Functional characterization of Pseudouridine synthase I and Y-box-binding protein 3

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
Alexandra Iulia Vasile

Submitted 19.11.2015

1st Reviewer: Dr. Markus Landthaler
Berlin Institute for Medical Systems Biology (BIMSB)
Max Delbrück Center for Molecular Medicine (MDC) in the Helmholtz Association
D-13125 Berlin, Germany
Tel.: +49 30 9406 3026
Email: markus.landthaler@mdc-berlin.de

2nd Reviewer: Prof. Dr. Udo Heinemann
Max Delbrück Center for Molecular Medicine (MDC) in the Helmholtz Association
D-13125 Berlin, Germany
and
Chemistry and Biochemistry Institute,
Freie Universität Berlin,
D-14195 Berlin, Germany
Tel.: +49 30 9406-3420
Email: heinemann@mdc-berlin.de

Date of defense: 18.03.2016

ACKNOWLEDGEMENTS

I would like first to thank my supervisor, Dr. Markus Landthaler for the opportunity to pursue my scientific interests in the form of a doctoral thesis in his group and for the supervision of my doctoral projects. The challenges and valuable discussions we had have made my PhD journey a very interesting experience from which I am very grateful to be able to learn and benefit from in my next career steps.
I would like to thank my Freie Universität supervisor, Prof. Dr. Udo Heinemann, for his examination of the thesis and his kindly advices for finishing my PhD thesis.

I would have to thank my dear colleagues, Kerstin, Lea, Mathias, and Yasuhiro with whom I have shared my challenging journey and who supported and helped me in every possible way. An important acknowledgement goes to Miha and Emanuel who patiently and critically reviewed my thesis and helped with valuable comments and discussions as well as performing important parts of the bioinformatic analysis for the two projects. Also, I would like to thank Dr. Guido Mastrobuoni for performing the mass-spectrometry and Prof. Dr. Wei Chen for the sequencing platform.

I am forever grateful to all my friends who have infinitely listened, supported and encouraged me to overcome every obstacle and look forward to positive outcomes. I thank Salah, Alex, Andreas, Kate, Elena, Anja, Branko, Anna, and many more friends for never giving up believing in me and for sharing great times during my stay here.

Last but not least, I would like to thank my family who has been the best moral support and source of encouragement anyone could possibly wish for. Without the continous and precious conversations with my mother I would have never come so far in pursuing my dreams which makes me endlessly grateful for everything she has offered me.

STATEMENT OF CONTRIBUTIONS

The work presented here is the result of two projects in collaborations with members of our institute.

Within the first project, the study of Pseudouridine synthase I, the experiments were designed and performed by myself and supervised by Dr. Markus Landthaler. The PAR-CLIP pipeline analysis was performed by Dr. Markus Schüler supervised by Dr. Christoph Dieterich (BIMSB, MDC, currently Max Planck Institute for Biology of Ageing). The computational analysis of PUS1-ψ-seq overlap was performed by Dr. Altuna Akalin.

Within the second project, the study of Y-box binding protein 3, the experiments were designed and performed by myself and supervised by Dr. Markus Landthaler. YBX3 PAR-CLIP pipeline analysis was performed by Dr. Christoph Dieterich and by Dr. Miha Milek. YBX3 motif analysis was performed by Dr. Emanuel Wyler, and YBX3 proteomics analysis was performed by Dr. Miha Milek. Dr. Guido Mastrobuoni from the lab of Dr. Stefan Kempa carried out mass-spectrometry run and ran MaxQuant. Sequencing runs were carried out by Mirjam Feldkamp and Claudia Langnick from the laboratory of Prof. Dr. Wei Chen (BIMSB, MDC).

ABBREVIATIONS

ψ	Pseudouridine
4SU	4-thiouridine
6SG	6-thioguanosine
3’ UTR	3’ untranslated region
5’ UTR	5’ unstranslated region
Ala/Pro	Alanine/Proline
Asp	Aspartic acid
CCHC	CysCysHisCys type zinc finger domain
cDNA	Complementary DNA
CDS	Coding DNA Sequence
CLIP	Crosslinking and immunoprecipitation
circRNA	Circular RNA
CMCT	N-Cyclohexyl-N’-(2-morpholinoethyl)carbodiimide methyl-p toluensulfonate
CRAC	Crosslinking and analysis of cDNAs
CSD	Cold shock domain
DC	Dyskeratosis congenita
dsRBD	Double-stranded RNA binding domain
ESC	Embryonic stem cell
FPKM	Fragments per kilobase of transcript per million mapped reads
HITS-CLIP	High-throughput sequencing CLIP
iCLIP	Individual nucleotide resolution CLIP
KH	K homology domain
LPS	Lipopolysaccharide
mRNP	Messenger RNA ribonucleoprotein
ncRNA	Non-coding RNA
NMR spectroscopy	Nuclear Magnetic Resonance spectroscopy
PAR-CLIP	Photoactivatable-ribonucleoside-enhanced CLIP
Psi	Pseudouridine
PsiU domain	Pseudouridine domain
pSILAC	Pulsed Stable Isotope Labelling with Amino Acids in Cell Culture
PTGR	Post-transcriptional gene regulation
PTRO	Post-transcriptional RNA operons
PUM-HD	Pumilio homology domain
RBP	RNA binding protein
RBPome	RNA binding proteome
RIP-chip	RNA immunoprecipitation and microarray analysis
RIP-seq	RNA immunoprecipitation and high-throughput sequencing

RNP1/2	Ribonucleoprotein RNA binding motif 1/2
RRM	RNA recognition motif
RPKM	Reads per kilobase of transcript per million mapped reads
RT	Reverse transcription
SELEX	Systematic evolution of ligands by exponential enrichment
TF	Transcription factor

INTRODUCTION

Post-transcriptional gene regulation during mRNA life cycle

Post-transcriptional gene regulation is a dynamic process describing the life cycle of an mRNA molecule from transcription, processing, nucleocytoplasmic transport, and localization, to translation and decay. This sequence of events is tightly regulated and interconnected at all levels by different classes of RNA-binding proteins, microRNAs and ncRNAs (Fabian et al., 2010; Gerstberger et al., 2014; Komili and Silver, 2008). The coupling mechanisms that allow the coordination between transcription and the co-transcriptional processing events (Figure 1) as capping, splicing, 3'-end processing and RNA editing include functionally related regulatory proteins to drive these processes in a organized manner (Bentley, 2014; Kanitz and Gerber, 2010; Müller-McNicoll and Neugebauer, 2013). The messenger RNA ribonucleoprotein (mRNP) complexes coordinate multiple mRNAs and constantly remodel in order to assist the mRNAs in the following post-transcriptional steps (Keene, 2007; Lunde et al., 2007). Moreover,

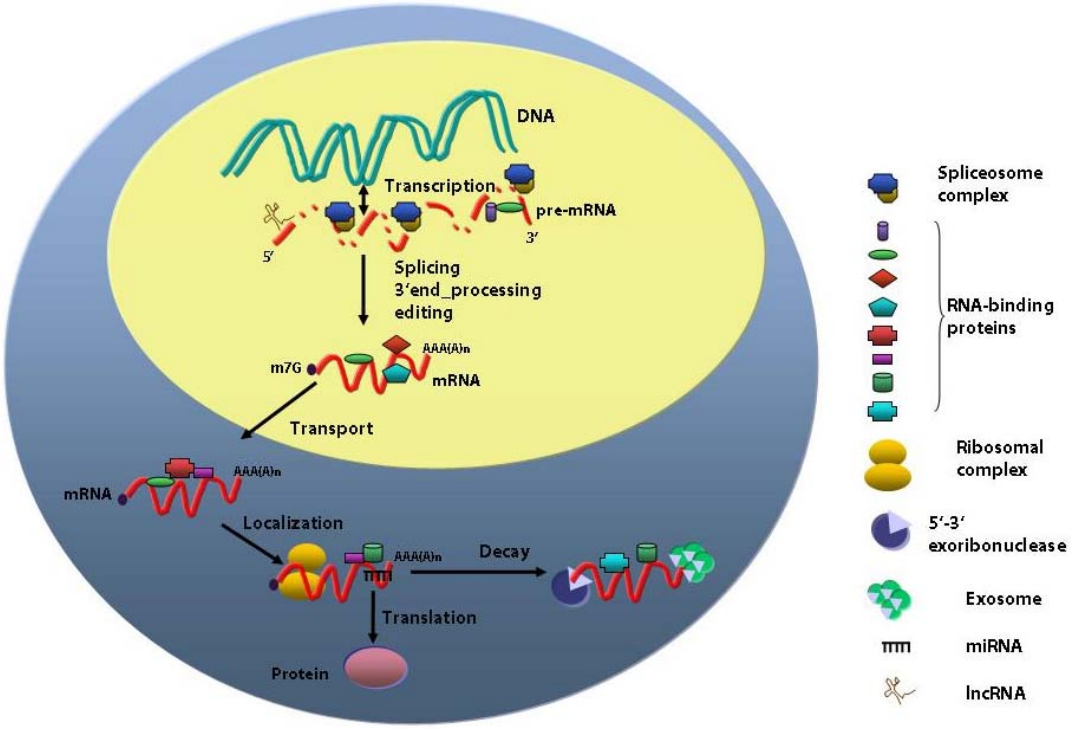


Figure 1. Cellular life-cycle of the mRNA. The complex life from birth to death of eukaryotic mRNA is a highly dynamic process which starts from transcription in the nucleus to translation and decay in the cytoplasm. The co-transcriptional processing involves already complex mixture of RNA-binding

proteins, involving cap-binding proteins, splicing complexes, RNA editing, cleavage and polyadenylation proteins besides the myriad of transcription elongation factors. Post-transcriptionally, RBPs are involved in the formation of ribonucleoprotein complexes or RNPs which escort the messenger RNAs throughout their complex lifetimes including cytoplasmic translocation, localization, translation and degradation by exoribonucleases and exosomes. These processes are highly interconnected and dynamically regulated by constantly remodeling RNPs that function coordinately in each step of the gene expression regulation. Besides RBPs, long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) participate in post-transcriptional regulatory mechanisms to drive gene expression specifically in each cellular context *(Moore, 2005)*.

a model in which RNA decay takes place co-translationally has been illustrated in a new study (Pelechano et al., 2015). In this model, a 5'-3' transcript degradation follows the last translating ribosome and is believed to affect most of the genes.

Post-transcriptional gene regulation (PTGR) has been suggested as a major influence factor of protein abundance in several studies where protein steady-state levels were found to vary from 20 to 30 fold (Greenbaum et al., 2003; Gygi et al., 1999). In addition, when mRNA levels were compared to protein copy numbers, the coefficient of determination (R^2) was 0.41, which means that 40% of the variance in protein levels could be explained by mRNA levels. These data suggest that mRNA stability has a smaller role compared to the transcriptional rates in controlling protein levels. When accounting for the translational rate constants the correlation coefficient was 0.95, suggesting that the translation rates have the dominant role for the protein levels control, while the protein degradation influence was rather small. (Schwanhausser et al., 2011).

RNA-binding proteins and their roles in PTGR

A recent review presented a census of 1,542 manually curated RBPs analyzed for their interactions with different classes of RNA, their evolutionary conservation, their abundance and their tissue-specific expression patterns (Gerstberger et al., 2014). The majority of these RBPs were ubiquitously expressed, typically at higher levels than average cellular proteins. Although RBPs and TFs are encoded by a similar number of genes (1,542 and 1,704 genes, respectively), the cumulative abundance of RPKM expression levels of RBPs contributed up to 20% of the expressed, protein-coding transcriptome, whereas TFs constituted only up to 3% by transcript abundance. The post-transcriptional regulatory code has been reevaluated by the most recent studies, which quantified transcript abundance and kinetic rates at the splice junction level in

pre-mRNA and mRNA, suggesting that only one third of genes are subjected to a dynamic post-transcriptional regulation while two thirds use transcriptional regulatory strategies in gene expression patterns (Rabani et al., 2014). Interestingly, pulsed SILAC-based mass spectrometry and LPS (lipopolysaccharide) stimulation of dendritic cells have revealed a new model where dynamic cellular responses are driven by changes in transcription rates while protein production or degradation rates are responsible for the change in the preexisting proteome comprised of proteins with basic cellular functions (Jovanovic et al., 2015).

Individually, RNA-binding proteins have been studied in high-throughput methods aim at characterizing protein-RNA interactions (Cook et al., 2015; Lunde et al., 2007). The systematic analysis of post-transcriptional regulation has not only revealed the genome-wide mRNAs associated with a particular RBP but also how these functionally related mRNAs can be associated in PTROs (post-transcriptional RNA operons) (Morris et al., 2010). It is widely accepted that the dynamic association of RBPs with RNA molecules influences their processing, localization, half-life and translation rate. One aspect of research has been to uncover the combinatorial effects of different functional protein domains and RNA *cis*-acting elements in selecting specifically their targets but having as well a diverse range of substrates (Lunde et al., 2007). The study of the most common RNA-binding domains like RRM (RNA recognition motif), KH domain (K homology), dsRBD (double-stranded RNA binding domain), or PUM-HD (Pumilio homology domain) have contributed to the understanding of RNA-protein interactions (Cook et al., 2015). However, knowing the structure and specificity of an RNA binding protein could not fully evoke the entire landscape of an RBPome. Therefore, methods to study the protein RNA-interactions have been developed throughout the time and comprised both *in vitro* and *in vivo* approaches. While SELEX (systematic evolution of ligands by exponential enrichment) (Ellington and Szostak, 1990; Tuerk and Gold, 1990), RNA compete (Ray et al., 2009; Ray et al., 2013) and RNA Bind-n-Seq (Lambert et al., 2014) determine the specificity of RBPs in a cell-free system thus escaping the role of other interacting factors or cellular conditions, *in vivo* methods have tried to overcome these biases. The early approaches RIP-chip (RNA immunoprecipitation and microarray analysis) (Tenenbaum et al., 2000) and RIP-seq (RIP combined with high-throughput sequencing) can identify the RNAs bound to immunoprecipitated RBPs but do not indicate the exact binding sites on their target RNAs and are highly contaminated with unspecific bound RNAs. Thus, crosslinking the proteins to the RNA by UV irradiation prior to immunoprecipitation would “freeze” in place the RNA-protein interactions through covalent bonds and could capture *in vivo* biological or functional interactions that occur

between RNA and RBPs within a few Ångstroms distance, permitting, in addition, a more stringent washing step in the immunoprecipitation protocol (Hockensmith et al., 1986). CLIP (crosslinking and immunoprecipitation) methods have advanced the study of protein-RNA interactions especially after the inclusion of high-throughput sequencing (HITS-CLIP, iCLIP, CRAC) (Milek et al., 2012) where the binding site could be identified at near-nucleotide resolution due to an error of the reverse-transcriptase at the crosslinked residue (HITS-CLIP) or due to a termination in reverse transcription (iCLIP) (Konig et al., 2010; Zhang and Darnell, 2011). To further improve the sensitivity of this method and to determine the exact precise location of the "amino acid tag" which crosslinks to the RNA, a photoactivatable nucleoside analogue was introduced such as 4-thiouridine (4SU) or 6-thioguanosine (6SG) (Hafner et al., 2010b). In addition to labeling of cell culture *in vivo* with 4SU or 6SG, PAR-CLIP (photoactivatable ribonucleoside enhanced crosslinking and immunoprecipitation) employs UV crosslinking at 365nm of the living cells or organisms which enables a more efficient protein-RNA crosslink between protein and ribonucleoside (Hafner et al., 2010b). The analysis of the sequencing reads obtained from PAR-CLIP take advantage of the T-to-C transition during the reverse transcription of the crosslinked uridine permitting the identification of the RNA-binding site at nucleotide resolution. Although several biases have been reported including the enrichment of RNA targets containing U-rich motifs (Kishore et al., 2011), and the background fraction of reads resulting from non-specific crosslinking (Friedersdorf and Keene, 2014b), these biases could be overcome when using another modified residue such as 6SG or sequencing additional negative controls to account for the non-specific binding.

The study of gene expression regulation has not only focused on characterizing the RNA-binding sites of specific RNA-binding proteins but have also used complementary approaches to characterize the landscape of the RNA-bound proteome transcriptome-wide. The identification of around 800 RNA-binding proteins in human embryonic kidney cells (HEK293), HeLa cells or yeast and around 500 in mouse ESCs, using mass spectrometry and oligo(dT) pulldown, has contributed to the impressive RBPome and suggested the magnitude of post-transcriptional gene regulation (Baltz et al., 2012a; Castello et al., 2012a; Kwon et al., 2013; Mitchell et al., 2012). The reverse approach, profiling the protein occupancy transcriptome-wide by sequencing the oligo(dT) pulldown RNA fragments after UV crosslinking available cellular RBPs, would pinpoint the exact footprints of the RNA-binding proteins in a global manner (Baltz et al., 2012a). Interestingly, applying this method to different cell lines or different cellular conditions could reveal biologically functional patterns of RNA binding proteins by

analyzing their differential protein occupancy profile as for example between HEK293 and MCF7 cell lines (Munschauer et al., 2013; Schueler et al., 2014).

Combining different methods to study RNA-protein interactions not only at the individual level but also as a cumulative effect on RNA metabolism and global gene expression can ultimately generate a comprehensive data set that would integrate RNA processing, translation and stability into a complex image of the cellular mRNA life-cycle (Cook et al., 2014; Geisberg et al., 2014; Ingolia et al., 2011).

Post-transcriptional regulation by RNA editing and modification

One of the processing steps of pre-mRNA to mature mRNA is RNA editing (Gott and Emeson, 2000; Gray, 2012). RNA editing was first described in unicellular protozoa (*Trypanosoma brucei*) as an insertion of uridylate (U) residues in mitochondrial mRNA that were not encoded in the DNA sequence (Benne et al., 1986). Since then, RNA editing was identified in various forms present in all RNA species from unicellular protozoa to mammals, either as insertion or deletion of nucleotides, but also conversion from one nucleotide to another through deamination (Aphasizhev and Aphasizheva, 2011; Gerber and Keller, 2001). More recently other modifications in RNAs such as methylation (Dominissini et al., 2012; Meyer et al., 2012) or isomerization of uridine (U) to pseudouridine (ψ) have been described as important editing steps in mRNA processing (Carlile et al., 2014; Schwartz et al., 2014b; Zhao and He, 2015). The functional role of mRNA editing has been characterized in relation to recoding the genetic information from DNA or to alteration in RNA structure with potential effects on RNA translation. Editing of rRNA is important in preserving the architecture of the secondary structure for ribosome assembly and translation efficiency (Jack et al., 2011b). RNA editing has been demonstrated as a functional tool in tRNA tertiary structure formation and loss of tRNA editing, especially mitochondrial tRNA, is associated with several mitochondrial diseases (Umeda et al., 2005). Cytidine (C) to uridine (U) conversion in mRNA by the RNA editing enzyme APOBEC1 is able to change a sense codon into a stop codon leading to a truncated apoB protein that has a different function in the small intestine (Rosenberg et al., 2011). Cytidine deamination by APOBEC proteins is used also as an anti-viral defense mechanism (Refsland et al., 2010). Adenosine (A) to inosine (I) editing is probably one of the most investigated categories of RNA editing. A to I conversion has been described in mRNA as functionally important due to changes in RNA secondary structure by editing of Alu sequences, as well as changes in the amino acid sequence of the encoded proteins.

This could lead to different protein conformational changes or could have implications in translation fidelity, premature stop codon formation or splicing (Solomon et al., 2013) as well as circRNA biogenesis (Ivanov et al., 2015; Nishikura, 2009; Yeo et al., 2010).

Other types of RNA modifications involve the addition of different “group tags” in form of methylation or thiolation (Fu et al., 2013; Niu et al., 2013; Wang et al., 2011). Moreover RNA methylation has been recently highlighted by an increasing number of studies as a functionally important modification in mRNA with a frequency of one N6-methyladenosine per gene (Dominissini et al., 2012; Meyer et al., 2012). N6-methyladenosine can be recognized by enzymes that specifically add the “tag” (modifiers), read the “tag” (readers) and finally can remove the “tag” (erasers). METTL3 and METTL14 have the methylation activity, while FTO and ALBKH5 are the two demethylases capable of removing this group from adenosine residues (Jia et al., 2011; Liu et al., 2014; Zheng et al., 2013). In between methylation and demethylation of m6A there are two other newly described proteins, YTHDF1 and YTHDF2, which function as RNA methylation “readers”, either by actively promoting protein synthesis through interactions with translation machinery, or by reducing the stability of target transcripts, respectively (Wang et al., 2014; Wang et al., 2015). This complex regulation of a single RNA modification illustrates a dynamic and important function of RNA editing enzymes in the mRNA life cycle.

Described as the most frequent modification in RNA, pseudouridine (ψ), a C5-glycosyl isomer of uridine (U), was first observed in 1951 (Cohn and Volkin, 1951) but its identity was demonstrated later in 1957 (Davis and Allen, 1957). Among over 109 RNA nucleotide modifications present in all species of RNA (Machnicka et al., 2013), pseudouridine was until recently studied only in rRNA, tRNA and ncRNAs (Bykhovskaya et al., 2004b; Ofengand et al., 1995; Yu et al., 2011; Zhao et al., 2004). The coupling of old techniques for pseudouridine mapping together with next generation sequencing platforms has offered new insights into the widespread distribution of pseudouridine in mRNA, in unicellular eukaryotes as well as in humans (Carlile et al., 2014; Lovejoy et al., 2014; Schwartz et al., 2014b; Zhao and He, 2015). Global pseudouridine mapping across the transcriptome of HEK293 cells, used the method called “ψ-seq mapping” which is based on the CMCT chemical derivatization of RNA followed by a mild alkaline treatment (Bakin and Ofengand, 1993). This chemical “tagging” of pseudouridine coupled with reverse transcription and deep sequencing could reveal the putative pseudouridine positions as stops of reverse transcriptase at the CMC-ψ bulky adducts. The study above revealed several hundred pseudouridines present mostly in mRNA, but also tRNA and ncRNA. Due to the specific enrichment of polyA mRNA rRNA pseudouridylation sites could not be determined. Moreover, the

different ψ sites were found to be depended on different pseudouridine synthases expressed in HEK293 cells or yeast, either part of the H/ACA box snRNA guided complex or RNA-independent pseudouridine synthases. The most recent study focusing on the distribution of pseudouridines in RNA highlighted a possible new layer of post-transcriptional regulation through the presence of 2,084 Psi/ψ sites within 1,929 human transcripts (Li et al., 2015a). In this study, CeU-seq (CMC-enriched pseudouridine sequencing) was employed as a variation of CMCT derivatization coupled with specific enrichment of CMCT-adducts by biotin conjugation via click chemistry. Using mass spectrometry pseudouridylation was quantified to a ψ/U ratio of approximately 0.2-0.4%, which is in the same range as the m6A/A ratio in mRNA (0.1-0.5%) (Dominissini et al., 2012; Jia et al., 2011; Zheng et al., 2013). In addition, four new sites were experimentally validated with a CMCT-independent method, the site-specific cleavage and radioactive labeling followed by ligation-assisted extraction and thin-layer chromatography (SCARLET) (Liu et al., 2013a). An interesting characteristic of pseudouridylation in mRNA was highlighted in all three recent studies: this RNA modification is dynamically regulated and highly dependent on cellular conditions. Various cellular stresses such as heat shock, serum deprivation or oxidative stress were increasing or decreasing the level of pseudouridylation in mRNA. This suggests a functional role of pseudouridine in response to stimuli-specific stresses and might indicate the importance of such RNA modification in the post-transcriptional regulation. An unexpected aspect of this study was the finding that ψ-profiles from two different mouse tissues indicated a tissue-specific pseudouridylation at several sites, although the distribution pattern along mRNA was similar between the tissues, as well as, between mice and human (Li et al., 2015a).

Pseudouridine synthases structure and function

After the discovery and experimental detection of pseudouridine in RNA, studies starting three decades earlier have been able to identify the enzymes responsible for this type of modification (Johnson and Soll, 1970). In vitro studies have shown that pseudouridylation is also specific to particular RNAs and this is achieved by specific ψ synthases which do not require cofactors for catalysis (Cortese et al., 1974). Pseudouridine synthases from bacteria were classified into five distinct families (Arena et al., 1978) which are conserved from yeast to human (Green et al., 1982; Samuelsson and Olsson, 1990). The families took the names of the *E. coli* enzymes: RluA, RsuA, TruA, TruB and TruD and present three specific conserved motifs that are

part of the catalytic domain in RluA and RsuA families yet the TruB family shares only two of these motifs (Koonin, 1996). A sixth family was recently described for the pseudouridine synthase 10 (PUS10) (McCleverty et al., 2007), which is found only in Archaea and higher Eukaryotes and contains besides the catalytic PsiU domain a THUMP domain (THioUridine synthases, RNA Methyltransferases and Pseudouridine synthases) unique to PUS10 family (Becker et al., 1997; Roovers et al., 2006). PUS10 is able to specifically catalyze the pseudouridylation of ψ55 in tRNA (Gurha and Gupta, 2008; Kamalampeta et al., 2013). Although members of the different families do not share common sequences, all the pseudouridine synthases present a similar structure with a core containing an eight-stranded mixed β-sheet and a catalytically active site-cleft (Hamma and Ferre-D'Amare, 2006). The active site is characterized by a conserved consensus sequence with the most important residue being the catalytically essential aspartic acid with a function in the enzymatic mechanism of isomerization. The RNA-independent pseudouridine synthases recognize and flip the uridine base using a sequence-specific binding such as the TruB tRNA pseudouridine synthase, yet other pseudouridine synthases such as the TruA, is capable of recognizing structurally similar RNAs and appears to be the only pseudouridine synthase which functions as a dimer. The mechanism of RNA recognition and base flipping for the different pseudouridine synthases depends on the presence of two peripheral domains, a thumb-like protrusion and a forefinger-like loop which play an important role in RNA binding. Flipping out the uridine that undergoes isomerization is a rather complex process involving three conserved polar amino acids including the catalytic aspartate. This mechanism has been deduced from the observation of these enzymes interacting with a 5-fluorouridine (f^5U) that is able to inhibit the catalytic activity of some of the pseudouridine synthases such as TruA and RluA but not TruB (Gu et al., 1999). Gu and colleagues proposed that the catalytic Asp makes a Michael-type attack on the C6 glycosidic bond , allowing the rotation of the detached base along the ester bond with the amino acid (Gu et al., 1999). Another hypothesis that would support the surprising ability of the different Psi synthases to isomerize f^5U-containing RNA, propose that the catalytic Asp makes a nucleophilic attack on the anomeric position of the ribose, rather than attacking the nucleobase (Huang et al., 1998). The clearly defined mechanism that would validate the pseudouridylation process for all the pseudouridine synthases given their high structural similarities has not yet been elucidated.

In spite of the singular mechanism of pseudouridylation, characteristic to most of the pseudouridine synthases, one member of the TruB family, the Cbf5 in yeast, or the human homologue, Diskerin (DKC1), functions by association with a guide H/ACA box snoRNA and three other accessory proteins, Gar1, Nhp2 (L7Ae, in Archaea) and

Nop10. These proteins form a complex which is able to isomerize uridines in a site-directed manner. DKC1 has been reported to act as well in an RNA-independent manner (Tillault et al., 2015). The different H/ACA snoRNAs guide the Cbf5/DKC1 complex to a specific site which is complementary with the sequence in the snoRNA bulge of the double helices contained in the structure (Heiss et al., 1998). This RNP complex is the only one known to act as an RNA-dependent pseudouridine synthase. Moreover, it can associate with the RNA-telomerase component which contains a box H/ACA sequence and functions in maintaining telomerase stability (Tercjak-Recko et al.). Mutations in dyskerin have been detected in the X-linked form of the human bone marrow failure syndrome known as dyskeratosis congenita (DC) while mutations of the RNA component TERC, were involved in the autosomal form of DC (Vulliamy et al., 2001).

Biological functions of pseudouridine synthases

First of the evident roles of pseudouridine comes from the fact that the uridine isomer presents an extra hydrogen bond donor at the N1 position exposed to the major groove. This hydrogen bond donor may interact with water molecules or form hydrogen bonds in order to stabilize the RNA secondary and tertiary structure through stacking interactions (Davis, 1995). Pseudouridines are found in positions across specific RNAs that are conserved between species. These positions detected in spliceosomal U small nuclear RNAs, rRNAs or tRNAs are conserved due to their important roles in splicing, ribosome assembly and translation (Jack et al., 2011b; Lane et al., 1995; Yu et al., 2011). Lack of pseudouridylation in mitochondrial tRNAs is often associated with mitochondrial myopathies (Gutgsell et al., 2000; Patton et al., 2005b), while rRNA pseudouridylation defects may lead to nonsense translational suppression. Ribosomal ligand binding and translational fidelity can also be affected by the loss of pseudouridylation in the peptidyl-transferase and decoding centers of the ribosome (Ejby et al., 2007; Jack et al., 2011a). The pseudouridines found in spliceosomal small nuclear RNAs as for example the human U2 snRNA which contains 13 ψ sites, are located in functionally important regions and may contribute to pre-mRNA splicing (Adachi and Yu, 2014; Yu et al., 2011; Yu et al., 1998). U2 snRNA pseudouridylation at three positions were found to be necessary for splicing during the formation of the early spliceosomal E complex (Donmez et al., 2004). The human telomerase complex formed by a reverse transcriptase (TERT) and a non-coding telomerase RNA component (TER/TERC) is affected in patients with dyskeratosis congenita due to a

defect in the DKC1 H/ACA snoRNA guided pseudouridine synthase. TERC is pseudouridylated at several positions of which two of them are found in the highly conserved and functionally important P6.1 hairpin and L6.1 loop, inside the CR4-CR5 domain (Chen et al., 2002). Pseudouridylated P6.1 stem-loop adopts a different conformation and is significantly more stable than the unmodified structure (Kim et al., 2010). Three of the pseudouridylated sites were found at U306 and U307, and one ψ site at position U179 in the P3 stem of the pseudoknot (Schwartz et al., 2014a). Overall, pseudouridine presence in the P6.1 hairpin of TERC has experimentally proven to slightly attenuate telomerase activity but slightly increase its processivity *in vitro*, suggesting that these modifications could influence human telomerase activity via TERT-TERC or TERC-TERC interactions (Kim et al., 2010).

Although four new studies have now accomplished the transcriptome-wide identification of pseudouridines in mRNA, the ψ landscape has not yet been linked to a precise molecular and biological function. Nevertheless, the observation that stress and growth conditions leads to an increase in pseudouridylation across all species investigated, suggests an important role that this modification might play in the cellular metabolism. The four studies in the past year revealing that pseudouridylation is widespread across the transcriptome and moreover, that it can be modulated by environmental stimuli such as heat-shock, starvation or oxidative stress, highlights this post-transcriptional modification as one of the significant RNA editing processes (Carlile et al., 2014; Li et al., 2015a; Lovejoy et al., 2014; Schwartz et al., 2014b). The major finding of these studies revealed the presence of pseudouridine in mRNA in a larger quantity than in ncRNAs (Table 1). Reportedly, heat-shock and H_2O_2 exposure increased ψ levels by 40-50%. Knockdown experiments of each of the most known pseudouridine synthases were able to pin-point specific ψ sites to their respective modifying enzyme.

Table 1. Overview of the studies on the identification of pseudouridine sites across the transcriptome.

Study	Organism	Total Ψ sites	mRNA sites	Number of genes	ncRNA sites	PUS1 dependent sites
Pseudo-seq Carlile et al., Nature, 2014	yeast	416	260	238	156	71
	HeLa cells	**108**	**96**	89	12	-
Ψ-seq Schwartz et al., Cell, 2014	yeast	328	185	-	143	10
	HEK293 cells	**396**	**353**	-	43	**84***
PSI-seq Lovejoy et al., PLOS One, 2014	yeast rep#1	103	56	-	47	-
	yeast rep#2	335	150	-	185	-
CeU-seq Li et al., Nat.Chem.Biol., 2015	HEK293 cells	**2084**	**1889**	1357	195	**77**
	mouse brain	1741	1659	-	82	-
	mouse liver	1543	1490	-	53	-

*number of sites which were not associated with the other known pseudouridine synthases

With a more comprehensive overview of pseudouridine modification landscape, the research into the molecular functions of mRNA pseudouridylation is of uttermost importance. The question remains of how specificity is acquired among the different pseudouridine synthases and how that correlates with a function in post-transcriptional mRNA regulation such as regulation of splicing, translation, structure and stability as well as degradation.

Pseudouridine synthase I

Among the distinct classes of pseudouridine synthases including RNA-dependent and RNA-independent, pseudouridine synthase I (PUS1) is an important modifying enzyme found in yeast as well as in humans. Phylogenetically, PUS1 is the homolog of the *E-coli* pseudouridine synthase TruA, although the sequence identity between the two is only 24%. The mechanism of action seems to involve no specific sequence motif preference on the target RNAs in contrast to the other pseudouridine synthases, but the several conserved pseudouridylation positions in tRNAs such as U27 and U28 in the anticodon stem loop (ASL) (Hellmuth et al., 2000) and U44 in U2 snRNA (Adachi and Yu, 2014), suggests a recognition mode based on structural RNA elements. Another unique aspect of the human PUS1 in comparison to its bacterial homologue which functions as a dimer, is that both human and yeast PUS1 have been characterized as monomers in crystal structures (Arluison et al., 1999; Czudnochowski

et al., 2013). PUS1 has a similar structure with its bacterial homolog TruA, however, PUS1 contains in addition a unique C-terminal insert represented by the α8, α9 and α10 helices. This insert prevents the human PUS1 to dock the tRNA in the same orientation as observed for the *E-coli* TruA (Hur and Stroud, 2007). In contrast to TruA, the two helices α8 and α9 form a vast electropositive surface that accommodates a tRNA molecule in the opposite direction and also confers higher target selectivity than the other pseudouridine synthases. The catalytic domain of PUS1 consists of 10 anti-parallel β-sheets, four of which are found in the N-terminal region and six in the C-terminal end. The 10 helices flank the N-terminal and C-terminal β-sheet structure, with the unique C-terminal helices α8-α10 protruding outside of the active cleft (Figure 2A). A number of loops are decorating the site cleft region with two distinct structures called “forefinger loop” formed by L1-L2 and “thumb loop” formed by L6-L7, respectively. These two loops are also found in other pseudouridine synthases although not entirely conserved (Hamma and Ferre-D'Amare, 2006). The catalytically essential Asp118 is found in the center of the active site cleft, part of the L2-3 loops. There are five conserved amino acid residues in the active site of all pseudouridine synthases, Asp118, Tyr173, Arg267, Ile266, and Leu305, which participate in binding the substrate. Interestingly, mutagenesis of these residues showed that many of the substitutions maintained PUS1 catalytic activity to a certain degree. This translates into a reduced activity on the positions 27 and 28 on tRNA to a complete loss of activity for other positions or other RNAs (Sibert et al., 2008). The missense mutation of the universally conserved Arg116 to tryptophan (R116W) and the nonsense mutation of Glu220 (E220X) are the molecular basis of mitochondrial myopathy and sideroblastic anemia (Syndrome and Report), which is a rare autosomal recessive oxidative phosphorylation disorder localized to skeletal muscle and bone marrow (Bykhovskaya et al., 2004a; Fernandez-Vizarra et al., 2007; Patton et al., 2005a). Besides the mutation of the catalytically essential D118, which renders inactive the enzymatic activity of PUS1, mutations of the conserved R116, which is suppose to intercalate into the RNA stem to flip the uridine base, were not essential for the PUS1 activity on the tRNA U27 and U28, however it inhibited the modification at additional sites. R116W mutation affects PUS1 activity probably due to the steric or charge constraints imposed by the bulky tryptophan which cannot intercalate into the stem and therefore could interfere with the enzymatic activity (Sibert et al., 2008). Moreover, experiments showing that the active site residues, excluding the catalytic Asp118 and the Arg116, are not essential for the activity. However, these residues participate in the conformational change required to interact with the tRNA in an induced-fit substrate binding. Together with the forefinger loop, the thumb loop and the α8-α10 C-terminal helices, the human PUS1 catalytic

core is able to bind and orient a tRNA molecule in a different mode than its bacterial homologue TruA (Figure 2B). This is also due to the large positively charged area found in PUS1, which not only prevents oligomerization, but can direct the tRNA structure in an ~180° angle to the orientation of TruA-tRNA bound substrate, making contacts with the anticodon stem-loop and TψC-loop. Human PUS1, unlike its yeast homologue or its relative PUS10, does not need a Zn^{2+} ion as a cofactor, which indicates the different recognition and binding specificity of PUS1 for modifying different positions in the same RNA targets or for a broader RNA substrate specificity (Czudnochowski et al., 2013).

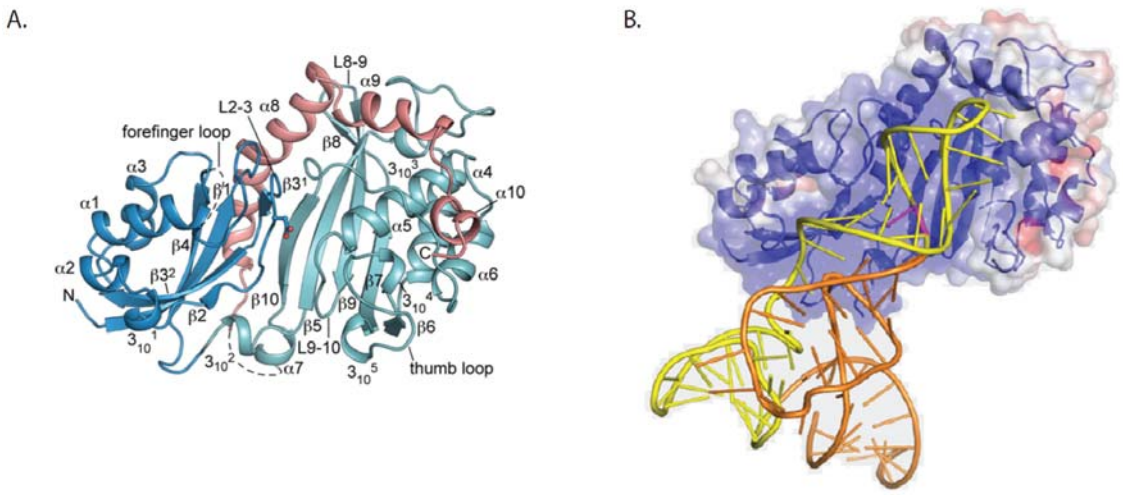


Figure 2. Structure of PUS1 catalytic domain. A. Representation of PUS1 crystal structure (rod form) with the N- and C-terminal domains (dark blue and green respectively), C-terminal extension helices α8-10 (salmon) and the catalytic Asp118 in ball-and-stick form, in the center of the cleft. (Cartoon representation) B. Stereo plot of $tRNA^{Phe}$ docked to apo-PUS1 form I (hexagonal plate). PUS1 structure surface is colored depending on its electrostatic potential. The tRNA substrate is colored orange with purple colored substrate base and yellow colored minimal substrate. (*Adapted from (Czudnochowski et al., 2013))*

PUS1 RNA-binding and biological function

Past findings on yeast correlated Pus1p activity with nuclear localization, multisubstrate activity on different RNA species such as snRNA and tRNA, and oligomerization in absence of a tRNA substrate (Massenet et al., 1999; Motorin et al., 1998; Simos et al., 1996). In contrast, the human PUS1 was identified as one of the mutated genes responsible for the mitochondrial myopathy and sideroblastic anemia, MLASA, a rare autosomal recessive disorder (Bykhovskaya et al., 2004a; Fallah, 2007).

Besides the missense mutation linked to this disorder (R116W), another nonsense mutation was later identified (E220X) in two patients, which can lead to a truncated protein. Resulting lack of pseudouridylation in mitochondrial and cytoplasmic tRNAs is restricted to muscles and erythroid cells, suggesting that compensatory ribosomal biogenesis pathways might prevent the phenotypic defect across all tissues (Bykhovskaya et al., 2007). In addition, the same study suggested that pseudouridylation of the non-tRNA targets might explain the affected oxidative phosphorylation and energy metabolism in muscle cells. Besides PUS1 mutations, MLASA has been more recently associated with mutations in the tyrosyl-tRNA synthetase (YARS2), leading to a decrease in mitochondrial protein synthesis and mitochondrial respiratory chain dysfunction (Riley et al., 2010; Shahni et al., 2013).

Despite the lack of studies for the human PUS1, several studies have focused on the mouse mPus1 pseudouridylation. One study of mouse mPus1, which shares 92% identity with human PUS1 (Chen and Patton, 1999) has shown that in comparison to the yeast scPus1p, mPus1 can modify an additional position in tRNA, but it does not have pseudouridylation activity on several other positions in the yeast tRNA (Behm-Ansmant et al., 2006). Moreover, both yeast and mouse pseudouridine synthase I were able to modify *in vitro* and *in vivo* the position 1 of cytoplasmic $tRNA^{Arg}$ (ACG). One study has shown that mPus1 can modify the long non-coding RNA steroid receptor co-activator (SRA), a co-activator of the retinoic acid receptor γ (mRARγ) (Zhao et al., 2004). Post-transcriptional modification of SRA at several positions in S91 melanoma cells, stimulate its proper folding and activate its function. The mPus1p-SRA protein-RNA complex is then able to bind to the DNA-binding domain of the nuclear receptor (NR) at the hormone response element (HRE) on the gene in a ligand-independent manner. In the presence of a ligand, mPus1p-SRA complex is able to enhance transcriptional activation of the nuclear receptor (Zhao et al., 2004). SRA activation by PUS1 pseudouridylation at position U206, in breast and prostate cancer cells, can be inhibited by a small hairpin structure, STR5, found in the conserved core of the SRA RNA. Expression of a synthetic STR5 disrupted the PUS1-SRA-steroid receptor pathway by inhibiting SRA pseudouridylation. Lack of SRA pseudouridylation resulted in repression of the ERα (estrogen receptor α) and AR (androgen receptor) dependent transactivation of the target genes. This finding could be of major interest in cancer therapeutics where SRA competitive inhibition could be achieved in steroid sensitive cancer cells (Ghosh et al., 2012). Interestingly, another mouse pseudouridine synthase mPus3 was also found to modify SRA, although mPus3 could not stimulate steroid receptor activity although mPus3 could pseudouridylate different positions in SRA

(Zhao et al., 2007). The human PUS1 was also able to bind and modify the U206 in the minimal H7 fragment of the SRA RNA (Huet et al., 2014).

Despite the fact that PUS1 was firstly described as a tRNA modifying enzyme, the study identifying 2084 ψ sites in HEK293T cells, associated only 77 PUS1-dependent ψ sites in PUS1$^{-/-}$ cells generated using CRISPR/Cas9 (Table 1) (Li et al., 2015a). Schwartz and colleagues determined 312 ψ sites in HEK293 cells using either a computational estimation based on the consensus sequence for TRUB1 and PUS7, or after DKC1 knockdown. This association of pseudouridine sites with their putative cognate pseudouridine synthases leaves a remaining of 84 ψ sites out of the total 396 ψ sites, which might be due to PUS1 acitivty. Based on the new transcriptome-wide maps of RNA pseudouridylation sites the web server PPUS: (http://lyh.pkmu.cn/ppus/) was created, employing a support vector machine (SVM) as the classifier and using nucleotides around the ψ sites as the features. PPUS is able to predict specific ψ sites for PUS1, PUS4 and PUS7 in yeast and PUS4 in human (Li et al., 2015b).

The myriad of publications referring to the yeast, mouse or even bacterial PUS1 homologues, offers a solid base for investigating the human PUS1. Only a handful of studies have actually investigated the influence of PUS1 in human health and disease and mostly only support the pathological observation of MLASA patients carrying the homozygous mutation R116W. The mechanisms by which the hPUS1 affects RNA metabolism and the effects on mRNA expression, processing and translation are not yet fully understood; therefore a more comprehensive study on PUS1 molecular function could provide new insights into these questions.

In the light of these findings, we were interested to identify the transcriptome-wide human PUS1 direct mRNA targets and the RNA-binding sites which could serve as basis for further investigation into the functional role of PUS1-dependent pseudouridylation on mRNA and in the post-transcriptional regulatory events.

Cold shock domain proteins and their biological functions

Cold shock domain proteins are one of the most conserved classes of proteins that are found from bacteria to vertebrates throughout all three kingdoms. This family of proteins contains a highly conserved domain, the cold shock domain (CSD), which is composed of two consensus RNA-binding motifs, the ribonucleoprotein 1 or RNP1 (K/N-G-F/Y-G-F-I/V) and ribonucleoprotein 2 or RNP2 (V-F-V-H-F) (Eliseeva et al., 2011) common to the RNA-recognition motif (RRM) domain (Chaikam and Karlson, 2010; Landsman, 1992; Manival et al., 2001). Bacterial cold shock proteins are small

proteins, with 65-70 amino acids, which can bind single stranded DNA and single stranded RNA. Bacterial CspA (cold shock protein A) and CspB (cold shock protein B) contain only one cold shock domain which has 43% sequence identity with the CSD of eukaryotic proteins (Schindelin et al., 1994b; Schindelin et al., 1993). Eukaryotic cold shock domain proteins have a modular design composed of one or multiple cold shock domains plus additional C-terminal domain. The evolutionary conserved structure of a cold shock domain which is structurally and functionally similar to the S1 domain, belongs to the superfamily of oligosacharride-/oligonucleotide-binding (OB) fold (Mihailovich et al., 2010b; Sommerville, 1999). The OB fold is a β-barrel closed structure composed of five anti-parallel β-strands with three consecutive β-strands (β1-3) on one side and a corresponding loop of 16 residues which bridges the β1-3 strands to the β4 and β5 strands. The β4-5 strands are packed against the β1-3 at an angle of approximately 90° forming an asymmetric L-shaped molecule (Schindelin et al., 1994a; Schindelin et al., 1993). The RNA-binding motifs RNP1 and RNP2 are found on different, juxtaposed, adjacent β-strands within the CSD (Newkirk et al., 1994; Schindelin et al., 1994a). The positively charged surface and the aromatic residues in the RNP1 and RNP2 are responsible for ssDNA binding through base stacking and hydrogen bonding but with a solvent exposed sugar-phosphate backbone which prevents the CSPs to distinguish between ribose and deoxyribose (Max et al., 2007). The study of the *Bacillus subtillis* CspB crystal structure in complex with two short RNA oligonucleotides containing a stretch of pyrimidines, revealed that *Bacillus subtillis* CspB associates in a monomeric form with RNA in contrast to the DNA binding dimer complex (Max et al., 2006; Sachs et al., 2012). It was also reported that the affinity for ssRNA was significantly lower than the affinity for ssDNA probably due to the substitution of thymine bases, which have an additional methyl group, by uracil and not due to the sugar-phosphate backbone interactions. The cold shock proteins in bacteria are involved in the cold stress adaptation and function as RNA chaperones or in transcription anti-termination. Bacterial CSPs are not only induced in a temperature-dependent manner but they are also present under normal conditions and can regulate different biological processes. In *E. coli* the rapid shift from 37°C to 10°C induces around 26 genes transiently, from which around 14 cold shock proteins, although the overall protein synthesis is markedly reduced (Horn et al., 2007; Schindelin et al., 1994a). Among these proteins induced upon cold shock, CspA is expressed up to a level of 13% of the total protein content, with an increased production of at least 100-fold. Nevertheless, studies on deletion of all four *E. coli* CSPs, CspA, CspB, CspG, and CspE showed that the cold shock proteins in bacteria have redundant functions (Xia et al., 2001). Interestingly, one study suggests that despite the CCAAT Y-box sequence in

its promoter, CspA does not regulate its own expression at the level of transcription but presumably by binding to its own mRNA and acting as an RNA chaperone (Tanabe et al., 1992). CspA was also observed to downregulate its own expression by binding to its 5' UTR, and destabilizing the secondary structure of its mRNA, an effect which was observed also for other mRNAs (Bae et al., 1997).

Cold shock domain proteins are found also in plants, although the knowledge about their structure and function is significantly reduced. The typical architecture of a CSD protein in plants resembles more the LIN28 (lin-28 homolog A (C. elegans)) cold shock domain protein. These plant CSD proteins contain two types of nucleic acid modules, an N-terminal cold shock domain and a variable number of C-terminal retroviral-like CCHC zinc fingers interspersed by glycine-rich regions (Chaikam and Karlson, 2010). The highly conserved cold shock domain plus the additional CCHC zinc fingers are suggested to function in specific nucleic acid binding and protein binding (Karlson et al., 2002). Different plant species have between two to seven cold shock domain gene families. One of the first cold shock domain protein characterized in plants was the wheat cold shock protein WCSP1 which is upregulated in response to cold and functions in cold acclimation. The WCSP1 contains a glycine-rich domain region interspersed with three C-terminal CCHC zinc fingers and is able to melt double-stranded nucleic acids by binding to DNA and RNA *in vitro* and *in vivo* (Nakaminami et al., 2006; Nakaminami et al., 2005). Four cold shock domain proteins were identified in *Arabidopsis*, the AtCSP1-AtCSP4 (CSDP1, AtGRP2), which are also upregulated during cold acclimation and function in freezing tolerance, flowering and embryo development (Kim et al., 2009). Plant cold shock domain proteins studies have not yet provided a clear function of specific CSD proteins in mRNA regulation.

Cold shock domain proteins in *Xenopus* are abundant in the oocytes and their roles were described in translational repression by "masking" the maternal mRNA (Murray et al., 1991). It was found that the frog germ cell-specific Y-box protein 2 - FRGY2 cold shock domain is required for sequence-specific interactions with mRNA but not for translational repression (Bouvet et al., 1995). N- and C-terminal regions also facilitate FRGY2 RNA binding but the non-specific C-terminal tail domain interaction is essential for translational silencing. Moreover, the presence of FRGY2 recognition elements within mRNA can potentially enhance translational repression (Matsumoto et al., 1996). In *C. elegans* Lin28 (abnormal cell lineage 28), a small cold shock domain protein contains a CSD and two additional CCHC zinc fingers forming the zinc-knuckle domain (ZKD). Lin28 pleiotropic functions include differentiation of embryonic stem cells and participation in developmental transitions during early stages. One of the important functions of the Lin28 protein is inhibition of let-7 microRNA biogenesis

(Newman et al., 2008; Piskounova et al., 2008; Viswanathan et al., 2008; Wulczyn et al., 2007). The crystal structure of the RNA binding domains of Lin28 in apo form and in nucleotide bound-forms revealed that the interaction with its known target, the pre-*let*-7 microRNA, depends on the CSD ability to remodel the terminal loop of pre-*let*-7 in order to facilitate the sequence-specific ZKD binding to the conserved GGAG motif (Loughlin et al., 2012; Mayr and Heinemann, 2013; Mayr et al., 2012; Nam et al., 2011). PAR-CLIP experiments with the human LIN28 in HEK293 cells identified a large set of approximately 3000 mRNA targets and 9500 binding sites found mostly in the 3' UTR and CDS, leading to an increase in protein levels (Hafner et al., 2013b). An interesting study focused on the role of another cold shock domain protein family in *C. elegan*s, the Y-box binding proteins CEYs (*C. elegans* YBPs), which include four members, CEY1-4, complementing the other well-known CSD-containing protein Lin28 (Arnold et al., 2014). Although they share the highly conserved cold shock domain, the CEYs contain additionally glycine-rich stretches such as the RG/RGG repeats, suggesting the resemblance with the other invertebrate Y-box binding proteins (Salvetti et al., 1998; Thieringer et al., 1997). CEY proteins were found to be involved in the formation of large polysomes in the soma, thereby leading to a decrease in polysomes together with an increase in monosomes and disomes in absence of these proteins but without observable effects for *C. elegans* development and homeostasis (Arnold et al., 2014). A couple of studies describe the *Drosophila melanogaster* Y-box protein, Yps (Ypsilon Schachtel), as a potential regulator of the *oskar* mRNA localization and translation in an antagonistic relationship with Orb protein, a known enhancer of *oskar* mRNA translation and localization (Mansfield et al., 2002; Thieringer et al., 1997). Two members of the cold shock domain proteins have been described in planarians as well as in one of the most common parasites, *Schistosoma mansoni*, the DjY1 and SMYB1 respectively (Dias et al., 2014; Franco et al., 1997; Salvetti et al., 1998).

The mouse Y-box cold shock domain proteins MSY1 (YBX1 human homologue), MSY2 (YBX2 human homologue) and MSY3/4 (YBX3 human homologue) are cold shock domain proteins found in the germ line, with MSY2 accounting for 2% of the total protein of the oocyte (Yu et al., 2001). In later developmental stages, MSY2 is reduced which is suggesting a role in maternal mRNA translational regulation. MSY1 deficient mice die late in embryogenesis (E18.5 to post-natal day 1) but MSY4 is not required for survival. However, MSY4 expression is preventing spermatocyte apoptosis and fertility defects in males. Interestingly, MSY1/MSY4 double-knockout leads to a much earlier death in mouse embryogenesis, around E8.5 to E11.5 (Lu et al., 2006). These observations indicate that MSY1 and MSY4 share important cellular functions but might also control independent processes. Using polysome fractionation it was found that

MSY4 is associated with mRNPs in the cytoplasm of germ cells, suggesting a role in storing repressed messages, while yeast three-hybrid system showed that MSY4 specifically binds to a site 37 nucleotides downstream in the 3' UTR of the protamine 1, Prm1, mRNA (Davies et al., 2000). Subsequently, it was reported that the translationally repressed Prm1 mRNA is specifically bound at the 5'-UCCAUCA-3' sequence in the 3' UTR, by both MSY2 and MSY4 (Giorgini et al., 2001b). Mutagenesis of this site determined a consensus sequence represented by U(/A/C)C(/A)**CA**U(/C)**C**A(/C/U) and defined as the Y-box recognition site (YRS). Moreover, in vivo reporter assay indicated that the YRS disruption could alleviate the Prm1-like translational repression of the reporter gene. These studies suggest a common role of the mouse homologues of Y-box binding proteins as translational repressors which can bind specifically to their target mRNAs. An interesting association between MSY3 with the zonula occludens-1 (ZO-1) protein was found at the oligodendrocyte and astrocyte gap junctions in the mouse brain which might indicate a localized translational regulation of the two tight junction factors (Penes et al., 2005).

Cold shock domain proteins in human

The complexity of the human organism has allowed cellular responses and organ responses to rapidly and efficiently adapt to changes in environmental conditions. While single-cell organisms respond directly through cellular changes to external factors, higher organisms utilize different strategies to reduce the effects on cellular level by local and global regulatory systems as for example temperature homeostasis in mammals (Horn et al., 2007). Although two proteins expressed under cold temperatures have been described in humans, the cold-inducible RNA-binding protein (CIRBP) and the RNA-binding motif 3 (RBM3), they do not contain a cold shock domain (Liu et al., 2013b). Moreover, the human cold shock domain proteins do not have a role in cold-shock adaptation but they seem to be involved in responses to other cellular stresses (Lindquist et al., 2014a).

Mammalian cold shock domain proteins present a structure which contains one or multiple cold shock domains, such as the upstream of N-Ras/CSDE1 protein, or a combination of the CSD and CCHC zinc-finger motifs. Although the cold shock domain is an ancient highly conserved structure, the evolution of specific domain combinations can be appreciated from the phylogenetic distribution of the related architecture proteins. Therefore, if a protein has a combination of domains which is very restricted in the phylogenetic tree, this would indicate a more recent origin evolved probably by

duplication, combination or loss of domains from older architectures (Mihailovich et al., 2010a). The InterPro database for protein sequence analysis and classification shows forty protein isoforms containing a cold shock domain/OB-fold in *Homo sapiens*, including the Y-box proteins, YBX1 and YBX3, CSDE1/UNR, CSDC2/PIPPin, LIN28A and LIN28B, and the calcium-regulated heat stable protein 1 (CARHSP1) (Mitchell et al., 2015). The cold shock domain in human is highly similar to the ancient CSD, with 43% identity to the bacterial CSD, 80% identity to the Xenopus and 96% identity to the mouse CSD amino acid sequence (Eliseeva et al., 2011). The five anti-parallel β-strands form the β-barrel structure, with the β1 strand wrapping around the β-barrel due to a conserved glycine residue which is conferring the structural flexibility. The OB-fold length is varying between 70 and 150 amino acids (Mihailovich et al., 2010a), with 65 amino acids for YBX1 or YBX3, and 71 amino acids for YBX2. The conserved RNP1 and RNP2 motifs are found as well in the β2 and β3 strands, respectively, with the ligand binding site on the left face of the β-barrel (Figure 3A). The CSD/ZKD in the LIN28 protein has a conserved function in humans, where it can inhibit the biogenesis of *let-7* miRNA family in embryonic stem cells thus preventing their differentiation (Mayr et al., 2012). The crystal structure of human LIN28 revealed that binding of the CSD to the terminal loop of *pre-let-7* can induce a conformational change of the Dicer cleavage site and subsequently, facilitates the GGAG sequence-specific binding of the ZKD (Mayr and Heinemann, 2013; Mayr et al., 2012). Besides the function in development and stem cell programming, the human LIN28A and LIN28B were found associated with polysomes bound mRNAs in the cytoplasm of HEK293 cells (Hafner et al., 2013a). Moreover, the binding to ssRNA within U-rich elements stabilized target mRNAs and significantly increased protein abundance of top targets including its own mRNA and cell cycle regulators. Interestingly, *let-7* miRNA family members were selectively regulated, suggesting the sequence-specific binding of LIN28. Hafner and colleagues showed that the LIN28A and LIN28B have a binding preference towards exonic regions, 3' UTR, and CDS, with an overlap of 60% between the two LIN28 family members. LIN28A and LIN28B have a common structure but only 73% sequence identity at the protein level. Furthermore, LIN28 protein not only regulates miRNA biogenesis, but is a target itself of *miR125*, the mammalian homologue of *lin-4* miRNA, during neuronal differentiation (Rybak et al., 2008; Wu and Belasco, 2005).

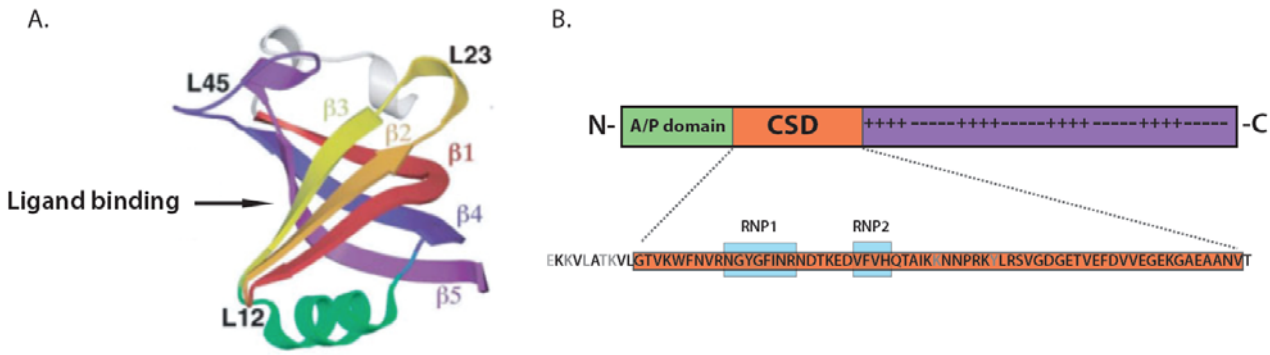


Figure 3. Structure of OB-fold and cold shock domain Y-box binding protein. A. The structure of the typical OB-fold domain. The colored β1-β5 strands are indicated individually with the variable loops colored in black. Left side arrow is indicating the ligand-binding surface. (*adapted from (Mihailovich et al., 2010a)*). B. Typical structure of a CSD containing Y-box binding protein. N-terminal domain (in green) is an Ala/Pro rich domain, while the C-terminal domain (in purple) is characterized by alternative clusters of positively and negatively charged amino acids of around 30 amino acids per cluster. The cold shock domain (in orange) in the middle is highlighted below as the conserved sequence containing the RNP1 and RNP2 motifs (in blue).

The cold shock domain containing E1 protein, CSDE1,or Upstream of N-ras (UNR), has a unique architecture among CSD proteins, with a five cold shock domain structure, which have a conservation between 59%-71% among higher eukaryotes (Mihailovich et al., 2010a). The CSDE1 protein was found in human, mouse and rat cells although it does not have a homologue in yeast or *C. elegans*. Nevertheless, homozygous mice lacking CSDE1 were not viable and did not develop further than the embryonic day 10 (Ray et al., 2015). The UNR name of this protein is suggesting the position of the CSDE1 gene, which is approximately 150 bp upstream of the N-ras locus and this position is conserved across multiple species (Jeffers et al., 1990). This head to tail gene tandem was investigated for a possible mechanism of transcriptional interference due to the same transcriptional orientation of the two genes. However, the study reported only an increase of 20-65% in expression level of N-ras across different tissues, in a heterozygous mouse model and suggested that expression of CSDE1 is not a major determinant of the transcriptional interference (Boussadia et al., 1997). CSDE1 was shown to interact with RNA in vitro and its intracellular localization has suggested a role in mRNA metabolism. SELEX experiments have identified two consensus high affinity (K_d of ~ 10 nM) sequence motifs, the AAGUA/G or AACG following an upstream purine-rich stretch. This motif, however, is different than the Y-box factors consensus motif indicating that CSD proteins can bind to a variety of sequences (Triqueneaux et al., 1999). The twenty-fold higher affinity of the first CSD1

domain in comparison to the fifth CSD5 domain together with the enhanced stability suggests an evolution of CSDE1 by repeated duplication of a one-CSD containing protein (Goroncy et al., 2010). Several studies on the *Drosophila* CSDE1 (UNR) reported cooperative molecular interactions in the complex regulation the male-specific lethal 2 mRNA - *msl2*, which encodes a limiting component of the *Drosophila* dosage compensation complex (DCC). The sex-lethal protein, SXL, forms a corepressor complex together with CSDE1, on *msl2* 3' UTR mRNA. By interacting with PABP (poly(A) binding protein), which is a translational activator, CSDE1-SXL complex is able to repress *msl2* translation by inhibiting the small ribosomal subunit recruitment to mRNA (Duncan et al., 2009; Hennig et al., 2014). CSDE1 has been described in various studies as an important cap-independent translational activator by binding to IRESes (internal ribosome entry site). This indicates that CSDE1 could function as an IRES trans-acting factor (ITAF) thereby regulating several viral and cellular IRESes besides its own 5' UTR IRES (Brown and Jackson, 2004; Dormoy-Raclet et al., 2005; Mitchell et al., 2003).

Y-box family of cold shock domain proteins

The human Y-box family of proteins is formed by the YBX1, YBX2 and YBX3 also known as YB-1/DBPB, CONTRIN and CSDA/DBPA/ZONAB respectively. The three members of the Y-box family share a common central cold shock domain with more than 90% amino acid sequence identity (Eliseeva et al., 2011).The Y-box factors were first observed as major components of cytoplasmic mRNPs, although later were described as DNA-binding proteins (Dbps) that recognize the inverted CCAAT Y-box in the major histocompatibility complex (MHC) class II genes promoters (Didier et al., 1988; Sakura et al., 1988). The structure of the Y-box binding proteins is differing from the other cold shock domain proteins by the unique N-terminal Ala/Pro rich domain combined with the C-terminal basic-acidic repeats, in addition to the central CSD containing the RNP1 and RNP2 motifs (Figure 3B). Solution structure of eukaryotic YBX1 cold shock domain using high-resolution NMR spectroscopy showed the conserved structural features of the OB-fold with a DNA-binding site represented by a patch of positively charged and aromatic residues on the surface of the β-barrel (Kloks et al., 2002). However, the cold shock domain has a weak affinity and specificity of DNA binding suggesting that the other domains may play an additional role in binding of DNA by YBX1. The restricted phylogenetic distribution of this domain architecture suggested that the Y-box factors might have arisen at the origin of chordate evolution,

while the association of the CSD with other domains indicated the multiplicity of functional structures (Mihailovich et al., 2010a). Besides the YBX2 which is found only in the germ cells and linked to possible male spermatogenic defects and fertility impairment (Deng et al., 2008; Hammoud et al., 2009), YBX1 and YBX3 have been linked to a wide range of cellular functions from DNA transcriptional regulation to translational regulatory functions (Eliseeva et al., 2011; Mihailovich et al., 2010a). YBX1 and YBX3 were first identified from a λgt11 cDNA expression library screen using double-stranded oligonucleotides containing the Y and X-boxes promoter region of the major histocompatibility complex class II genes. DNA probes containing an enhancer region of the human EGFR (epidermal growth factor receptor) gene and the promoter region of the human ERBB2 (erb-b2 receptor tyrosine kinase 2) gene were also used to identify the YBX1 and YBX3 proteins (Didier et al., 1988; Sakura et al., 1988).

Despite the fact that both YBX1 an YBX3 have been described as DNA-binding proteins using *in vitro* experiments, additional studies rather suggest a functional role in post-transcriptional regulation. YBX1 is demonstrated to have a higher preference to single-stranded DNA than to double-stranded, and a binding preference to diverse DNA sequences different than the Y-box sequence. An example of a controversial DNA target for YBX1 is the MDR1 (multidrug resistance) gene, which was initially thought to be transcriptionally activated by YBX1 and later shown that its interaction with DNA is dependent on formation of a complex with acetylated APE1 (APEX nuclease (multifunctional DNA repair enzyme) 1) and p300 (Hu et al., 2000; Sundseth et al., 1997). Moreover, the nuclear transcription factor NF-Y has been frequently characterized to bind the Y-box sequence promoter regions with a significantly higher affinity ($K_D \sim 10^{-11}$) than YBX1 ($K_D \sim 10^{-9}$) and also to be the main factor to regulate MDR1 promoter activity (Dolfini and Mantovani, 2013; Kim and Sheffery, 1990). In addition, experimental approaches for testing the *in vivo* properties of YBX1 did not integrate the regulation of mRNA stability, splicing or translation. Thereby a proposed hypothesis would be that YBX1 is rather transiently located in proximity of transcriptionally activated units but loaded on the nascent pre-mRNA (Dolfini and Mantovani, 2013) to regulate its packaging, transport, localization and translation.

Similarly, YBX3 (ZONAB, ZO-1- associated nucleic acid-binding protein) DNA-binding functions studies suggested opposite effects of YBX3 on the p21 gene promoter. ZO-1 (zonula occludens 1/ TJP1, tight junction protein 1) protein was described to interact with YBX3 and modulate ERBB2 (erb-b2 receptor tyrosine kinase 2) promoter in a cell density-dependent manner. Moreover, this study reported a very high affinity of YBX3 for the p21 promoter containing the CCAAT sequence using an *in*

vitro gel-shift assay (Balda and Matter, 2000b). In the follow-up experiment, it was demonstrated that p21 is regulated by YBX3 binding to its 3' UTR region and that depletion of YBX3 has no effect on the p21 promoter activity in a luciferase reporter assay (Nie et al., 2012).

The Y-box proteins have been associated with both translational activation and translational repression (Minich and Ovchinnikov, 1992). One of the first studies has found that YBX1 can be part of free mRNPs as well as polyribosomal mRNPs and has suggested a role in mRNA translation by melting the mRNA up to 60%, which could in turn affect the binding of other translational factors or regulatory proteins (Evdokimova et al., 1995). Subsequently, Edvokimova and colleagues have investigated the YBX1 translational activities in a cell-free translational system using and anti-YBX1 antibody immunodepleted rabbit reticulocyte lysates or incubated with purified recombinant rabbit YBX1. This system provided a model in which low levels of YBX1/mRNA ratio stimulated protein synthesis while an increase in YBX1/mRNA ratio gradually inhibited translation (Evdokimova et al., 1998).

Y-box proteins have also been correlated with cancer progression and aggressiveness and suggested as potential biological markers (Matsumoto and Bay, 2005). YBX1 was upregulated during prostate cancer tumor progression and could induce chromosomal instability in breast cancer development (Bergmann et al., 2005; Gimenez-Bonafe et al., 2004). YBX3 has been proposed as a prognostic marker in hepatocellular carcinoma where it was found significantly upregulated and localized to the nucleus which was an indicator of poor prognosis (Yasen et al., 2005; Yasen et al., 2012). However, YBX3 overexpression has also been suggested to inhibit tumor growth and lymph node metastasis by inhibiting the expression of the vascular endothelial growth factors, VEGFA and VEGFC, in both normoxic and hypoxia-induced NR-S1M murine squamous cell carcinoma cell line (Matsumoto et al., 2010). Coles and colleagues have reported YBX3 function as a repressor of the hypoxia-responsive region on the VEGF promoter (Coles et al., 2002). However, it was shown that VEGF expression is regulated by a complex containing a Y-box protein, presumably YBX1, and PTB (polypyrimidine tract binding protein). The multiprotein complex including PTB and YBX1 could bind to the VEGF 5' UTR and 3' UTR specific sequences and stabilize its mRNA (Coles et al., 2004b). The link between YBX3 and hypoxia has been investigated in the study of skeletal muscle angiogenesis where YBX3 upregulation repressed the activity of HIF1 and NFκB by binding to the hypoxia response element (HRE) and the nuclear factor κB-binding element (NFκB-BE) (Saito et al., 2011). Moreover, depletion of YBX3 increased the secretion of the angiogenic factors VEGFA, IL-6 (interleukin 6) and IL-8 (interleukin 8) in skeletal muscle cells suggesting a

potential therapeutic strategy for peripheral arterial disease. An interesting study reported recently that a specific class of tRNA-derived fragments (tRFs) are induced under hypoxic stress and YBX1 displacement from the 3' UTR leads to suppression of multiple oncogenic transcripts (Goodarzi et al., 2015). Using asLNAs (anti-sense locked nucleic acids) and synthetic RNA mimetics, Goodarzi and colleagues demonstrated that the destabilization of pro-oncogenic transcripts which were stabilized by YBX1 can reduce the growth and cancer cell invasion.

A recent review underlined a multitude of roles for the cold shock domain and Y-box proteins in inflammatory diseases (Lindquist et al., 2014b). YBX1 was found to be secreted in a non-classical pathway and induce migration and cell proliferation. It was suggested as well that YBX1 posseses an oligomerization/amyloid-like fibril formation activity (Frye et al., 2009; Guryanov et al., 2012). The reported functions of extracellular YBX1 could indicate a possible therapeutic target by small molecule inhibitors which could prevent its oligomerization as well as binding to specific receptors (Raffetseder et al., 2011).

Y-box binding protein 3 and its cellular functions

The YBX3 protein was first isolated by Sakura and colleagues from a λgt11 cDNA library using the EGFR enhancer and ERBB2 promoter as DNA probes (Sakura et al., 1988). Subsequently, Kudo and colleagues characterized the full lenght YBX3 gene using the λ phage genomic clones (Kudo et al., 1995). YBX3 gene was assigned to chromosome 12p3.1, expanding a 24-kb genomic region, and comprising 10 exons, with the cold-shock domain encoded by exons 2-5, and an alternative exon 6. It was found that YBX3 had the highest expression in skeletal muscle and heart, but could be detected in both nucleus and cytoplasm. Early studies that paralleled the discovery of YBX3 gene reported that YBX3 also termed NF-GMb (nuclear factor, granulocyte-macrophage), could bind specifically the GM-CSF (granulocyte-macrophage colony stimulating factor) gene promoter and repress its transcription (Coles et al., 1996; Coles et al., 1994). However, it was observed that overexpression of YBX3/YBX1 complex leads to the repression of a promoter which lacks the YBX3 binding sites suggesting the existence of two different mechanisms in which YBX3 does not directly target gene promoters. Nevertheless, several studies examined the RNA-binding preferences of the different domains of a Y-box binding protein showing that both the cold-shock domain and the C-terminal domain are involved in RNA-binding (Ladomery and Sommerville, 1994). Despite the previous described functions of YBX3 including

the regulation of ERBB2 expression, it has been postulated that the cold-shock domain complex including YBX3 and PTB could regulate VEGFA mRNA through 5' UTR and 3' UTR binding (Coles et al., 2004b). Interestingly, using *in vivo* approaches such as CLIP-seq in HeLa cells, the transcriptome-wide mapping of PTB binding sites did not identify a PTB binding site at the specific PTB 3' UTR motif on VEGFA mRNA reported by Coles and colleagues using *in vitro* experiments (Xue et al., 2009). RALA (v-ral simian leukemia viral oncogene homolog A (ras related)), a Ras GTPase family member interacting with YBX3 at high cellular density in a GTP-dependent manner resulted in transcriptional derepression mediated by YBX3 in a reporter assay containing the ERBB2 promoter (Frankel et al., 2005). Cancer research studies have also associated the upregulation of YBX3 mRNA level with advanced stages of hepatocellular carcinoma (Wang et al., 2009; Yasen et al., 2005). Within gastric cancer tissues, YBX3 silencing increased the expression of E-cadherin, decreased expression of cyclin D1, APC (adenomatous polyposis coli) and β-catenin thus enhancing chemosensitivity to 5-fluorouracil (Wang et al., 2009; Yasen et al., 2005). However, different qRT-PCR control housekeeping genes used for normalization revealed that YBX3 expression levels were not increased in samples from clear cell renal cell carcinoma (ccRCC). Therefore, when grading different stages of ccRCC, YBX3 expression was found to be higher in low-grade tumors than high-grade tumors (Dupasquier et al., 2014). In 2010, Sears and colleagues reported an interesting finding in chronic myeloid leukemia (CML) Bcr-Abl (ABL1, ABL proto-oncogene 1, non-receptor tyrosine kinase) dependent transformation (Sears et al., 2010b). This study shows, for the first time, that YBX3 is phosphorylated by AKT1 (v-akt murine thymoma viral oncogene homolog 1) at Ser134 in non-CML cells but Bcr-Abl activation results in MEK-dependent (MAP2K7, mitogen-activated protein kinase kinase 7) RSK phospshorylation of YBX3 Ser134 in both CML cell lines and CML patients primary progenitor cell. YBX3 phosphorylation leads to an increase in proliferation that is negatively regulated by expressing the Ser134Ala mutant. These results indicate that RSK- or AKT1-dependent phosphorylation of YBX3 presumably modulate YBX3 function on translation in CML and Bcr-Abl leukemias. Moreover, YBX3 Ser134 phosphorylation may represent a druggable alternative target to PI3K/AKT1 inhibition in leukemogenesis. YBX3 has been investigated in relation to its tight junction modulators, the ZO-1, ZO-2 and ZO-3 (zonula occludens 1) proteins, which function redundantly to prevent YBX3 translocation to the nucleus and activation of cell proliferation targets (Spadaro et al., 2014). CDKN1A/p21 (cyclin-dependent kinase inhibitor 1A) is one of the factors involved in different aspects of cell cycle regulation including apoptosis and cell death, with both positive and negative functions. The p21 expression is regulated

by multiple different pathways at both transcriptional and post-transcriptional levels, however, studies have not yet revealed how different signals are integrated in the overall p21 response (Jung et al., 2010). Highlighting this question, one study showed that YBX3 is an up-regulator of CDKN1A message during stress activation thus leading to a decrease in apoptosis and suggesting a role for YBX3 in cell survival pathways (Nie et al., 2012). This study suggested that YBX3 role as post-transcriptional regulator is mediated by ARHGEF2/GEF-H1 (Rho/Rac guanine nucleotide exchange factor (GEF) 2) activation upon cellular stresses induced by TNFα, EGFR or Erk signaling. However, YBX3 described as a potential up-regulator in stress conditions for one mRNA candidate would not explain all the cellular changes downstream this target, unless stress would trigger YBX3 to exert a global function on the transcriptome. Therefore, YBX3 studies are currently lacking an overview of the potential global functional role in regulating multiple targets on post-transcriptional level and exerting an overall effect on cellular responses.

In summary, the available data suggests that YBX3 has diverse transcriptiona and post-transcriptional cellular functions. Given the fact that one of the key features is the cytoplasmic localization and a high binding preference towards cellular RNAs, we were interested in elucidating its regulatory function at the system-wide level by identifying its direct mRNA targets as well as the molecular function at the post-transcriptional level.

AIMS OF THE THESIS

I. Identification of the transcriptome-wide binding sites of Pseudouridine synthase 1 (PUS1)

RNA editing is an important aspect of RNA metabolism, with over 100 known modifications in various RNA species. Pseudouridylation (U->ψ) was previously thought to be found only in non-coding RNA, however, it has been recently uncovered that pseudouridine is wide-spread across the transcriptome. Pseudouridine synthase I is an RNA-independent RNA-editing enzyme that catalyzes the isomerization of uridine to pseuoduridine and it was discovered as the only ψ-synthase enriched in three independent mRNA-bound proteome studies. In light of this finding, we aimed to characterize the *in vivo* RNA binding sites of PUS1 by asking the following biologically relevant questions:

1. What is the PUS1 binding pattern across the transcriptome of HEK293 cells?

2. What are the most predominant RNA targets of PUS1 (tRNA, mRNA)?

3. Can we detect PUS1 pseudouridylation sites in tRNA and mRNA from HEK293 cells?

4. Does PUS1 catalyze the formation of ψ in mRNA and can we identify these positions at nucleotide resolution?

In order to answer these questions we set up the experimental design outlined in Figure 4.

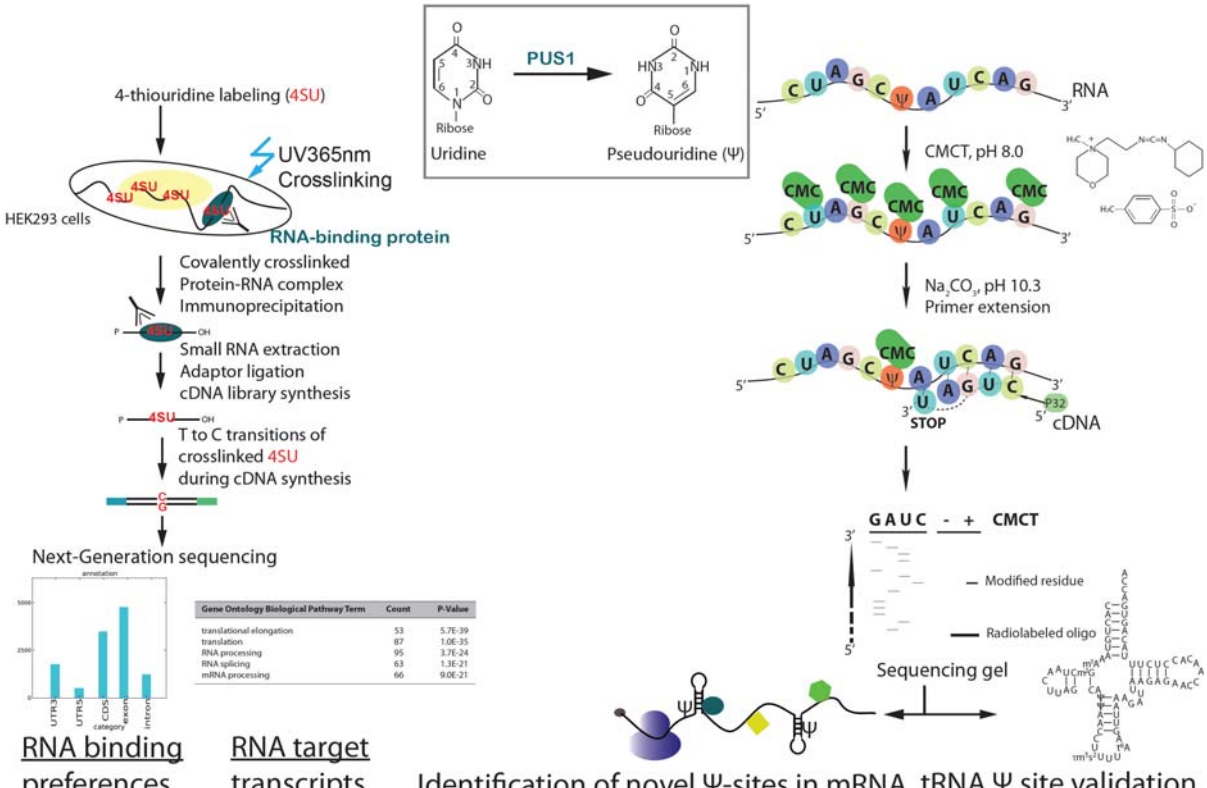


Figure 4. Experimental outline of the PUS1 study. We set up to map PUS1 binding sites across the transcriptome by employing the PAR-CLIP method (Hafner et al., 2010b) in HEK293 cells. Cells expressing FLAG/HIS/HA-PUS1 were labeled with 4SU and crosslinked at UV365nm. The RNA-protein crosslinked complexes were immunoprecipitated and RNA was extracted, ligated and reverse transcribed into a cDNA library which was submitted to next-generation sequencing. Due to the T-C conversion of the crosslinked uridine during reverse transcription we can determine the RNA binding sites across the transcriptome at nucleotide resolution as well as indentify PUS1 target transcripts. We used a CMCT-based protocol (Bakin and Ofengand, 1998) in combination with primer extension to be able to map the exact position of pseudouridines present in previously described tRNA substrate or to identify *de novo* putative pseudouridine sites among the mRNA PUS1 PAR-CLIP binding sites.

II. Y-box binding protein 3 (YBX3) transcriptome-wide mapping and functional characterization of mRNA binding sites

YBX3 protein is characterized by a highly conserved cold shock domain containing two RNP motifs with high affinity to DNA or RNA sequences, and a basic-acidic C-terminal domain which could participate in protein-protein interactions as well as nucleic acid binding. Described as a multifunctional DNA/RNA binding protein, YBX3 has been suggested to act in post-transcriptional regulation of several target mRNA possibly by binding to the 3' UTR regions, however the molecular mechanism has not been yet uncovered. We aimed at characterizing YBX3 transcriptome-wide binding sites and and investigate YBX3 functional role along the following key questions:

1. What is the transcriptome-wide YBX3 binding pattern?

2. Does YBX3 bind to a consensus RNA recognition motif on its target transcripts?

3. Does YBX3 depletion influence the protein synthesis of its target mRNAs?

4. What are the YBX3 protein interaction partners do they play a role in the post-transcriptional regulation by YBX3?

In order to answer these questions we set up the following experimental design described in Figure 5.

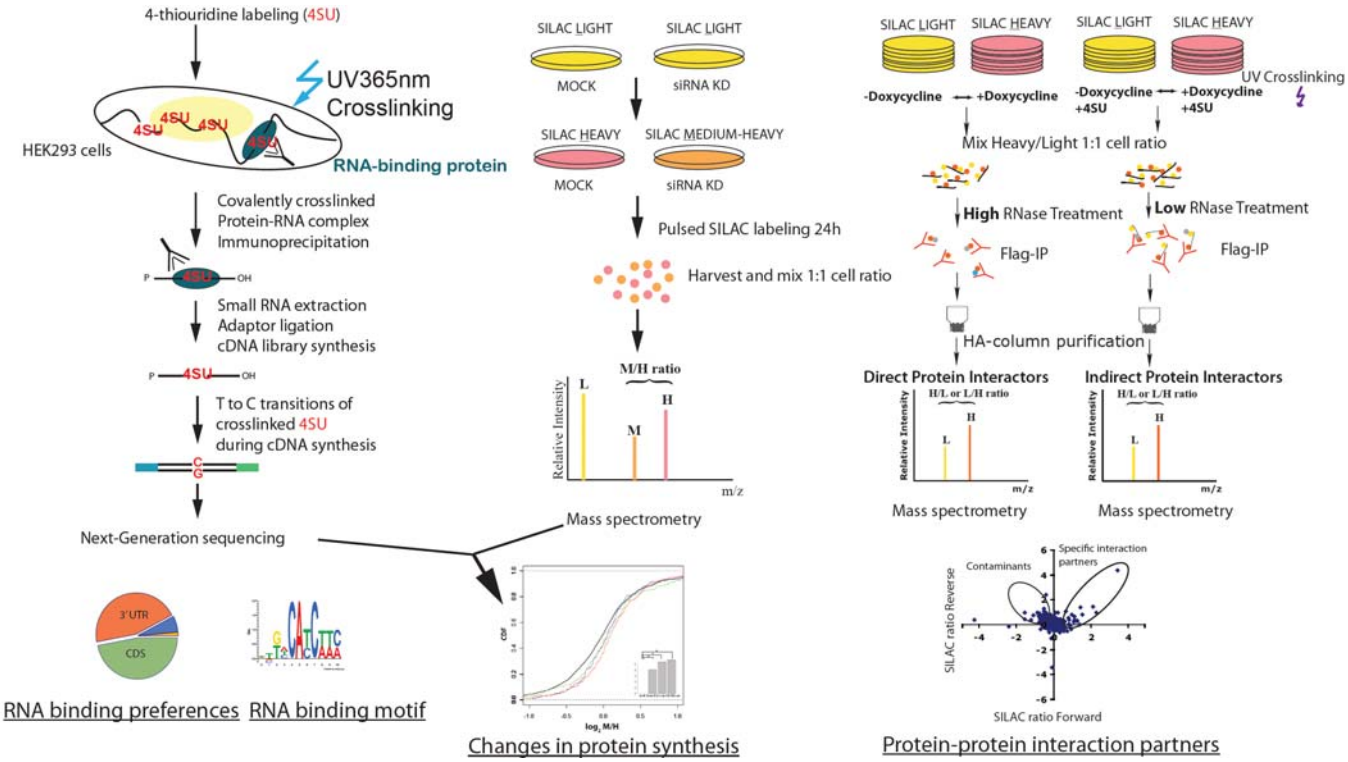


Figure 5. Experimental outline of the YBX3 study. We set up to map YBX3 binding sites across the transcriptome by employing the PAR-CLIP method (Hafner et al., 2010b) in HEK293 cells. Due to the T-C conversion of the crosslinked uridine during reverse transcription we can determine the RNA binding sites across the transcriptome at nucleotide resolution, indentify YBX3 binding preferences and YBX3 RNA binding motif in the targeted transcripts. Using a pulsed SILAC based mass-spectrometry and quantitative proteomics (Schwanhäusser et al., 2011) we are able to quantify changes in protein synthesis and determine YBX3 effects on the translational output of its target mRNAs. With a similar SILAC approach, we used light and heavy labeling of YBX3 expressing or non-induced HEK293 cells in combination with 4SU labeling and UV crosslinking in order to capture direct and RNA-dependent protein interactors of YBX3. The identification of protein-protein interaction factors contributed to the understandanding YBX3 functional role in post-transcriptional regulation.

The studies described in this thesis aimed to gain a deeper insight into the molecular mechanisms of post-transcriptional gene regulation exhibited by PUS1 and YBX3. Such insights represent a first step towards understanding the functional relevance of these two important proteins, which may be relevant in several disease-related phenotypes and therefore constitute the basis for future medical applicability of the expected results obtained by this thesis.

MATERIALS AND METHODS

Antibodies

anti-FLAG (Sigma, F1804)
anti-HA.11. (Covance, 16B12, mouse monoclonal)
anti-PUS1 (Thermo Scientific, PA5-29108, rabbit polyclonal IgG)
anti-YBX3 (Bethyl/Biomol, A303-070A rabbit polyclonal IgG)
anti-Vinculin (Sigma, V9131, monoclonal mouse clone hVIN-1, ascites fluid)
Polyclonal Goat Anti-Mouse Immunoglobulins/HRP (DAKO, P0447)
Polyclonal Goat Anti-Rabbit Immunoglobulins/HRP (DAKO, P0448)

Oligonucleotides

PAR-CLIP small RNA cloning adaptors

3’ DNA barcoded adapters:
NBC1 5’ AppTCTAAAATCGTATGCCGTCTTCTGCTTG3InvdT 3’
NBC2 5’ AppTCTCCCATCGTATGCCGTCTTCTGCTTG3InvdT 3’
NBC3 5’ AppTCTGGGATCGTATGCCGTCTTCTGCTTG3InvdT 3’
NBC4 5’ AppTCTTTTATCGTATGCCGTCTTCTGCTTG3InvdT 3’
NBC5 5’ AppTCTCACGTCGTATGCCGTCTTCTGCTTG3InvdT 3’
NBC6 5’ AppTCTCCATTCGTATGCCGTCTTCTGCTTG3InvdT 3’
NBC7 5’ AppTCTCGTATCGTATGCCGTCTTCTGCTTG3InvdT 3’
NBC8 5’ AppTCTCTGCTCGTATGCCGTCTTCTGCTTG3InvdT 3’

3’ DNA non-barcoded adapter:
NN-oR3 5’ AppNNTCGTATGCCGTCTTCTGCTTG3InvdT 3’

5’ RNA adapter:
5’ rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC 3’

5’ RNA NN-adapters:
NN-oR5 5’ GUUCAGAGUUCUACAGUCCGACGAUCNN 3’
R5BC1 5’ NN GUUCAGAGUUCUACAGUCCGACGAUCCCCANN 3’

R5BC2 5’ NN GUUCAGAGUUCUACAGUCCGACGAUCCACGNN 3’
R5BC3 5’ NN GUUCAGAGUUCUACAGUCCGACGAUCCGUANN 3’
R5BC4 5’ NN GUUCAGAGUUCUACAGUCCGACGAUCCUGCNN 3’
R5BC5 5’ NN GUUCAGAGUUCUACAGUCCGACGAUCCCAUNN 3’
R5BC6 5’ NN GUUCAGAGUUCUACAGUCCGACGAUCGGGANN 3’

PAR-CLIP DNA oligonucleotides:

3’ PCR primer 5’ CAAGCAGAAGACGGCATACGA 3‘
5’ PCR primer 5‘ AATGATACGGCGACCACCGACAGGTTCAGAGTTCTAC AGTCCGA 3‘

qRT-PCR DNA oligonucleotides:

PUS1 Fw: 5’ GTGTGGCTGATTGACGACAT 3’
PUS1 Rv: 5’ CACATCTGTTCTTGGAGTTAAACC 3’
YBX3 Fw: 5’ CTTCACAAGATGGCAAAGAGG 3’
YBX3 Rv: 5’ GAGCCTGGTGTTACTCAGCA 3’

3’ UTR cloning DNA oligonucleotides (**XhoI/NotI restriction site**):

ARF1
5’ACGC C**TCGAG** ACG CGA CCC CCC TCC CTC TCA 3’
5’ CCGC **GCGGCCGC** GAA ATA GTT AAG AGA TTT TAT TCT 3’
TROVE2
5’ ACGC **CTCGAG** CCA TAA GCA GCA GCA CGA TCC 3’
5’ CCGC **GCGGCCGC** CCT CAC AAT GAA GGA TAC TTG CTC 3’
PARP1
5’ ACGC **CTCGAG** GAG AGG TAG CCG AGT CAC ACC 3’
5’ CCGC **GCGGCCGC** GAA GGC ATC TGC ATT TTT AAT CGA G 3’
CDC25A
5’ ACGC **CTCGAG** CCA GCA GCA GCC CAA GCT TC 3’
5’ CCGC **GCGGCCGC** GGG TTC AAG ATC TTT TAT TTT CAG AGC 3’
ARHGEF2
5’ ACGC **CTCGAG** GTG CCC CAC TGA AGA ACA TTA C 3’
5’ CCGC **GCGGCCGC** CAG TAC CTT ATG GAG TTC CAG G 3’
FAM98A
5’ ACGC **CTCGAG** GGC TAC CGA ACC TTA CAT TTT GC 3’
5’ CCGC **GCGGCCGC** GAA CTT AAC AGA TAC TTT ATT GCT CAC 3’

PDCD10
5’ ACGC **CTCGAG** GTA TAT GTT AAG AGA TGT ACT TCT CAG 3’
5’ CCGC **GCGGCCGC** GTA AAC TAT TTA TTG AAT ATT TGC CAC C 3’
MTOR
5’ ACGC **CTCGAG** GAG GCC CAG ATG TGC CCA TC 3’
5’ CCGC **GCGGCCGC** GAT GTC ATT TAT TGG CAC AAA AAT TAT TC 3’

Plasmid constructs and cloning

The PUS1 CDS was obtained from the hORFeome database as a pDONR221 vector which was subsequently recombined into a pFRT/TO/FLAG/HIS/HA DEST vector (Invitrogen/Thermo) using the Gateway recombination system and LR Recombinase (Invitrogen/Thermo). The destination vector encoding human PUS1 coding sequence (NM_001002019.2) was transfected into Flp-In T-REx HEK293 cells using Lipofectamine 2000 (Landthaler et al., 2008).
The coding sequence of YBX3 was obtained from the hORFeome database as a pDONR221 vector. The pDONR221 plasmid was recombined in the destination vector pFRT/TO/FLAG/HIS/HA DEST (Invitrogen/Thermo) using Gateway recombination system (Invitrogen/Thermo). The pFRT/TO/FLAG/HIS/HA YBX3 construct (NM_003651.4) was transfected into Flp-In T-REx HEK293 cells (Introgen/Thermo) using Lipofectamine 2000 (Invitrogen/Thermo) (Landthaler et al., 2008). Due to the Gateway recombination system we were able to use the TO/CMV promoter in order to induce FLAG/HIS/HA YBX3 expression in the stably-transfected cells using doxycycline.

Cell lines and culture conditions

Flp-In T-REx HEK293 (Invitrogen/Thermo) cells were grown in D-MEM high glucose media containing 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 (Gibco), 100 µg/ml zeocin (Invivogen, USA) and 15 µg/ml blasticidin (Invivogen, USA). Cell lines stably expressing FLAG/HIS/HA-tagged PUS1 or FLAG/HIS/HA-tagged YBX3 were generated by co-transfection of pFRT/TO/FLAG/HIS/HA construct with pOG44 plasmid (Invitrogen/Thermo) using the manufacturer’s protocol. Cells were selected by exchanging zeocin with 100 µg/ml hygromycin. Expression of epitope-tagged protein was induced by addition of 1 µg/ml doxycycline 12 to 16 hours before harvesting or

crosslinking. Expression of epitope-tagged PUS1 or epitope-tagged YBX3 was validated by western blot analysis using anti-HA.11 mouse monoclonal antibody and HRP polyclonal anti-mouse/anti-rabbit immunoglobulins (DAKO).

Western Blot analysis

In order to validate the expression or knockdown efficiency of PUS1, YBX3, FLAG/HIS/HA-tagged PUS1 or FLAG/HIS/HA-tagged YBX3 in the Flp-In HEK293 cells, we performed western blot using total cell lysates or immunoprecipitated protein samples incubated in 2X SDS-PAGE loading buffer (160 mM Tris-HCl, pH 6.8, 20 % (v/v) Glycerol, 4 % (w/v) SDS, 200 mM DTT, 0.2 % (w/v) Bromophenol blue) followed by SDS-PAGE gel separation. Total protein quantification was determined using Bradford reagent and BSA (Thermo Scientific) as a standard for calibration curve. Protein samples were loaded on 10% or 12% denaturing SDS-PAGE gel and resolved in 10X SDS-PAGE Running Buffer (250 mM Tris base, 1.92 M glycine, 1% SDS, H2O to 1000ml, pH 8.3). Proteins were transferred onto nitrocellulose membrane (GE Healthcare) using a semi-dry transfer apparatus (Bio-Rad) at 20V. For blocking we used 5% non-fat milk solution in TBST buffer (0.02% Tween-20, 20 mM Tris-HCl, 150 mM NaCl) for 1h to overnight at 4° C. We incubated with primary antibody 1:1000 (1 µg/ml) in TBST for 3 h to overnight followed by incubation with polyclonal anti-mouse or anti-rabbit HRP-conjugated secondary antibody (Dako) 1:2000 in TBST, for 1-2 h at room temperature. Vinculin antibody was used with a 1:4000 dilution in TBST (Sigma-Aldrich). Using the ECL detection reagent (GE Healthcare) we developed the membrane and visualized the protein bands using ImageQuant LAS 4000 digital imaging system (GE Healthcare) and quantified the bands using Multi Gauge v. 3.2 software (FujiFilm).

PAR-CLIP

PAR-CLIP experiment was performed as described (Hafner et al., 2010b) with the following modifications. Flp-In HEK293 cells stably expressing FLAG/HIS/HA-tagged PUS1 or FLAG/HIS/HA-tagged YBX3 were grown overnight in SILAC media containing SILAC-DMEM high glucose (PAA) supplemented with 10% (v/v) dialyzed fetal bovine serum (Sigma-Aldrich), 1% (v/v) 2 mM L-glutamine (PAA), 0.05 mg/mL Lysine (Sigma-Aldrich) and 0.03 mg/mL Arginine (Sigma-Aldrich), labeled with 100 µM 4SU (ChemGenes) or 200 µM 6SG (Sigma-Aldrich) and induced with doxycycline (1 µg/ml).

After 12-16h of labeling cells were crosslinked with 365nm UV light (UV Stratalinker, Stratagene) and lysed in NP-40 lysis buffer (50 mM HEPES-KOH, pH 7.5, 150 mM KCl, 2 mM EDTA-NaOH, pH 8.0, 0.5% (v/v) NP40 substitute, 0.5 mM dithiothreitol (DTT) fresh, complete EDTA-free protease inhibitor cocktail fresh). The cleared cell lysate was partially digested with RNaseT1 (20 U/µl) (Fermentas) or RNaseI (1 U/µl) (Ambion) and FLAG/HIS/HA-tagged YBX3 or FLAG/HIS/HA-tagged PUS1, respectively, was immunoprecipitated using anti-FLAG antibody (0.25 mg/ml) (Sigma, F1804) conjugated to Protein G Dynabeads (Invitrogen) for 1 h on rotating wheel at 4° C. Beads were then treated with 0.5 U/µl calf intestinal phosphatase (Giorgi et al., 2009) (Henneberg et al., 2010) for 1 h at 37°C and crosslinked RNAs-protein complexes were labeled with radioactive gamma-ATP (P-32) 9,25MBq (NEG 502A, NEN) and T4 polynucleotide kinase (T4 PNK) (Fermentas). PUS1 RNA-protein complexes were separated on SDS-PAGE, and then transferred to a Whatman 100% Pure Nitrocellulose membrane (Protran BA85) and the complex migrating at a molecular weight of ~47kDa, was excised and Proteinase K digested for 1 h at 55°C. Immunoprecipitated RNA was eluted using phenol-chloroform extraction and precipitated overnight in 100% EtOH. YBX3 RNA-protein complexes were cut directly from the SDS-PAGE gel at the molecular weight of ~ 60kDa and electroeluted in D-Tube Dialyzer Midi Tube with 1X MOPS SDS running buffer (NuPAGE MOPS SDS Running Buffer (20X), Invitrogen). The electroeluate was treated with 2 mg/ml Proteinase K (Roche) for 1 h at 55°C. Immunoprecipitated RNA was recovered by phenol-chloroform extraction and ethanol precipitation. Small RNAs were ligated to a bar-coded 3' adapter DNA oligonucleotide and a 5' adapter RNA oligonucleotide. cDNA synthesis was performed using SuperScript III reverse transcriptase and a 3' PCR primer. PCR amplification of the library was done after a test PCR to identify the optimal number of cycles per library using the PAR-CLIP 3' and 5' PCR primers.

siRNA-mediated knockdown

We used two siRNA duplexes (Sigma) to efficiently knockdown PUS1 protein levels in Flip-In HEK293 cells. Cells overexpressing FLAG/HIS/HA-PUS1 or non-expressing cells were transfected with siRNAs targeting the coding region of PUS1, using Lipofectamine RNAiMAX (Invitrogen). Knockdown efficiency was tested using Dual-Glo Luciferase system (Promega), qRT-PCR and Western Blot. For the Dual-Glo Luciferase assay PUS1 coding sequence was recombined from pENTR4 vector into psiCHECK2 vector using LR Recombinase (Invitrogen). psiCHECK2 PUS1 plasmid

was co-transfected together with siRNA in Flp-In HEK293 cells using Lipofectamine 2000. We perfomed PUS1 siRNAs knockdown using the following sequences:
siRNAduplex1, CGCACAACUUCCACAAUUU[dT][dT],
AAAUUGUGGAAGUUGUGCG [dT][dT]
siRNAduplex2, GUCGGGUCCUCACAAUUCA[dT][dT],
UGAAUUGUGAGGACCCGAC [dT][dT] (Sigma).
YBX3 cDNA was recombined into a psiCHECK2 (Promega) vector by LR recombination from the pDONR221 vector. The psiCHECK2 plasmid was transfected into HEK293 Flp-In TRex cells using Lipofectamine 2000 (Invitrogen) and siRNAs (Sigma) targeting YBX3 coding sequence were co-transfected with the psiCHECK2 plasmid. Luciferase activity was measured using the Dual-Glo Luciferase Assay system (Promega) with a TECAN infinite M200 reader and YBX3 knockdown was quantified by the ratio of the relative intensities of Renilla/Firefly luciferase compared to mock transfected cells. We performed YBX3 siRNA knockdown using the following siRNAs:
siRNAduplex1,AGACGUGGCUACUAUGGAA[dT][dT],
UUCCAUAGUAGCCACGUCU [dT][dT],
siRNAduplex2,AAAUCGAAAUGACACCAAA[dT][dT],
UUUGGUGUCAUUUCGAUUU [dT][dT],
siRNAduplex3,UGGAGAGGCAGAAGAUAAA[dT][dT],
UUUAUCUUCAGCCUCUCCA [dT][dT] (Sigma).

Pulsed SILAC and mass spectrometry

For the pulsed SILAC transfection, cells were grown in SILAC “light” DMEM containing non-labeled aminoacids 84 mg/l $^{12}C_6{}^{14}N_4$ L-arginine and 146 mg/l $^{12}C_6{}^{14}N_2$ L-lysine. siRNAs were transfected 24h after seeding a 6-well plate with $2.5x10^4$ cells/ml, and using Lipofectamine RNAiMAX (Invitrogen). siRNA transfection was done at 60nM final concentration, whereas mock controls were treated with the transfection reagent only. After 24h of knockdown, SILAC “light” media was exchanged with SILAC “heavy” or “medium-heavy” containing 84 mg/l $^{13}C_6{}^{15}N_4$ L-arginine and 146 mg/l $^{13}C_6{}^{15}N_2$ L-lysine or 84 mg/l $^{13}C_6$-L-arginine (Sigma-Aldrich) and 146 mg/l D4-L-lysine (Cambridge Isotope Laboratories), respectively (Ong and Mann, 2006; Selbach et al., 2008). After the 24 h of pulsed labeling cells were harvested, counted and combined into equal amounts of siRNA and mock-transfected cells. Proteins were digested with LysC and trypsin overnight, desalted and fractionated by isoelectric focusing on a microrotofor Cell device (Biorad). Resulting peptides from each fraction were desalted through

STAGE Tips[67] and analyzed by LC-MS/MS on LTQ-Orbitrap Velos (Thermo Scientific) mass spectrometer by Guido Mastruoboni (in the group of Stefan Kempa, MDC). Raw data was analyzed using MaxQuant software (v1.2.2.5), in order to identify peptide sequences present in the samples with a 1% false discovery rate. The knockdown efficiency was examined by real-time quantitative PCR using primers designed specifically for YBX3 (described at qRT-PCR DNA oligonucleotdies section). For further analysis, only proteins with two SILAC ratios were considered and H/M ratios of two biological replicates were averaged. For further analysis, only proteins with two SILAC ratios were considered, which represented 1892 proteins, and H/M ratios of two biological replicates were averaged.

Quantitative real-time PCR

Total RNA was isolated from the knockdown and mock transfected cells using miRNeasy MiniKit (QIAGEN). Single stranded cDNA was synthesized from total RNA using 18nt and 19nt oligodT primers mix and SuperScript III Reverse Transcriptase (Invitrogen) using manufacturer's instructions. Real-time PCR was performed using Power SYBR Green PCR master mix (Applied Biosystem) as described by the manufacturer on the StepOne Real-Time PCR System (Applied Biosystem). PUS1 qRT-PCR primers were desgined using Roche online tool (http://lifescience.roche.com/shop/products/universal-probelibrary-system-assay-design) and synthesized by Eurofins.

PAR-CLIP analysis

PAR-CLIP analysis was performed using a computational pipeline described previously (Lebedeva et al., 2011a). Illumina sequencing reads were processed towards adapter removal using FAR 2.0 (the flexible adapter remover: http://sourceforge.net/projects/flexbar/) (Dodt et al., 2012) and aligned using BWA 0.6.2 (Li and Durbin, 2009) to the human genome sequence assembly hg18 or hg19 or to a set of reference sequences RefSeq pre-mRNA (Kent et al., 2002). The unique alignments were collected into a pileup file using SAMtools 0.1.8 (Li and Durbin, 2009) which was subsequently analyzed using a set of custom scripts that creates read clusters as continuous stretches of read coverage on one strand. The read clusters are characterized by two different quality measures, the nucleotide transition events (T to C in case of 4SU libraries, G to A in case of 6SG libraries) and the cluster score as a

measure of nucleotide variation. Clusters aligning antisense to the annotated direction of transcripts were disregarded and used to determine a 5 % false positive threshold. The resulting bed file includes clusters from PAR-CLIP libraries which were stringently filtered and pooled from different PAR-CLIP experiments into a consensus or conservative set (analysis performed by Miha Milek in the group of Markus Landthaler). The consensus set includes reads from different experiments which were counted for T to C or G to A conversions, and additionally, reads from two out of three experiments were supporting the cluster. For the stringent filtered conservative set the reads and conversions from all three experiments have to support the cluster. The most frequent T to C conversion within cluster has been annotated as the preferred crosslinking site and provides a nucleotide resolution profile of the binding sites.

Computational analysis of the YBX3 binding motif

To find possible YBX3 binding motifs, Emanuel Wyler (in the group of Markus Landthaler, MDC) has analysed the 40 nucleotide long CSDA 4SU conservative cluster set sequences for word enrichement. Occurences of 7-mers were counted and ranked by enrichment over equidistribution. From the top motifs, a consensus sequence and a corresponding position weight matrix (pwm) was deduced. This pwm was then used to score all 3'-UTRs in the hg19 Refseq annotation. The pwm was used to score all 3'-UTRs in the hg19 Refseq annotation. We assumed that the U/A only motifs constitute artifacts due to the overrepresentation of these bases in 3'-UTRs. We further performed a motif search in all mRNAs and all 3' UTRs and calculated for each a motif score. We ranked our list based on motif scores and we further analyzed the top 1000 genes according to 3' UTR score and top 1000 FPKM T-C normalized transcripts identified in PAR-CLIP.

Identification of YBX3 protein-protein interactors

YBX3 protein interaction partners were detected using a SILAC proteomics approach described in (Hubner et al., 2010; Hubner and Mann, 2011) and optimized in our laboratory to identify RNA-dependent interactions and direct RNA-independent interactions (Gregersen et al., 2014a). HEK293 Flp-In cells overexpressing FLAG/HIS/HA YBX3 were labelled with SILAC "light" $^{12}C_6$ $^{14}N_4$ L- arginine and $^{12}C_6$ $^{14}N_2$ L-lysine (Sigma-Aldrich) and SILAC "heavy" media containing $^{13}C_6$ $^{15}N_4$ L-arginine and $^{13}C_6$ $^{15}N_2$ L-lysine and induced overnight with 1 µg/ml doxycycline. For complete incorporation of labelled amino acids, cells were split for 5-7 passages in SILAC media

before performing the experiment. To be able to identify specific interaction partners from contaminants we set up a label-swap experiment, where overexpressing cells were labelled with SILAC “light” and non-induced cells with SILAC “heavy”. To identify YBX3 indirect protein-protein interactions that might interact through RNA we used SILAC labeling in combination with 100 µM 4-thiouridine RNA labeling and UV crosslinking. We used aproximatively $5x10^7$ per each experiment, which were combined into 1:1 heavy with light cell number ratios. We lysed the cell pellets using the NP-40 lysis buffer for the RNA-dependent interactions and NP-40 lysis buffer plus benzonase (1 U/µl) and Mg^{2+} (1mM) for the RNA-independent interactions, for 30 minutes on ice, and subsequently centrifuge at 15000g for 20 minutes at 4°C. The supernatant was further treated with a low RNase T1 concentration (20U/ul) for 5’ at 22°C for the RNA-dependent interactions or with a high RNase T1 concentration (50U/ul) for 5’ at 37°C for RNA-independent interactions. For both RNA-dependent and independent experiments we perfomed a double-affinity immunoprecipitation, using ANTI-FLAG M2 Magnetic Beads (Sigma) followed by 3X FLAG peptide (Sigma-Aldrich) elution and subsequently HA-purification using uMACS Anti-HA MicroBeads and uMACS columns (Miltenyi Biotec). After application to the column, immunoprecipitated proteins were washed three times with 800 µl ice-cold Wash Buffer I (150mM NaCl, 50mM TRIS-HCl pH 7.5, 5% glycerol, 0.05% NP-40) followed by two times 500 µl ice-cold Wash Buffer II (50mM NaCl, 50mM TRIS-HCl, pH 7.5, 5% glycerol). Immunopurified proteins were in-column digested with 150 ng trypsin in 25 µl 2M urea in 100 mM Tris-HCl pH 7.5, 1 mM DTT at room temperature for 30 min, eluted two times with 50 µl 2M urea in 100 mM Tris-HCl pH 7.5, 5 mM iodoacetamide. Digestion was stopped the next day by adding 1 µl trifluoroacetic acid and peptides were purified on C-18 Stage Tips and reconstituted in 15 µl of 0.5 % acetic acid in water (Rappsilber et al., 2003). The samples were analyzed by LC-MS/MS on Thermo LTQ Velos Orbitrap mass spectrometer. Raw data was analyzed with the MaxQuant (v1.2.2.5) by Guido Mastruoboni (in the group of Stefan Kempa, MDC) proteomics pipeline (Cox and Mann, 2008) with peptide assignments false discovery rate of 0.01.

Cloning of 3’ UTRs into the psiCHECK2 vector

I used a dual-luciferase based-assay to test the functionality of the 3’ UTRs of candidate genes identified by PAR-CLIP and pSILAC protemics. The 3’UTRs of the candidate genes were PCR-amplified using oligos flanking the 5’ and 3’ end of each UTR and containing XhoI/NotI (Fermentas) restriction sites (in bold, described at the 3’

UTR cloning DNA oligonucleotides section). The PCR amplicons were restriction digested with the specific restriction enzymes and ligated into the psiCHECK2 vector (Promega) using T4 DNA Ligase (Fermentas).

psiCHECK2 3'UTR dual luciferase reporter assay

PCR-amplified fragments of the 3' UTRs of candidate genes were cloned behind a Renilla luciferase reporter gene of the psiCHECK2 vector (Promega) using a restriction/ligation reaction. YBX3 expression was silenced using siRNAduplex3 tested previously (Sigma) at a final concentration of 60nM. Transfection was performed using Lipofectamine RNAiMAX (Invitrogen) and DMEM (Gibco) in 2x10-cm plates seeded with $1x10^5$ cells/ml. After a 24 hours siRNA knockdown the cells were split to a 96-well plate at a density of $3x10^5$ cells/ml, in a 50 µl volume/well. After 24h the psiCHECK2 vectors containing 3' UTRs or empty control were transfected using Lipofectamine2000 (Invitrogen) and Optimem (Gibco). 24 hours after psiCHECK transfection, we measured the luciferase activity using Dual-Glo Luciferase System (Promega) protocol and TECAN infinite M200 reader for luminescence measurement. Ratios of Renilla/Firefly luciferase activity were measured for each 3'UTR candidate and normalized to the ratios of Renilla/Firefly luciferase activity of the empty vector.

CMCT treatment of total RNA

Total RNA was isolated from Flp-In HEK293 cells using mRNeasy kit (Qiagen) and final RNA concentration and purity was measured by NanoDrop analysis. Total RNA was submitted to three different treatments, with one sample non-CMCT treated, 0.167 M CMCT treated (+CMCT) and 1 M CMCT treated (++CMCT) for 10 minutes at 37°C. 30ug total RNA was first denatured in EDTA 5 mM at 90°C for 3 minutes, followed by adding either CMCT buffer (-CMCT), 0.167 M CMCT (+CMCT) or 1 M CMCT (++CMCT) (Sigma-Aldrich) solution prepared in CMCT buffer (BEU buffer) containing 50 mM Bicine, 4 mM EDTA, 7 M Urea. CMCT treatment was performed at 37°C for 2 minutes in -CMCT/+OH conditions or for 10 minutes for the +CMCT and ++CMCT conditions. After several rounds of washing and precipitation of treated/non-treated RNA, the hydrolysis of CMC-U and CMC-G adducts is performed by incubating the RNA pellets with a mild alkaline treatment in 50 mM Na_2CO_3 pH 10.3, at 37°C for 4 hours. All samples were subsequently washed, precipitated and resuspended in RNase

free H2O at a concentration of 1 µg/ul and used further for determining position specific pseudouridine modifications in RNA using a primer extension assay.

Primer extension assay

Primer extension is based on the reverse transcription stops at an RNA bulky adduct or stable secondary structure and can be visualized using radiolabeled olignucleotides and primer extension on a sequencing gel electrophoresis. For tRNA pseudouridylation sites we used a previously published reverse transcription primer for assaying ψ-sites in human mitochondrial $tRNA^{Lys}$ (UUU) (Patton et al., 2005b):
5' GAGTTGGTTCTCTTAATC 3' (Eurofins).
For the PAR-CLIP CCT3 mRNA target candidate we designed a primer in vicinity of the preferred crosslinking site within the CCT3 mRNA PAR-CLIP binding site:
5' GTGCATTGAAGTAGCCTTGC 3'.
For the tRNA primer extension we used 1 µg/ul to 1.5 µg/ul total RNA extracted from Flp-In HEK293 cells. The RNA was derivatized using CMCT and reverse transcribed using the radiolabeled primer. For mapping the known positions of each of the four RNA nucleotides downstream of the primer, we added dideoxynucleotides (0.5 mM, Jena Bioscience) in the "poisoned" primer extension reaction of non-derivatized RNA (Alfonzo et al., 1999). For mapping the pseudouridine position in CCT3 gene we used a sequencing ladder synthesized using the protocol described in the Sequenase Version 2.0 PCR Product Sequencing kit (Affymetrix USB). We radiolabeled 500 ng of oligo (Eurofins) in 1.5 ul T4 PNK buffer (Fermentas) , 1.5 ul T4 PNK (Fermentas) and 2 µl gamma-ATP (P-32) 9,25MBq (NEG 502A, NEN/PerkinElmer) and incubated 30 minutes at 37°C. Labeled oligos were purified on a 15% 7.5M Urea-PAA gel (SequaGel - UreaGel System EC-833, National Diagnostics) followed by extraction and precipitation overnight at -20°C using 100% EtOH and 0.5 µl Glycoblue. The radiolabeled oligos were rehydrated in MilliQ H_2O. Primer extension of the treated/non-treated RNA was performed using 1 µl total/modified RNA, 1 µl radiolabeled oligo, and 2 µl HB buffer 5X (Hybridization Buffer) containing 250 mM Tris-HCl pH 8.3, 300 mM NaCl and 50 mM DTT and H_2O to a hybridization volume of 10 µl. The hybridization mix was incubated at 70°C for 10 minutes and slowed down to 30°C for 10 minutes followed by 2 minutes on ice. Primer extension reaction was initiated by adding 1 µl PPE buffer 10X (Primer Extension Buffer) containing 500 mM Tris-HCl, 600 mM NaCl, 100 mM DTT and 60 mM MgCl2, 1 µl dTNPs (0.5 mM), 7.5 U/µl SuperScript III Reverse Transcriptase (Invitrogen) and H2O to a total volume of 20ul. The primer

extension reaction mix was incubated at 55°C for 50 minutes and the reaction was stopped by adding 20 µl formamide mix (80%(w/v) deionized formamide, 1 mg/ml xylene cyanol, 1 mg/ml bromphenol blue, 10 mM EDTA pH 8.0). Before loading on the gel the primer extension products were denatured at 95°C for 1 minute and placed on ice. For the “poisoned” primer extension to map the first positions of each of the four nucleotides, the primer extension reaction was performed excluding one dNTP at a time and replacing it with a corresponding ddNTP, for example for a “G-stop”, the mix would contain 2ul of dATP (0.5 mM), 2ul dTTP (0.5 mM), 2ul dGTP (0.5 mM), and 2ul ddCTP (0.5 mM).

Sequencing Ladder

In order to obtain a sequencing ladder we used the Affymetrix USB sequenase kit which contains a specific polymerase for manual DNA sequencing, derived from a T7 polymerase but with no 3’-5’ exonuclease activity and lower sensitivity to secondary structure. Using the designated primers for CCT3 mRNA template, we amplified a 130 bp PCR product using KOD Hot Start Polymerase (Novagen), and the RT primer extension oligonucleotides:
RT-Reverse: 5’ GTGCATTGAAGTAGCCTTGC 3’
Forward: 5’ ACGGCAGTTCTGCTACTGCG 3’.
The resulting PCR products were purified using the Agencourt AMPure XP beads (Beckman Coulter) according the manufacturer’s protocol. We used 20 to 30 ng of purified PCR product for the manual Sanger sequencing reaction, which was pre-treated with Exonuclease I (Exo I) and Shrimp Alkaline Phosphatase (SAP) according to the protocol described in the kit, in order to remove excess of dNTPs and primers from the PCR reaction. The PCR product was denatured at 100 °C for 2-3 minutes and annealed to the RT oligonucleotide by placing it on ice for 5 minutes. The labeling reaction was performed using the Labeling Mix 5X, 10 µCi P32 ATP (PerkinElmer) and the Sequenase polymerase. The sequencing reaction mix was distributed to pre-warmed termination mixes containing either ddA, ddC, ddT, or ddG and further incubated for 10 minutes at 37°C. The reaction was stopped with the Stop Solution included in the kit. The resulting ladder will contain four samples, each corresponding to one of the four DNA oligonucleotides.

Sequencing gels

The sequencing ladder together with the primer extension products of CMCT treated RNA were separated on a 10% 7.5 M Urea-PAA sequencing gel. For the manual sequencing system we prepared 150 ml gel mix using the SequaGel system (National Diagnostics) and cast the gel using the 35x45cm sequencing plates and a shark comb (Bio-Rad). The glass plates were previously prepared by coating with a layer of Gel Repel sequencing glass plate coating. Fractionation was achieved by loading 2-5 µl of sample in each well of the sequencing gel and perform an electrophoresis in TBE buffer (89mM TRIS-borate pH 8.3, 2mM Na2EDTA) at 40W for 3 hours. For the electrophoresis we used an Owl S3S Aluminum Backed Sequencing System, 35x45cm (Thermo Scientific). The gel was transferred to filter paper (Filterbogen FP 0859, 580x580 mm, 61g/qm, Th. Geyer), covered with Saran foil and vacuum dried at 80°C on the SG210D SpeedGel dryer (Thermo Electron Corporation) for 30 minutes. After cooling down to room temperature the dried gel was exposed to a phosphoimager screen (Amersham Bioscience Hypercassette) overnight. The phosphoimager screen was scanned using a FujiFilm-FLA7000 reader and the image was analyzed using MultiGauge v3.2 (FujiFilm).

RESULTS I

PAR-CLIP identifies thousands of PUS1 binding sites in the transcriptome

In order to characterize PUS1 binding sites at transcriptome-wide level we employed PAR-CLIP (PhotoActivatable Ribonucleoside enhanced Crosslink and Immunoprecipitation) (Hafner et al., 2010b). I first generated a cell line that stably expressed PUS1, by transfecting a pFRT/TO/HIS/FLAG/HA PUS1 construct into Flp-In T-REx HEK293 cells. This allowed recombinant PUS1 to be integrated into a single genomic Flp recombinase site. We tested the expression of recombinant PUS1 in HEK293 cells, using doxycycline at 1 µg/ml (Figure 6A). Using an anti-PUS1 antibody we could observe a similar expression level of epitope-tagged PUS1 compared to the endogenous PUS1 protein (Figure 6B). PUS1 has two different isoforms, one with a molecular weight of 37 kDa and the second with a molecular weight of 44 kDa. We could observe both isoforms as the two lower bands on the western blot using anti-PUS1 antibody. When the expression of recombinant PUS1 was induced, we noticed a partial overlap with the background band below 55 kDa although this had a similar intensity to the one in the non-induced lane (Figure 6B). Taking advantage of the PAR-CLIP method which uses 4SU labeling of nascent RNAs and UV crosslinking at 365 nm, we were able to immunoprecipitate protein-RNA crosslinked complexes using anti-FLAG antibody. The immunopurified PUS1-RNA crosslinked complexes were radio-labeled and the pshosphorimage indicated a crosslinked protein-RNA complex at the expected molecular size (Figure 6C).

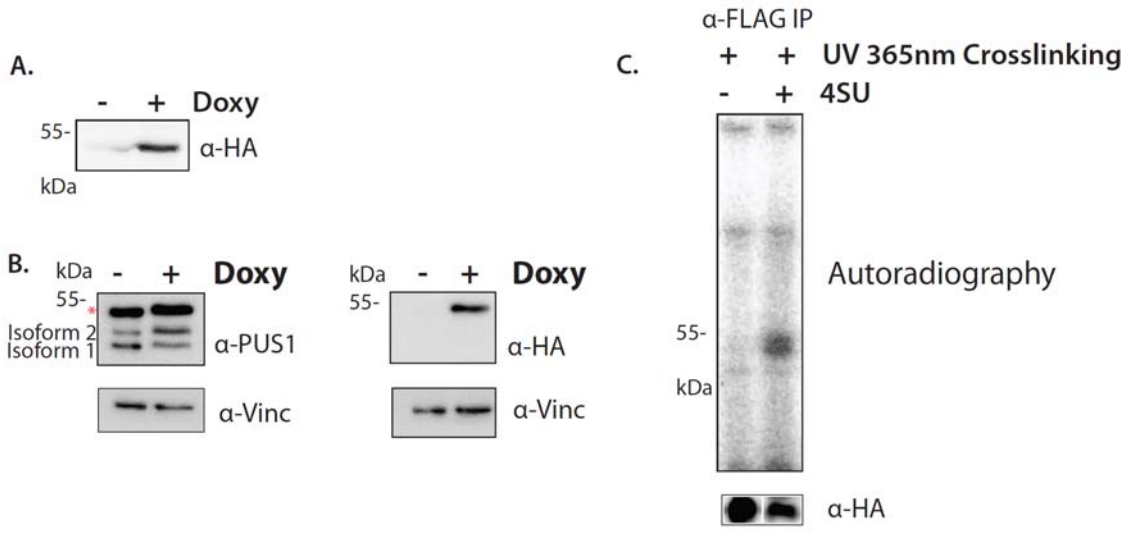


Figure 6. Recombinant PUS1 can be expressed in Flp-In HEK293 cells and crosslinks to RNA *in vivo*. (A) Western Blot of the SDS-PAGE gel for the doxycycline induced expression of the recombinant FLAG/HIS/HA-PUS1 protein probed with anti-HA antibody. (B) PUS1 expression was also tested in non-induced and induced cells with an anti-PUS1 antibody which revealed the two PUS1 isoforms (isoform1 37kDa, isoform2 44kDa). A strong unspecific band just below 55kDa (red asterisk) overlapped with the doxycycline induced epitope-tagged PUS1 at 50 kDa (left panel). The intensity of the recombinant PUS1 band suggested that the levels of overexpression were similar to endogenous levels. The right panel shows the same western blot probed with anti-HA antibody. The lower panels show the loading control probed with anti-vinculin antibody. (C) Autoradiography of the SDS-PAGE gel detected the radiolabeled protein-RNA crosslinked immunoprecipitated complexes at the molecular size of 50 kDa in the extracts from the 4SU labeled cells only, due to the 4SU-enhanced crosslinking efficacy. The lower panel shows the western blot of the immunoprecipitated complexes probed with the anti-HA antibody.

The RNA fragments extracted from these complexes were adapter ligated and reverse transcribed into a cDNA library. The library was submitted for Next-Generation Sequencing on an Illumina platform. The T to C transition in sequencing reads which are present at the crosslinked uridine, permitted the identification of RNA-binding sites at nucleotide resolution. The sequencing reads were mapped to the genome and analyzed using a fully computational PAR-CLIP pipeline (as described in the Materials and Methods). Briefly, the sequencing reads were processed in order to remove the adapters, and aligned to the human genome assembly hg18. The unique alignments were clustered together and the resulting clusters were scored using a 5% false positive rate as a cutoff. Filtered clusters were used to annotate the binding sites in the targeted transcripts. The PAR-CLIP sequencing reads had a high number of T-C transitions and a high number of edited reads in comparison to the perfectly matched reads (Figure 7A) suggesting that the majority of sequence reads were recovered from PUS1-RNA crosslinked complexes. The computational analysis of the PAR-CLIP libraries identified 9,911 and 11,541 raw clusters, respectively. The clusters were filtered to 6,506 clusters for the first replicate and 8,161 clusters for the second biological replicate (Table 2), which mapped to 2,571 genes and 3,526 genes, respectively. We performed a consensus and conservative analysis for the two biological replicates for which clusters require reads from both libraries or transitions from both libraries, respectively. The consensus analysis identified 8,338 filtered clusters in 3,123 genes, while the conservative analysis identified 6,087 filtered clusters in 2,516 genes (Table 2).

Table 2. PAR-CLIP libraries summary statistics.

Cluster Set	Libraries	Mapped reads	*Kept reads	Raw clusters	Filtered clusters
PUS1_consensus	2			11602	8338
PUS1_conservative	2			7442	6087
PUS1_replicate_1	1	2.8 M	0.4 M	9911	6506
PUS1_replicate_2	1	8.9 M	0.6 M	11541	8161

*Uniquely mapped reads kept for further analysis

The PUS1 binding sites mapped predominantly to mRNA coding sequences with over 2,000 clusters identified in the CDS. Around 1,500 clusters mapped to 3' UTR regions as well, while 5' UTR binding sites were represented by approximately 200 clusters. When comparing the exon versus intron coverage of PAR-CLIP clusters, over 3,000 binding sites mapped to exons, while intronic binding sites were aproximately three times less frequent in PUS1 data (Figure 7B).

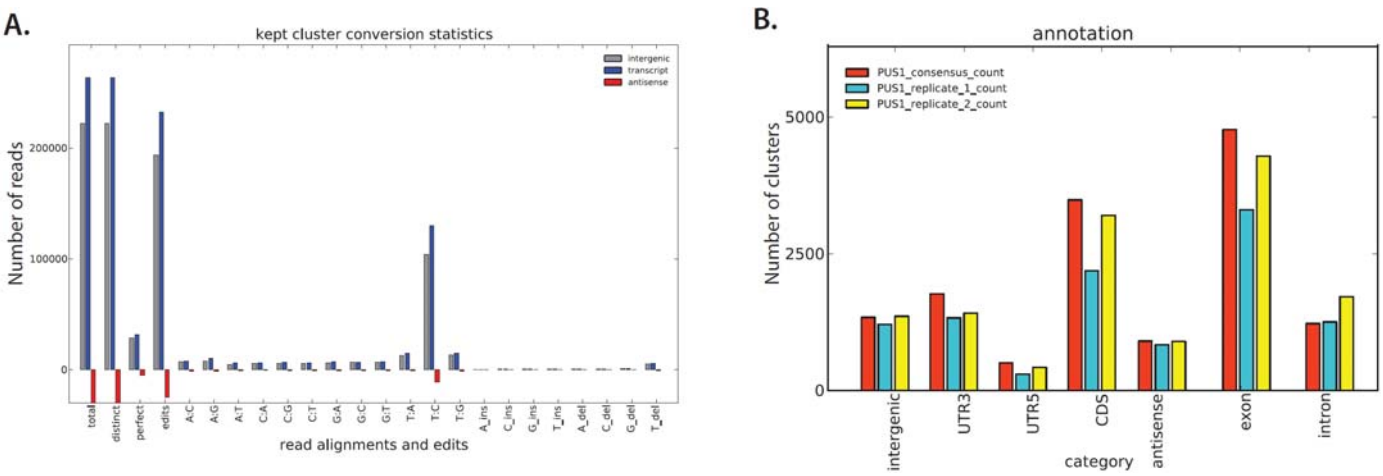


Figure 7. PAR-CLIP library statistics and annotation to different transcript regions shows PUS1 binding preferences. (A) The quality of the PAR-CLIP library is shown by the high number of reads with T-C transitions from the total number of reads as well as a high number of edited reads compared to perfect match reads which were recovered from the PUS1-RNA crosslinked complexes. (B) PUS1 binding sites shows a higher preference for coding sequence and a lower preference for 3'UTRs or 5'UTRs. Three times to five times more clusters were annotated to exons than to intronic regions.

Several known pseudouridine sites in tRNALys (UUU) (Sibert and Patton, 2011) and U6 snRNA have been described previously (Czerwoniec et al., 2009; Dunin-Horkawicz et al., 2006; Machnicka et al., 2013; Yu et al., 2011). These sites could be also found in the RNA modification database (http://modomics.genesilico.pl/). In order to confirm the presence of PUS1 reads and T-C transitions in these transcripts we aligned the

sequenced PAR-CLIP reads to a collection of tRNAs, rRNAs and snRNAs. By using the Integrative Genomics Viewer (IGV) (Robinson et al., 2011; Thorvaldsdottir et al., 2013) to visualize the PUS1 PAR-CLIP read coverage, we observed that position U27 and U28, which are pseudouridylated in mitochondrial tRNALys (UUU), as well as positions U31, U40, and U86, which are pseudouridylated in U6 snRNA (Wu et al., 2011; Yu et al., 2011) were indeed bound by PUS1 (Figure 8A,B). Moreover, the reads obtained from the PUS1 PAR-CLIP mapping to U6 snRNA contained a high number of T-C transitions at the U31, U40, U85 positions which were previously detected as pseudouridine sites on U6 snRNA with the exception of U85 which is next to the known pseudouridylated U86 (Figure 8B). In contrast, PUS1 reads mapping to the tRNALys (UUU) contain a high number of T-C transitions at the U47 and U66 positions which differ from the known pseudouridine positions at U27 and U28 (Figure 8A), but that could be in proximity of the pseudouridylated sites when the cloverleaf structure becomes the tertiary "L-shaped" tRNA structure. These results show that our PAR-CLIP experiment detects known sites of pseudouridylation and possibly potential novel sites.

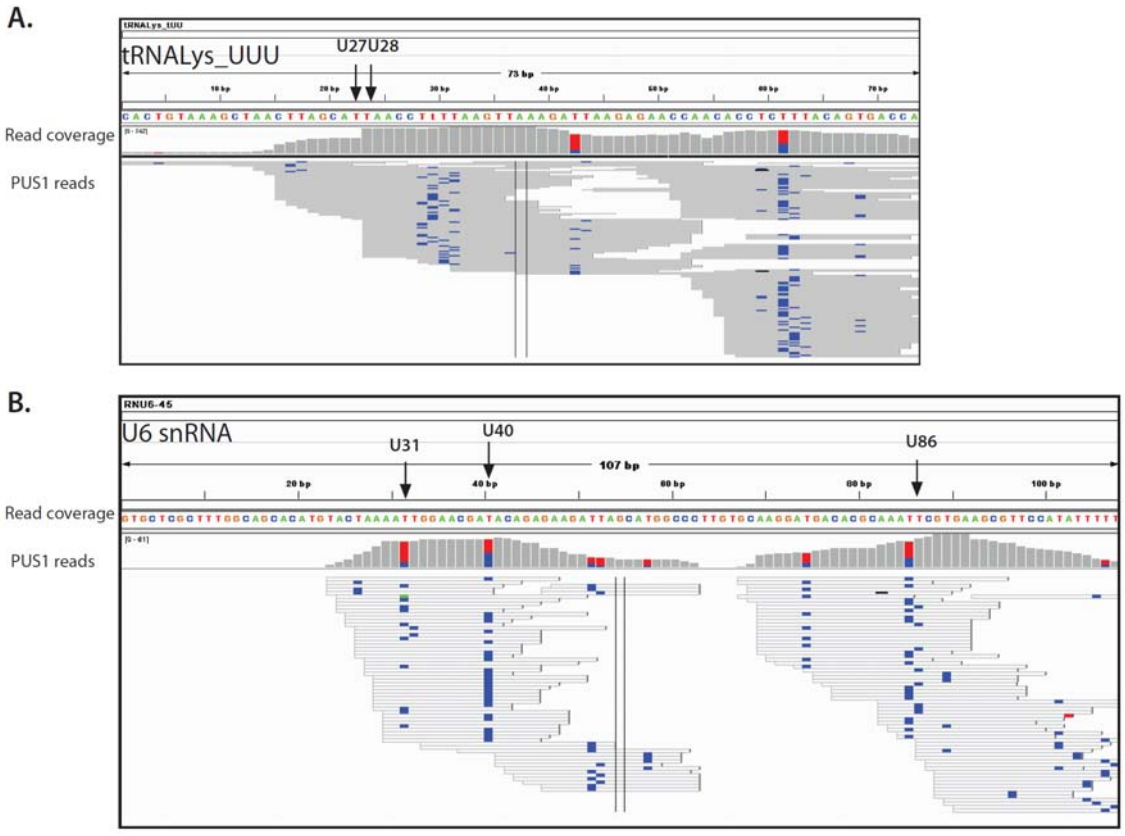


Figure 8. PUS1 binding sites mapped to tRNA and snRNA sequences. (A) PUS1 PAR-CLIP coverage on the tRNALys (UUU) sequence shows a distribution of reads mapping to the tRNA with a high

number of T-C transitions at position U47 in the variable loop and U66 in the T stem (TψC stem) of the tRNA clover leaf structure. PUS1 reads mapped to the pseudouridylated sites U27 and U28 as well (black arrows), but crosslinking events had lower frequency at these positions. (B) PUS1 reads also mapped to spliceosomal snRNAs, exemplified here by U6snRNA, with a high number of T-C conversions at the known pseudouridylated sites U31, U40 and at U85 next to the known pseudouridine site at U86 (black arrows).

The widespread distribution of PUS1 binding sites in mRNA prompted for the investigation of functional annotations of the target transcripts. In order to gain insight into the biological function of the PUS1 targets we used DAVID annotational tool (Huang da et al., 2009a, b) to find enrichment of gene ontology terms within the top 1000 PUS1 targets. The Gene Ontology term enrichment clustered PUS1 target genes according to the biological processes and revealed different set of genes involved in translational elongation, translation, RNA processing, RNA splicing, and mRNA processing (Table 3).

Table 3. Enrichment of Gene Ontology Biological Pathway terms for top 1000 target genes.

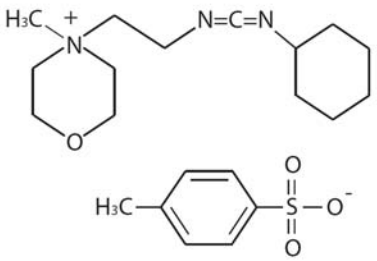
Gene Ontology Biological Pathway Term	Count	P-Value
translational elongation	53	5.7E-39
translation	87	1.0E-35
RNA processing	95	3.7E-24
RNA splicing	63	1.3E-21
mRNA processing	66	9.0E-21

In summary we have identified over 6,000 PUS1 binding sites within 2,500 genes as a filtered conservative set, with the vast majority of binding sites residing in the exonic regions over intronic regions. PUS1 binding pattern indicated a high preference for coding sequence and a lower preference for 3' UTRs. The reads recovered from PUS1-RNA crosslinked complexes mapped to known pseudouridylated tRNA sequences and spliceosomal RNA species with crosslinked uridines overlapping known pseudouridylation sites. Therefore, unlike previously PUS1 described functions, our PAR-CLIP results show that PUS1 can bind to thousands of mRNAs and prefers coding regions suggesting a regulatory function in gene expression.

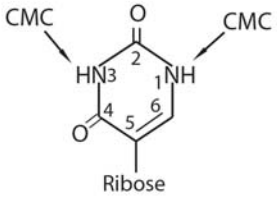
PUS1 known pseudouridylation sites are detected in tRNA

Using a similar protocol based on the CMCT treatment of RNA and primer extension (Bakin and Ofengand, 1998; Motorin et al., 2007) we set to identify the specific pseudouridine positions in PUS1 mRNA targets. We first intended to establish the method therefore we used tRNALys (UUU) as proof of principle. The important function of pseudouridines in tRNA has been discussed extensively and two sites in mitochondrial transfer RNAs, ψ27 and ψ28 have been correlated with mitochondrial myopathy and sideroblastic anemia or MLASA (Patton et al., 2005b). The CMCT compound has the property to react with U residues, ψ residues, and G residues (Figure 9A). CMCT profiling is based on the property of pseudouridine to react with CMCT and form a stable adduct, N_3-CMC-pseudouridine (N_3-CMC-ψ), which cannot be hydrolyzed under mild alkaline conditions in contrast to the CMC-U or CMC-G adducts. When we applied the optimized CMCT derivatization and primer extension method to map the two pseudouridine positions ψ27 and ψ28 in human mitochondrial tRNALys (UUU), we performed in parallel a poisoned primer extension assay on underivatized RNA (Figure 9C). The aim was to map the positions of the first A, U, C and G downstream of the reverse transcription primer using dideoxynucleotides which block the RT on the RNA template. This is particularly useful if the assayed region for pseudouridine identification is short and the ψ sites are presumably known. The primer extension analysis confirmed the presence of the two pseudouridine residues at position 27 and 28 of the human mitochondrial tRNALys (UUU). Two bands were found only in the CMCT-treated RNA, but missing from the non-CMCT RNA sample (Figure 9D).

A.

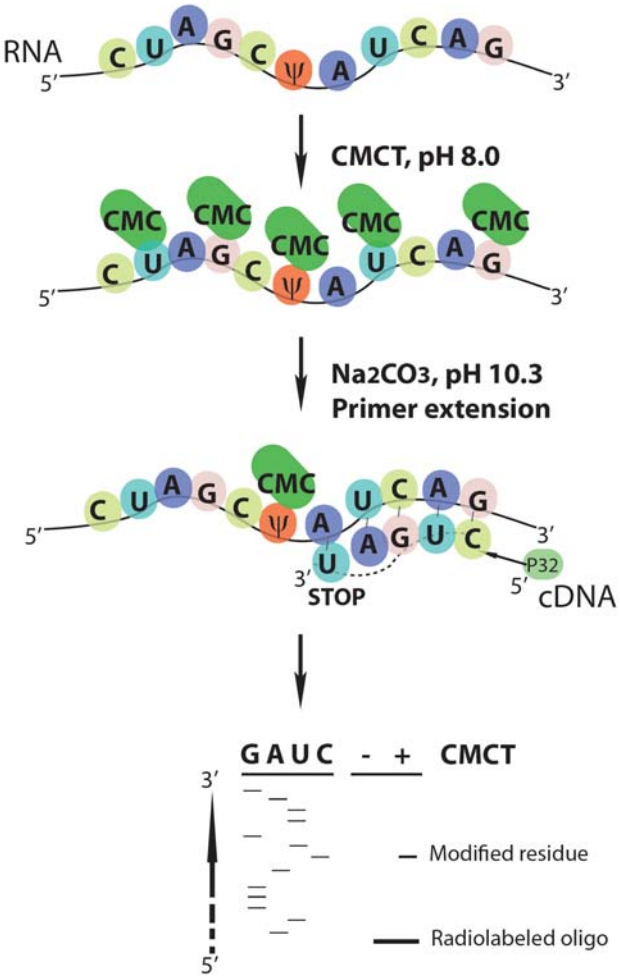


N-Cyclohexyl-N′-(2-morpholinoethyl)carbodiimide methyl-p-toluensulfonate

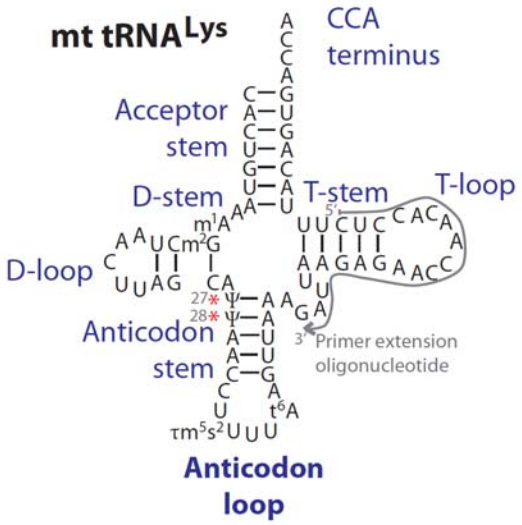


Pseudouridine

B.



C.



D.

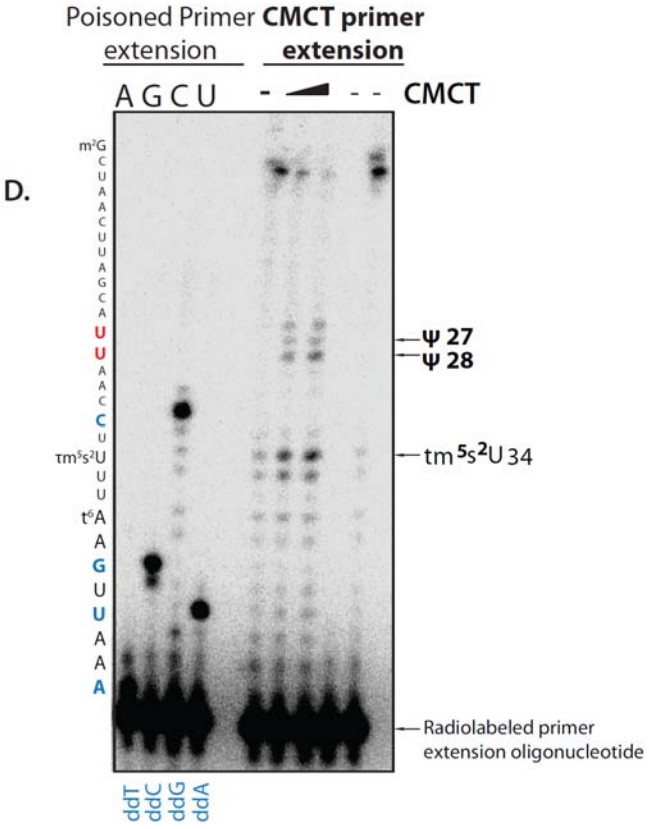


Figure 9. CMCT treatment and primer extension identifies pseudouridine sites in tRNALys (UUU). (A) CMCT structure as a carbodiimide derivate, the N-Cyclohexyl-N'-(2-morpholinoethyl)carbodiimide methyl-p-toluensulfonate (upper panel), and pseudouridine N1 and N3 CMC derivatization positions (lower panel) . (B) Schematic representation of CMCT profiling for pseudouridine detection involves a chemical treatment step of total RNA at pH 8.0, followed by a mild alkaline treatment at pH 10.3 to remove weak covalent modifications of G residues and U residues but not ψ residues. CMC- ψ bulky adducts lead to a stop in the reverse transcription primer extension step which can be visualized by gel electrophoresis. The primer extension products can be aligned to a sequencing ladder and the exact stops at putative pseudouridine sites can be identified by comparing the non-CMCT treated samples with the CMCT treated samples. (C) The canonical cloverleaf tRNA secondary structure of the mitochondrial tRNALys (UUU) (adapted from (Suzuki and Nagao, 2011)) containing the two ψ27 and ψ28 positions in the anticodon stem (marked with asterisks). (D) Phosphorimage of the primer extension sequencing gel for tRNALys (UUU). Two strong stops can be observed corresponding to the ψ27 and ψ28 only in the CMCT positive lanes, and a third band upstream, due to previously described "stuttering" events in the primer extension. Downstream of ψ27 and ψ28 positions in tRNALys (UUU), a stop in the reverse transcription lanes could be observed which might be due to the hypermodified τm5s2U34 (5-taurinomethyl-2-thiouridine) present in the anticodon loop. In parallel, the poisoned primer extension lanes synthesized using dideoxynucleotides show the position of the first T, C, G and A residues (marked in blue) downstream of the radiolabeled oligonucleotide.

Together with the reverse transcriptions stops at these sites, we observed an additional stop upstream of the ψ sites described as reverse transcriptase "stuttering" events, which takes place when the reverse transcriptase reads through the bulky adduct before halting. Downstream of the ψ sites, a strong stop could be detected on the primer extension electrophoresis, which is likely the hypermodified tm^5s^2U (5-taurinomethyl-2-thiouridine) at position 34 in the tRNALys (UUU) (Suzuki et al., 2002; Umeda et al., 2005). This stop was present in both the non-CMCT and CMCT treated RNA sample, which indicate the reverse transcriptase halting at other modified or hypermodified residues in the RNA template.

We applied a CMCT-based method to detect two physiologically important pseudouridine sites described previously in the mitochondrial tRNALys and confirmed for the first time the presence of pseudouridines in PUS1 tRNALys binding sites in HEK293 cells. This set a basis for further analysis of putative pseudouridine sites in mRNA targets of the PUS1 RNA modification enzyme.

CMC-ψ profile of CCT3 mRNA reveals two reverse transcriptase stops but does not confirm pseudouridine modification

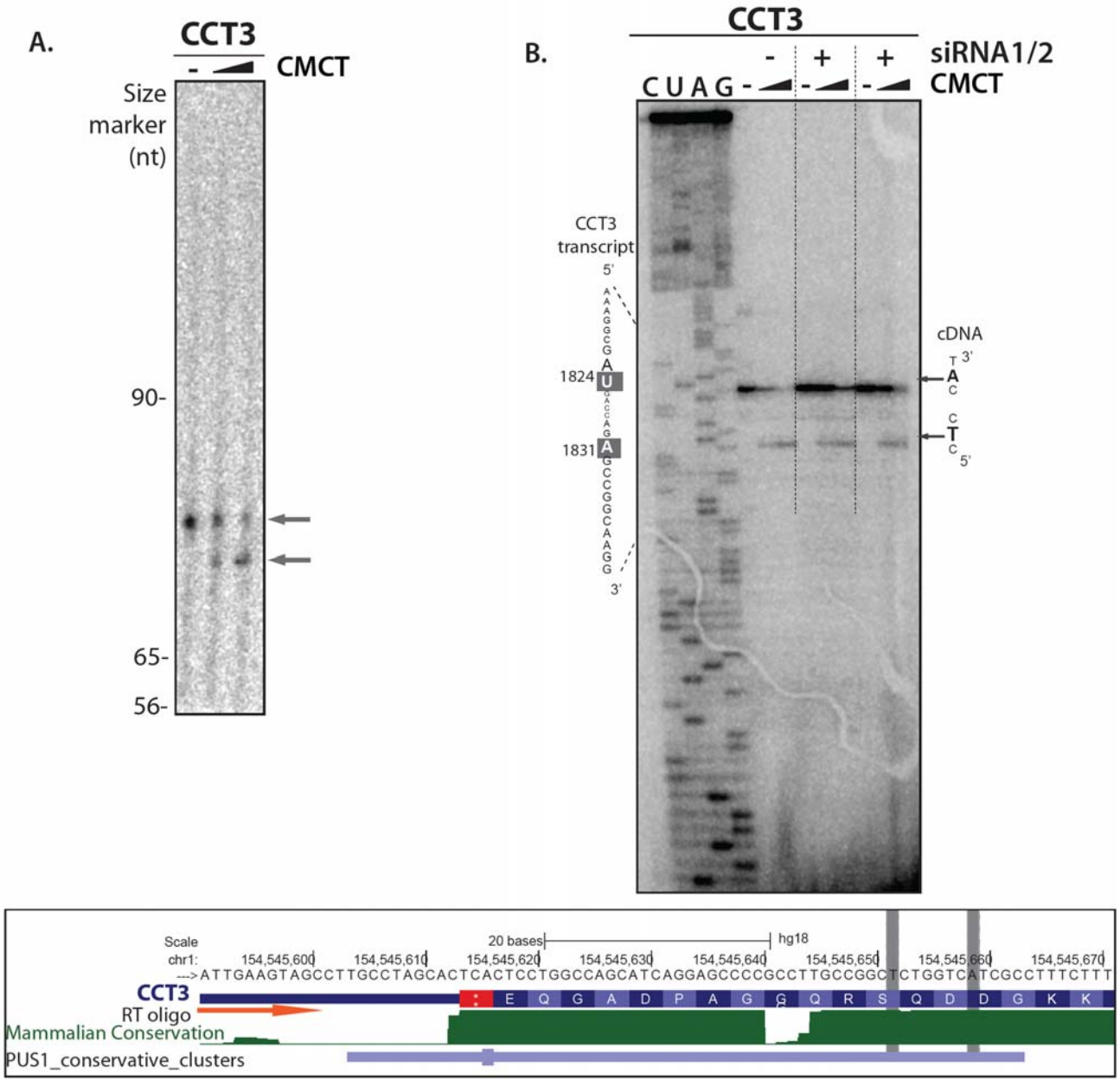
A few recent studies have revealed the presence of pseudouridines extensively in mRNAs as well as rRNAs, tRNAs and snRNAs (Table 1) (Carlile et al., 2014; Lovejoy et al., 2014; Schwartz et al., 2014b). PUS1 mRNA targets found in the PAR-CLIP experiment indicated direct PUS1 binding to mRNA, however the exact pseudouridylation sites were not detected by this method. We therefore set to identify the presence of pseudouridine sites within PUS1 binding sites using and independent approach.

Pseudouridine detection at nucleotide resolution using the primer extension assay is based on the sequencing gel electrophoresis readout of strong reverse transcriptase stops 3' to the pseudouridine site and possibly also as an additional stop at the site itself due to the "stuttering" of the reverse transcriptase (Bakin and Ofengand, 1993; Bakin and Ofengand, 1998; Denman et al., 1988) (Figure 9B). However, several points make the analysis difficult, such as the fact that the reverse transcriptase also pauses at strong secondary structures, or stops at other naturally modified residues. CMCT derivatization and primer extension coupled with gel electrophoresis has been performed only for rRNA or tRNA. We thus followed up on the CCT3 (chaperonin containing TCP1, subunit 3 (gamma)) mRNA, one of the top 100 PUS1 targeted transcripts. We tested the last exon region of the transcript (hg18; chr1:154,545,587-154,545,718), where a PUS1 binding site is located (Figure 10A), for putative pseudouridine sites. The CMCT derivatization and primer extension for the CCT3 region investigated showed a reverse transcriptase stop in the CMCT treated samples before the A1831 position of CCT3 mRNA (Figure 10A and 10B). Interestingly, we could also observe an additional strong reverse transcriptase block before the U1824 position, in absence of CMCT treatment as well as in the CMCT-treated lanes (Figure 10A and 10B). This strong stop diminished in intensity in the CMCT-treated lanes. The presence of a reverse transcriptase stop before U1824 stop was observed in three different different biological replicates (Supplementary Figure S5C). The observed results are inconsistent with expected block upon modification with CMCT, therefore we could not confirm the presence of a pseudouridine modification in the CCT3 mRNA. The block of the reverse transcriptase primer extensions could be due to several unrelated reasons some of them being discussed later on (see Discussion section).

In the light of these results, the investigation of pseudouridylation sites in PUS1 mRNA targets using an individual CMCT-based primer extension assay enabled the

identification of two different reverse transcription stops in CCT3 mRNA, however these require a further examination.

In order to investigate whether the specific PUS1 bound CCT3 region contains a pseudouridine residue that is dependent on PUS1 modification, we performed an additional assay where depleted levels of PUS1 were established using two different siRNAs, followed by CMCT derivatization and individual mRNA pseudouridine profiling. We first tested various siRNA duplexes with a Luciferase reporter assay containing the CDS of PUS1. The two different siRNA duplexes could independently reduce PUS1 protein levels down to 30% (Supplemental Figure S1). Knockdown was confirmed by qRT-PCR and western blot analysis (Figure 10C).



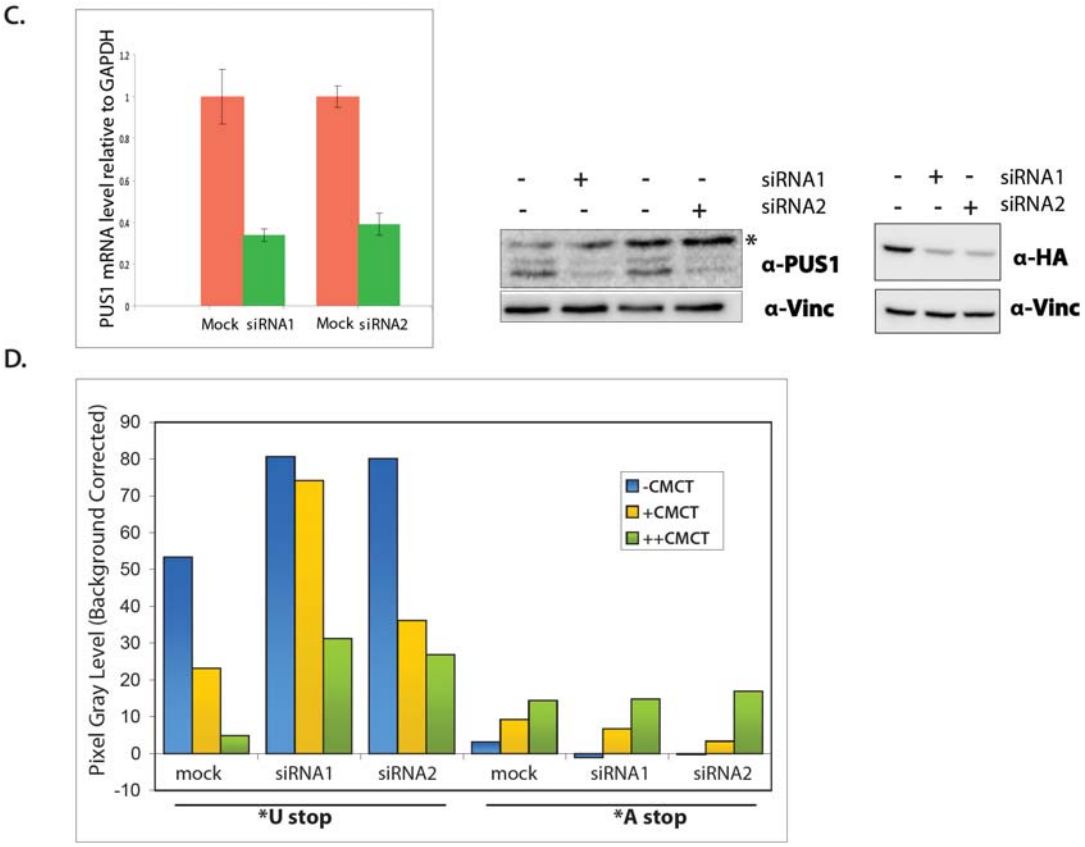


Figure 10. CMC-ψ profile of PUS1 CCT3-bound exon reveals the presence of two reverse transcriptase stops. (A) The autoradiography of CCT3 primer extension shows two stops in the reverse transcription of the CMCT treated samples (grey arrows). (B) Electrophoresis of the individual primer extension CMC-ψ profile of CCT3 mRNA next to a sequencing ladder using the same primer shows the same pattern as in (A) and indicates the position of the first reverse transcriptase stop before the A1831 (grey arrow) and the position of the upstream stop before the U1824 (grey arrow). The individual CMC-ψ profile and primer extension of PUS1 siRNA1 and PUS1 siRNA2 knockdown samples reveal an increased rate of reverse transcriptase stop before the U1824. The stop before the A1831 did not change on the autoradiography. Lower panel shows the UCSC browser visualization of the PUS1 cluster on CCT3 gene and the position of the oligonucleotide used for reverse transcription (orange arrow). The U1824 and A1831 are highlighted in grey. (C) PUS1 knockdown efficiency tested for the two siRNAs by qRT-PCR and western blot probed with PUS1 antibody for the endogenous protein or with an anti-HA antibody. (D) Densiometry of pixel gray levels of the reverse transcription stops measured an increase in signal intensity of siRNA knockdown samples before the U1824 in the CMCT-treated (yellow and green bars) or non-treated samples (blue bars) (*U stop) comparing to mock transfection. However, the signal intensity of the A1831 stop did not show notable changes between the mock and the PUS1 knockdown conditions (*A stop).

We next extracted total RNA from PUS1 knockdown in HEK293 cells and performed CMCT derivatization in order to determine whether PUS1 depletion affects the CCT3

CMC-ψ profile. In contrast to the expected results, the individual CMC-ψ profile of the CCT3 mRNA from PUS1 knockdown cells revealed an increase in signal intensity of up to 6.3 fold of the reverse transcriptase stop rate before the U1824 observed in both siRNA knockdown experiments (Figure 10B and 10D) which was not observed for the A1831 position upon PUS1 depletion (Figure 10D). Taken together, the presence of a reverse transcriptase stop in the absence of CMCT treatment before the U1824 and the increase of the stop rate in the PUS1 knockdown does not confirm a pseudouridine site but perhaps indicates possibly other RNA modifications.

In summary, our individual mRNA CMC-ψ profiles revealed the presence of two different reverse transcription stops in the PUS1 binding site on CCT3 mRNA, before the A1831 and before the U1824 position, respectively. Despite the expected pattern following PUS1 depletion, we observed an increase in signal intensity before the U1824, thus complementary methods to determine RNA modifications should be used in order to further investigate the primer extension results for the CCT3 binding site. In addition, other PUS1 RNA-binding sites on different transcripts should be investigated.

PUS1 binding site in the human non-coding RNA TERC is close to known pseudouridine site in folded RNA structure

A recent study has described two putative pseudouridine sites in the human non-coding RNA telomerase component, TERC, in HEK293 cells and in human fibroblasts from patients with dyskeratosis congenita (DC). We, therefore, became interested to reveal whether PUS1 could bind as well to TERC in the same region as the reported pseudouridine sites. The two putative pseudouridine sites were identified at position U179 and U307 in the non-coding RNA component TERC. Interestingly, the peak at position U307 was decreased in DC patients, while U179 pseudouridylation level did not change (Schwartz et al., 2014b). In our PAR-CLIP analysis, we identified one PUS1 binding site mapping to TERC, although this was supported by a low number of reads, probably due to the low expression levels of TERC. This cluster overlapped with a region where an evolutionarily conserved 77 nucleotide secondary structure (chr3:170965360-170965436) is predicted by EvoFold v.2 (Parker et al., 2011; Pedersen et al., 2006). This conserved structure contains the putative pseudouridylated U179 reported in the study above (Figure 11A).

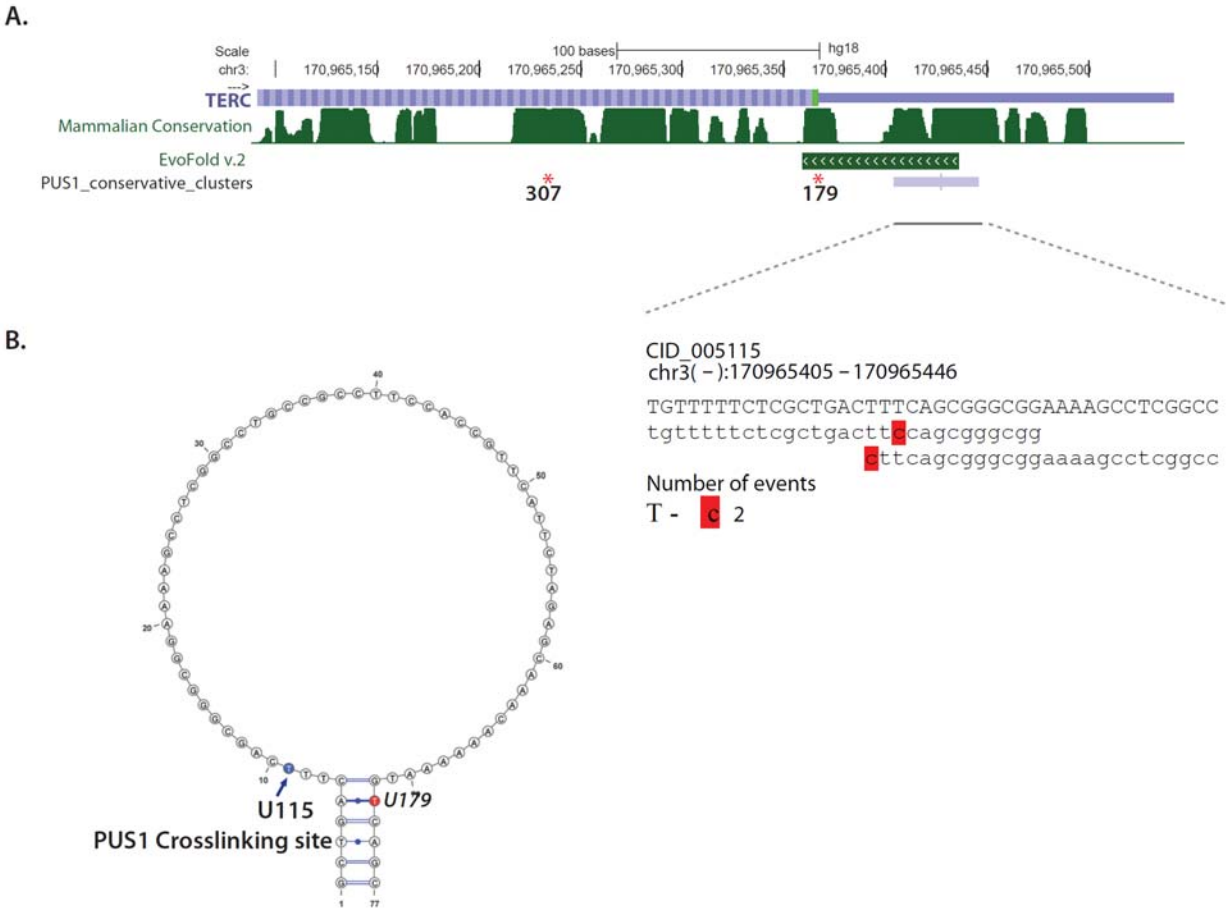


Figure 11. PUS1 cluster is located in an evolutionary conserved secondary structure of human non-coding RNA telomerase component TERC. (A) TERC (Telomerase RNA Component) ncRNA contains a predicted conserved secondary structure across vertebrate species in the 5' region of this transcript. One PUS1 PAR-CLIP binding site overlapped with the conserved structure of TERC although the binding site was supported by few reads with just two T-C transitions inside the predicted structure. (B) Folding of the conserved structural element present in TERC RNA, with the U179 putative pseudouridine site (Schwartz et al., 2014b) highlighted in red, and the PUS1 crosslinked uridine at U115, could indicate that PUS1 might be in contact with the putative pseudouridine site.

The most frequently crosslinked U in the PUS1 binding site in TERC was U115, situated at the 5' end of this structured element, while the ψ179 was found at the 3' end of the sequence. Although these sites are not in close proximity in the linear sequence, the folded structure brings the PUS1 crosslinking site and ψ179 in close proximity (Figure 11B), suggesting the possible involvement of PUS1 in the editing mechanism at this site.

In conclusion, the PUS1 binding site found in TERC non-coding RNA together with the two newly reported potential novel pseudouridine sites, suggest a possible involvement of PUS1 activity presumably at the U179 pseudouridylation site. In addition, detection

of a binding site in the lowly expressed TERC RNA in close proximity to a pseudouridine site serves as a proof of principle for the PAR-CLIP method in identifying possible true interactions.

RESULTS II

YBX3 PAR-CLIP in HEK293 cells identifies transcriptome-wide RNA binding sites

In order to study another putative RNA-binding protein and its interaction with mRNA we employed PAR-CLIP (Photoactivatable Ribonucleoside Enhanced Crosslinking and Immunoprecipitation), the method described above (Hafner et al., 2010b) in HEK293 cells. Flp-In HEK293 cells stably expressing an epitope-tagged version were tested for expression levels induced by doxycycline (1 µg/ml). The levels of overexpressed YBX3 were compared with the endogenous levels by western blot probed with either anti-HA or anti-YBX3 antibody against the endogenous protein (Figure 12A). The FLAG/HIS/HA-YBX3 protein was expressed at similar levels with endogenous YBX3. Both wild-type and epitope-tagged YBX3 present a double band on the SDS-PAGE gel, which is presumably due to post-translational modificatios, as YBX3 is known to have several phosphorylation and acetylation sites (Coles et al., 2005; Sears et al., 2010a). The recombinant YBX3, which runs at around 65 kDa on the SDS-PAGE, showed a partial overlap with the wild-type YBX3 observed at around 60 kDa size marker, however the overexpressed YBX3 had comparable intensity to the band corresponding to the endogenous protein. We performed a 4SU and 6SG 12 hours labelling of HEK293 cells and induced the expression of the FLAG/HIS/HA-YBX3, as previously described in the PAR-CLIP protocol (Hafner et al., 2010b). After crosslinking at UV 365nm the cells were lysed and RNA-protein complexes were immunopurified using the FLAG antibody (Figure 12B).

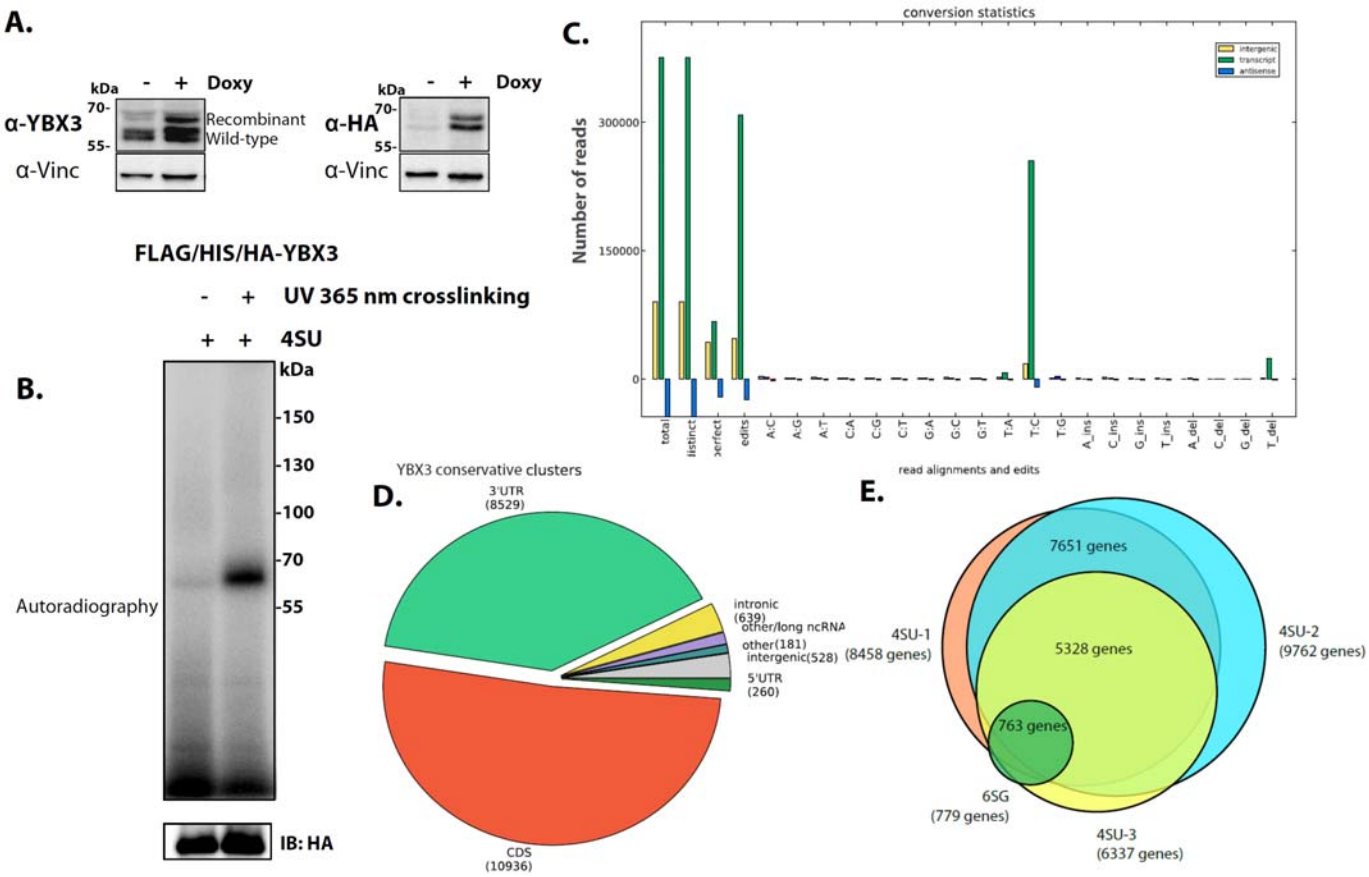


Figure 12. Recombinant YBX3 can be expressed in Flp-In HEK293 cells and binds to thousands of mRNAs with binding preference to 3' UTRs and CDSs. Western Blot of the SDS-PAGE gel of the doxycycline induced recombinant FLAG/HIS/HA-YBX3 protein probed with anti-YBX3 antibody (left panel) and anti-HA antibody (right panel) which revealed a similar expression level for the wild-type YBX3 and recombinant YBX3 in HEK293 cells. Both the endogenous YBX3 and the epitope-tagged YBX3 present two different bands on the western blot (around 60 kDa for YBX3 and around 65 kDa for FLAG/HIS/HA-YBX3) presumably due to the presence of post-translational modifications. (B) Autoradiography of the SDS-PAGE gel showing the radiolabeled protein-RNA crosslinked complexes at around 65 kDa. YBX3-RNA crosslinked complexes are observed in the 4SU lane only, due to photoactivatable 4SU stabilization of covalent bonds formed between protein and RNA upon UV crosslinking. The lower panel shows the western blot for the FLAG-IP of YBX3 complexes, probed with anti-HA antibody. (C) PAR-CLIP library statistics shows T-C transitions for the 4SU library. The mapped reads showed a high number of T-C transitions, with approximately 250,000 T-C containing reads from a total number of approximately 300,000 edited reads. This indicated that the majority of YBX3 PAR-CLIP reads was recovered from crosslinked RNA and can support further processing of YBX3 reads into clusters. (D) YBX3 binding sites showed a binding preference equally distributed towards 3' UTRs and coding sequences (8,000 and 10,000 conservative clusters, respectively). A low number of clusters were annotated to intronic regions (3 %) and 5' UTRs (1.5 %). (E) Venn diagram of the YBX3 PAR-CLIP target genes identified in the 4SU biological replicates and 6SG library showed a high overlap of up to

95% between the replicates with 5,328 genes identified in all three 4SU libaries, and 763 genes out of 779 commonly enriched between the 4SU and 6SG libraries.

Three YBX3 4SU and one 6SG biological replicates were generated and the sequencing reads were analyzed using a consensus and a conservative approach (Table 4, as described in Materials and Methods). The three 4SU PAR-CLIP libraries were filtered into 35,047, 58,290, and 19,065 sequence clusters mapping to 8,458, 9,762, and 6,337 genes, respectively, while the 6SG library constituted 1,277 sequence clusters mapping to 779 genes annotated in the hg19 human genome assembly. The consensus analysis required reads from two out of three libraries and generated 82,965 filtered clusters within 11,071 genes, while the conservative set analysis which required T-C transitions from all three libraries generated 21,479 sequence clusters within 6,452 genes. In addition, we performed an analysis between the 4SU libraries and 6SG library and generated a consensus set (reads from three out of four libraries) of 35,581 filtered clusters. The conservative analysis (T-C transitions from four out of four libraries) between the 4SU and 6SG libraries resulted in 811 filtered clusters (Table 4).

Table 4. PAR-CLIP libraries summary statistics.

Cluster Set	Libraries	Mapped reads	*Kept reads	Raw clusters	Filtered clusters
YBX3_4SU_consensus	3			84937	82965
YBX3_4SU_conservative	3			21728	21479
YBX3_4SU_6SG_consensus	4			36244	35581
YBX3_4SU_6SG_conservative	4			823	811
YBX3_4SU_replicate1	1	1.95 M	0.4 M	35849	35047
YBX3_4SU_replicate2	1	1.56 M	0.51 M	59351	58290
YBX3_4SU_replicate3	1	2.1 M	0.27 M	19531	19065
YBX3_6SG	1	1.16 M	0.15 M	1865	1277

*Uniquely mapped reads kept for further analysis.

The 4SU libraries recovered a high number of reads containing T-C transitions (250,000 out of a total of 400,000 uniquely mapped contributing to the kept reads) while for the 6SG library a lower number of reads contained G-A transitions typical for the 6SG labeling (2,000 out of a total of 9,000 uniquely mapped reads contributing to the kept clusters), probably due to the lower efficiency of 6SG incorporation or the weaker 6SG crosslinking efficiency (Figure 12C). Further analysis revealed that YBX3 putative binding sites are found in CDS, 3' UTR, introns and 5' UTR gene regions according to RefSeq hg19 annotations (Figure 12D). The YBX3 4SU consensus set identified 38,036 binding sites in the coding sequence, 30,876 binding sites in the 3' UTR and 5,961 binding sites in introns. The YBX3 conservative binding sites

comprised 10,936 clusters in the coding sequence, 8,529 clusters in the 3' UTR and 639 clusters in intronic regions. In the 6SG library the distribution of binding sites followed the one in the 4SU libraries although with a lower number of clusters representative for each gene region: 765 clusters in CDS and 311 clusters in the 3' UTR. The distribution of binding sites resulting from the consensus and conservative analysis showed an equal preference of YBX3 binding to 3' UTR and CDS regions. A further aspect of the PAR-CLIP analysis is the reproducibility between the biological replicates; therefore I used a Venn diagram representation of the target mRNAs to show how many mRNAs were commonly detected in the three 4SU PAR-CLIP data sets (Figure 12E). We observed a high overlap between the biological replicates using 4SU labelling which shared 5,328 genes. In addition, when comparing with the 6SG library the overlap was over 95% of the total number of genes (763 genes out of 779 total gene targets) (Figure 12E).

In summary, PAR-CLIP identified a high confident set of over 21,000 biding sites in 6,000 YBX3 target mRNAs, distributed equally between CDS and 3' UTR which showed YBX3 binding preference towards these transcript regions. The high overlap between the 4SU libraries or between 4SU and 6SG libraries suggested a good reproducibility between PAR-CLIP biological replicates.

YBX3 binding sites distribution along the transcript regions

We computed the density of YBX3 T-C transitions from the PAR-CLIP dataset in each of the transcript regions across all the transcripts. The systematic analysis of the density of T-C transitions normalized to the length of the transcript specific regions revealed a uniform distribution along the transcript and the CDS (Figure 13A). Taken separately, the T-C density profile along the 3' UTRs showed a high peak towards the 5' end indicating an increase in YBX3 binding at the beginning of the 3' UTR. Interestingly, the 5' UTRs normalized distribution profile presented a high peak towards the 3' end. However, taking into consideration the shorter length of the 5' UTRs and the significantly lower number of binding sites in the 5' UTRs compared with CDS and 3' UTRs, this higher binding frequency at the end of the 5' UTR is negligible. The strong binding preference to 3' UTRs which are mediating diverse regulatory functions such as stability or degradation suggests a role of YBX3 in post-transcriptional reglation.

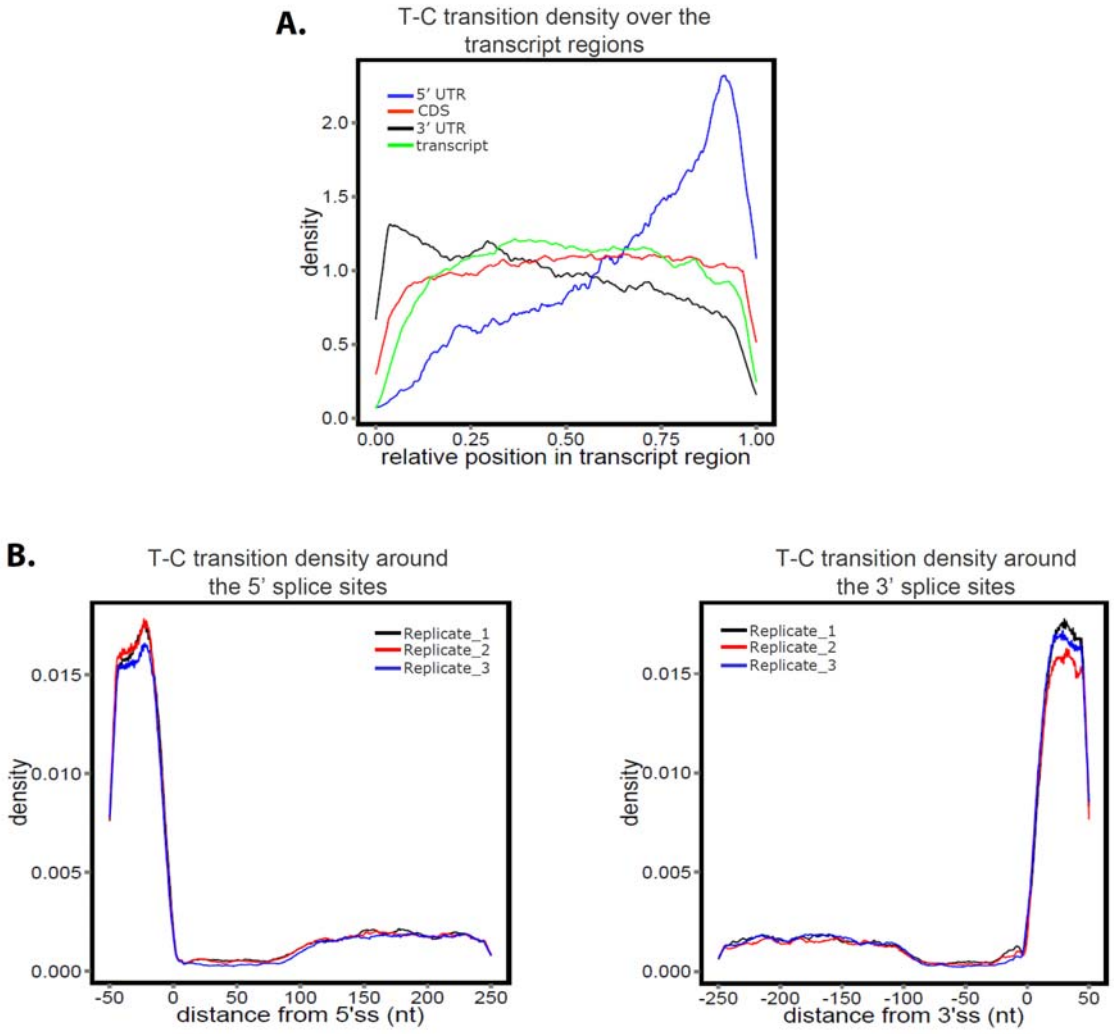


Figure 13. YBX3 T-C transition density shows a higher 3′ UTR 5′-end binding preference and no binding at the splice sites. (A) The density of T-C transitions relative to the position in transcript region across all the targeted transcripts including 5′ UTR, coding sequence and 3′ UTR. The T-C density profile of YBX3 along the coding sequence (red line) and for the whole transcript (green line) showed a uniform distribution. The 3′ UTR T-C density profile (black line) suggested a higher binding preference towards the 5′-end, while the 5′ UTR T-C density (blue line) showed a higher density at the 3′-end, although the 5′ UTR average length and the significantly reduced number of clusters in 5′ UTRs could be a determinant of this pattern. Therefore, the 5′ UTR T-C density is not necessarily suggesting a certain binding preference towards the end of this region. (B) The T-C density profile of YBX3 binding sites around the splice sites for the 5′ splice sites (left panel) and 3′ splice sites (right panel) indicated that YBX3 does not bind directly to exon-intron junctions. However, YBX3 T-C density was high 20-50 nucleotides upstream or downstream into the exons and 150 nucleotides downstream or upstream from the splice site into the introns.

According to several studies on another member of the Y-box family, YBX1 has been reported to promote splicing of CD44 (Stickeler et al., 2001). In this light, we proposed

to test whether YBX3 could bind to splice sites as well. We have analyzed the T-C density profile around the splice sites and observed that the 5' and 3' splice sites of intron-exon junctions across target transcripts were not highly bound by YBX3. The T-C distribution around the splice sites showed that YBX3 binds the exons in proximity of the splice sites but does not overlap with them. Additionally YBX3 shows a rather low preference towards the introns as we deduced from the comparatively lower number of T-C transitions comparing to the exons (Figure 13B). Nevertheless, this would not necessarly exclude a role of YBX3 involvement in splicing regulation.

The YBX3 T-C transition density as a measure of binding preference showed that YBX3 is uniformly present on the coding sequence of the transcripts with a higher preference towards the 5'-end of the 3' UTRs. YBX3 does not bind directly to splice sites although YBX3 is highly bound 20 to 50 nucleotides upstream or downstream splice junctions within the exons of targeted mRNAs. Taken together, these results suggest that YBX3 could be involved in the dynamic regulation of gene expression.

YBX3 target transcripts encode proteins involved in cell cycle

The set of genes encoding conservative YBX3-bound mRNA transcripts were submitted to DAVID functional annotation tool and clustered into gene ontology terms in order to evaluate the functions and biological processes these targets are involved in. The top 1000 enriched target genes in the 4SU conservative set depending on T-C score suggested that YBX3 PAR-CLIP targets encode proteins involved in mRNA metabolic process, mRNA processing, cellular macromolecular complex subunit organization, cell cycle, chromosome organization and RNA splicing (Table 5).

Table 5. Enrichment of Gene Ontology Biological Pathway terms for top 1000 target genes.

Gene Ontology Biological Pathway Term	Count	P-Value
mRNA metabolic process	65	1.50E-16
mRNA processing	58	3.00E-15
cellular macromolecular complex subunit organization	61	7.30E-15
cell cycle	98	1.60E-14
chromosome organization	72	3.40E-14

Interestingly, when using a different data set generated by normalizing the 4SU T-C score to the expression FPKM values, the target mRNAs encode proteins involved in transcription, cell cycle, regulation of transcription and chromosome organization,

however this gene ontology terms had lower p-values than the mRNA processes clustered genes (Table 6).

Table 6. Enrichment of Gene Ontology Biological Pathway terms for top 1000 FPKM normalized target genes.

Gene Ontology Biological Pathway Term	Count	P-Value
transcription	188	1.70E-11
cell cycle	87	7.40E-10
regulation of transcription	214	1.00E-09
chromosome organization	60	1.50E-08
transcription, DNA-dependent	40	4.20E-07
RNA biosynthetic process	40	6.00E-07

It appears that the gene ontology term analysis could be used to estimate a possible functional characteristic of a target gene list but only additional perturbation experiments would provide a more reliable estimation of functional versus non-functional targets in order to make generalized validated assumptions. Nevertheless, the enrichment of transcripts encoding proteins involved in the cell cycle pathway suggested that YBX3 could be involved in the regulation of cell cycle which is in line with previously described functions of YBX3 (Nie et al., 2012; Sears et al., 2010a; Sourisseau et al., 2006; Spadaro et al., 2014).

pSILAC mass-spectrometry quantitative proteomics shows an overall increase in protein synthesis as a result of YBX3 depletion

In order to assess YBX3 role in regulation of its target mRNAs, we depleted YBX3 by siRNA-mediated knockdown in HEK293 cells. We used luciferase reporter assay to test three different siRNAs for the knockdown efficiency (Supplementary Figure S2A). Using a reporter assay containing the YBX3 CDS we observed that siRNAs reduce the levels of the reporter gene compared to mock control down to 20%. The reporter assay was followed by YBX3 knockdown in HEK293 cells using the same siRNAs. Transfection of each of the three siRNAs reduced the YBX3 mRNA level also down to 20% compared to mock transfection measured by qRT-PCR (Figure 14B). YBX3 protein levels upon siRNA knockdown were tested using the Flp-In HEK293 cell line stably expressing YBX3. Western blot quantification of the SDS-PAGE gel showed a significant reduction in recombinant YBX3 protein levels (Figure 14B). We have further asked whether YBX3 binding to its targets may impact the translational output by

applying pulsed stable Isotope Labelling with Amino Acids in Cell Culture (pSILAC) quantitative proteomics in mock and YBX3 siRNA knockdown cells.

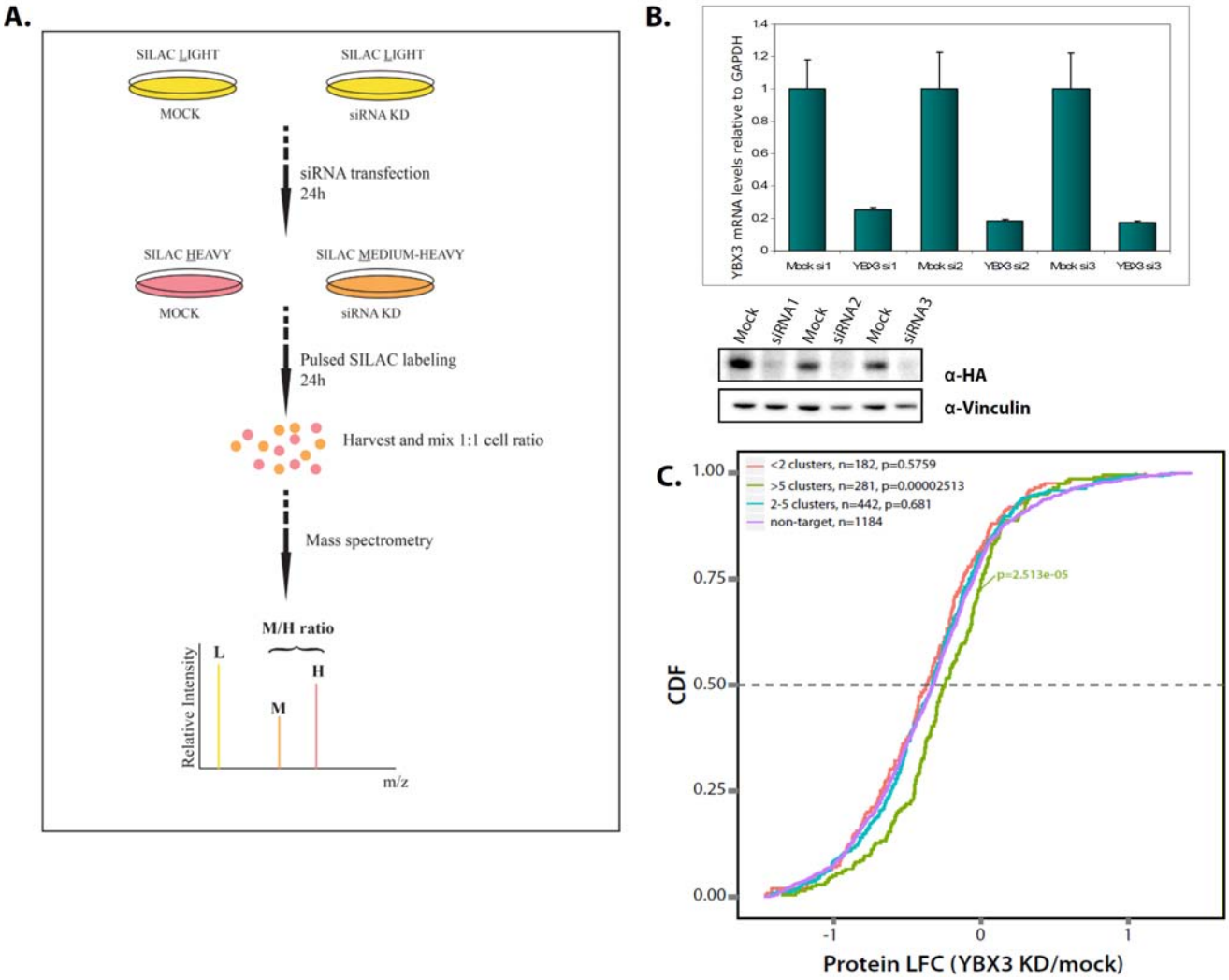


Figure 14. YBX3 siRNA-mediated knockdown and pSILAC quantitative proteomics identifies functional YBX3-mRNA interactions and shows changes in protein synthesis. (A) YBX3 depletion and pulsed SILAC mass-spectrometry experimental setup. YBX3 siRNA transfected or mock HEK293 cells were labeled with light amino acids. After 24h the cells were pulsed labeled with SILAC heavy (H) and medium-heavy (M), respectively, for another 24h. The cells were harvested, mixed on a 1:1 cell count ratio and submitted mass-spectrometry (LC-MS/MS) for quantitative proteomics. The medium to heavy-medium SILAC ratios (M/H) normalized to SILAC light intensities were used to analyze the newly synthesized protein output between the two conditions, mock and knockdown. (B) Quantification of YBX3 mRNA levels by qRT-PCR as well as western blot quantification of recombinant YBX3 protein levels upon YBX3 depletion suggested that each of the three siRNAs is able to efficiently reduce YBX3 mRNA and protein levels. (C) The cumulative distribution fraction plot analysis of the log2 fold change of M/H SILAC ratios showed that the top PAR-CLIP target transcripts normalized by expression level (n=281) had significant changes in protein synthesis (p-value = 2.513e-05). These

mRNA targets which had over 5 binding sites per transcript (green line) had increased protein synthesis upon YBX3 depletion compared to the non-targets (n=1184, purple line). The YBX3 target transcripts with 2-5 binding sites per transcript (blue line) or with less than 2 binding sites per transcript (red line) did not show significant changes in protein synthesis. The effect of YBX3 knockdown suggests a derepression of YBX3 target transcripts or a stabilitzation together with an increase in translation.

Using a SILAC 24h pulse labelling (pSILAC), with SILAC Heavy (H) and SILAC Medium-Heavy (M) amino acids, we were able to quantify changes in protein synthesis between the two conditions, mock and knockdown, independent of the initial pool of proteins (Figure 14A). Around 3,300 proteins were quantified in our experiment, from which the number of proteins with an M/H SILAC ratio in all three siRNA knockdowns was around 2,000. In addition, when comparing the proteomics set (only proteins with SILAC ratios in all three siRNA experiments) with the consensus and conservative PAR-CLIP dataset, only 800 proteins and 1,000 proteins, respectively, overlapped between PAR-CLIP mRNA targets and proteomics. This could be explained either by the differences in quantification and dynamic range between next-generation sequencing and quantitative mass-spectrometry or by the labelling efficiency using the pSILAC method. The normalized M/H SILAC ratios (normalized for the light M/L and H/L ratios) showed the effects of the YBX3 depletion on translation. We have analyzed the YBX3 role on its target mRNAs using a cumulative distribution function plot of log2 fold change of knockdown (M) to mock (H) SILAC ratios. When comparing the SILAC ratios (M/H) of the PAR-CLIP conservative targets, we observed that the changes in protein synthesis upon YBX3 depletion were not significant for the mRNA targets with less than two binding sites per transcript (n=182, p-value = 0.5759) and between two and five clusters per transcript (n=442, p-value = 0.681) compared to the non-targets (n=1184) (Figure 14C). In contrast, when we compared the YBX3 top 281 conservative PAR-CLIP target transcripts, with more than five binding sites per transcript, to the non-targets, we found that these mRNAs had increased protein synthesis upon YBX3 depletion (p-value = 2.513e-05) (Figure 14C). This result showed that the top YBX3 targeted mRNAs were less repressed and more engaged in translation in the absence of YBX3. The increase in protein synthesis suggests that YBX3 could have a negative impact mRNA stability or translation, a role that has been previously suggested for YBX3 by several individual *in vitro* studies (Coles et al., 2004b; Coles et al., 2002; Evdokimova et al., 1998).

In summary, the pulsed SILAC based mass-spectrometry and quantitative proteomics identified changes in protein synthesis between YBX3 knockdown and mock conditions.

The significant increase in protein synthesis of the top YBX3 target mRNAs suggested a global functional role of YBX3 in gene expression either by influencing mRNA stability or mRNA degradation or rather translational output.

YBX3 RNA binding motif analysis identifies a consensus sequence containing the conserved motif CAUC

The bioinformatic analysis undertaken by Dr. Emanuel Wyler, was performed using cERMIT, an efficient computational tool for motif discovery based on genome-wide quantitative regulatory evidence such as PAR-CLIP (Georgiev et al., 2010b). In our analysis, we took 40 nucleotide sequences within the YBX3 conservative binding sites set, in order to find possible YBX3 binding motifs. We counted and ranked by enrichment over equidistribution 7-mers found in the 40-nucleotide long sequences. From these 7-mers, we deduced the consensus sequence [T/A][T/A/C][G/T][A/T/C]**CA**[T/C]**C**[T/A][T/A][C/A] containing the motif CAUC (Figure 15A), which was previously

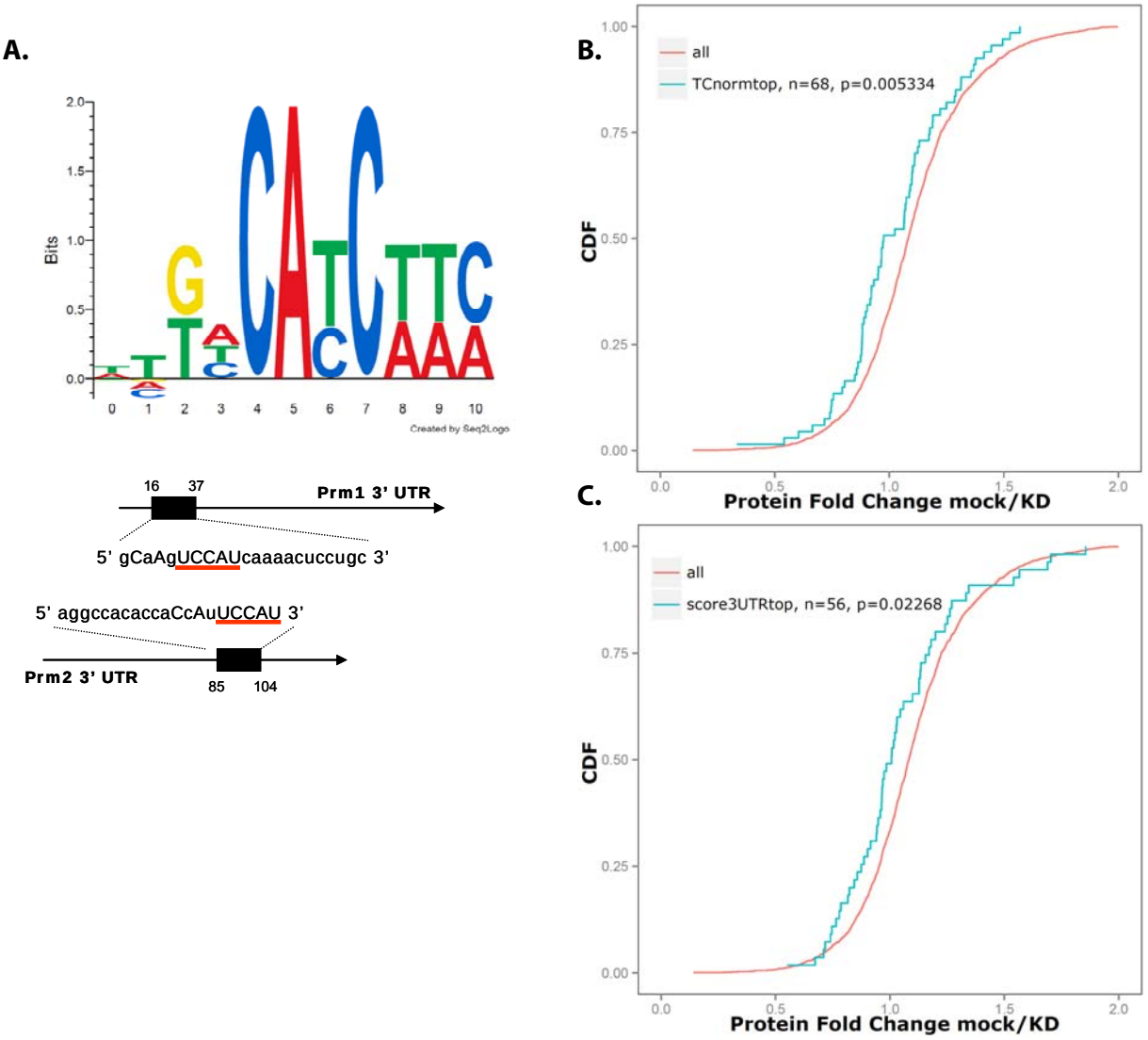


Figure 15. RNA binding motif search identifies a consensus motif in YBX3 binding sites which uncovers a subset of the YBX3 regulated mRNA targets. (A) The consensus motif sequence logo of the top enriched 10-mer sequences in the PAR-CLIP data set analyzed using cERMIT computational tool (Georgiev et al., 2010a) showed YBX3 pyrimidine-rich RNA recognition sequence with the central CAUC conserved motif (upper panel). The motif present in the YBX3 target transcripts has been previously described in an *in vitro* study on the YBX3 mouse homologue for the Prm1 and Prm2 mRNA, in the 3' UTR region of these transcripts (*adapted from (Giorgini et al., 2001a)*) (lower panel). (B) The cumulative distribution fraction plot of the log2 fold change calculated for the protein H/M (mock/knockdown) SILAC ratios of the YBX3 top 600 FPKM normalized T-C ranked target transcripts according to the motif score. The top motif scored YBX3 target transcripts (n=68, blue line) show a significant increase in protein synthesis (p-value = 0.005334) after YBX3 knockdown (shift to the left) relative to all proteins (n=1892) (red line). (C) The cumulative distribution fraction plot of the log2 fold change calculated for the protein H/M (mock/knockdown is reversed due to the H/M SILAC ratios used in the bioinformatic analysis) SILAC ratios of the top 427 transcripts according to the 3' UTR motif score (n=56, blue line).The proteins containing a YBX3 RNA binding motif in their transcripts show higher protein synthesis in YBX3 depleted cells reinforcing the possible role of YBX3 in translational repression.

suggested by *in vitro* experiments with the mouse homologue MSY3/4 (Giorgini et al., 2001a). A further investigation of the YBX3 binding sites revealed a set of the eight most enriched 7-mer sequences containing two preferred crosslinking sites (Supplemental Figure S2B).

Next, we performed an analysis of the top 600 FPKM normalized T-C ranked PAR-CLIP target transcripts. Our analysis of the YBX3 top 600 mRNA targets, ranked by motif score, of which 68 had a SILAC ratio, showed that these proteins had lower H/M SILAC ratios (mock/knockdown) than all the proteins (n=1892). The cumulative distribution fraction plot of the protein log fold change showed a shift to the left, indicating that YBX3 top motif scored mRNA targets were positively regulated at translational level upon siRNA knockdown (p-value = 0.005) (Figure 15B). The low number of proteins with SILAC ratios for the top YBX3 targets containing the consensus motif could be explained by the stringent filtering of the proteome dataset, as well as the low mass-spectrometry coverage. Furthermore, when we analyzed the top 3' UTR motif scored genes (n=427), of which 56 had SILAC ratios (mock/knockdown), we observed once again a significant shift to the left compared to all the proteins (p=0.02), suggesting that YBX3 depletion induces an increase in protein synthesis of the mRNAs with a high 3' UTR motif score (Figure 15C). The gene ontology functional annotation suggested that these targets are involved in various

RNA processing pathways, including splicing and mRNA metabolic processes, but also response to DNA damage and DNA repair (Supplemental Table S1).

An analysis of the average FPKM values of YBX3 mRNA targets (44.91) having the consensus motif, showed this was lower than the average FPKM values of the mRNAs without a motif (59.29) as well as the average FPKM values of all target mRNAs (58.48) indicating that the mRNA targets which do not fullfil the motif are much higher expressed and thus the interaction could be presumably non-physiological. Moreover, these results suggest that the identification of YBX3 motif which is deduced from the PAR-CLIP dataset could be used to determine which are the target transcripts destabilized by YBX3. Taken together, the motif search and analysis of the pulsed SILAC proteomics which show that the top RNA motif scored 3'UTR targets have higher protein synthesis upon YBX3 knockdown indicate that YBX3 has a repressive effect on gene expression.

YBX3 3' UTR targets show changes in luciferase activity in a dual luciferase reporter assay

With the observed global protein changes in YBX3 depleted cells, I further set to validate several PAR-CLIP identified targets which were affected by YBX3 knockdown, using a dual Renilla/Firefly luciferase reporter assay. While YBX3 showed equal preference of binding to the coding regions and 3' UTRs, we aimed to evaluate the effect of YBX3 depletion on translation using a dual luciferase reporter assay and confirm the regulatory functions of YBX3 on individual 3' UTRs containing YBX3 RNA binding motif.

We PCR-amplified full-length 3' UTRs of several candidate targets from PAR-CLIP dataset and cloned each 3' UTR through a restriction/ligation reaction, behind the Renilla luciferase gene in a dual reporter construct. The reporter constructs containing the candidate 3' UTRs or the empty vector were transfected in mock and YBX3 knockdown cells in order to measure the Renilla luciferase activity normalized to the Firefly luciferase activity, as an output for translational regulation (Figure 16A).

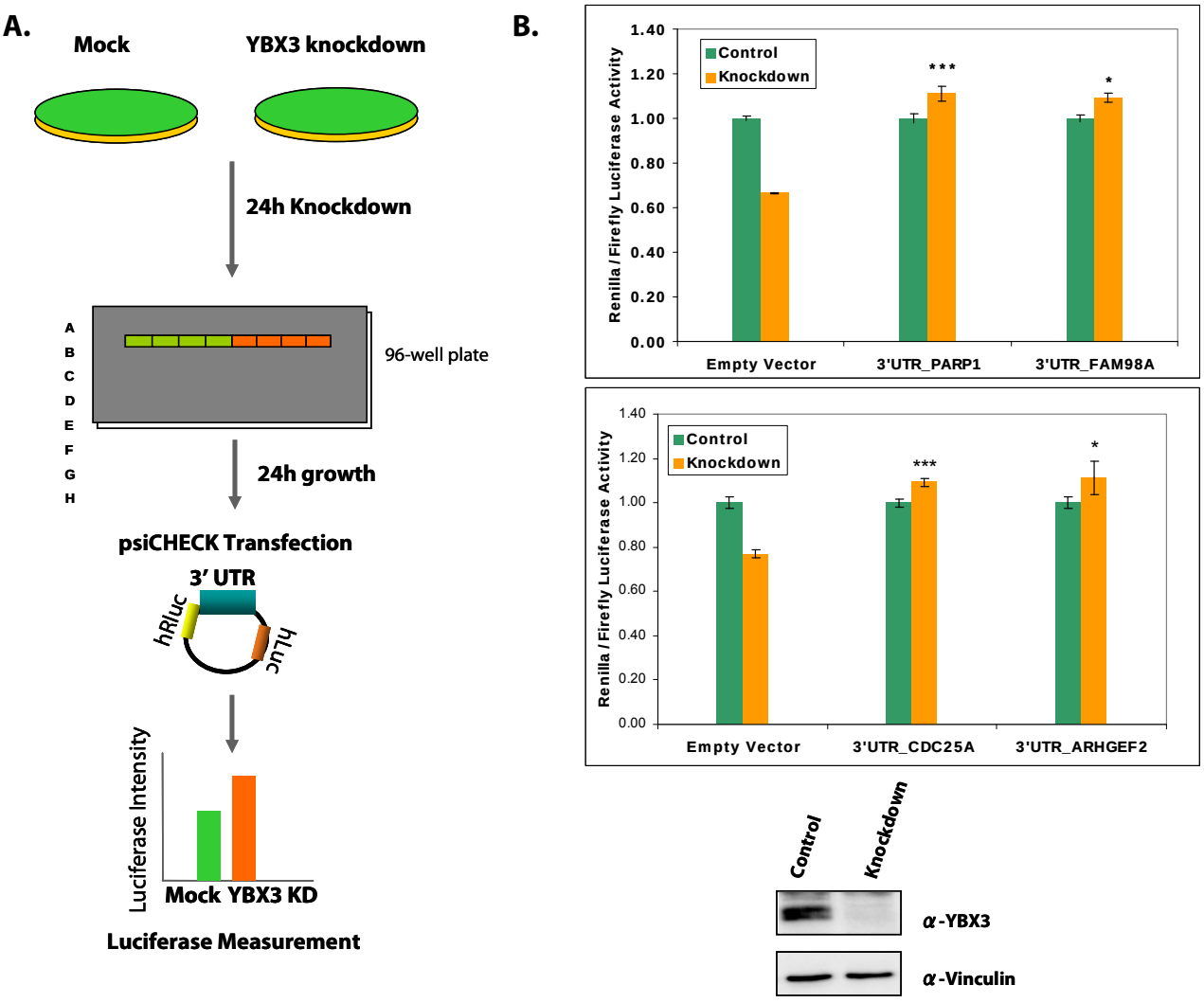


Figure 16. Dual-luciferase reporter assay validation of the YBX3 role on translation. (A) Experimental setup of the luciferase dual-reporter assay using mock and siRNA knockdown conditions. After a 24h RNAi transfection, the cells were seeded in a 96-well plate and grown for another 24h before being transfected with the empty vector or 3' UTR candidate dual-luciferase psiCHECK vector. The next day the luciferase intensity was measured for the two conditions and the Renilla/Firefly ratio measured the reporter activity as an output of 3' UTR mediated translational regulation. (B) The reporters containing the 3' UTRs of PARP1, FAM98A, CDC25A, and ARHGEF2 had an increase in luciferase intensity (normalized to the empty vector) in the knockdown cells comparing to control (first and second panel). The efficient knockdown of YBX3 endogenous levels was confirmed by western blot (lower panel) probed with anti-YBX3 antibody. The luciferase activity increase indicated that YBX3 depletion had positively affected the translation of the candidate 3' UTRs constructs, thus confirming the role of YBX3 in translational regulation.

The luciferase activity in cells transfected with the reporter constructs containing the 3'UTR of PARP1 (poly (ADP-ribose) polymerase 1), FAM98A (family with sequence

similarity 98, member A), CDC25A (cell division cycle 25A), and ARHGEF2 (Rho/Rac guanine nucleotide exchange factor (GEF) 2) was measured and normalized to the empty vector (Figure 16B). The luciferase assay showed a modest but significant increase in luciferase levels for the 3' UTR candidates suggesting a role in 3' UTR mediated gene expression regulation of YBX3. In order to asses the functionality of the YBX3 binding sites in the 3' UTR sequence of the tested candidates we aim to further perform a mutagenesis of the motif the 3' UTR reporter constructs and compare with the previous observations.

In summary, YBX3 3' UTR target transcripts containing the YBX3 consensus motif were tested in order to confirm the effects on mRNA expression detected in the mass-spectrometry quantitative proteomics. We found that several 3' UTR candidates had higher luciferase activity in the absence of YBX3 suggesting a potential implication of YBX3 in gene expression possibly via 3' UTR binding.

YBX3 RNA-independent and RNA-dependent protein-protein interactions indicate possible interaction with translation initiation factors

Further investigation into YBX3 molecular function took our attention towards possible YBX3 interacting factors which could contribute to a better understanding of YBX3 biological role. Using a co-immunoprecipitation SILAC-based method optimized in our laboratory (Gregersen et al., 2014b), and combined with UV crosslinking and RNase treatment, we set to identify eventual protein-protein interactions that could act synergistically or antagonistically with YBX3, in an RNA-dependent or independent manner (Figure 16A). We used two different conditions, where we UV crosslinked 4SU labelled HEK293 cells and used a mild RNase treatment in order to capture YBX3 interactions bound in proximity on the same RNA, or we immunoprecipitated YBX3 using anti-FLAG antibody and used a high RNase/benzonase treatment in order to capture direct protein-protein interactions (Supplementary Figure S3A, S3B). For each of the direct/ RNA-independent (high RNase conditions) and indirect/ RNA-dependent approaches (mild RNase conditions) we used two SILAC heavy amino acids (H) and light amino acids (L), including swap labelling, of YBX3-induced and non-induced cells in order to quantify enriched proteins against background contaminants (Figure 17A). We mixed the cell pellets containing both H and L labelled proteins, in one sample, right after harvesting in order to capture the most confident interactions. While the YBX3 indirect protein-protein interaction assay could "freeze" in place the RNA-dependent partners due to croslinking, the direct protein-protein interaction experiment

could capture only stable protein complexes. We identified several RNA-dependent YBX3 protein interactors, know RNA binding proteins with equal binding preference to CDS and 3' UTR regions, amongst which IGF2BP family, IGF2BP1, IGF2BP2 and IGF2BP3 (insulin-like growth factor 2 mRNA binding protein 1,2 and 3) (Figure 17B). Interestingly, IGF2BP1 and IGF2BP3, HNRNPQ (SYNCRIP- synaptotagmin binding, cytoplasmic RNA interacting protein) and YBX1 have been found previously to co-immunoprecipitate with YBX3 in an SILAC-based *in vitro* RNA-pull down experiment using PTPN13 mRNA 3' UTR and JMJD1C CDS sequence (jumonji domain containing 1C) (Scheibe et al., 2012).

When we performed the direct protein-protein interaction experiment, due to our double affinity stringent conditions and probably due to weak or transient interactions of YBX3 with other proteins in a stable complex, we have not obtained SILAC ratios (log2 H/L, log2 L/H) greater than 1 (Supplementary Figure S3C). Nevertheless, the unique peptide counts quantified by mass-spectrometry in the forward against the swap labelling showed an enrichment in eukaryotic translation initiation factors EIF3A, EIF4G1, EIF4G3, EIF3C, EIF5B, EIF3D, as well as several eukaryotic translation elongation factors EEF1A1, EEF1G, EEF1D, EEF2 (Figure 17C). EIF3A, a translation initiation factor which is involved in 40S ribosomal unit binding was one of the most enriched YBX3 direct protein-protein interaction partner with a high number of unique peptide counts. This might suggest YBX3 interaction with translation initiation factors but further co-immunoprecipitation experiments are necessary to confirm these observations. The ARHGEF2 enrichment among the direct interactors is the Rho/Rac guanine nucleotide exchange factor 2 which has been previously reported in two studies to interact with YBX3.

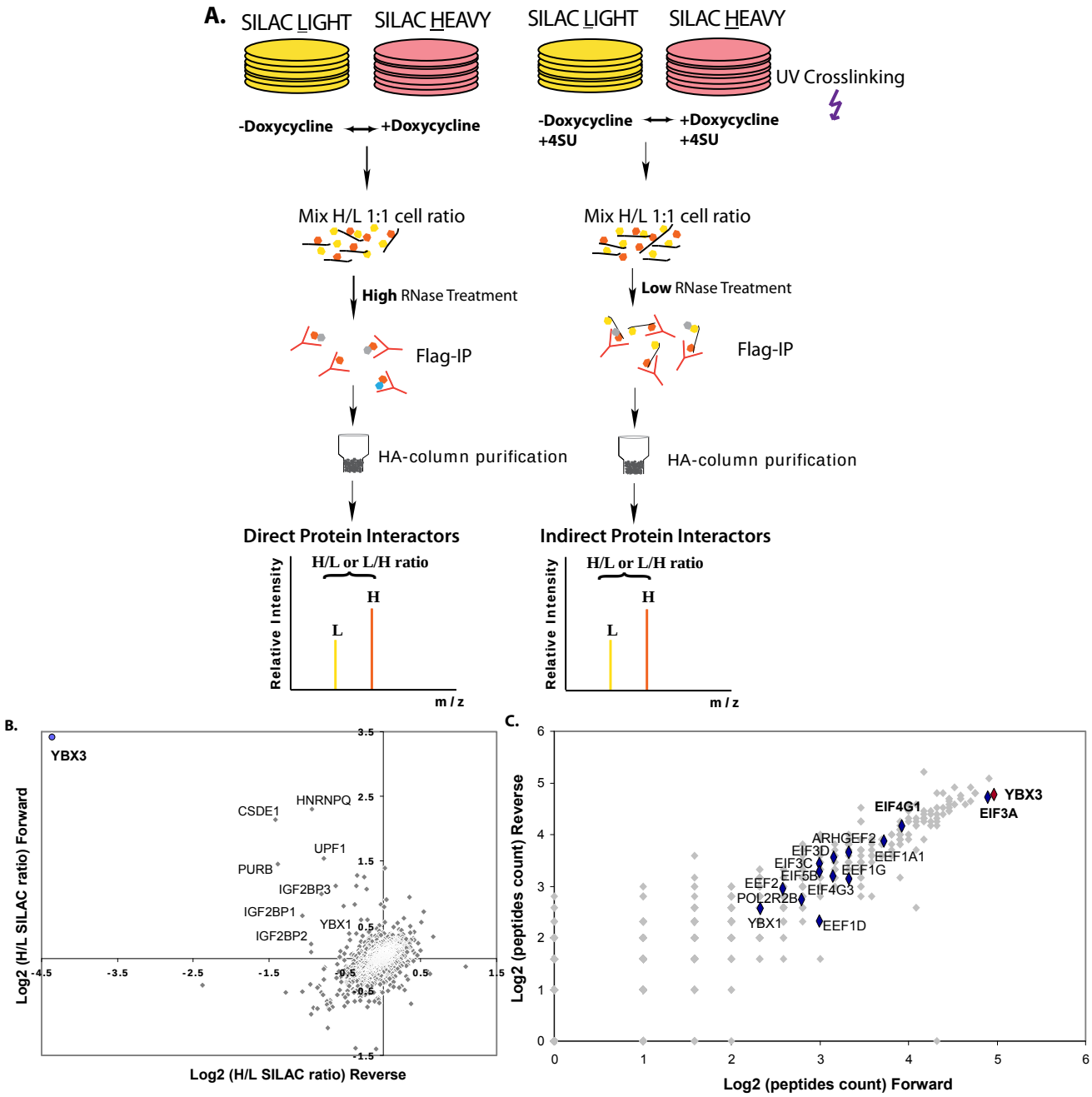


Figure 17. YBX3 protein-protein interactions captures additional 3′ UTR binding proteins and translation initiation factors. (A) SILAC based protein-protein interaction assay using the light and heavy labeled aminoacids were performed with a high RNase or with a low RNase concentration treatment combined with 4SU labeling and UV crosslinking. In order to identify the specific interaction partners from the contaminants we also performed a SWAP labelling experiment. YBX3 co-immunoprecipitation was performed using a double affinity purification including FLAG-IP and HA-column purification. The protein sample was digested to peptides and submitted to mass-spectrometry. The H/L and L/H SILAC ratios were used to show enrichment of protein interactors. (B) The SILAC H/L ratios plot of the forward and the swap experiment revealed several known RNA binding proteins which also have a binding preference towards coding sequences and 3′ UTR regions

such as IGF2BP1, IGF2BP3, HNRNPQ and YBX1. (C) YBX3 protein-protein interactions assay for the identification of factors directly interacting with YBX3 did not result in experimental quantification of H/L SILAC ratios higher than 1 (log2 H/L). Nevertheless, H/L unique peptide count ratios plot of the forward against the reverse experiment showed that several translation initiation and translation elongation factors were commonly enriched between the two experiments with high peptide counts. The EIF3A, EIF4G1, EIF4G3, EIF3C, EIF5B and EIF3D translation initiation factors were detected together with several translation elongation factors.

In conclusion, the YBX3 protein-protein interaction assay has revealed an expected RNA-dependent interaction with known RNA-binding proteins which are also known to bind the 3' UTR regions and coding sequences. The co-immunoprecipitation conditions we used to detect direct protein-protein interactions showed a high SILAC unique peptides count for translation initiation and translation elongation factors which co-purified with YBX3. Although the interaction of YBX3 twith these EIF proteins has not yet been validated, the Rho GTPase ARHGEF2 found in our purification was shown in previous studies to interact with YBX3 and modulate its transcriptional/translational activities.

DISCUSSION

Our approach to studying the function of PUS1 had arisen from the previously published (Baltz et al., 2012b) RNA-bound proteome identification in HEK293 cells where this enzyme was found to be enriched in all three biological replicates. Surprinsingly, PUS1 was the only pseudouridine synthase that was found in both HeLa cells RBPome (Castello et al., 2012b) and the global analysis of the yeast mRNPs (Mitchell et al., 2012), suggesting the biological importance of this pseudouridine synthase. Applying the PAR-CLIP method (Photo-Activatable Ribonucleoside Enhanced Crosslinking and Immunoprecipitation) (Hafner et al., 2010b) permitted the evaluation of PUS1 RNA-binding properties and identification of thousands of binding sites in protein coding genes. Therefore, I will discuss further the transcriptome-wide profiling of PUS1 binding sites and mRNA targets, validation of two biologically important tRNA PUS1-dependent pseudouridylation sites and position-specific mapping of a PUS1 putative pseudouridylation sites in individual mRNAs. I would like to highlight as well the future implications of pseudouridine modifications and suggest possible new directions in uncovering PUS1 molecular function in post-transcriptional regulation.

In the second part of my thesis, the characterization of another RNA binding protein, the Y-box-binding protein 3 or YBX3 (also known as CSDA/ZONAB/DbpA), we initiated the study based on our finding of YBX3 as part of the RNA-bound proteome together with several studies suggesting the multiple role of YBX3 in ssDNA and ssRNA binding with possible implications in cancer (Balda and Matter, 2000a; Dupasquier et al., 2014; Nie et al., 2012; Sears et al., 2010a). I will discuss further the transcriptome-wide profile of YBX3 RNA-binding sites, the YBX3 depletion effect on translation, the putative interactions with translation initiation factors and the validation of the YBX3 3' UTR mediated translational repression on YBX3 target candidates.

PUS1 transcriptome-wide profile of RNA binding sites

In order to characterize PUS1 function as an RNA-binding protein and RNA-editing enzyme, we started with identification of the transcriptome-wide binding sites by PAR-CLIP (Hafner et al., 2010a). Utilizing PAR-CLIP we obtained a consensus set (an analysis considering reads from both biological replicates) of 8,338 binding sites in 3,123 genes. When we applied a more stringent filter (analysis considering T-C

conversions from both biological replicates), we obtained a high confidence set of 6,087 filtered clusters in 2,516 genes (Table 2). In comparison with the published studies which used the CMCT-Psi (CMC-ψ or ψ-seq) profiling which could indentify pseudouridine modifications across the transcriptome (Li et al., 2015a; Schwartz et al., 2014a), our PAR-CLIP identified binding sites is exceeding by several fold the total number of pseudouridine sites (396 ψ sites in HEK293, 2,084 ψ sites in HEK293T, respectively). This could be due to several reasons, including the fact that PAR-CLIP recovers both physiological and non-physiological protein-RNA interactions (Friedersdorf and Keene, 2014a), thus PUS1 binding sites might not all contain pseudouridine sites, indicating a possible other biological function of this protein. Another reason could be that the two methods have different protocols which lead to different readouts. PAR-CLIP captures interactions which take place *in vivo* between the exogenous PUS1 and the target transcripts at nucleotide resolution, which might also be biased towards the most highly expressed targets. The CMCT-Psi method requires a chemical modification of total RNA combined with deep sequencing. CMCT profile could also contain false positives or false negatives generated by non-specific reverse transcriptase stops. The ψ-ratio count calculated from deep-sequencing reads could therefore be a source of variability between the studies above. In addition, CMCT profiles could not distinguish between pseudouridine sites catalyzed by RNA-dependent mechanisms (such as DKC1 pseudouridine synthase) or RNA-independent mechanisms (such as PUS1, PUS7 stand-alone pseudouridine synthases). Therefore, these studies have employed either bioinformatic analysis or when available, RNAi and CRISPR/Cas9 knockout methods, to associate the pseudouridine sites with the respective pseudouridine synthases. Schwartz and colleagues concluded that from the 396 ψ sites in HEK293 cells, a number of 84 ψ sites could be the modified by other putative pseudouridine synthases than DKC1, PUS4 and PUS7. In lack of a similar CMCT global profile, we then attempted to compare the pseudouridine sites with our PUS1 PAR-CLIP clusters in order to estimate the level of overlap between PUS1 mRNA targets and all the pseudouridine containing mRNAs (Schwartz et al., 2014b). We overlapped 30 nucleotides around ψ sites identified in HEK293 cells by ψ-seq (Schwartz et al., 2014a) with PUS1 PAR-CLIP binding sites. The analysis revealed a set of 32 mRNAs containing ψ sites which were also identified as PUS1 mRNA targets by PAR-CLIP (Supplementary Table S2). This does not show an impressive overlap but taking into account we are comparing two different methods, we could suspect that PUS1 has indeed an alternative function or the ψ-Seq methods developed so far are underestimating the true number of pseudouridine sites. Li and colleagues associated only 77 ψ-sites (only 72 in mRNA) out of 2,084 ψ-sites with PUS1, in HEK293T cells (Li

et al., 2015a). The overlap between the 64 mRNA transcripts containing 72 PUS1-dependent ψ-sites with the PUS1 PAR-CLIP target mRNAs revealed that only 25 of the 64 mRNA transcripts were also bound by PUS1 in our PAR-CLIP dataset. This could suggest once more the possibility of a different function of PUS1, perhaps in stabilizing secondary structures, similar to its function in tRNAs and rRNAs, although without performing an pseudouridylation activity. When we compared all the genes containing ψ-sites in the CeU-Seq profile with the PUS1 PAR-CLIP consensus and conservative set of targets, we found a common set of 425 (31%) and 351 (26%) genes, respectively (Supplementary Figure S4). Although this might be surprising, expanding the comparison between the ψ-seq and CeU-Seq, which used the similar CMCT-profile methods, revealed that only around 30% ψ-containing genes of the first set were found in the second and they represent less than 1/10 of the total identified CeU-Seq genes (Supplementary Figure S4) suggesting that different experimental setups (CMCT treatment versus CMCT enriched RNA) could generate different datasets. However, these studies have yet not accounted for a possible redundancy of the pseudouridine synthases in modifiying the same position, therefore, the number of pseudouridine sites modified by each pseudouridine synthase might be underestimated. It has been suggested that the ψ55 in tRNA, for example, could be edited by two different pseudouridine synthases PUS10 and DKC1/Cbf5 (Gurha and Gupta, 2008; Kamalampeta and Kothe, 2012; Tillault et al., 2015). Moreover, the entire landscape of pseudouridine modifications could still be uncovered by the development of new methodologies including 4SU pulse labeling, or mass-spectrometry in combination with next-generation sequencing which could improve detection and presumably reveal a higher level of pseudouridylation.

Distribution of PUS1 binding sites across the transcriptome and PUS1 pseudouridylation sites in tRNA and TERC ncRNA

PAR-CLIP analysis of the transcript distribution of PUS1 binding sites (Figure 7B) revealed that PUS1 has a high binding preference towards 3' UTRs but also towards CDS where the presence of binding sites in exons is markedly increased. Although PUS1 function was previously only related to tRNA and UsnRNA pseudouridylation, affecting splicing and translation (Behm-Ansmant et al., 2006; Bykhovskaya et al., 2007; Chen and Patton, 2000), the preference of PUS1 binding in exons or 3' UTRs could suggest a possible direct role in mRNA metabolism. In addition, PUS1 activity

has not been correlated with sequence specificity but rather with a higher preference towards structured RNA substrates (Sibert and Patton, 2012).

Considering that pseudouridylation of tRNA and U snRNA have important phenotypic implications in either mitochondrial diseases or splicing deficiency we observed a high PUS1 read coverage at known pseudouridine sites in tRNALys (positions U27 and U28) and U6 snRNA (positions U31, U40, and U86) (Figure 8A and 8B) indicating that PAR-CLIP can detect previous reported PUS1 targets and identify binding sites containing known pseudouridine residues. We next took a further step to investigate if the functionally important U27 and U28 pseudouridine sites in tRNALys that were previously correlated with MLASA are indeed pseudouridylated in HEK293 cells (Bykhovskaya et al., 2004a; Patton et al., 2005b). The CMCT profile of tRNALys (UUU) confirmed the presence of the two ψ sites in the CMCT treated samples, and indicated the presence of presumably the hypermodified tRNA uridine residue, the τm5s2U34 (5-taurinomethyl-2-thiouridine) at position 34 in the tRNA anticodon loop (Suzuki et al., 2002; Umeda et al., 2005). This result suggested that our optimized CMCT protocol could further reveal putative pseudouridylation sites in mRNA.

Previous studies suggested that defects of the RNA-guided pseudouridine synthase DKC1/dyskerin and also of its associated non-coding RNA telomerase component, TERC, are correlated with dyskeratosis congenita (Mitchell et al., 1999). Two of the six proposed pseudouridine positions were found at U307 and U179 in the P3 stem of the TERC pseudoknot (Schwartz et al., 2014a). In fibroblasts from dyskeratosis congenita patients, it was observed a reduction in pseudouridine levels at U307 but not at U179, suggesting that U179 could be modified by a different pseudouridine synthase. Interestingly, our PUS1 PAR-CLIP profile revealed one conservative binding site near the U179 pseudouridine site, which also overlapped with an evolutionary conserved secondary structure predicted by EvoFold 2.0 (Parker et al., 2011; Pedersen et al., 2006). Despite of the non-overlap between PUS1 cluster and U179, when folded into a secondary structure the PUS1 crosslinking site at U115 re-locates in close proximity with the pseudouridine at U179 situated at the 3’ end of the conserved element (Figure 11A,B). Taking into consideration this observation together with the PUS1 binding site near the U179 and PUS1 affinity for secondary structures (Sibert and Patton, 2012), we could speculate the PUS1 activity at U179. Due to the experimental studies which showed that pseudouridine presence in TERC increases the processivity *in vitro* of the telomerase through TERT-TERC or TERC-TERC interactions (Kim et al., 2010), it would be of interest to validate PUS1 and DKC1 putative pseudouridylation sites at U179 and U307, respectively, and also determine whether the two editing sites are redundant or independent between the two enzymes.

CMC-ψ profile on individual PUS1 mRNA targets

Following the confirmation of known PUS1-dependent tRNA modifications, we set to identify putative pseudouridine sites in mRNA, within the PUS1 binding sites identified by PAR-CLIP. In this case we chose three of the PUS1 target transcripts, and applied the reverse transcription primer extension assay on CMCT derivatized total RNA using specific primers (Supplemental Figure S1A and S1B). The individual mRNA CMCT profile showed several reverse transcriptase stops in each of the PUS1 tested targets (Supplemental Figure S1A). The CCT3 mRNA region bound by PUS1, indicated two stops of the reverse transcriptase before an adenosine at position 1831 and a uridine at position 1824 (Figure 10B). Due to the fact that inosine can also be derivatized by CMCT treament (Ho and Gilham, 1971) and that we observed a reverse transcriptase stop before the A1831 in the CMCT-treated lanes only, we could speculate that A1831 might be the result of A-I editing thefore its derivatization could lead to a reverse transcriptase stop at this position. The strong stop in both the non-treated and CMCT-treated lanes before the U1824 residue (Figure 10B, Supplemental Figure S5C) does not confirm a pseudouridine site. However, the PUS1 knockdown revealed a notable increase of the reverse transcriptase stop rate at U1824 in the knockdown compared to the mock cells (Figure 10D). Although this is contrary to what we should expect, one explanation could be that the reverse transcriptase is halting before a bulky residue such as an RNA hypermodification similar to the ones often present in rRNAs or tRNAs (Ofengand and Bakin, 1997). If indeed this would be the case, the increase in the intensity of the stop before U1824 in PUS1 knockdown conditions could be explained either by an increase in mRNA expression levels or the redundant activity of another pseudouridine synthase such as PUS7. Interestingly, the PPUS web server for pseudouridine sites (Li et al., 2015b) has predicted several yeast PUS1 ψ-sites for CCT3 mRNA and two PUS7 sites. One of the PUS1 predicted sites for the yeast CCT3 mRNA is found at the correspondent U1824 position in the human CCT3 transcript (Supplementary Figure S6A and S6B). Similarly to mock lanes, the reverse transcriptase stops observed at a higher CMCT concentration had lower intensities which could not be explained by the definition of CMCT derivatization method.

In these lines of evidence, our results showed a reverse transcriptase stop before the U1824 residue in the human CCT3 mRNA which had an unexpected increase when PUS1 levels were reduced by RNAi-mediated knockdown. We could not attribute this directly to a PUS1 modification site, therefore further investigation is necessary to detect whether this is the case of a different type of RNA modifcation. The specific *in vivo* RAP (RNA antisense purification) enrichment of CCT3 mRNA followed by LC/MS

mass-spectrometry (Cao and Limbach, 2015) could be used to identify the presence, for example, of a methylpseudouridine similarly to the identification of a 3-methylpseudouridine at position 1915 in 23s rRNA (Kowalak et al., 1996). Whether PUS1 depletion is affecting the local structure of the CCT3 mRNA-bound region could be tested using smFRET (single molecule FRET), that is exquisitely sensitive to conformational changes at single molecule level (Kobitski et al., 2008). Ultimately, one major interest represents the PUS1 activity on CCT3 stability or translation, therefore using qRT-PCR to test CCT3 expression in presence of absence of PUS1, or reporter assays to evaluate translation efficiency could indicate a possible connection between pseudouridylation and mRNA metabolism.

In the light of these results, it can be concluded that PUS1 can bind to a large number of mRNA transcripts besides tRNA and snRNA, it can modify tRNA and possibly mRNA and ncRNA, but further investigation would is necessary to uncover PUS1 functional role in mRNA metabolism. Together with the recent advances on pseudouridine profiling we can look forward to shed light on PUS1 RNA-binding function that could either change mRNA structure for RNA recognition by other RNA-binding proteins or affect translation, splicing and mRNA half-life. It cannot be ruled out that pseudouridylation could act *in vivo* in order to possibly alter the genetic coding post-transcriptionally thus a robust tool to evaluate global Ψ dynamics and changes can contribute to revealing the underlying mechanisms (Figure 18) (Zhao and He, 2015).

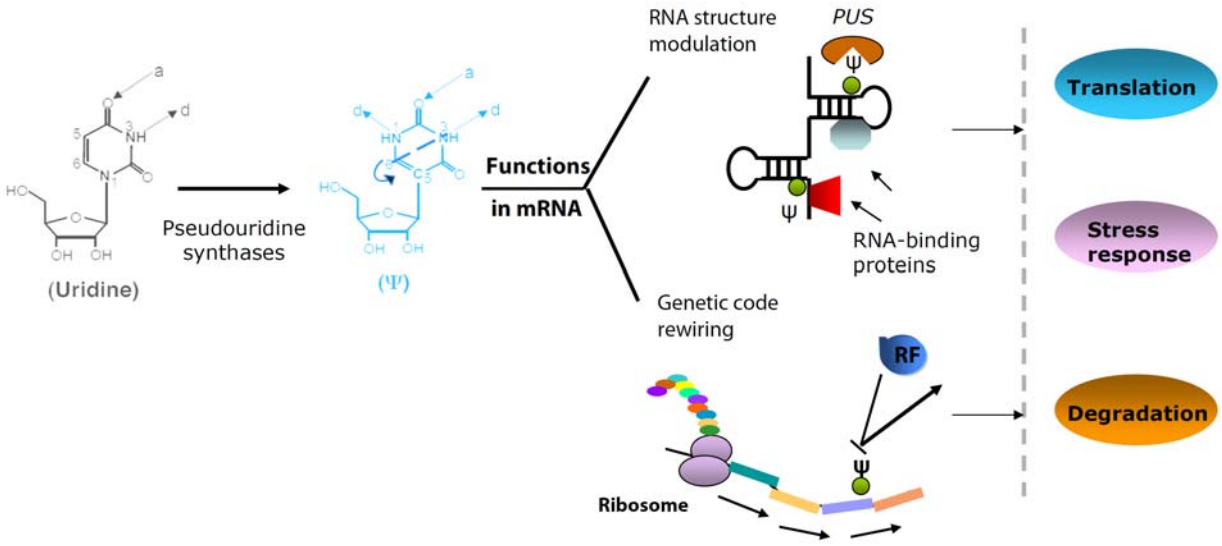


Figure 18. Possible functional roles for RNA pseudouridylation. As RNA secondary structure is closely related to multiple aspects of mRNA metabolism and function, it is possible that pseudouridylation induces structural changes and thus affects functions presumably by altering the protein occupancy profile around these sites. Besides its function in mediating nonsense-to-sense codon conversion by ribosomes, pseudouridylation could also introduce post-transcriptional genetic

recoding, thus diversifying the proteome (*Adapted from (Zhao and He, 2015)*). The presence of pseudouridine in mRNA could therefore impact all levels of post-transcriptional regulation, including translation, stress response and degradation.

YBX3 transcriptome-wide characterization of RNA-binding sites

In our study of the Y-box binding protein 3 (YBX3), we applied PAR-CLIP (Photoactivatable Ribonucleoside Enhanced Crosslinking and Immunoprecipitation) (Hafner et al., 2010a) method in order to identify the RNA-binding sites in the transcriptome of HEK293 cells. YBX3 identification of RNA-binding sites revealed a high number of binding sites, between 19,000 and 58,000 filtered clusters mapping to 6,000-9,000 genes recovered from the 4SU libraries, while the 6SG library counted 1,200 clusters in approximately 800 genes (Table 4). Interestingly, the high overlap between the 4SU PAR-CLIP targets showed a good reproducibility between the biological replicates with over 5,000 genes commonly enriched. Moreover, 95% of the 6SG targets were enriched in all three 4SU libraries suggesting that 4SU labeling does not introduce a bias in target genes enrichment (Figure 12E). The over 20,000 high confidence binding sites indentified in 6,000 genes (Table 4) suggested large target diversity for YBX3, therefore he functionality of such a high number of target transcripts was tested by pSILAC mass-spectrometry based proteomics further in order to evaluate YBX3 regulatory activity of its target mRNAs.

YBX3 binding across the transcript regions showed an equally distributed preference for 3' UTRs and CDS regions and to a lesser extent towards the 5' UTR and introns (Figure 12D, 13A). In line with this observation, two reports indicate YBX3 binding preference to mRNA 3' UTR region. One study has suggested a role of YBX3 in mRNA stability by binding to the VEGF 3' UTR and 5' UTR in a complex with PTB and YBX1 (Coles et al., 2004a). Another study involving the YBX3 mouse homologue MSY4 showed a function in translational repression by binding to the 3' UTR region of Prm1 mRNA containing a conserved RNA consensus motif (Davies et al., 2000; Giorgini et al., 2001a). Furthermore, density of the T-C conversions profile in the different transcript regions (Figure 13A) showed that YBX3 binds uniformly to the CDS but has a higher binding preference towards the beginning of the 3' UTRs.

The next question we aimed to answer in characterizing YBX3-binding sites across the transcriptome was the function of YBX3 target transcripts. A gene ontology term analysis of the top 1000 YBX3 target transcripts revealed that the bound mRNAs encode proteins involved in mRNA processing, cell cycle and chromosome

organization (Table 5 and Table 6). The enrichment of cell cycle genes in the gene ontology analysis is supporting the role of YBX3 in cell cycle regulation, which has been suggested by the study of p21 3' UTR mediated binding of YBX3 (Nie et al., 2012). The study using YBX3 overexpression and cDNA arrays reported that the expression of cell cycle genes such as PCNA and cyclin D1 as well as ERBB2 cancer related gene was affected by YBX3 overexpression (Sourisseau et al., 2006).

YBX3 depletion and global effects on protein synthesis of target mRNAs

In our characterization of YBX3 regulatory functions on its target mRNAs, we performed YBX3 siRNA knockdown couped with pulsed SILAC and mass-spectrometry based proteomics (Lebedeva et al., 2011b; Schwanhäusser et al., 2011). The LC-MS (liquid chromatography and mass-spectrometry) quantification of the protein synthesis changes as an output of translational regulation detected around 3,300 proteins from which around 2,000 were quantified with a SILAC ratio across all three siRNA experiments. We have further performed an analysis of the proteomics dataset in comparison to the PAR-CLIP top target transcripts in order to asses the YBX3 depletion effects on global protein levels. The cumulative distribution function plot of the log2 fold change of knockdown/mock SILAC ratios showed significant differences in protein synthesis of the top YBX3 targeted transcripts with more than five binding sites per mRNA (Figure 14C). The observation that YBX3 top target transcripts (n=281, p-value = 2.513e-05) had a higher level of protein than the non-targeted transcripts might indicate that YBX3 depletion is releasing its mRNA targets or affecting their stability. In addition, we performed a validation of the 3' UTR target mRNAs using a dual luciferase reporter assay and measured the luciferase activity of the 3' UTR reporter constructs in order to evaluate changes in luciferase activity as a measure of translational output. We have determined slight increases in luciferase production for several 3' UTRs derived from the PARP1, FAM98A, CDC25A and ARHGEF2 mRNAs (Figure 16B). The increased luciferase activities indicate that YBX3 depletion supports the observed global effects on protein synthesis assessed by the pulsed SILAC mass-spectrometry quantitative proteomics. Nevertheless, a mutagenesis experiment of the YBX3 RNA-recognition motif present in the 3' UTRs of these candidates, could indicate the functional relationship between the YBX3 and its mRNA targeted transcripts.

In the context of the published studies about the multifunctionality of YBX3 and in general of Y-box binding proteins (Matsumoto and Wolffe, 1998; Mihailovich et al., 2010a), our results show that high confidence mRNA targets bound by YBX3 are

affected at the gene expression level. Further experiments to show whether this effect takes place via mRNA stability or decay through or through a translational repression mechanism are necessary to determine YBX3 role in post-transcriptional regulation. Moreover, in the context of studies which have reported YBX3 transcriptional function, *in vivo* approaches using transcriptional or translational inhibitors, such as actinomycin D or cycloheximide could be used to delineate YBX3 function in transcription and/or translation. A few studies focusing on single mRNA targets have described the translational repression or mRNA stabilization function of YBX3. It was observed that YBX3 in mouse can bind to a 5' end specific sequence of the Prm1 3' UTR containing a conserved sequence and also that it appears to be associated with mRNPs in polysome fractionation consistent with a repression of Prm1 message (Davies et al., 2000). However, a global functional characterization of YBX3 has not yet been reported.

YBX3 RNA binding motif and validation of motif-containing mRNA targets

Using cERMIT, a computational tool designed for the search of motif based on genome-wide quantitative regulatory evidence (Georgiev et al., 2010b) we identified a set of RNA binding motifs and derived a consensus sequence for YBX3 binding sites. We filtered the mass-spectrometry identified proteome and used for further analysis a set of 1892 proteins which had SILAC ratios in two out of three siRNA experiments, and at least two peptides identified per protein. Within the YBX3-binding sites identified by PAR-CLIP we derived the following consensus sequence [T/A][T/A/C][G/T][A/T/C]**CA**[T/C]**C**[T/A][T/A][C/A] which contains the highly conserved CAUC motif (Figure 15A). The identification of this motif is in line with the conserved mouse homologue MSY3/4 RNA motif found in the Prm1 3' UTR sequence (Giorgini et al., 2001a) which might suggest a conservation of RNA binding specificities across species. Interestingly, the top FPKM normalized YBX3 target transcripts ranked according to the motif score, showed a significant increase in protein synthesis upon YBX3 depletion compared to all quantified proteins ($p=0.005$) (Figure 15B). In addition, the top 3' UTR motif scored transcripts had also a significant increase in protein levels compared with all the SILAC quantified proteins ($p=0.02$) (Figure 15C). The increase in protein synthesis of a subset of YBX3 targets containing a motif in their 3' UTR are functionally relevant for YBX3 regulation and support further the YBX3 involvement in gene expression regulation. In addition, the YBX3 consensus RNA motif present in the top motif ranked transcripts could be used to determine which mRNAs are affected at protein synthesis level by YBX3 depletion.

YBX3 putative protein-protein interactions with factors involved in translation initiation

We have set to identify YBX3 protein-protein interactions that could be direct binders or interact indirectly on the same mRNA. Our protocol for protein-protein interaction in RNA-dependent and RNA-independent manners (Gregersen et al., 2014b), permitted the identification of putative interactions partners by SILAC labeling and mass-spectrometry. The co-immunoprecipitation of RNA-dependent protein interactions (Figure 17A) resulted in the identification of several putative interaction partners such as the IGF2BP family of proteins, CSDE1, HNRNPQ, and YBX1, which are known to be general RNA binding proteins thus it does not clearly indicate a possible mechanism for YBX3 function (Figure 17B). Nevertheless, we did confirm the interactions described by Scheibe and colleagues, who found an RNA-dependent complex containing IGF2BP1, IGF2BP3, HNRNPQ and YBX1 (Scheibe et al., 2012) that co-immunoprecipitated on the same 3' UTR short fragment derived from the PTPN13 mRNA as well as for the JMJD1C specific CDS sequence. In contrast to what we expected, our YBX3 direct protein-protein interaction assay for direct binders did not give log2 SILAC ratios higher than 1 suggesting that YBX3 is involved in weak or transient interactions (Supplementary Figure S3C). We thus evaluated putative direct interactors using the unique peptide counts for each protein quantified in both forward and reverse experiment. The peptide counts enrichment revealed a number of translation initiation and translation elongation factors, from which the most enriched one was the EIF3A eukaryotic translation initiation factor (Figure 17C). Together with other members of EIF3 family as well as EIF4G1, and EIF4G3, YBX3 putative interaction with these complexes might suggest a possible role in translational regulation. In the initiation of translation events, the EIF4E complex binds to the cap-binding mRNA complex while EIF3 binds to the 40S ribosomal subunit preventing the association of the large subunit before the mRNA is ready for translation (Jackson et al., 2010). EIF4G is a large scaffolding protein which is known to interact with the poly(A) binding protein PABP, which is found on the poly(A) tail of the mRNA, promoting the circularization and activation of the message (Hinnebusch and Lorsch, 2012). In addition to the effect observed on protein synthesis of its mRNA targets upon YBX3 depletion, and the possible interaction with the translation initiation factors in the circularized mRNA, we could hypothesize that YBX3 might be implicated in translation at the initiation stage. Interestingly, two other RNA binding proteins, annotated as cold-shock domain proteins, CSDE1 and YBX1 have been implicated in 3' UTR mediated translation inhibition via the interaction with either PABP (Duncan et al., 2009) or

EIF4E (Svitkin et al., 2009). In *Drosophila,* female viability depends on the repression of msl2 mRNA via the interaction of CSDE1-SXL complex with the 3' UTR. YBX1 was reported to inhibit PABP-dependent translation, where PABP would stimulate eIF4F activity, by competition with eIF4G for mRNA binding. In addition, we also observed a high peptide count for the Rho/Rac guanine nucleotide exchange factor (GEF) 2, ARHGEF2, which was previously reported to interact with YBX3 in order to stimulate its transcriptional or translational activities (Nie et al., 2009; Nie et al., 2012). These findings could indicate a function of YBX3 in translational repression by interacting with the translation initiation complexes. Further validation as well as an examination of the possible synergistic or antagonistic nature of these putative interactions could offer more insights into the YBX3 molecular mechanism in post-transcriptional and translational regulation.

In conclusion, we have perfomed, for the first time, a transcriptome-wide profile for the Y-box-binding protein 3, and we have tested the effect of YBX3 depletion on protein synthesis using a pulsed SILAC-based quantitative proteomics. We have observed significant changes in protein synthesis of YBX3 top PAR-CLIP target transcripts following YBX3 depletion which suggested a functional role in gene expression. We have detected a conserved RNA sequence motif in YBX3 binding sites and established that the motif presence in the 3' UTR of mRNAs determines its fate in regard to YBX3 fuction. YBX3 protein-protein interaction experiments suggested a possible role in combination with translation initiation factors but further validation such as co-immunoprecipitations are necessary in order to establish the YBX3 molecular mechanism. To define a possible role in translation, several approaches could be employed including polysome profile which could be used to asses the YBX3 effect on mRNP loading of its target mRNAs. Using CRISPR/Cas9 technology we could also take future directions of investigating the effect of YBX3 mutations, such as the AKT1 phosphorylation site which could possibly affect YBX3 regulatory functions of its mRNA target transcripts.

CONCLUSIONS AND OUTLOOK

In this thesis, I have shown that two putative RNA-binding proteins have transcriptome-wide binding characteristics and are involved in a large set of mRNA transcript targets recognition. PUS1 has been described as a tRNA putative editing enzyme, however, we have shown that it can bind thousands of mRNAs. We have identified over 6,000 binding sites in over 2,500 mRNAs using the PAR-CLIP method, but an RNAi mediated knockdown combined with pSILAC proteomics could give an insight into the PUS1 function on the protein synthesis of its target mRNAs. Also the global effect of pseudouridylation on PUS1 targets needs to be established with methods to measure mRNA half-life or translational efficiency. We demonstrated that a disease related tRNA pseudouridylation site is edited in HEK293 cells using an optimized CMCT derivation and primer extension assay and observed a strong reverse transcription stop on CCT3 mRNA. We detected a binding site in proximity of two reported pseudouridine sites in the functionally important TERC non-coding RNA associated with the telomerase. One of the reported pseudouridine sites was associated with an RNA-depdendent pseudouridine synthase, DKC1. The presence of two different pseudouridine associated sites on the same ncRNA might suggest a possible interaction between the two modifiying enzymes. Thus, it would be of interest to characterize PUS1 protein-protein interactions, in RNA-dependent or RNA-independent conditions. Ultimately, PUS1 pseudouridylation on mRNA could primarily affect the local secondary structure and indirectly the binding of other regulatory RNA-binding proteins. Once we could establish this, we can further understand how the dynamic pseudouridylation could impact mRNA metabolism.

The Y-box binding protein 3, previously described as a dual transcriptional/translational regulator has shown a high binding preference for mRNA, in our *in vivo* experiments, with thousands of binding sites and hundreds of functional targets. We have shown that YBX3 binding to a subset of functional targets is explained by an RNA binding motif present in the 3' UTR of these targets. We suspect YBX3 binds to 3' UTRs and by possible interactions with translation initiation factors to interfere at the initiation step. Mutagenesis of the consensus motif in 3' UTRs of these bound transcripts could validate these functional YBX3 cis-elements. Further experiments such as co-immunoprecipitations are required to confirm the interactions identified by SILAC based mass-spectrometry. Together these approaches will allow an insight into the molecular mechanism of YBX3 and uncover its global regulatory functions in post-transcriptional regulation.

SUMMARY

Studies of RNA-binding proteins have revealed a large set of factors that assemble into ribonucleoprotein particles and regulate each step of the mRNA life cycle from transcription to translation and decay. In these processes, mRNA editing is considered an important step. Amongst different RNA editing enzymes, pseudouridine synthase 1 (PUS1) is a member of the RNA-independent pseudouridine synthase family and has been detected in the mRNA-bound proteome studies in human and yeast. PUS1 function had been previously associated with tRNA modification and translation efficiency. However, pseudouridylation has not been understood in the context of regulation of stability, translation or degradation of mRNA. In this thesis, I have applied PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) in combination with next-generation sequencing in order to identify PUS1 RNA binding sites across the transcriptome. We discovered that PUS1 has a high binding preference towards coding sequence regions of mRNA. Furthermore, we validated two known biologically relevant PUS1 pseudouridine sites in tRNALys (UUU) using an N-Cyclohexyl-N'-(2-morpholinoethyl)carbodiimide methyl-p-toluensulfonate (CMCT)-treatment in combination with primer extension. We applied the same method to identify individual pseudouridine sites in PUS1-bound transcripts and detected a reverse transcriptase stop in the last exon of CCT3 mRNA.

The Y-box binding protein 3 is one of the RNA-binding proteins with multiple functions bearing a highly conserved cold shock domain involved in nucleic acid binding. The multifunctional roles of YBX3 have been correlated with transcription and translation, however, the molecular mechanism of post-transcriptional regulation has not been uncovered. Using PAR-CLIP we found thousands of YBX3 binding sites in a large number of mRNA target transcripts. YBX3 binding pattern revealed an equal preference towards 3' UTRs and CDSs. We analyzed YBX3 RNA sequence preferences and identified a consensus motif among the top target transcripts. To asses the influence of YBX3 on protein synthesis, we applied a pSILAC (pulsed Stable Isoptope Labeling with Amino Acids in Cell Culture) based quantitative proteomics. We found that translation of YBX3 top mRNA targets is increased upon YBX3 depletion. In addition, proteins from top bound transcripts, containing the YBX3 consensus motif in their 3' UTR, were significantly more abundant upon YBX3 depletion, suggesting that YBX3 could have a potential role in translational repression or mRNA stability. The 3' UTRs of several candidates were validated using a dual luciferase reporter assay combined with YBX3 knockdown. We investigated YBX3 function in a protein-protein interaction assay using SILAC and mass-spectrometry and found that YBX3 interacts

with other RNA-binding proteins and several translation initiation factors, suggesting a potential role in translational regulation.

In summary, the study of two different RNA-binding proteins and their associated mRNA target transcripts indicated that both PUS1 and YBX3 might play important functional roles in post-transcriptional regulation. A further investigation could provide insights at which stage these proteins are indispensable and influence in a global manner mRNA stability or translation.

ZUSAMMENFASSUNG

Bisherige Studie über RNA-bindende Proteine haben eine große Anzahl von Faktoren beschrieben, die in Ribonukleoprotein-Komplexe jeden Schritt im Leben einer Boten-RNA (mRNA) regulieren, von der Transkription hin zu Translation und Abbau. Das Editieren von RNA, d.h. die Veränderung einzelner Nukleotide nach der Transkription, ist dabei ein wichtiger Schritt. Neben anderen Enzyme spielt dabei die Pseudouridin-Synthase 1 (PUS1), die in Hefe und menschlichen Zellen gefunden wurde, eine Rolle. PUS1 wurde bisher mit der Editierung von Transfer-RNA und der Regulierung der Translationseffizienz assoziiert. Über die Funktion von PUS1 in der Regulierung der Stabilität von mRNA sowie deren Translation ist bisher nichts bekannt.

In dieser Doktorarbeit verwendete ich PAR-CLIP (Photoactivierbares-Ribonucleosid-Gestütztes Crosslinking und Immunoprecipitation) in Kombination mit Hochdurchsatzsequenzierung, um die Bindungsstelle von PUS1 im Transkriptom zu identifizieren. Dies hat gezeigt, das PUS1 vor allem im kodierenden Teil der mRNA bindet. Zudem konnten wir, mittels N-Cyclohexyl-N'-(2-morpholinoethyl)carbodiimide methyl-p-toluensulfonate (CMCT)-Behandlung kombiniert mit reverser Transkription, zwei physiologisch relevante PUS1-abhängige Pseudouridine in tRNALys(UUU) validieren. Dieselbe Methode haben wir darauf für die Identifizierung einzelner Pseudouridine in von PUS1 gebundenden Transkripten verwendet, und eine mögliche solche Modifikation im letzten Exon der CCT3-mRNA gefunden.

Ein anderes RNA-bindendes Protein ist das Y-Box bindende Protein 3 (YBX3). YBX3 hat verschiedene Funktionen und besitzt eine evolutionär hoch konservierte Kälteschock-Domäne. YBX3 ist involviert in Transkription und Translation von mRNA, die molekularen Mechanismen sind aber noch unbekannt. Mittels PAR-CLIP konnten wir tausende Bindungsstellen im Transkriptom identifizieren, gleichermaßen im kodierenden Teil und den untranslatierten Regionen am 3'-Ende der mRNAs (3'-UTRs). Die Analyse der gebundenen Sequenzen zeigte ein definiertes Sequenz-Motiv, das von YBX3 erkannt wird.

Um den Einfluss von YBX3 auf die Proteinmenge zu bestimmen, haben wir pSILAC (pulsed Stable Isoptope Labeling with Amino Acids in Cell Culture) angewandt. Bei dieser auf Massenspektrometrie beruhenden Methoden werden mit schweren Isotopen markierte Aminosäuren eingesetzt, um Änderungen der Proteinmengen quantifizieren zu können. In diesem Experiment haben wir beobachtet, dass die Depletierung von YBX3 die Mengen der Proteine, die von den von YBX3 gebundenen mRNAs translatiert wird, erhöht. Insgesamt waren die Proteine, deren mRNAs das YBX3-Sequenzmotiv in ihren 3'-UTRs beinhalten, erhöht. Dies deutet darauf hin, dass YBX3

einen negativen Einfluss auf Translation und/oder mRNA-Stabilität hat. Validiert wurde dies mittels Luciferase-Reporter-Experimenten kombiniert mit YBX3-Depletierung. In Proteinbindungsexperimenten zeigte sich, dass YBX3 indirekt mit anderen RNA-bindenden Proteinen und Translationsinitiationsfaktoren interagiert.

Zusammengefasst haben unsere Experimente dargelegt, dass PUS1 und YBX3 wichtige Funktionen in der post-transkriptionellen Regulation aufweisen. Weitere Studien können nun eingrenzen, wie diese Proteine auf globale Art und Weise in die Translation und mRNA-Stabilität eingreifen.

REFERENCES

Adachi, H., and Yu, Y.T. (2014). Insight into the mechanisms and functions of spliceosomal snRNA pseudouridylation. World J Biol Chem *5*, 398-408.

Alfonzo, J.D., Blanc, V., Rubio, M.A.T., Este, A.M., and Simpson, L. (1999). C to U editing of the anticodon of imported mitochondrial tRNA Trp allows decoding of the UGA stop codon in Leishmania tarentolae. EMBO J *18*, 7056-7062.

Aphasizhev, R., and Aphasizheva, I. (2011). Uridine insertion/deletion editing in trypanosomes: a playground for RNA-guided information transfer. Wiley interdisciplinary reviews RNA *2*, 669-685.

Arena, F., Ciliberto, G., Ciampi, S., and Cortese, R. (1978). Purification of pseudouridylate synthetase I from Salmonella typhimurium. Nucleic Acids Res *5*, 4523-4536.

Arluison, V., Batelier, G., Ries-Kautt, M., and Grosjean, H. (1999). RNA:pseudouridine synthetase Pus1 from Saccharomyces cerevisiae: oligomerization property and stoichiometry of the complex with yeast tRNA(Phe). Biochimie *81*, 751-756.

Arnold, A., Rahman, M.M., Lee, M.C., Muehlhaeusser, S., Katic, I., Gaidatzis, D., Hess, D., Scheckel, C., Wright, J.E., Stetak, A.*, et al.* (2014). Functional characterization of C. elegans Y-box-binding proteins reveals tissue-specific functions and a critical role in the formation of polysomes. Nucleic Acids Res *42*, 13353-13369.

Bae, W., Jones, P.G., and Inouye, M. (1997). CspA, the major cold shock protein of Escherichia coli, negatively regulates its own gene expression. J Bacteriol *179*, 7081-7088.

Bakin, A., and Ofengand, J. (1993). Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. Biochemistry *32*, 9754-9762.

Bakin, A.V., and Ofengand, J. (1998). Mapping of pseudouridine residues in RNA to nucleotide resolution. Methods Mol Biol *77*, 297-309.

Balda, M.S., and Matter, K. (2000a). The tight junction protein ZO-1 and an interacting transcription factor regulate ErbB-2 expression. The EMBO journal *19*, 2024-2033.

Balda, M.S., and Matter, K. (2000b). The tight junction protein ZO-1 and an interacting transcription factor regulate ErbB-2 expression. EMBO J *19*, 2024-2033.

Baltz, A.G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M.*, et al.* (2012a). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell *46*, 674-690.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M.*, et al.* (2012b). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Molecular cell *46*, 674-690.

Becker, H.F., Motorin, Y., Planta, R.J., and Grosjean, H. (1997). The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs. Nucleic Acids Res *25*, 4493-4499.

Behm-Ansmant, I., Massenet, S., Immel, F., Patton, J.R., Motorin, Y., and Branlant, C. (2006). A previously unidentified activity of yeast and mouse RNA:pseudouridine synthases 1 (Pus1p) on tRNAs. Rna *12*, 1583-1593.

Benne, R., Burg, J.V.D., Boom, J.V., and Tromp, M.C. (1986). Major transcript of the frameshifted coxll from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell *46*, 819-826.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat Rev Genet *15*, 163-175.

Bergmann, S., Royer-Pokora, B., Fietze, E., Jurchott, K., Hildebrandt, B., Trost, D., Leenders, F., Claude, J.C., Theuring, F., Bargou, R.*, et al.* (2005). YB-1 provokes breast cancer through the induction of chromosomal instability that emerges from mitotic failure and centrosome amplification. Cancer Res *65*, 4078-4087.

Boussadia, O., Amiot, F., Cases, S., Triqueneaux, G., Jacquemin-Sablon, H., and Dautry, F. (1997). Transcription of unr (upstream of N-ras) down-modulates N-ras expression in vivo. FEBS Lett *420*, 20-24.

Bouvet, P., Matsumoto, K., and Wolffe, A.P. (1995). Sequence-specific RNA recognition by the Xenopus Y-box proteins. An essential role for the cold shock domain. J Biol Chem *270*, 28297-28303.

Brown, E.C., and Jackson, R.J. (2004). All five cold-shock domains of unr (upstream of N-ras) are required for stimulation of human rhinovirus RNA translation. J Gen Virol *85*, 2279-2287.

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

Bykhovskaya, Y., Casas, K., Mengesha, E., Inbal, A., and Fischel-Ghodsian, N. (2004b). Missense mutation in pseudouridine synthase 1 (PUS1) causes mitochondrial myopathy and sideroblastic anemia (MLASA). American journal of human genetics *74*, 1303-1308.

Bykhovskaya, Y., Mengesha, E., and Fischel-Ghodsian, N. (2007). Pleiotropic effects and compensation mechanisms determine tissue specificity in mitochondrial myopathy and sideroblastic anemia (MLASA). Mol Genet Metab *91*, 148-156.

Cao, X., and Limbach, P.A. (2015). Enhanced Detection of Post-Transcriptional Modifications Using a Mass-Exclusion List Strategy for RNA Modification Mapping by LC-MS/MS. Anal Chem *87*, 8433-8440.

Carlile, T.M., Rojas-Duran, M.F., Zinshteyn, B., Shin, H., Bartoli, K.M., and Gilbert, W.V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature *515*, 143-146.

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.* (2012a). Resource Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell, 1-14.

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.* (2012b). Resource Insights

into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell *149*, 1393-1406.

Chaikam, V., and Karlson, D.T. (2010). Comparison of structure, function and regulation of plant cold shock domain proteins to bacterial and animal cold shock domain proteins. BMB Rep *43*, 1-8.

Chatterjee, K., Blaby, I.K., Thiaville, P.C., Majumder, M., Grosjean, H., Yuan, Y.A., Gupta, R., and de Crecy-Lagard, V. (2012). The archaeal COG1901/DUF358 SPOUT-methyltransferase members, together with pseudouridine synthase Pus10, catalyze the formation of 1-methylpseudouridine at position 54 of tRNA. Rna *18*, 421-433.

Chen, J., and Patton, J.R. (1999). Cloning and characterization of a mammalian pseudouridine synthase. Rna *5*, 409-419.

Chen, J., and Patton, J.R. (2000). Mouse pseudouridine synthase 1: gene structure and alternative splicing of pre-mRNA. Biochem J *352 Pt 2*, 465-473.

Chen, J.L., Opperman, K.K., and Greider, C.W. (2002). A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. Nucleic Acids Res *30*, 592-597.

Cohn, W.E., and Volkin, E. (1951). Nucleoside-5'-Phosphates from Ribonucleic Acid. Nature *167*, 483-484.

Coles, L.S., Bartley, M.A., Bert, A., Hunter, J., Polyak, S., Diamond, P., Vadas, M.a., and Goodall, G.J. (2004a). A multi-protein complex containing cold shock domain (Y-box) and polypyrimidine tract binding proteins forms on the vascular endothelial growth factor mRNA. Potential role in mRNA stabilization. European Journal of Biochemistry *271*, 648-660.

Coles, L.S., Bartley, M.A., Bert, A., Hunter, J., Polyak, S., Diamond, P., Vadas, M.A., and Goodall, G.J. (2004b). A multi-protein complex containing cold shock domain (Y-box) and polypyrimidine tract binding proteins forms on the vascular endothelial growth factor mRNA. Potential role in mRNA stabilization. Eur J Biochem *271*, 648-660.

Coles, L.S., Diamond, P., Lambrusco, L., Hunter, J., Burrows, J., Vadas, M.A., and Goodall, G.J. (2002). A novel mechanism of repression of the vascular endothelial growth factor promoter, by single strand DNA binding cold shock domain (Y-box) proteins in normoxic fibroblasts. Nucleic Acids Res *30*, 4845-4854.

Coles, L.S., Diamond, P., Occhiodoro, F., Vadas, M.A., and Shannon, M.F. (1996). Cold shock domain proteins repress transcription from the GM-CSF promoter. Nucleic Acids Res *24*, 2311-2317.

Coles, L.S., Lambrusco, L., Burrows, J., Hunter, J., Diamond, P., Bert, A.G., Vadas, M.A., and Goodall, G.J. (2005). Phosphorylation of cold shock domain/Y-box proteins by ERK2 and GSK3beta and repression of the human VEGF promoter. FEBS Lett *579*, 5372-5378.

Coles, L.S., Occhiodoro, F., Vadas, M.A., and Shannon, M.F. (1994). A sequence-specific single-strand DNA binding protein that contacts repressor sequences in the human GM-CSF promoter. Nucleic Acids Res *22*, 4276-4283.

Cook, K.B., Hughes, T.R., and Morris, Q.D. (2014). High-throughput characterization of protein-RNA interactions. Briefings in Functional Genomics.

Cook, K.B., Hughes, T.R., and Morris, Q.D. (2015). High-throughput characterization of protein-RNA interactions. Brief Funct Genomics *14*, 74-89.

Cortese, R., Kammen, H.O., Spengler, S.J., and Ames, B.N. (1974). Biosynthesis of pseudouridine in transfer ribonucleic acid. J Biol Chem *249*, 1103-1108.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

Czerwoniec, A., Dunin-Horkawicz, S., Purta, E., Kaminska, K.H., Kasprzak, J.M., Bujnicki, J.M., Grosjean, H., and Rother, K. (2009). MODOMICS: a database of RNA modification pathways. 2008 update. Nucleic Acids Res *37*, D118-121.

Czudnochowski, N., Wang, A.L., Finer-Moore, J., and Stroud, R.M. (2013). In human pseudouridine synthase 1 (hPus1), a C-terminal helical insert blocks tRNA from binding in the same orientation as in the Pus1 bacterial homologue TruA, consistent with their different target selectivities. J Mol Biol *425*, 3875-3887.

Davies, H.G., Giorgini, F., Fajardo, M.A., and Braun, R.E. (2000). A sequence-specific RNA binding complex expressed in murine germ cells contains MSY2 and MSY4. Dev Biol *221*, 87-100.

Davis, D.R. (1995). Stabilization of RNA stacking by pseudouridine. Nucleic Acids Res *23*, 5020-5026.

Davis, F.F., and Allen, F.W. (1957). Ribonucleic acids from yeast which contain a fifth nucleotide. J Biol Chem *227*, 907-915.

Deng, Y., Zhang, W., Su, D., Yang, Y., Ma, Y., Zhang, H., and Zhang, S. (2008). Some single nucleotide polymorphisms of MSY2 gene might contribute to susceptibility to spermatogenic impairment in idiopathic infertile men. Urology *71*, 878-882.

Denman, R., Colgan, J., Nurse, K., and Ofengand, J. (1988). Crosslinking of the anticodon of P site bound tRNA to C-1400 of E.coli 16S RNA does not require the participation of the 50S subunit. Nucleic Acids Res *16*, 165-178.

Dias, S.R.C., Boroni, M., Rocha, E.a., Dias, T.L., de Laet Souza, D., Oliveira, F.M.S., Bitar, M., Macedo, A.M., Machado, C.R., Caliari, M.V.*, et al.* (2014). Evaluation of the Schistosoma mansoni Y-box-binding protein (SMYB1) potential as a vaccine candidate against schistosomiasis. Frontiers in genetics *5*, 174-174.

Didier, D.K., Schiffenbauer, J., Woulfe, S.L., Zacheis, M., and Schwartz, B.D. (1988). Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box. Proc Natl Acad Sci U S A *85*, 7322-7326.

Dodt, M., Roehr, J.T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology (Basel) *1*, 895-905.

Dolfini, D., and Mantovani, R. (2013). Targeting the Y/CCAAT box in cancer: YB-1 (YBX1) or NF-Y? Cell Death Differ *20*, 676-685.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M.*, et al.* (2012).

Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201-206.

Donmez, G., Hartmuth, K., and Luhrmann, R. (2004). Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. Rna *10*, 1925-1933.

Dormoy-Raclet, V., Markovits, J., Jacquemin-Sablon, A., and Jacquemin-Sablon, H. (2005). Regulation of Unr expression by 5'- and 3'-untranslated regions of its mRNA through modulation of stability and IRES mediated translation. RNA Biol *2*, e27-35.

Duncan, K.E., Strein, C., and Hentze, M.W. (2009). The SXL-UNR corepressor complex uses a PABP-mediated mechanism to inhibit ribosome recruitment to msl-2 mRNA. Mol Cell *36*, 571-582.

Dunin-Horkawicz, S., Czerwoniec, A., Gajda, M.J., Feder, M., Grosjean, H., and Bujnicki, J.M. (2006). MODOMICS: a database of RNA modification pathways. Nucleic Acids Res *34*, D145-149.

Dupasquier, S., Delmarcelle, A.S., Marbaix, E., Cosyns, J.P., Courtoy, P.J., and Pierreux, C.E. (2014). Validation of housekeeping gene and impact on normalized gene expression in clear cell renal cell carcinoma: critical reassessment of YBX3/ZONAB/CSDA expression. BMC Mol Biol *15*, 9.

Ejby, M., Sorensen, M.A., and Pedersen, S. (2007). Pseudouridylation of helix 69 of 23S rRNA is necessary for an effective translation termination. Proc Natl Acad Sci U S A *104*, 19410-19415.

Eliseeva, I.A., Kim, E.R., Guryanov, S.G., Ovchinnikov, L.P., and Lyabin, D.N. (2011). REVIEW Y Box Binding Protein 1 (YB 1) and Its Functions. Biochemistry Moscow *76*, 1402-1433.

Ellington, A.D., and Szostak, J.W. (1990). In vitro selection of RNA molecules that bind specific ligands. Nature *346*, 818-822.

Ero, R., Peil, L., Liiv, A., and Remme, J. (2008). Identification of pseudouridine methyltransferase in Escherichia coli. Rna *14*, 2223-2233.

Evdokimova, V.M., Kovrigina, E.A., Nashchekin, D.V., Davydova, E.K., Hershey, J.W., and Ovchinnikov, L.P. (1998). The major core protein of messenger ribonucleoprotein particles (p50) promotes initiation of protein biosynthesis in vitro. J Biol Chem *273*, 3574-3581.

Evdokimova, V.M., Wei, C.L., Sitikov, A.S., Simonenko, P.N., Lazarev, O.A., Vasilenko, K.S., Ustinov, V.A., Hershey, J.W., and Ovchinnikov, L.P. (1995). The major protein of messenger ribonucleoprotein particles in somatic cells is a member of the Y-box binding transcription factor family. J Biol Chem *270*, 3186-3192.

Fabian, M.R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. Annu Rev Biochem *79*, 351-379.

Fallah, R. (2007). MLASA SYNDROME: A CASE REPORT. Iran J Child Neurology, 47-50.

Fernandez-Vizarra, E., Berardinelli, A., Valente, L., Tiranti, V., and Zeviani, M. (2007). Nonsense mutation in pseudouridylate synthase 1 (PUS1) in two brothers affected by

myopathy, lactic acidosis and sideroblastic anaemia (MLASA). J Med Genet *44*, 173-180.

Franco, G.R., Garratt, R.C., Tanaka, M., Simpson, A.J., and Pena, S.D. (1997). Characterization of a Schistosoma mansoni gene encoding a homologue of the Y-box binding protein. Gene *198*, 5-16.

Frankel, P., Aronheim, A., Kavanagh, E., Balda, M.S., Matter, K., Bunney, T.D., and Marshall, C.J. (2005). RalA interacts with ZONAB in a cell density-dependent manner and regulates its transcriptional activity. EMBO J *24*, 54-62.

Friedersdorf, M.B., and Keene, J.D. (2014a). Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. Genome biology *15*, R2-R2.

Friedersdorf, M.B., and Keene, J.D. (2014b). Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. Genome Biol *15*, R2.

Frye, B.C., Halfter, S., Djudjaj, S., Muehlenberg, P., Weber, S., Raffetseder, U., En-Nia, A., Knott, H., Baron, J.M., Dooley, S.*, et al.* (2009). Y-box protein-1 is actively secreted through a non-classical pathway and acts as an extracellular mitogen. EMBO Rep *10*, 783-789.

Fu, Y., Jia, G., Pang, X., Wang, R.N., Wang, X., Li, C.J., Smemo, S., Dai, Q., Bailey, K.a., Nobrega, M.a.*, et al.* (2013). FTO-mediated formation of N(6)-hydroxymethyladenosine and N(6)-formyladenosine in mammalian RNA. Nature communications *4*, 1798-1798.

Geisberg, Joseph V., Moqtaderi, Z., Fan, X., Ozsolak, F., and Struhl, K. (2014). Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast. Cell *156*, 812-824.

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

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

Gerber, a.P., and Keller, W. (2001). RNA editing by base deamination: more enzymes, more targets, new mysteries. Trends in biochemical sciences *26*, 376-384.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nature Reviews Genetics.

Ghosh, S.K., Patton, J.R., and Spanjaard, R.a. (2012). A Small RNA Derived from RNA Coactivator SRA Blocks Steroid Receptor Signaling via Inhibition of Pus1p-Mediated Pseudouridylation of SRA: Evidence of a Novel RNA Binding Domain in the N-Terminus of Steroid Receptors. Biochemistry *51*, 8163-8172.

Gimenez-Bonafe, P., Fedoruk, M.N., Whitmore, T.G., Akbari, M., Ralph, J.L., Ettinger, S., Gleave, M.E., and Nelson, C.C. (2004). YB-1 is upregulated during prostate cancer tumor progression and increases P-glycoprotein activity. Prostate *59*, 337-349.

Giorgi, A., Di Francesco, L., Principe, S., Mignogna, G., Sennels, L., Mancone, C., Alonzi, T., Sbriccoli, M., De Pascalis, A., Rappsilber, J.*, et al.* (2009). Proteomic

profiling of PrP27-30-enriched preparations extracted from the brain of hamsters with experimental scrapie. Proteomics *9*, 3802-3814.

Giorgini, F., Davies, H.G., and Braun, R.E. (2001a). MSY2 and MSY4 bind a conserved sequence in the 3' untranslated region of protamine 1 mRNA in vitro and in vivo. Mol Cell Biol *21*, 7010-7019.

Giorgini, F., Davies, H.G., and Braun, R.E. (2001b). MSY2 and MSY4 Bind a Conserved Sequence in the 3′ Untranslated Region of Protamine 1 mRNA In Vitro and In Vivo MSY2 and MSY4 Bind a Conserved Sequence in the 3 J Untranslated Region of Protamine 1 mRNA In Vitro and In Vivo. *21*.

Goodarzi, H., Liu, X., Nguyen, Hoang C.B., Zhang, S., Fish, L., and Tavazoie, Sohail F. (2015). Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. Cell *161*, 790-802.

Goroncy, A.K., Koshiba, S., Tochio, N., Tomizawa, T., Inoue, M., Watanabe, S., Harada, T., Tanaka, A., Ohara, O., Kigawa, T.*, et al.* (2010). The NMR solution structures of the five constituent cold-shock domains (CSD) of the human UNR (upstream of N-ras) protein. J Struct Funct Genomics *11*, 181-188.

Gott, J.M., and Emeson, R.B. (2000). Functions and mechanisms of RNA editing. Annu Rev Genet, 499-531.

Gray, M.W. (2012). Evolutionary Origin of RNA Editing. Biochemistry.

Green, C.J., Kammen, H.O., and Penhoet, E.E. (1982). Purification and properties of a mammalian tRNA pseudouridine synthase. J Biol Chem *257*, 3045-3052.

Greenbaum, D., Colangelo, C., Williams, K., and Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol *4*, 117.

Gregersen, L.H., Schueler, M., Munschauer, M., Mastrobuoni, G., Chen, W., Kempa, S., Dieterich, C., and Landthaler, M. (2014a). Article Contributing to UPF1 mRNA Target Degradation by Translocation along 3 0 UTRs. Molecular Cell *54*, 1-13.

Gregersen, L.H., Schueler, M., Munschauer, M., Mastrobuoni, G., Chen, W., Kempa, S., Dieterich, C., and Landthaler, M. (2014b). MOV10 Is a 5' to 3' RNA helicase contributing to UPF1 mRNA target degradation by translocation along 3' UTRs. Mol Cell *54*, 573-585.

Gu, X., Liu, Y., and Santi, D.V. (1999). The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA. Proc Natl Acad Sci U S A *96*, 14270-14275.

Gurha, P., and Gupta, R. (2008). Archaeal Pus10 proteins can produce both pseudouridine 54 and 55 in tRNA. Rna *14*, 2521-2527.

Guryanov, S.G., Selivanova, O.M., Nikulin, A.D., Enin, G.A., Melnik, B.S., Kretov, D.A., Serdyuk, I.N., and Ovchinnikov, L.P. (2012). Formation of amyloid-like fibrils by Y-box binding protein 1 (YB-1) is mediated by its cold shock domain and modulated by disordered terminal domains. PLoS ONE *7*, e36969.

Gutgsell, N., Englund, N., Niu, L., Kaya, Y., Lane, B.G., and Ofengand, J. (2000). Deletion of the Escherichia coli pseudouridine synthase gene truB blocks formation of pseudouridine 55 in tRNA in vivo, does not affect exponential growth, but confers a strong selective disadvantage in competition with wild-type cells. Rna *6*, 1870-1881.

Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. Mol Cell Biol *19*, 1720-1730.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M.*, et al.* (2010a). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. CELL *141*, 129-141.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M.*, et al.* (2010b). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129-141.

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

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

Hamma, T., and Ferre-D'Amare, A.R. (2006). Pseudouridine synthases. Chem Biol *13*, 1125-1135.

Hammoud, S., Emery, B.R., Dunn, D., Weiss, R.B., and Carrell, D.T. (2009). Sequence alterations in the YBX2 gene are associated with male factor infertility. Fertil Steril *91*, 1090-1095.

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

Hellmuth, K., Grosjean, H., Motorin, Y., Deinert, K., Hurt, E., and Simos, G. (2000). Cloning and characterization of the Schizosaccharomyces pombe tRNA:pseudouridine synthase Pus1p. Nucleic Acids Res *28*, 4604-4610.

Hengesbach, M., Voigts-Hoffmann, F., Hofmann, B., and Helm, M. (2010). Formation of a stalled early intermediate of pseudouridine synthesis monitored by real-time FRET. RNA (New York, NY) *16*, 610-620.

Henneberg, B., Swiniarski, S., Sabine, B., and Illenberger, S. (2010). A conserved peptide motif in Raver2 mediates its interaction with the polypyrimidine tract-binding protein. Exp Cell Res *316*, 966-979.

Hennig, J., Militti, C., Popowicz, G.M., Wang, I., Sonntag, M., Geerlof, A., Gabel, F., Gebauer, F., and Sattler, M. (2014). Structural basis for the assembly of the Sxl-Unr translation regulatory complex. Nature *515*, 287-290.

Hinnebusch, A.G., and Lorsch, J.R. (2012). The mechanism of eukaryotic translation initiation: new insights and challenges. Cold Spring Harb Perspect Biol *4*.

Ho, N.W., and Gilham, P.T. (1971). Reaction of pseudouridine and inosine with N-cyclohexyl-N'-beta-(4-methylmorpholinium)ethylcarbodiimide. Biochemistry *10*, 3651-3657.

Hockensmith, J.W., Kubasek, W.L., Vorachek, W.R., and von Hippel, P.H. (1986). Laser cross-linking of nucleic acids to proteins. Methodology and first applications to the phage T4 DNA replication system. J Biol Chem *261*, 3512-3518.

Horn, G., Hofweber, R., Kremer, W., and Kalbitzer, H.R. (2007). Structure and function of bacterial cold shock proteins. Cell Mol Life Sci *64*, 1457-1470.

Hu, Z., Jin, S., and Scotto, K.W. (2000). Transcriptional activation of the MDR1 gene by UV irradiation. Role of NF-Y and Sp1. J Biol Chem *275*, 2979-2985.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Huang, L., Pookanjanatavip, M., Gu, X., and Santi, D.V. (1998). A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. Biochemistry *37*, 344-351.

Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. The Journal of cell biology *189*, 739-754.

Hubner, N.C., and Mann, M. (2011). Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC). Methods (San Diego, Calif) *53*, 453-459.

Huet, T., Miannay, F.A., Patton, J.R., and Thore, S. (2014). Steroid receptor RNA activator (SRA) modification by the human pseudouridine synthase 1 (hPus1p): RNA binding, activity, and atomic model. PLoS ONE *9*, e94610.

Hur, S., and Stroud, R.M. (2007). How U38, 39, and 40 of many tRNAs become the targets for pseudouridylation by TruA. Mol Cell *26*, 189-203.

Ingolia, Nicholas T., Lareau, Liana F., and Weissman, Jonathan S. (2011). Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. Cell, 1-14.

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M., Dieterich, C.*, et al.* (2015). Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. Cell Rep *10*, 170-177.

Jack, K., Bellodi, C., Landry, D.M., Niederer, R.O., Meskauskas, A., Musalgaonkar, S., Kopmar, N., Krasnykh, O., Dean, A.M., Thompson, S.R.*, et al.* (2011a). rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. Mol Cell *44*, 660-666.

Jack, K., Bellodi, C., Landry, Dori M., Niederer, Rachel O., Meskauskas, A., Musalgaonkar, S., Kopmar, N., Krasnykh, O., Dean, Alison M., Thompson, Sunnie R., *et al.* (2011b). rRNA Pseudouridylation Defects Affect Ribosomal Ligand Binding and Translational Fidelity from Yeast to Human Cells. Molecular Cell *44*, 660-666.

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.

Jeffers, M., Paciucci, R., and Pellicer, A. (1990). Characterization of unr; a gene closely linked to N-ras. Nucleic Acids Res *18*, 4891-4899.

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.G., *et al.* (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol *7*, 885-887.

Johnson, L., and Soll, D. (1970). In vitro biosynthesis of pseudouridine at the polynucleotide level by an enzyme extract from Escherichia coli. Proc Natl Acad Sci U S A *67*, 943-950.

Jovanovic, M., Rooney, M.S., Mertins, P., Przybylski, D., Chevrier, N., Satija, R., Rodriguez, E.H., Fields, A.P., Schwartz, S., Raychowdhury, R., *et al.* (2015). Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens. Science *347*, 1259038.

Jung, Y.S., Qian, Y., and Chen, X. (2010). Examination of the expanding pathways for the regulation of p21 expression and activity. Cell Signal *22*, 1003-1012.

Kamalampeta, R., Keffer-Wilkes, L.C., and Kothe, U. (2013). tRNA binding, positioning, and modification by the pseudouridine synthase Pus10. J Mol Biol *425*, 3863-3874.

Kamalampeta, R., and Kothe, U. (2012). Archaeal proteins Nop10 and Gar1 increase the catalytic activity of Cbf5 in pseudouridylating tRNA. Sci Rep *2*, 663.

Kanitz, A., and Gerber, A.P. (2010). Circuitry of mRNA regulation. Wiley Interdiscip Rev Syst Biol Med *2*, 245-251.

Karlson, D., Nakaminami, K., Toyomasu, T., and Imai, R. (2002). A cold-regulated nucleic acid-binding protein of winter wheat shares a domain with bacterial cold shock proteins. J Biol Chem *277*, 35248-35256.

Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. Nature reviews Genetics *8*, 533-543.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res *12*, 996-1006.

Kim, C.G., and Sheffery, M. (1990). Physical characterization of the purified CCAAT transcription factor, alpha-CP1. J Biol Chem *265*, 13362-13369.

Kim, M.H., Sasaki, K., and Imai, R. (2009). Cold shock domain protein 3 regulates freezing tolerance in Arabidopsis thaliana. J Biol Chem *284*, 23454-23460.

Kim, N.K., Theimer, C.A., Mitchell, J.R., Collins, K., and Feigon, J. (2010). Effect of pseudouridylation on the structure and activity of the catalytically essential P6.1 hairpin in human telomerase RNA. Nucleic Acids Res *38*, 6746-6756.

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.

Kloks, C.P., Spronk, C.A., Lasonder, E., Hoffmann, A., Vuister, G.W., Grzesiek, S., and Hilbers, C.W. (2002). The solution structure and DNA-binding properties of the cold-shock domain of the human Y-box protein YB-1. J Mol Biol *316*, 317-326.

Kobitski, A.Y., Hengesbach, M., Helm, M., and Nienhaus, G.U. (2008). Sculpting an RNA conformational energy landscape by a methyl group modification--a single-molecule FRET study. Angew Chem Int Ed Engl *47*, 4326-4330.

Komili, S., and Silver, P.a. (2008). Coupling and coordination in gene expression processes: a systems biology view. Nature reviews Genetics *9*, 38-48.

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. Nat Struct Mol Biol *17*, 909-915.

Koonin, E.V. (1996). Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. Nucleic Acids Res *24*, 2411-2415.

Kowalak, J.A., Bruenger, E., Hashizume, T., Peltier, J.M., Ofengand, J., and McCloskey, J.A. (1996). Structural characterization of U*-1915 in domain IV from Escherichia coli 23S ribosomal RNA as 3-methylpseudouridine. Nucleic Acids Res *24*, 688-693.

Kudo, S., Mattei, M.G., and Fukuda, M. (1995). Characterization of the gene for dbpA, a family member of the nucleic-acid-binding proteins containing a cold-shock domain. Eur J Biochem *231*, 72-82.

Kwon, S.C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nat Struct Mol Biol *20*, 1122-1130.

Ladomery, M., and Sommerville, J. (1994). Binding of Y-box proteins to RNA: involvement of different protein domains. Nucleic Acids Res *22*, 5582-5589.

Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P.a., and Burge, C.B. (2014). RNA Bind-n-Seq: Quantitative Assessment of the Sequence and Structural Binding Specificity of RNA Binding Proteins. Molecular cell *54*, 887-900.

Landsman, D. (1992). RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain. Nucleic Acids Res *20*, 2861-2864.

Landthaler, M., Gaidatzis, D., Rothballer, A., Chen, P.Y., Soll, S.J., Dinic, L., Ojo, T., Hafner, M., Zavolan, M., and Tuschl, T. (2008). Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. RNA *14*, 2580-2580.

Lane, B.G., Ofengand, J., and Gray, M.W. (1995). Pseudouridine and O2'-methylated nucleosides. Significance of their selective occurrence in rRNA domains that function in ribosome-catalyzed synthesis of the peptide bonds in proteins. Biochimie *77*, 7-15.

Lebedeva, S., Jens, M., Theil, K., Schwanhäusser, B., Selbach, M., Landthaler, M., and Rajewsky, N. (2011a). Transcriptome-wide Analysis of Regulatory Interactions of the RNA-Binding Protein HuR. Molecular cell.

Lebedeva, S., Jens, M., Theil, K., Schwanhäusser, B., Selbach, M., Landthaler, M., and Rajewsky, N. (2011b). Transcriptome-wide Analysis of Regulatory Interactions of the RNA-Binding Protein HuR. Molecular cell, 1-13.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, X., Zhu, P., Ma, S., Song, J., Bai, J., Sun, F., and Yi, C. (2015a). Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nat Chem Biol.

Li, Y.H., Zhang, G., and Cui, Q. (2015b). PPUS: a web server to predict PUS-specific pseudouridine sites. Bioinformatics.

Lindquist, J.a., Brandt, S., Bernhardt, A., Zhu, C., and Mertens, P.R. (2014a). The role of cold shock domain proteins in inflammatory diseases. Journal of molecular medicine (Berlin, Germany).

Lindquist, J.A., Brandt, S., Bernhardt, A., Zhu, C., and Mertens, P.R. (2014b). The role of cold shock domain proteins in inflammatory diseases. J Mol Med (Berl) *92*, 207-216.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., *et al.* (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol *10*, 93-95.

Liu, N., Parisien, M., Dai, Q., Zheng, G., He, C., and Pan, T. (2013a). Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. Rna *19*, 1848-1856.

Liu, Y., Hu, W., Murakawa, Y., Yin, J., Wang, G., Landthaler, M., and Yan, J. (2013b). Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation. Sci Rep *3*, 2054.

Loughlin, F.E., Gebert, L.F., Towbin, H., Brunschweiger, A., Hall, J., and Allain, F.H. (2012). Structural basis of pre-let-7 miRNA recognition by the zinc knuckles of pluripotency factor Lin28. Nat Struct Mol Biol *19*, 84-89.

Lovejoy, A.F., Riordan, D.P., and Brown, P.O. (2014). Transcriptome-Wide Mapping of Pseudouridines: Pseudouridine Synthases Modify Specific mRNAs in S. cerevisiae. PloS one *9*, e110799-e110799.

Lu, Z.H., Books, J.T., and Ley, T.J. (2006). Cold shock domain family members YB-1 and MSY4 share essential functions during murine embryogenesis. Mol Cell Biol *26*, 8410-8417.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nature reviews Molecular cell biology *8*, 479-490.

Machnicka, M.A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M., *et al.* (2013). MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res *41*, D262-267.

Manival, X., Ghisolfi-Nieto, L., Joseph, G., Bouvet, P., and Erard, M. (2001). RNA-binding strategies common to cold-shock domain- and RNA recognition motif-containing proteins. Nucleic Acids Res *29*, 2223-2233.

Mansfield, J.H., Wilhelm, J.E., and Hazelrigg, T. (2002). Ypsilon Schachtel, a Drosophila Y-box protein, acts antagonistically to Orb in the oskar mRNA localization and translation pathway. Development *129*, 197-209.

Massenet, S., Motorin, Y., Lafontaine, D.L., Hurt, E.C., Grosjean, H., and Branlant, C. (1999). Pseudouridine mapping in the Saccharomyces cerevisiae spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. Mol Cell Biol *19*, 2142-2154.

Matsumoto, G., Yajima, N., Saito, H., Nakagami, H., Omi, Y., Lee, U., and Kaneda, Y. (2010). Cold shock domain protein A (CSDA) overexpression inhibits tumor growth and lymph node metastasis in a mouse model of squamous cell carcinoma. Clin Exp Metastasis *27*, 539-547.

Matsumoto, K., and Bay, B.-h. (2005). Significance of the Y-box proteins in human cancers. Journal of Molecular and Genetic Medicine, 11-17.

Matsumoto, K., Meric, F., and Wolffe, A.P. (1996). Translational repression dependent on the interaction of the Xenopus Y-box protein FRGY2 with mRNA. Role of the cold shock domain, tail domain, and selective RNA sequence recognition. J Biol Chem *271*, 22706-22712.

Matsumoto, K., and Wolffe, A.P. (1998). Gene regulation by Y-box proteins: coupling control of transcription and translation. Trends Cell Biol *8*, 318-323.

Max, K.E., Zeeb, M., Bienert, R., Balbach, J., and Heinemann, U. (2006). T-rich DNA single strands bind to a preformed site on the bacterial cold shock protein Bs-CspB. J Mol Biol *360*, 702-714.

Max, K.E., Zeeb, M., Bienert, R., Balbach, J., and Heinemann, U. (2007). Common mode of DNA binding to cold shock domains. Crystal structure of hexathymidine bound to the domain-swapped form of a major cold shock protein from Bacillus caldolyticus. FEBS J *274*, 1265-1279.

Mayr, F., and Heinemann, U. (2013). Mechanisms of Lin28-mediated miRNA and mRNA regulation--a structural and functional perspective. Int J Mol Sci *14*, 16532-16553.

Mayr, F., Schutz, A., Doge, N., and Heinemann, U. (2012). The Lin28 cold-shock domain remodels pre-let-7 microRNA. Nucleic Acids Res *40*, 7492-7506.

McCleverty, C.J., Hornsby, M., Spraggon, G., and Kreusch, A. (2007). Crystal structure of human Pus10, a novel pseudouridine synthase. Journal of molecular biology *373*, 1243-1254.

Meyer, B., Wurm, J.P., Kotter, P., Leisegang, M.S., Schilling, V., Buchhaupt, M., Held, M., Bahr, U., Karas, M., Heckel, A.*, et al.* (2011). The Bowen-Conradi syndrome protein Nep1 (Emg1) has a dual role in eukaryotic ribosome biogenesis, as an essential assembly factor and in the methylation of Psi1191 in yeast 18S rRNA. Nucleic Acids Res *39*, 1526-1537.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Resource Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3 0 UTRs and near Stop Codons. Cell, 1635-1646.

Mihailovich, M., Militti, C., Gabaldon, T., and Gebauer, F. (2010a). Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. Bioessays *32*, 109-118.

Mihailovich, M., Militti, C., Gabaldón, T., and Gebauer, F. (2010b). Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. BioEssays : news and reviews in molecular, cellular and developmental biology *32*, 109-118.

Milek, M., Wyler, E., and Landthaler, M. (2012). Transcriptome-wide analysis of protein-RNA interactions using high-throughput sequencing. Semin Cell Dev Biol *23*, 206-212.

Minich, W.B., and Ovchinnikov, L.P. (1992). Role of cytoplasmic mRNP proteins in translation. Biochimie *74*, 477-483.

Mitchell, A., Chang, H.Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S*., et al.* (2015). The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res *43*, D213-221.

Mitchell, J.R., Wood, E., and Collins, K. (1999). A telomerase component is defective in the human disease dyskeratosis congenita. Nature *402*, 551-555.

Mitchell, S.A., Spriggs, K.A., Coldwell, M.J., Jackson, R.J., and Willis, A.E. (2003). The Apaf-1 internal ribosome entry segment attains the correct structural conformation for function via interactions with PTB and unr. Mol Cell *11*, 757-771.

Mitchell, S.F., Jain, S., She, M., and Parker, R. (2012). Global analysis of yeast mRNPs. Nature Structural & Molecular Biology, 1-9.

Moore, M.J. (2005). From birth to death: the complex lives of eukaryotic mRNAs. Science *309*, 1514-1518.

Morris, A.R., Mukherjee, N., and Keene, J.D. (2010). Systematic analysis of posttranscriptional gene expression. Wiley Interdiscip Rev Syst Biol Med *2*, 162-180.

Motorin, Y., Keith, G., Simon, C., Foiret, D., Simos, G., Hurt, E., and Grosjean, H. (1998). The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. Rna *4*, 856-869.

Motorin, Y., Muller, S., Behm-Ansmant, I., and Branlant, C. (2007). Identification of modified residues in RNAs by reverse transcription-based methods. Methods in enzymology *425*, 21-53.

Müller-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.

Munschauer, M., Schueler, M., Dieterich, C., and Landthaler, M. (2013). High-resolution profiling of protein occupancy on polyadenylated RNA transcripts. Methods (San Diego, Calif).

Murray, M.T., Krohne, G., and Franke, W.W. (1991). Different forms of soluble cytoplasmic mRNA binding proteins and particles in Xenopus laevis oocytes and embryos. J Cell Biol *112*, 1-11.

Nakaminami, K., Karlson, D.T., and Imai, R. (2006). Functional conservation of cold shock domains in bacteria and higher plants. Proc Natl Acad Sci U S A *103*, 10122-10127.

Nakaminami, K., Sasaki, K., Kajita, S., Takeda, H., Karlson, D., Ohgi, K., and Imai, R. (2005). Heat stable ssDNA/RNA-binding activity of a wheat cold shock domain protein. FEBS Lett *579*, 4887-4891.

Nam, Y., Chen, C., Gregory, Richard I., Chou, James J., and Sliz, P. (2011). Molecular Basis for Interaction of let-7 MicroRNAs with Lin28. Cell *147*, 1080-1091.

Newkirk, K., Feng, W., Jiang, W., Tejero, R., Emerson, S.D., Inouye, M., and Montelione, G.T. (1994). Solution NMR structure of the major cold shock protein (CspA) from Escherichia coli: identification of a binding epitope for DNA. Proceedings of the National Academy of Sciences of the United States of America *91*, 5114-5118.

Newman, M.A., Thomson, J.M., and Hammond, S.M. (2008). Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. Rna *14*, 1539-1549.

Nie, M., Aijaz, S., Leefa Chong San, I.V., Balda, M.S., and Matter, K. (2009). The Y-box factor ZONAB/DbpA associates with GEF-H1/Lfc and mediates Rho-stimulated transcription. EMBO Rep *10*, 1125-1131.

Nie, M., Balda, M.S., and Matter, K. (2012). Stress- and Rho-activated ZO-1-associated nucleic acid binding protein binding to p21 mRNA mediates stabilization, translation, and cell survival. Proc Natl Acad Sci U S A *109*, 10897-10902.

Nishikura, K. (2009). Functions and Regulation of RNA Editing by ADAR Deaminases. Annual review of biochemistry, 1-29.

Niu, Y., Zhao, X., Wu, Y.-S., Li, M.-M., Wang, X.-J., and Yang, Y.-G. (2013). N(6)-methyl-adenosine (m(6)A) in RNA: An Old Modification with A Novel Epigenetic Function. Genomics, proteomics & bioinformatics *11*, 8-17.

Ofengand, J., and Bakin, A. (1997). Mapping to Nucleotide Resolution of Pseudouridine Residues in Large Subunit Ribosomal RNAs from Representative Eukaryotes , Prokaryotes , Archaebacteria , Mitochondria and Chloroplasts. J Mol Biol *266*, 246-268.

Ofengand, J., Bakin, A., Wrzesinski, J., Nurse, K., and Lane, B.G. (1995). The pseudouridine residues of ribosomal RNA. Biochem Cell Biol *73*, 915-924.

Ong, S.E., and Mann, M. (2006). A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat Protoc *1*, 2650-2660.

Parker, B.J., Moltke, I., Roth, A., Washietl, S., Wen, J., Kellis, M., Breaker, R., and Pedersen, J.S. (2011). New families of human regulatory RNA structures identified by comparative analysis of vertebrate genomes. Genome Res *21*, 1929-1943.

Patton, J.R., Bykhovskaya, Y., Mengesha, E., Bertolotto, C., and Fischel-Ghodsian, N. (2005a). Mitochondrial myopathy and sideroblastic anemia (MLASA): missense

mutation in the pseudouridine synthase 1 (PUS1) gene is associated with the loss of tRNA pseudouridylation. J Biol Chem *280*, 19823-19828.

Patton, J.R., Bykhovskaya, Y., Mengesha, E., Bertolotto, C., and Fischel-Ghodsian, N. (2005b). Mitochondrial myopathy and sideroblastic anemia (MLASA): missense mutation in the pseudouridine synthase 1 (PUS1) gene is associated with the loss of tRNA pseudouridylation. The Journal of Biological Chemistry *280*, 19823-19828.

Pedersen, J.S., Bejerano, G., Siepel, A., Rosenbloom, K., Lindblad-Toh, K., Lander, E.S., Kent, J., Miller, W., and Haussler, D. (2006). Identification and classification of conserved RNA secondary structures in the human genome. PLoS Comput Biol *2*, e33.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread Co-translational RNA Decay Reveals Ribosome Dynamics. CELL *161*, 1400-1412.

Penes, M.C., Li, X., and Nagy, J.I. (2005). Expression of zonula occludens-1 (ZO-1) and the transcription factor ZO-1-associated nucleic acid-binding protein (ZONAB)-MsY3 in glial cells and colocalization at oligodendrocyte and astrocyte gap junctions in mouse brain. The European journal of neuroscience *22*, 404-418.

Piskounova, E., Viswanathan, S.R., Janas, M., LaPierre, R.J., Daley, G.Q., Sliz, P., and Gregory, R.I. (2008). Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. J Biol Chem *283*, 21310-21314.

Purta, E., Kaminska, K.H., Kasprzak, J.M., Bujnicki, J.M., and Douthwaite, S. (2008). YbeA is the m3Psi methyltransferase RlmH that targets nucleotide 1915 in 23S rRNA. Rna *14*, 2234-2244.

Rabani, M., Raychowdhury, R., Jovanovic, M., Rooney, M., Stumpo, D.J., Pauli, A., Hacohen, N., Schier, A.F., Blackshear, P.J., Friedman, N.*, et al.* (2014). High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. CELL *159*, 1698-1710.

Raffetseder, U., Rauen, T., Boor, P., Ostendorf, T., Hanssen, L., Floege, J., En-Nia, A., Djudjaj, S., Frye, B.C., and Mertens, P.R. (2011). Extracellular YB-1 blockade in experimental nephritis upregulates Notch-3 receptor expression and signaling. Nephron Exp Nephrol *118*, e100-108.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal Chem *75*, 663-670.

Ray, D., Kazan, H., Chan, E.T., Peña 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.

Ray, S., Catnaigh, P.O., and Anderson, E.C. (2015). Post-transcriptional regulation of gene expression by Unr. Biochem Soc Trans *43*, 323-327.

Refsland, E.W., Stenglein, M.D., Shindo, K., Albin, J.S., Brown, W.L., and Harris, R.S. (2010). Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. Nucleic acids research *38*, 4274-4284.

Riley, L.G., Cooper, S., Hickey, P., Rudinger-Thirion, J., McKenzie, M., Compton, A., Lim, S.C., Thorburn, D., Ryan, M.T., Giege, R.*, et al.* (2010). Mutation of the mitochondrial tyrosyl-tRNA synthetase gene, YARS2, causes myopathy, lactic acidosis, and sideroblastic anemia--MLASA syndrome. Am J Hum Genet *87*, 52-59.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Roovers, M., Hale, C., Tricot, C., Terns, M.P., Terns, R.M., Grosjean, H., and Droogmans, L. (2006). Formation of the conserved pseudouridine at position 55 in archaeal tRNA. Nucleic Acids Res *34*, 4293-4301.

Rosenberg, B.R., Hamilton, C.E., Mwangi, M.M., Dewell, S., and Papavasiliou, F.N. (2011). Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. Nature structural & molecular biology.

Rybak, A., Fuchs, H., Smirnova, L., Brandt, C., Pohl, E.E., Nitsch, R., and Wulczyn, F.G. (2008). A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. Nat Cell Biol *10*, 987-993.

Sachs, R., Max, K.E.a., Heinemann, U., and Balbach, J. (2012). RNA single strands bind to a conserved surface of the major cold shock protein in crystals and solution. RNA (New York, NY) *18*, 65-76.

Saito, Y., Nakagami, H., Azuma, N., Hirata, S., Sanada, F., Taniyama, Y., Morishita, R., Kaneda, Y., and Sasajima, T. (2011). Critical roles of cold shock domain protein A as an endogenous angiogenesis inhibitor in skeletal muscle. Antioxid Redox Signal *15*, 2109-2120.

Sakura, H., Maekawa, T., Imamoto, F., Yasuda, K., and Ishii, S. (1988). Two human genes isolated by a novel method encode DNA-binding proteins containing a common region of homology. Gene *73*, 499-507.

Salvetti, A., Batistoni, R., Deri, P., Rossi, L., and Sommerville, J. (1998). Expression of DjY1, a protein containing a cold shock domain and RG repeat motifs, is targeted to sites of regeneration in planarians. Dev Biol *201*, 217-229.

Samuelsson, T., and Olsson, M. (1990). Transfer RNA pseudouridine synthases in Saccharomyces cerevisiae. J Biol Chem *265*, 8782-8787.

Scheibe, M., Butter, F., Hafner, M., Tuschl, T., and Mann, M. (2012). Quantitative mass spectrometry and PAR-CLIP to identify RNA-protein interactions. Nucleic Acids Res *40*, 9897-9902.

Schindelin, H., Jiang, W., Inouye, M., and Heinemann, U. (1994a). Crystal structure of CspA, the major cold shock protein of Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America *91*, 5119-5123.

Schindelin, H., Jiang, W., Inouye, M., and Heinemann, U. (1994b). Crystal structure of CspA, the major cold shock protein of Escherichia coli. Proc Natl Acad Sci U S A *91*, 5119-5123.

Schindelin, H., Marahiel, M.A., and Heinemann, U. (1993). Universal nucleic acid-binding domain revealed by crystal structure of the B. subtilis major cold-shock protein. Nature *364*, 164-168.

Schueler, M., Munschauer, M., Gregersen, L.H., Finzel, A., Loewer, A., Chen, W., Landthaler, M., and Dieterich, C. (2014). Differential protein occupancy profiling of the mRNA transcriptome. Genome biology *15*, R15-R15.

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.

Schwanhäusser, 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.

Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., Leon-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S.*, et al.* (2014a). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. CELL *159*, 148-162.

Schwartz, S., Bernstein, Douglas A., Mumbach, Maxwell R., Jovanovic, M., Herbst, Rebecca H., León-Ricardo, Brian X., Engreitz, Jesse M., Guttman, M., Satija, R., Lander, Eric S.*, et al.* (2014b). Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. Cell, 1-15.

Sears, D., Luong, P., Yuan, M., Nteliopoulos, G., Man, Y.K., Melo, J.V., and Basu, S. (2010a). Functional phosphoproteomic analysis reveals cold-shock domain protein A to be a Bcr-Abl effector-regulating proliferation and transformation in chronic myeloid leukemia. Cell Death Dis *1*, e93.

Sears, D., Luong, P., Yuan, M., Nteliopoulos, G., Man, Y.K.S., Melo, J.V., and Basu, S. (2010b). Functional phosphoproteomic analysis reveals cold-shock domain protein A to be a Bcr-Abl effector-regulating proliferation and transformation in chronic myeloid leukemia. Cell death & disease *1*, e93-e93.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58-63.

Shahni, R., Wedatilake, Y., Cleary, M.A., Lindley, K.J., Sibson, K.R., and Rahman, S. (2013). A distinct mitochondrial myopathy, lactic acidosis and sideroblastic anemia (MLASA) phenotype associates with YARS2 mutations. Am J Med Genet A *161A*, 2334-2338.

Sibert, B.S., Fischel-Ghodsian, N., and Patton, J.R. (2008). Partial activity is seen with many substitutions of highly conserved active site residues in human Pseudouridine synthase 1. Rna *14*, 1895-1906.

Sibert, B.S., and Patton, J.R. (2011). Pseudouridine synthase 1: a site-specific synthase without strict sequence recognition requirements. Nucleic acids research, 1-12.

Sibert, B.S., and Patton, J.R. (2012). Pseudouridine synthase 1: a site-specific synthase without strict sequence recognition requirements. Nucleic Acids Res *40*, 2107-2118.

Simos, G., Tekotte, H., Grosjean, H., Segref, A., Sharma, K., Tollervey, D., and Hurt, E.C. (1996). Nuclear pore proteins are involved in the biogenesis of functional tRNA. EMBO J *15*, 2270-2284.

Solomon, O., Oren, S., Safran, M., Deshet-Unger, N., Akiva, P., Jacob-Hirsch, J., Cesarkas, K., Kabesa, R., Amariglio, N., Unger, R.*, et al.* (2013). Global regulation of alternative splicing by adenosine deaminase acting on RNA (ADAR). RNA (New York, NY).

Sommerville, J. (1999). Activities of cold-shock domain proteins in translation control. Bioessays *21*, 319-325.

Sourisseau, T., Georgiadis, A., Tsapara, A., Ali, R.R., Pestell, R., Matter, K., and Balda, M.S. (2006). Regulation of PCNA and cyclin D1 expression and epithelial morphogenesis by the ZO-1-regulated transcription factor ZONAB/DbpA. Mol Cell Biol *26*, 2387-2398.

Spadaro, D., Tapia, R., Jond, L., Sudol, M., Fanning, A.S., and Citi, S. (2014). ZO proteins redundantly regulate the transcription factor DbpA/ZONAB. J Biol Chem *289*, 22500-22511.

Stickeler, E., Fraser, S.D., Honig, a., Chen, a.L., Berget, S.M., and Cooper, T.a. (2001). The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. The EMBO journal *20*, 3821-3830.

Sundseth, R., MacDonald, G., Ting, J., and King, A.C. (1997). DNA elements recognizing NF-Y and Sp1 regulate the human multidrug-resistance gene promoter. Mol Pharmacol *51*, 963-971.

Suzuki, T., and Nagao, A. (2011). Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. Annu Rev Genet *45*, 299-329.

Suzuki, T., Wada, T., Saigo, K., and Watanabe, K. (2002). Taurine as a constituent of mitochondrial tRNAs: new insights into the functions of taurine and human mitochondrial diseases. EMBO J *21*, 6581-6589.

Svitkin, Y.V., Evdokimova, V.M., Brasey, A., Pestova, T.V., Fantus, D., Yanagiya, A., Imataka, H., Skabkin, M.A., Ovchinnikov, L.P., Merrick, W.C.*, et al.* (2009). General RNA-binding proteins have a function in poly(A)-binding protein-dependent translation. EMBO J *28*, 58-68.

Syndrome, M., and Report, C. (2007). ARTICLE CASE REPORT. 47-50.

Tanabe, H., Goldstein, J., Yang, M., and Inouye, M. (1992). Identification of the promoter region of the Escherichia coli major cold shock gene, cspA. J Bacteriol *174*, 3867-3873.

Tenenbaum, S.A., Carson, C.C., Lager, P.J., and Keene, J.D. (2000). Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc Natl Acad Sci U S A *97*, 14085-14090.

Tercjak-Recko, M., Luczynski, W., Bernatowicz, P., Zalewski, G., Rembinska, M., Lachowska, U., Recko, P., Suchon, P., Czaban, M., Sokal, J.*, et al.* (2012). [Polymorphism rs9939609 of FTO gene is related to the body mass index in children from Podlaskie voievodship]. Med Wieku Rozwoj *16*, 53-60.

Thieringer, H.A., Singh, K., Trivedi, H., and Inouye, M. (1997). Identification and developmental characterization of a novel Y-box protein from Drosophila melanogaster. Nucleic Acids Res *25*, 4764-4770.

Thomas, S.R., Keller, C.A., Szyk, A., Cannon, J.R., and Laronde-Leblanc, N.A. (2011). Structural insight into the functional mechanism of Nep1/Emg1 N1-specific pseudouridine methyltransferase in ribosome biogenesis. Nucleic Acids Res *39*, 2445-2457.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform *14*, 178-192.

Tillault, A.S., Fourmann, J.B., Loegler, C., Wieden, H.J., Kothe, U., and Charpentier, B. (2015). Contribution of two conserved histidines to the dual activity of archaeal RNA guide-dependent and -independent pseudouridine synthase Cbf5. Rna *21*, 1233-1239.

Triqueneaux, G., Velten, M., Franzon, P., Dautry, F., and Jacquemin-Sablon, H. (1999). RNA binding specificity of Unr, a protein with five cold shock domains. Nucleic Acids Res *27*, 1926-1934.

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505-510.

Umeda, N., Suzuki, T., Yukawa, M., Ohya, Y., Shindo, H., and Watanabe, K. (2005). Mitochondria-specific RNA-modifying enzymes responsible for the biosynthesis of the wobble base in mitochondrial tRNAs. Implications for the molecular pathogenesis of human mitochondrial diseases. J Biol Chem *280*, 1613-1624.

Viswanathan, S.R., Daley, G.Q., and Gregory, R.I. (2008). Selective blockade of microRNA processing by Lin28. Science *320*, 97-100.

Vulliamy, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M., Mason, P.J., and Dokal, I. (2001). The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. Nature *413*, 432-435.

Wang, F., Liu, M., Qiu, R., and Ji, C. (2011). The dual role of ubiquitin-like protein Urm1 as a protein modifier and sulfur carrier. Protein & cell *2*, 612-619.

Wang, G.R., Zheng, Y., Che, X.M., Wang, X.Y., Zhao, J.H., Wu, K.J., Zeng, J., Pan, C.E., and He, D.L. (2009). Upregulation of human DNA binding protein A (dbpA) in gastric cancer cells. Acta Pharmacol Sin *30*, 1436-1442.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G.*, et al.* (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117-120.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. CELL *161*, 1388-1399.

Wu, G., Yu, A.T., Kantartzis, A., and Yu, Y.-T. (2011). Functions and mechanisms of spliceosomal small nuclear RNA pseudouridylation. Wiley Interdisciplinary Reviews: RNA, n/a-n/a.

Wu, L., and Belasco, J.G. (2005). Micro-RNA regulation of the mammalian lin-28 gene during neuronal differentiation of embryonal carcinoma cells. Mol Cell Biol *25*, 9198-9208.

Wulczyn, F.G., Smirnova, L., Rybak, A., Brandt, C., Kwidzinski, E., Ninnemann, O., Strehle, M., Seiler, A., Schumacher, S., and Nitsch, R. (2007). Post-transcriptional regulation of the let-7 microRNA during neural cell specification. FASEB J *21*, 415-426.

Wurm, J.P., Meyer, B., Bahr, U., Held, M., Frolow, O., Kotter, P., Engels, J.W., Heckel, A., Karas, M., Entian, K.D.*, et al.* (2010). The ribosome assembly factor Nep1 responsible for Bowen-Conradi syndrome is a pseudouridine-N1-specific methyltransferase. Nucleic Acids Res *38*, 2387-2398.

Xia, B., Ke, H., and Inouye, M. (2001). Acquirement of cold sensitivity by quadruple deletion of the cspA family and its suppression by PNPase S1 domain in Escherichia coli. Mol Microbiol *40*, 179-188.

Xue, Y., Zhou, Y., Wu, T., Zhu, T., Ji, X., Kwon, Y.S., Zhang, C., Yeo, G., Black, D.L., Sun, H.*, et al.* (2009). Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. Mol Cell *36*, 996-1006.

Yasen, M., Kajino, K., Kano, S., Tobita, H., Yamamoto, J., Uchiumi, T., Kon, S., Maeda, M., Obulhasim, G., Arii, S.*, et al.* (2005). The up-regulation of Y-box binding proteins (DNA binding protein A and Y-box binding protein-1) as prognostic markers of hepatocellular carcinoma. Clin Cancer Res *11*, 7354-7361.

Yasen, M., Obulhasim, G., Kajino, K., Mogushi, K., Mizushima, H., Tanaka, S., Tanaka, H., Hino, O., and Arii, S. (2012). DNA binding protein A expression and methylation status in hepatocellular carcinoma and the adjacent tissue. Int J Oncol *40*, 789-797.

Yeo, J., Goodman, R.a., Schirle, N.T., David, S.S., and Beal, P.a. (2010). RNA editing changes the lesion specificity for the DNA repair enzyme NEIL1. Proceedings of the National Academy of Sciences of the United States of America *107*, 20715-20719.

Yu, A.T., Ge, J., and Yu, Y.-T. (2011). Pseudouridines in spliceosomal snRNAs. Protein & cell *2*, 712-725.

Yu, J., Hecht, N.B., and Schultz, R.M. (2001). Expression of MSY2 in mouse oocytes and preimplantation embryos. Biol Reprod *65*, 1260-1270.

Yu, Y.-t., Shu, M.-d., and Steitz, J.A. (1998). Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *17*, 5783-5795.

Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat Biotechnol *29*, 607-614.

Zhao, B.S., and He, C. (2015). Pseudouridine in a new era of RNA modifications. Cell Res *25*, 153-154.

Zhao, X., Patton, J.R., Davis, S.L., Florence, B., Ames, S.J., and Spanjaard, R.a. (2004). Regulation of nuclear receptor activity by a pseudouridine synthase through

posttranscriptional modification of steroid receptor RNA activator. Molecular cell *15*, 549-558.

Zhao, X., Patton, J.R., Ghosh, S.K., Fischel-Ghodsian, N., Shen, L., and Spanjaard, R.A. (2007). Pus3p- and Pus1p-dependent pseudouridylation of steroid receptor RNA activator controls a functional switch that regulates nuclear receptor signaling. Mol Endocrinol *21*, 686-699.

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. Mol Cell *49*, 18-29.

SUPPLEMENTARY INFORMATION

Supplementary Figure S1.

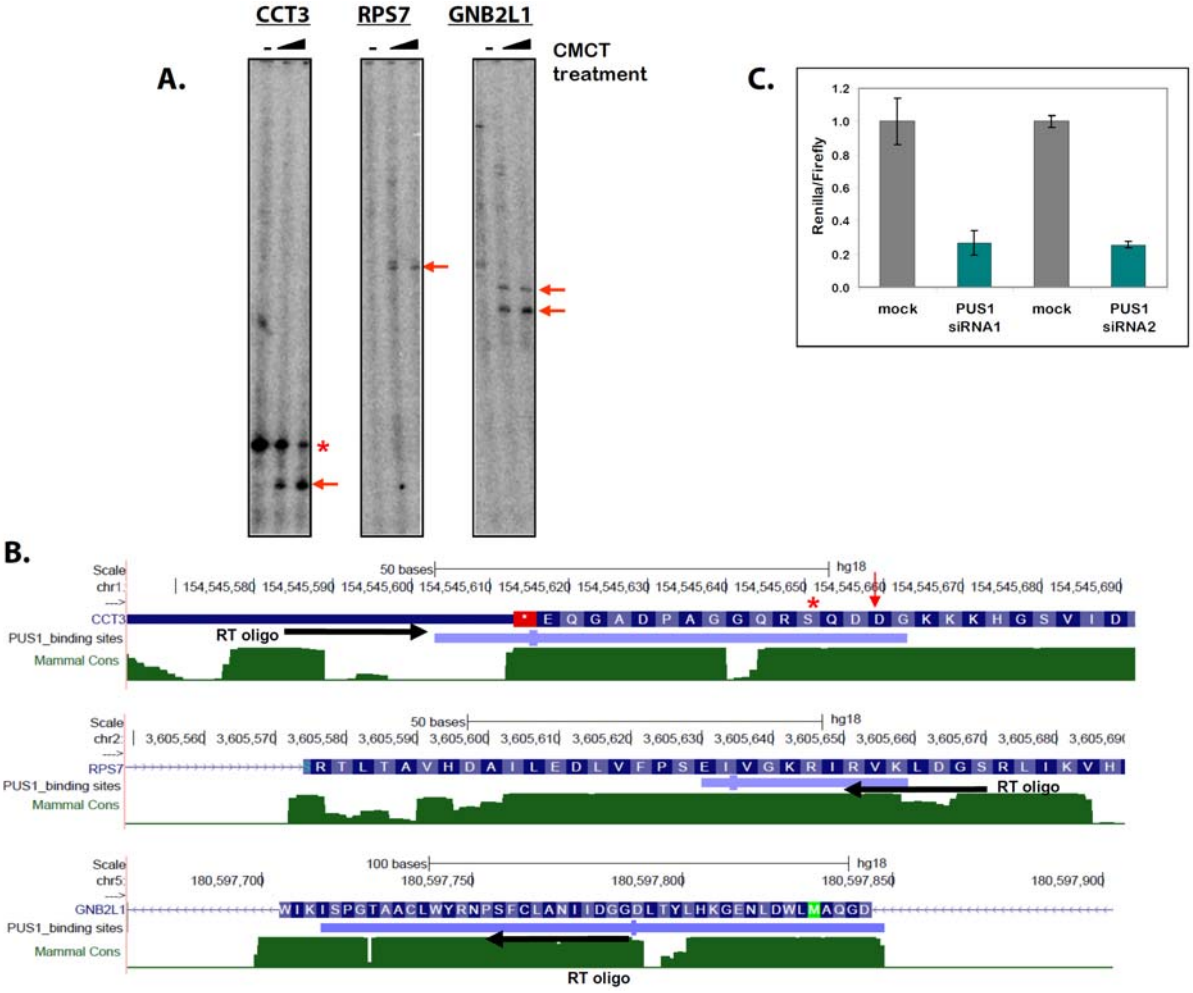


Figure S1. CMCT profile and UCSC genome browser view of CCT3, RPS7 and GNB2L1. (A) Autoradiogram of the sequencing gel for the CMCT treatment and primer extension of CCT3, RPS7 and GNB2L1 transcripts. Several stops of the reverse transcriptase could be observed for the assayed regions of these transcripts, with an additional stop for CCT3 at U1824 present in the non-treated sample (-CMCT, red asterisk) and two stops for the GNB2L1 transcript. (B) UCSC browser view of the regions of interest bound by PUS1 and assayed in the CMCT individual profile. The RT oligos (black arrows) under each gene indicate the position of the primer extension start, downstream or overlapping with PUS1 binding sites. (C) Luciferase assay for PUS1 siRNA test indicating the efficiency of two different siRNAs for PUS1 knockdown. The two siRNAs screen for PUS1 knockdown show a knockdown efficiency to 20-30% of the PUS1 protein level.

Supplementary Figure S2.

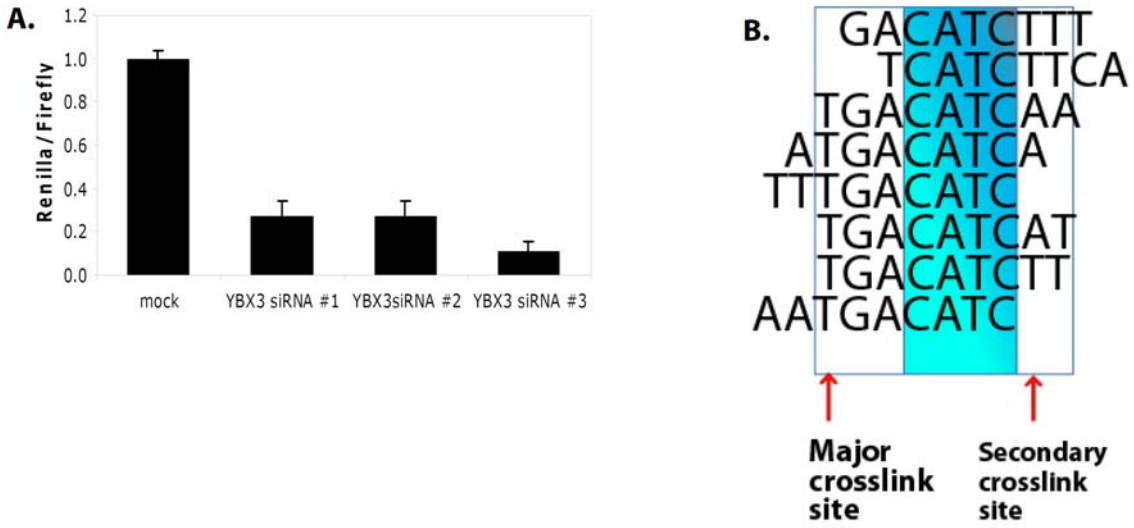


Figure S2. YBX3 luciferase assay for RNAi screen and YBX3 top eight RNA binding motifs. (A) RNAi screen using a dual luciferase assay for the YBX3 knockdown efficiency show that the tested siRNAs can reduce YBX3 protein levels to 10-25%. (B) Using a computational analysis to search for an RNA binding motif in YBX3 PAR-CLIP target transcripts we identified several preferred motifs, of which the top 8 YBX3 motifs contain the conserved 4-mer sequence CATC. The major crosslink site resides in the first position of the 9-mer motif sequence, while the secondary crosslink site is found right next to the CATC conserved sequence motif.

Supplementary Figure S3.

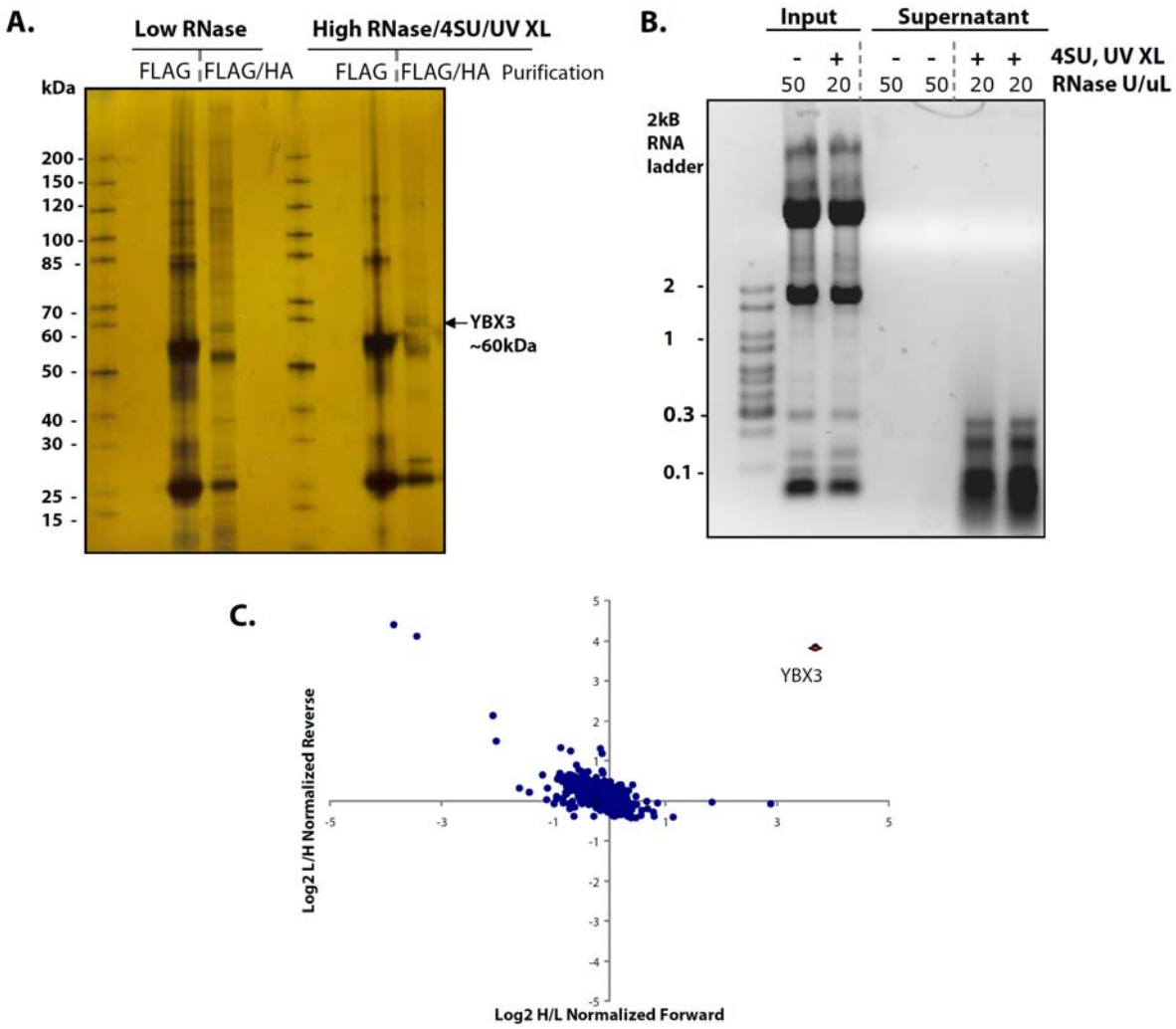


Figure S3. YBX3 protein-protein interaction assay. (A) Silver staining of the SDS-PAGE for the SILAC based protein-protein interaction experiment showing the RNA-dependent and RNA-independent putative YBX3 protein interactors. The FLAG-immunoprecipitation and the HA-purification protein elution were resolved on the SDS-PAGE gel. YBX3 appears at around 60 kDa and can be observed in both RNA-dependent and RNA-independent samples. (B) The agarose gel shows the RNA length size of the two conditions employed in our protein-protein interaction assay. The 50 U/µl RNase T1 treatment in the direct protein-protein interactions experiment shows full RNA digestion, while the 20 U/ µl of RNase T1 treatment in the indirect experiment shows a final RNA length of 300 bp. (C) Protein-protein interaction plot of H/L SILAC ratios for YBX3 direct protein interactions forward and reverse experiments. YBX3 can be observed enriched in the upper right corner while no direct interactors show increased SILAC ratios, presumably due to the transient nature of YBX3 protein interactions.

Supplementary Figure S4.

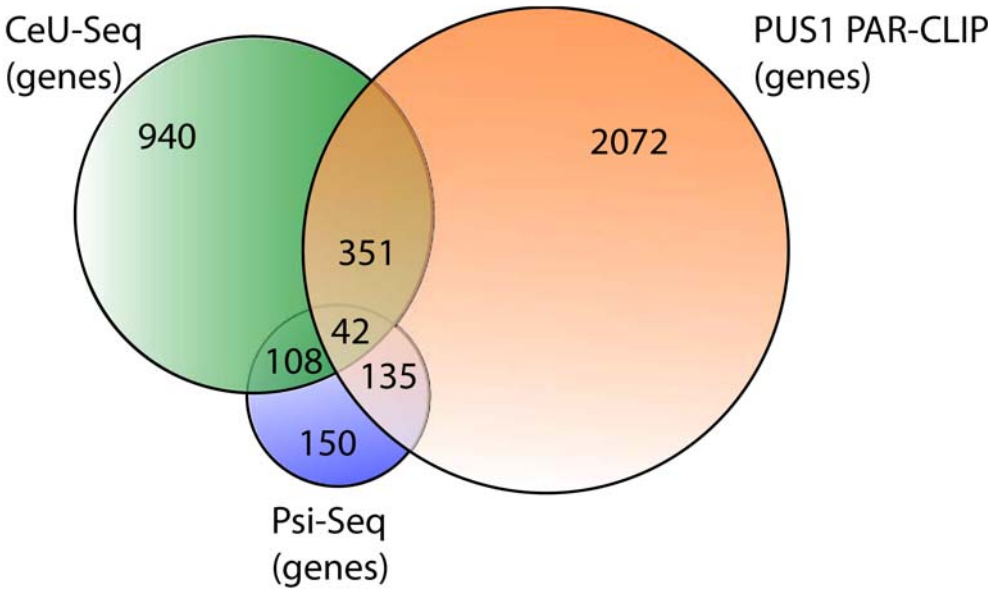


Figure S4. Overlap between the PUS1 PAR-CLIP target genes, Psi-Seq pseudouridine containing genes and CeU-Seq enriched pseudouridylated genes. The Venn-diagram shows the overlap between the PUS1 PAR-CLIP target genes in the 4SU conservative analysis set, the pseudouridine transcriptome-wide profile obtained with Psi-Seq (Schwartz et al., 2014b) and CeU-Seq (Li et al., 2015a). Around 108 genes were found commonly by the Psi-Seq and CeU-Seq pseudouridine profile in HEK293 and HEK293T cells, respectively, while 351 genes are shared between PUS1 PAR-CLIP dataset and CeU-Seq pulldown. A number of 42 genes are shared in all three studies. This comparison is showing the different approaches of pseudouridine identification. The differences reflect the variability in different experimental approaches as well as the lack of both fully comprehensive and highly sensitive method to capture this type of RNA modification.

Supplementary Figure S5.

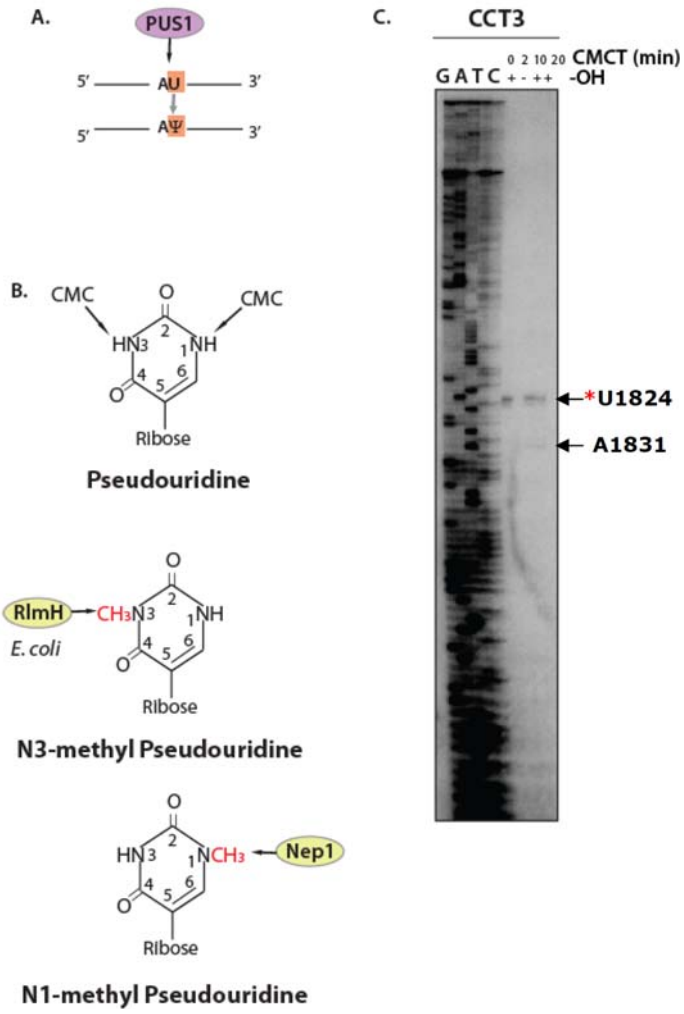


Figure S5. PUS1 preferred sequence, pseudouridine modifications and CCT3 primer extension replicate. (A) Pseudouridylation sites found in yeast across the transcriptome for PUS1 (Schwartz et al., 2014b) are frequently preceded by an A residue. PUS1 is known withouth strict sequence specificities, however, it appears to modify U residues downstream of adenosines. (B) Schematic representation of the pseudouridine CMC-ψ adduct formation positions and two modified pseudouridines N3-methyl pseudouridine and N1-methyl pseudouridine. On the side of each structure is indicated the enzyme responsible for the methylation of pseudouridines. The RlmH was found in E. coli, which methylates ψ at N3 position and Nep1 which is present in humans and can methylate the N1 position of pseudouridine. (C) CMCT-primer extension sequencing gel showing a replicate for CCT3 pseudouridine identification using an extra control, "+CMCT,-OH", which is excluding the alkaline treatment, in order to map the positions of uridines and guanosines. Due to early stops of the reverse transcriptase at uridine bases downstream the reverse transcriptase primer extension oligonucleotide, no stops are seen at the U1924 position.

Supplemental Figure S6.

A.

Species	PUS	GeneID	Position	M-Score
Yeast	PUS1	CCT3	205	8
Yeast	PUS1	CCT3	525	8
Yeast	PUS1	CCT3	564	9
Yeast	PUS1	CCT3	653	9
Yeast	PUS1	CCT3	867	9
Yeast	PUS1	CCT3	934	9
Yeast	PUS1	CCT3	1070	8
Yeast	PUS1	CCT3	1139	8
Yeast	PUS1	CCT3	1172	8
Yeast	PUS1	CCT3	1205	9
Yeast	PUS1	CCT3	1251	8
Yeast	PUS1	CCT3	1370	9
Yeast	PUS1	CCT3	1388	8
Yeast	PUS1	CCT3	1479	8
Yeast	PUS7	CCT3	209	9
Yeast	PUS7	CCT3	1325	9

B.

Figure S6. The PPUS server pseudouridine predicted positions in yeast CCT3 and alignment with the human CCT3 sequence. (A) The PPUS server for prediction of pseudouridine (Li et al., 2015b) based on experimental data obtained in previous studies has predicted several putative pseudouridine sites at different positions by different pseudouridine synthases. For PUS1 a pseudouridine site is predicted at position U1479 with a high prediction score (M-score). (B) The alignment with the human CCT3 region shows the predicted PUS1 pseudouridine position at U1479 corresponds to the U1824 stop in our CCT3 CMCT individual profile.

Supplementary Figure S7.

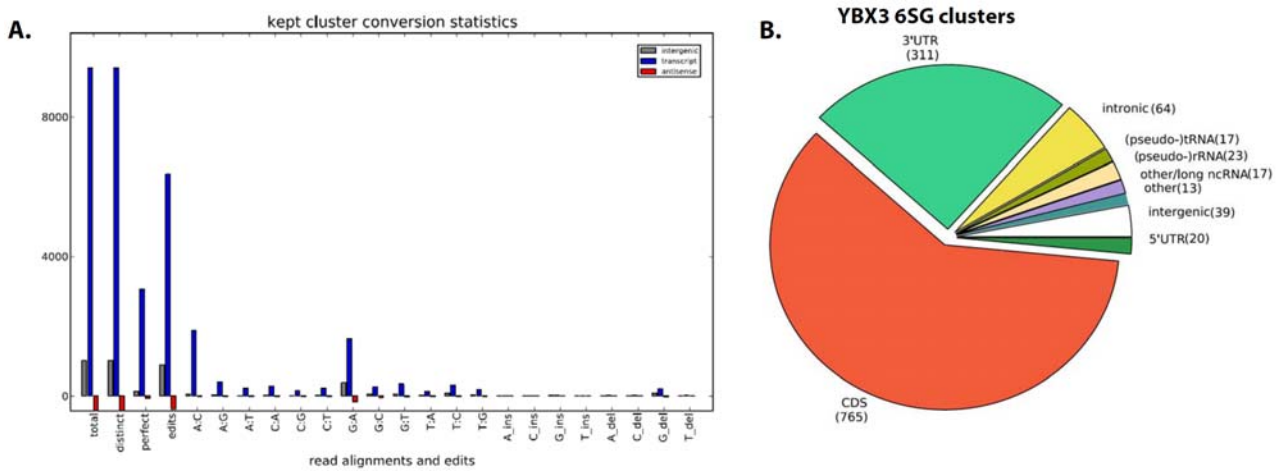


Figure S7. PAR-CLIP editing statistics for the YBX3 6SG library and YBX3 6SG clusters distribution. (A) The 6SG library for YBX3 shows the G-A conversion characteristic for 6SG labeled PAR-CLIP libraries. The read alignments and edits of the kept clusters show that the number of G-A transition reads is lower than the T-C transitions reads typical for the 4SU libraries (see Figure 12C). (B) YBX3 6SG clusters show a similar distribution preference of the binding sites to the 4SU libraries. These clusters were found in the coding sequence and 3' UTR regions, however, a lower number of binding sites were identified in the 6SG library.

Supplementary Table S1.

GO Terms for Biological Processes of the top YBX3 3' UTR motif scored targets.

Term	Count	P-Value
RNA splicing	11	5.70E-08
mRNA metabolic process	12	6.30E-08
mRNA processing	11	1.80E-07
RNA processing	13	4.00E-07
response to DNA damage stimulus	7	2.20E-03
nuclear mRNA splicing, via spliceosome	5	2.20E-03
RNA splicing, via transesterification reactions	5	2.20E-03
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	5	2.20E-03
DNA repair	6	3.50E-03
spliceosomal snRNP biogenesis	3	4.60E-03

Supplementary Table S2.

Gene list of the PUS1 binding sites and the Psi-Seq HEK293 pseudouridines overlap (hg18).

Gene Name	Chromosome	Position_Start	Position_End
RPL37A	chr2	217072345	217072375
ATP5E	chr20	57037246	57037276
FAM120A	chr9	95367411	95367441
MALAT1	chr11	65026953	65026983
chr1.tRNA49_GluTTC	chr1	159849169	159849199
CNBP	chr3	130371964	130371994
ERH	chr14	68916724	68916754
RNF167	chr17	4788925	4788955
SLC25A1	chr22	17545461	17545491
chr1.tRNA132_GlyCCC	chr1	16745023	16745053
chr1.tRNA99_ValCAC	chr1	147561292	147561322
chr1.tRNA59_GluCTC	chr1	247135108	247135138
chr12.tRNA12_AspGTC	chr12	123977846	123977876
chr14.tRNA2_LeuTAG	chr14	20163417	20163447
chr14.tRNA5_TyrGTA	chr14	20221327	20221357
chr14.tRNA6_ProTGG	chr14	20222053	20222083
chr16.tRNA17_LeuCAG	chr16	55891413	55891443
chr17.tRNA12_TrpCCA	chr17	19352124	19352154
chr5.tRNA9_LysCTT	chr5	180567400	180567430
MALAT1	chr11	65027383	65027413
chr1.tRNA49_GluTTC	chr1	159849129	159849159
MCL1	chr1	148816147	148816177
CDK2AP1	chr12	122311728	122311758
ERH	chr14	68916725	68916755
SNORD60 (U60)	chr16	2145067	2145097
RPL32 (ACA7, SNORA7A)	chr3	12856848	12856878
RPL13A (U32, SNORD32A)	chr19	54685046	54685076
	chr19	54685049	54685079
MALAT1	chr11	65025313	65025343
chr1.tRNA99_ValCAC	chr1	147561293	147561323
chr1.tRNA91_PseudoCCC	chr1	147946837	147946867
	chr1	147946836	147946866
chr1.tRNA45_GlyTCC	chr1	159767524	159767554
chr2.tRNA5_IleTAT	chr2	42891199	42891229
chr6.tRNA139_ValAAC	chr6	27726716	27726746