	regulation of pole plasm oskar mRNA localization	11	1.5	1.1E-05

GO:0008298	intracellular mRNA localization	18	2.5	1.3E-05
GO:0009057	macromolecule catabolic process	44	6.1	1.5E-05
GO:0060281	regulation of oocyte development	11	1.5	1.6E-05
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	26	3.6	1.6E-05
GO:0000398	nuclear mRNA splicing, via spliceosome	26	3.6	1.6E-05
GO:0006446	regulation of translational initiation	9	1.3	3.4E-05
GO:0031123	RNA 3'-end processing	11	1.5	3.4E-05
GO:0006323	DNA packaging	17	2.4	3.6E-05
GO:0017148	negative regulation of translation	12	1.7	4.0E-05
GO:0007059	chromosome segregation	24	3.3	4.1E-05
GO:0007309	oocyte axis specification	24	3.3	4.7E-05
GO:0031124	mRNA 3'-end processing	10	1.4	5.5E-05
GO:0048469	cell maturation	25	3.5	6.3E-05
GO:0007308	oocyte construction	24	3.3	7.1E-05
GO:0051301	cell division	34	4.7	7.3E-05
GO:0007316	pole plasm RNA localization	14	1.9	1.1E-04
GO:0048599	oocyte development	24	3.3	1.4E-04
GO:0051169	nuclear transport	17	2.4	1.4E-04
GO:0006913	nucleocytoplasmic transport	17	2.4	1.4E-04
GO:0007319	negative regulation of oskar mRNA translation	6	0.8	1.5E-04
GO:0030422	RNA interference, production of siRNA	6	0.8	1.5E-04
GO:0021700	developmental maturation	25	3.5	1.5E-04
GO:0016568	chromatin modification	22	3.1	1.8E-04
GO:0030423	RNA interference, targeting of mRNA for destruction	5	0.7	1.9E-04
GO:0009994	oocyte differentiation	25	3.5	2.1E-04
GO:0000280	nuclear division	26	3.6	2.5E-04
GO:0007097	nuclear migration	9	1.3	2.6E-04
GO:0016569	covalent chromatin modification	14	1.9	3.0E-04
GO:0016570	histone modification	14	1.9	3.0E-04
GO:0007318	pole plasm protein localization	6	0.8	3.1E-04
GO:0034660	ncRNA metabolic process	24	3.3	3.1E-04
GO:0019094	pole plasm mRNA localization	13	1.8	3.6E-04
GO:0006325	chromatin organization	28	3.9	4.1E-04
GO:0048285	organelle fission	26	3.6	4.3E-04
GO:0007067	mitosis	25	3.5	5.1E-04
GO:0000087	M phase of mitotic cell cycle	25	3.5	6.2E-04
GO:0040023	establishment of nucleus localization	9	1.3	6.8E-04
GO:0000070	mitotic sister chromatid segregation	12	1.7	7.8E-04
GO:0016571	histone methylation	8	1.1	8.6E-04
GO:0045995	regulation of embryonic development	13	1.8	9.2E-04

GO:0000819	sister chromatid segregation	12	1.7	9.4E-04
GO:0034984	cellular response to DNA damage stimulus	17	2.4	9.8E-04
GO:0007076	mitotic chromosome condensation	7	1.0	1.0E-03
GO:0008283	cell proliferation	22	3.1	1.0E-03
GO:0006281	DNA repair	19	2.6	1.1E-03
GO:0006267	pre-replicative complex assembly	5	0.7	1.2E-03
GO:0050658	RNA transport	14	1.9	1.2E-03
GO:0050657	nucleic acid transport	14	1.9	1.2E-03
GO:0051236	establishment of RNA localization	14	1.9	1.4E-03
GO:0044087	regulation of cellular component biogenesis	16	2.2	1.5E-03
GO:0043631	RNA polyadenylation	7	1.0	1.5E-03
GO:0000245	spliceosome assembly	7	1.0	1.5E-03
GO:0030952	establishment or maintenance of cytoskeleton polarity	9	1.3	1.5E-03
GO:0030951	establishment or maintenance of microtubule cytoskeleton polarity	9	1.3	1.5E-03
GO:0016325	oocyte microtubule cytoskeleton organization	9	1.3	1.5E-03
GO:0060284	regulation of cell development	22	3.1	1.5E-03
GO:0008595	determination of anterior/posterior axis, embryo	23	3.2	1.6E-03
GO:0007351	tripartite regional subdivision	23	3.2	1.6E-03
GO:0008104	protein localization	49	6.8	1.7E-03
GO:0051640	organelle localization	16	2.2	1.9E-03
GO:0033962	cytoplasmic mRNA processing body assembly	4	0.6	1.9E-03
GO:0046012	positive regulation of oskar mRNA translation	4	0.6	1.9E-03
GO:0051647	nucleus localization	9	1.3	1.9E-03
GO:0006900	membrane budding	7	1.0	2.0E-03
GO:0032508	DNA duplex unwinding	5	0.7	2.2E-03
GO:0032392	DNA geometric change	5	0.7	2.2E-03
GO:0006376	mRNA splice site selection	5	0.7	2.2E-03
GO:0007350	blastoderm segmentation	31	4.3	2.2E-03
GO:0015931	nucleobase, nucleoside, nucleotide and nucleic acid transport	14	1.9	2.5E-03
GO:0000578	embryonic axis specification	23	3.2	2.5E-03
GO:0032269	negative regulation of cellular protein metabolic process	12	1.7	2.7E-03
GO:0006414	translational elongation	7	1.0	2.8E-03
GO:0006338	chromatin remodeling	11	1.5	2.9E-03
GO:0009948	anterior/posterior axis specification	23	3.2	3.0E-03
GO:0051248	negative regulation of protein metabolic process	12	1.7	3.1E-03
GO:0046907	intracellular transport	40	5.6	3.4E-03
GO:0006418	tRNA aminoacylation for protein translation	11	1.5	3.4E-03
GO:0043039	tRNA aminoacylation	11	1.5	3.4E-03

GO:0043624	cellular protein complex disassembly	6	0.8	3.5E-03
GO:0006378	mRNA polyadenylation	6	0.8	3.5E-03
GO:0043241	protein complex disassembly	6	0.8	3.5E-03
GO:0006379	mRNA cleavage	6	0.8	3.5E-03
GO:0043038	amino acid activation	11	1.5	4.0E-03
GO:0042254	ribosome biogenesis	13	1.8	4.2E-03
GO:0006974	response to DNA damage stimulus	19	2.6	4.2E-03
GO:0051252	regulation of RNA metabolic process	74	10.3	4.4E-03
GO:0031053	primary microRNA processing	4	0.6	4.5E-03
GO:0048134	germ-line cyst formation	7	1.0	4.9E-03
GO:0035282	segmentation	34	4.7	5.1E-03
GO:0045451	pole plasm oskar mRNA localization	10	1.4	5.1E-03
GO:0009880	embryonic pattern specification	31	4.3	5.3E-03
GO:0008213	protein amino acid alkylation	8	1.1	5.4E-03
GO:0006479	protein amino acid methylation	8	1.1	5.4E-03
GO:0034470	ncRNA processing	16	2.2	5.9E-03
GO:0051656	establishment of organelle localization	14	1.9	5.9E-03
GO:0007616	long-term memory	7	1.0	6.2E-03
GO:0007613	memory	9	1.3	6.4E-03
GO:0070142	synaptic vesicle budding	6	0.8	6.7E-03
GO:0016185	synaptic vesicle budding from presynaptic membrane	6	0.8	6.7E-03
GO:0006399	tRNA metabolic process	16	2.2	7.2E-03
GO:0007040	lysosome organization	8	1.1	8.1E-03
GO:0007277	pole cell development	8	1.1	8.1E-03
GO:0034613	cellular protein localization	25	3.5	8.1E-03
GO:0032259	methylation	11	1.5	8.3E-03
GO:0033119	negative regulation of RNA splicing	4	0.6	8.4E-03
GO:0032984	macromolecular complex disassembly	6	0.8	8.8E-03
GO:0034623	cellular macromolecular complex disassembly	6	0.8	8.8E-03
GO:0034504	protein localization in nucleus	9	1.3	9.0E-03
GO:0009798	axis specification	27	3.8	9.5E-03
GO:0051168	nuclear export	8	1.1	9.7E-03
GO:0031023	microtubule organizing center organization	10	1.4	1.1E-02
GO:0007033	vacuole organization	8	1.1	1.2E-02
GO:0065004	protein-DNA complex assembly	10	1.4	1.2E-02
GO:0010558	negative regulation of macromolecule biosynthetic process	31	4.3	1.3E-02
GO:0033554	cellular response to stress	24	3.3	1.3E-02
GO:0070201	regulation of establishment of protein localization	8	1.1	1.4E-02
GO:0046822	regulation of nucleocytoplasmic transport	8	1.1	1.4E-02
GO:0016050	vesicle organization	8	1.1	1.4E-02

GO:0051223	regulation of protein transport	8	1.1	1.4E-02
GO:0006333	chromatin assembly or disassembly	11	1.5	1.4E-02
GO:0043623	cellular protein complex assembly	14	1.9	1.4E-02
GO:0000288	nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay	4	0.6	1.4E-02
GO:0010256	endomembrane organization	4	0.6	1.4E-02
GO:0070828	heterochromatin organization	4	0.6	1.4E-02
GO:0031327	negative regulation of cellular biosynthetic process	31	4.3	1.5E-02
GO:0009890	negative regulation of biosynthetic process	31	4.3	1.5E-02
GO:0032386	regulation of intracellular transport	8	1.1	1.6E-02
GO:0009060	aerobic respiration	8	1.1	1.6E-02
GO:0008105	asymmetric protein localization	8	1.1	1.6E-02
GO:0009952	anterior/posterior pattern formation	23	3.2	1.6E-02
GO:0007126	meiosis	26	3.6	1.6E-02
GO:0051327	M phase of meiotic cell cycle	26	3.6	1.6E-02
GO:0009894	regulation of catabolic process	6	0.8	1.8E-02
GO:0006353	transcription termination	3	0.4	1.8E-02
GO:0000018	regulation of DNA recombination	3	0.4	1.8E-02
GO:0051567	histone H3-K9 methylation	3	0.4	1.8E-02
GO:0007611	learning or memory	13	1.8	1.8E-02
GO:0006084	acetyl-CoA metabolic process	8	1.1	1.9E-02
GO:0043414	biopolymer methylation	9	1.3	1.9E-02
GO:0007306	eggshell chorion assembly	12	1.7	2.0E-02
GO:0006457	protein folding	17	2.4	2.0E-02
GO:0006886	intracellular protein transport	23	3.2	2.1E-02
GO:0006350	transcription	50	6.9	2.1E-02
GO:0007293	germarium-derived egg chamber formation	11	1.5	2.7E-02
GO:0030163	protein catabolic process	25	3.5	2.8E-02
GO:0032880	regulation of protein localization	8	1.1	2.9E-02
GO:0042306	regulation of protein import into nucleus	7	1.0	2.9E-02
GO:0022416	bristle development	10	1.4	2.9E-02
GO:0051321	meiotic cell cycle	26	3.6	3.0E-02
GO:0045879	negative regulation of smoothened signaling pathway	4	0.6	3.0E-02
GO:0042176	regulation of protein catabolic process	4	0.6	3.0E-02
GO:0034968	histone lysine methylation	4	0.6	3.0E-02
GO:0051235	maintenance of location	9	1.3	3.1E-02
GO:0007294	germarium-derived oocyte fate determination	6	0.8	3.2E-02
GO:0006997	nucleus organization	8	1.1	3.3E-02
GO:0010033	response to organic substance	17	2.4	3.3E-02
GO:0033157	regulation of intracellular protein transport	7	1.0	3.3E-02

GO:0051603	proteolysis involved in cellular protein catabolic process	23	3.2	3.4E-02
GO:0044257	cellular protein catabolic process	23	3.2	3.4E-02
GO:0050792	regulation of viral reproduction	3	0.4	3.4E-02
GO:0035087	RNA interference, siRNA loading onto RISC	3	0.4	3.4E-02
GO:0007278	pole cell fate determination	3	0.4	3.4E-02
GO:0048525	negative regulation of viral reproduction	3	0.4	3.4E-02
GO:0045071	negative regulation of viral genome replication	3	0.4	3.4E-02
GO:0031468	nuclear envelope reassembly	3	0.4	3.4E-02
GO:0045069	regulation of viral genome replication	3	0.4	3.4E-02
GO:0002168	instar larval development	9	1.3	3.5E-02
GO:0006351	transcription, DNA-dependent	18	2.5	3.5E-02
GO:0003002	regionalization	48	6.7	3.8E-02
GO:0006730	one-carbon metabolic process	12	1.7	3.8E-02
GO:0048515	spermatid differentiation	12	1.7	3.8E-02
GO:0030706	germarium-derived oocyte differentiation	6	0.8	3.8E-02
GO:0007279	pole cell formation	6	0.8	3.8E-02
GO:0030716	oocyte fate determination	6	0.8	3.8E-02
GO:0006099	tricarboxylic acid cycle	7	1.0	3.8E-02
GO:0051028	mRNA transport	7	1.0	3.8E-02
GO:0046356	acetyl-CoA catabolic process	7	1.0	3.8E-02
GO:0006342	chromatin silencing	15	2.1	3.9E-02
GO:0045814	negative regulation of gene expression, epigenetic	15	2.1	3.9E-02
GO:0008356	asymmetric cell division	10	1.4	3.9E-02
GO:0048232	male gamete generation	20	2.8	4.0E-02
GO:0007283	spermatogenesis	20	2.8	4.0E-02
GO:0030727	germarium-derived female germ-line cyst formation	4	0.6	4.0E-02
GO:0000059	protein import into nucleus, docking	5	0.7	4.2E-02
GO:0008103	oocyte microtubule cytoskeleton polarization	5	0.7	4.2E-02
GO:0016072	rRNA metabolic process	8	1.1	4.2E-02
GO:0032774	RNA biosynthetic process	18	2.5	4.2E-02
GO:0009109	coenzyme catabolic process	7	1.0	4.4E-02
GO:0032507	maintenance of protein location in cell	7	1.0	4.4E-02
GO:0051225	spindle assembly	7	1.0	4.4E-02
GO:0017145	stem cell division	9	1.3	4.8E-02
GO:0051187	cofactor catabolic process	7	1.0	5.0E-02
GO:0030717	karyosome formation	5	0.7	5.0E-02
GO:0051052	regulation of DNA metabolic process	5	0.7	5.0E-02
GO:0008154	actin polymerization or depolymerization	4	0.6	5.2E-02
GO:0048135	female germ-line cyst formation	4	0.6	5.2E-02
GO:0051297	centrosome organization	8	1.1	5.2E-02

GO:0006271	DNA strand elongation during DNA replication	3	0.4	5.4E-02
GO:0000289	nuclear-transcribed mRNA poly(A) tail shortening	3	0.4	5.4E-02
GO:0022616	DNA strand elongation	3	0.4	5.4E-02
GO:0048488	synaptic vesicle endocytosis	7	1.0	5.6E-02
GO:0051651	maintenance of location in cell	7	1.0	5.6E-02
GO:0033365	protein localization in organelle	13	1.8	5.7E-02
GO:0030707	ovarian follicle cell development	24	3.3	5.7E-02
GO:0034728	nucleosome organization	8	1.1	5.8E-02
GO:0007389	pattern specification process	49	6.8	5.9E-02
GO:0043543	protein amino acid acylation	5	0.7	6.0E-02
GO:0007143	female meiosis	10	1.4	6.1E-02
GO:0009891	positive regulation of biosynthetic process	19	2.6	6.2E-02
GO:0031328	positive regulation of cellular biosynthetic process	19	2.6	6.2E-02
GO:0006606	protein import into nucleus	7	1.0	6.3E-02
GO:0046843	dorsal appendage formation	7	1.0	6.3E-02
GO:0043066	negative regulation of apoptosis	8	1.1	6.4E-02
GO:0007127	meiosis I	8	1.1	6.4E-02
GO:0022411	cellular component disassembly	6	0.8	6.8E-02
GO:0051726	regulation of cell cycle	20	2.8	7.0E-02
GO:0051258	protein polymerization	5	0.7	7.0E-02
GO:0007304	chorion-containing eggshell formation	13	1.8	7.0E-02
GO:0051170	nuclear import	7	1.0	7.1E-02
GO:0045185	maintenance of protein location	7	1.0	7.1E-02
GO:0006461	protein complex assembly	20	2.8	7.3E-02
GO:0070271	protein complex biogenesis	20	2.8	7.3E-02
GO:0007286	spermatid development	11	1.5	7.4E-02
GO:0030703	eggshell formation	13	1.8	7.5E-02
GO:0006268	DNA unwinding during replication	3	0.4	7.7E-02
GO:0007312	oocyte nucleus migration during oocyte axis specification	3	0.4	7.7E-02
GO:0006998	nuclear envelope organization	3	0.4	7.7E-02
GO:0008407	bristle morphogenesis	8	1.1	7.8E-02
GO:0019941	modification-dependent protein catabolic process	20	2.8	8.1E-02
GO:0051172	negative regulation of nitrogen compound metabolic process	25	3.5	8.1E-02
GO:0045934	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	25	3.5	8.1E-02
GO:0045132	meiotic chromosome segregation	9	1.3	8.2E-02
GO:0043632	modification-dependent macromolecule catabolic process	20	2.8	8.4E-02
GO:0006897	endocytosis	28	3.9	8.8E-02
GO:0010324	membrane invagination	28	3.9	8.8E-02

GO:0010557	positive regulation of macromolecule biosynthetic process	16	2.2	9.0E-02
GO:0051253	negative regulation of RNA metabolic process	22	3.1	9.0E-02
GO:0002164	larval development	10	1.4	9.0E-02
GO:0033043	regulation of organelle organization	14	1.9	9.3E-02
GO:0006901	vesicle coating	4	0.6	9.5E-02
GO:0016183	synaptic vesicle coating	4	0.6	9.5E-02
GO:0006364	rRNA processing	7	1.0	9.6E-02
GO:0009950	dorsal/ventral axis specification	10	1.4	9.7E-02

Supplementary table S8: INTERPRO analysis (Drosophila melanogaster embryos)

INTERPRO ID	*Category*	*Count*	*%*	*p-value*
IPR012677	Nucleotide-binding, alpha-beta plait	56	7.78	3.7E-30
IPR000504	RNA recognition motif, RNP-1	57	7.92	8.4E-30
IPR001650	DNA/RNA helicase, C-terminal	31	4.31	2.0E-16
IPR012340	Nucleic acid-binding, OB-fold	24	3.33	3.4E-16
IPR014001	DEAD-like helicase, N-terminal	30	4.17	1.3E-15
IPR014021	Helicase, superfamily 1 and 2, ATP-binding	29	4.03	5.4E-15
IPR004087	K Homology	15	2.08	9.7E-10
IPR004088	K Homology, type 1	12	1.67	7.2E-08
IPR011545	DNA/RNA helicase, DEAD/DEAH box type, N-terminal	15	2.08	4.1E-07
IPR003100	Argonaute and Dicer protein, PAZ	7	0.97	5.8E-07
IPR002464	DNA/RNA helicase, ATP-dependent, DEAH-box type, conserved site	9	1.25	5.5E-06
IPR018111	K Homology, type 1, subgroup	9	1.25	9.3E-06
IPR018525	DNA-dependent ATPase MCM, conserved site	6	0.83	2.5E-05
IPR001208	DNA-dependent ATPase MCM	6	0.83	2.5E-05
IPR003165	Stem cell self-renewal protein Piwi	5	0.69	9.5E-05
IPR011989	Armadillo-like helical	15	2.08	9.5E-05
IPR000330	SNF2-related	8	1.11	9.7E-05
IPR000795	Protein synthesis factor, GTP-binding	8	1.11	1.5E-04
IPR001159	Double-stranded RNA binding	7	0.97	1.7E-04
IPR014720	Double-stranded RNA-binding-like	6	0.83	4.4E-04
IPR000571	Zinc finger, CCCH-type	8	1.11	5.6E-04
IPR004161	Translation elongation factor EFTu/EF1A, domain 2	7	0.97	5.8E-04
IPR002553	Clathrin/coatomer adaptor, adaptin-like, N-terminal	5	0.69	5.9E-04
IPR014014	RNA helicase, DEAD-box type, Q motif	9	1.25	9.7E-04
IPR001353	Proteasome, subunit alpha/beta	8	1.11	9.7E-04
IPR002423	Chaperonin Cpn60/TCP-1	6	0.83	1.1E-03
IPR017998	Chaperone, tailless complex polypeptide 1	5	0.69	1.1E-03
IPR011129	Cold shock protein	4	0.56	1.1E-03
IPR002059	Cold-shock protein, DNA-binding	4	0.56	1.1E-03
IPR000999	Ribonuclease III	4	0.56	1.1E-03
IPR002194	Chaperonin TCP-1, conserved site	5	0.69	1.9E-03
IPR005824	KOW	4	0.56	2.7E-03
IPR019775	WD40 repeat, conserved site	19	2.64	3.0E-03
IPR019781	WD40 repeat, subgroup	21	2.92	3.2E-03
IPR019782	WD40 repeat 2	20	2.78	3.4E-03
IPR008271	Serine/threonine protein kinase, active site	21	2.92	3.8E-03
IPR004038	Ribosomal protein L7Ae/L30e/S12e/Gadd45	4	0.56	5.1E-03
IPR017986	WD40 repeat, region	20	2.78	5.3E-03
IPR015943	WD40/YVTN repeat-like	24	3.33	5.7E-03
IPR000629	RNA helicase, ATP-dependent, DEAD-box, conserved site	7	0.97	6.4E-03
IPR001680	WD40 repeat	23	3.19	6.9E-03
IPR003594	ATP-binding region, ATPase-like	4	0.56	8.5E-03
IPR004365	Nucleic acid binding, OB-fold, tRNA/helicase-type	4	0.56	8.5E-03
IPR004160	Translation elongation factor EFTu/EF1A, C-terminal	4	0.56	8.5E-03

IPR002343	Paraneoplastic encephalomyelitis antigen	4	0.56	8.5E-03
IPR003395	RecF/RecN/SMC protein, N-terminal	4	0.56	8.5E-03
IPR017442	Serine/threonine protein kinase-related	21	2.92	8.6E-03
IPR000953	Chromo domain	6	0.83	1.1E-02
IPR002917	GTP-binding protein, HSR1-related	5	0.69	1.2E-02
IPR007502	Helicase-associated region	5	0.69	1.2E-02
IPR014717	Translation elongation factor EF1B/ribosomal protein S6	3	0.42	1.3E-02
IPR007647	RNA polymerase Rpb2, domain 5	3	0.42	1.3E-02
IPR015712	DNA-directed RNA polymerase, subunit 2	3	0.42	1.3E-02
IPR007641	RNA polymerase Rpb2, domain 7	3	0.42	1.3E-02
IPR006548	Splicing factor ELAV/HuD	3	0.42	1.3E-02
IPR007120	DNA-directed RNA polymerase, subunit 2, domain 6	3	0.42	1.3E-02
IPR007121	RNA polymerase, beta subunit, conserved site	3	0.42	1.3E-02
IPR007645	RNA polymerase Rpb2, domain 3	3	0.42	1.3E-02
IPR007644	RNA polymerase, beta subunit, protrusion	3	0.42	1.3E-02
IPR007642	RNA polymerase Rpb2, domain 2	3	0.42	1.3E-02
IPR019844	Cold-shock conserved site	3	0.42	1.3E-02
IPR006133	DNA-directed DNA polymerase, family B, exonuclease	3	0.42	1.3E-02
IPR019821	Kinesin, motor region, conserved site	6	0.83	1.3E-02
IPR001494	Importin-beta, N-terminal	5	0.69	1.5E-02
IPR002290	Serine/threonine protein kinase	15	2.08	1.7E-02
IPR017441	Protein kinase, ATP binding site	21	2.92	2.0E-02
IPR004095	TGS	3	0.42	2.5E-02
IPR015928	Aconitase/3-isopropylmalate dehydratase, swivel	3	0.42	2.5E-02
IPR001030	Aconitase/3-isopropylmalate dehydratase large subunit, alpha/beta/alpha	3	0.42	2.5E-02
IPR018136	Aconitase family, 4Fe-4S cluster binding site	3	0.42	2.5E-02
IPR000235	Ribosomal protein S7	3	0.42	2.5E-02
IPR002942	RNA-binding S4	3	0.42	2.5E-02
IPR015931	Aconitase/3-isopropylmalate dehydratase large subunit, alpha/beta/alpha, su	3	0.42	2.5E-02
IPR015932	Aconitase/3-isopropylmalate dehydratase large subunit, alpha/beta/alpha, su	3	0.42	2.5E-02
IPR015937	Aconitase-like core	3	0.42	2.5E-02
IPR000573	Aconitase A/isopropylmalate dehydratase small subunit, swivel	3	0.42	2.5E-02
IPR006134	DNA-directed DNA polymerase, family B, conserved region	3	0.42	2.5E-02
IPR006172	DNA-directed DNA polymerase, family B	3	0.42	2.5E-02
IPR006073	GTP1/OBG	4	0.56	2.5E-02
IPR001878	Zinc finger, CCHC-type	6	0.83	3.1E-02
IPR001752	Kinesin, motor region	6	0.83	3.1E-02
IPR000426	Proteasome, alpha-subunit, conserved site	4	0.56	3.3E-02
IPR003307	eIF4-gamma/eIF5/eIF2-epsilon	3	0.42	3.9E-02
IPR006630	RNA-binding protein Lupus La	3	0.42	3.9E-02
IPR002452	Alpha tubulin	3	0.42	3.9E-02
IPR002562	3'-5' exonuclease	3	0.42	3.9E-02
IPR016102	Succinyl-CoA synthetase-like	3	0.42	3.9E-02
IPR016157	Cullin, conserved site	3	0.42	3.9E-02
IPR005811	ATP-citrate lyase/succinyl-CoA ligase	3	0.42	3.9E-02
IPR016050	Proteasome, beta-type subunit, conserved site	4	0.56	4.2E-02

IPR011991	Winged helix repressor DNA-binding	11	1.53	4.7E-02
IPR014721	Ribosomal protein S5 domain 2-type fold	4	0.56	5.2E-02
IPR011709	Region of unknown function DUF1605	4	0.56	5.2E-02
IPR019559	Cullin protein, neddylation domain	3	0.42	5.6E-02
IPR006195	Aminoacyl-tRNA synthetase, class II, conserved region	4	0.56	6.3E-02
IPR000717	Proteasome component region PCI	4	0.56	6.3E-02
IPR000357	HEAT	6	0.83	6.7E-02
IPR001373	Cullin, N-terminal	3	0.42	7.5E-02
IPR018355	SPla/RYanodine receptor subgroup	3	0.42	7.5E-02
IPR016158	Cullin homology	3	0.42	7.5E-02
IPR013017	NHL repeat, subgroup	3	0.42	7.5E-02
IPR003029	Ribosomal protein S1, RNA binding domain	3	0.42	7.5E-02
IPR000910	High mobility group, HMG1/HMG2	5	0.69	8.2E-02
IPR000719	Protein kinase, core	21	2.92	8.8E-02
IPR016021	MIF4-like, type 1/2/3	3	0.42	9.6E-02
IPR009818	Ataxin-2, C-terminal	3	0.42	9.6E-02
IPR000315	Zinc finger, B-box	3	0.42	9.6E-02