DHX36 function in RNA G-quadruplex-mediated posttranscriptional gene regulation

Funktion von DHX36 in RNA G-Quadruplex-vermittelter posttranskriptioneller Genregulierung

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section: Biomedicine

submitted by

Markus Sauer

from

Aschaffenburg, Germany

Würzburg, 2019

Submitted on: ……

Office stamp

Members of the Promotionskomitee:

Chairperson:	**Prof. Dr. Manfred Gessler**
Primary Supervisor:	**Prof. Dr. Katrin Paeschke**
Supervisor (Second):	**Prof. Dr. Jörg Vogel**
Supervisor (Third):	**Prof. Dr. Utz Fischer**

Date of Public Defense: ……

Date of Receipt of Certificates: ……

Table of contents

1. Summary 1
2. Zusammenfassung 2
3. Introduction 3
3.1 Eukaryotic gene expression 3
3.1.1 The complex life of eukaryotic mRNAs 3
3.1.2 Regulation of eukaryotic gene expression 8
3.2 Structural aspects of RNA 14
3.2.1 Classical RNA structures and their functions 14
3.2.2 G-quadruplex structures 16
3.3 Proteins interacting with RNA 24
3.3.1 General aspects of RNA-binding proteins 24
3.3.2 G-quadruplex-interacting proteins 25
3.3.3 DEAH-box helicase DHX36 27
3.4 Aim of this thesis 31
4. Results 32
4.1 DHX36 is a cytoplasmic mRNA-binding protein 32
4.2 DHX36 binds defined G-rich sites on thousands of mRNAs 35
4.3 Binding sites of DHX36 can form G-quadruplexes *in vitro* 41
4.4 Loss of DHX36 results in increased target mRNA abundance 45
4.5 DHX36 decreases target mRNA stability 49
4.6 Target mRNA abundance is influenced by DHX36 helicase activity 51
4.7 Loss of DHX36 does not result in more translated protein 53
4.8 DHX36 increases translation efficiency of target mRNA 57
4.9 DHX36 affects translation efficiency of rG4-bearing target mRNAs 59
4.10 Connection of DHX36 and rG4s to cellular stress 62
5. Discussion 65
5.1 Characterization of DHX36 in HEK293 cells 66
5.2 RNA binding of DHX36 68
5.3 The RNA interactome of DHX36 69
5.4 The function of DHX36 on its target mRNAs 70
5.4.1 Abundance of DHX36 target mRNAs 70
5.4.2 Translation of DHX36 target mRNA 71
5.4.3 Sequestration of DHX36 target mRNAs 73
5.5 DHX36 and its role in rG4 biology 75
5.6 A model for DHX36 function 76
5.7 DHX36 relevance for life and concluding remarks 78
6. Material & Methodology 80
6.1 Material 80
6.1.1 General equipment 80
6.1.2 Consumables 80
6.1.3 Chemicals 80
6.1.4 Commercial systems 81
6.1.5 Strains and cell lines 82
6.1.6 Enzymes 84
6.1.7 Antibodies 85

6.1.8 Plasmids 86
6.1.9 Synthetic oligonucleotides 87
6.1.10 Buffers & solutions 90
6.1.11 Growth media & antibiotics 92
6.2 Methodology 94
6.2.1 Bacteria cell culture 94
6.2.2 Eukaryotic cell culture 95
6.2.3 DNA biochemistry 97
6.2.4 Protein biochemistry 101
6.2.5 RNA biochemistry 102
6.2.6 Molecular biological experiments 104
6.2.7 Systems-wide experiments 107
6.2.8 Fluorescence microscopy 112
6.2.9 Others 113
7. Appendix 114
7.1 Supplementary data 114
7.2 Abbreviations 119
7.2.1 Names 119
7.2.2 Scientific terms 119
7.2.3 Proteins, nucleic acids, and complexes 121
7.2.4 Reagents 123
7.2.5 Units 123
7.2.6 Dimensions 123
8. Bibliography 124
9. Table of Figures 138
10. Publications 139
11. Curriculum vitae 140
12. Acknowledgements 142
13. Affidavit 144

1. Summary

The expression of genetic information into proteins is a key aspect of life. The efficient and exact regulation of this process is essential for the cell to produce the correct amounts of these effector molecules to a given situation. For this purpose, eukaryotic cells have developed many different levels of transcriptional and posttranscriptional gene regulation. These mechanisms themselves heavily rely on interactions of proteins with associated nucleic acids. In the case of posttranscriptional gene regulation an orchestrated interplay between RNA-binding proteins, messenger RNAs (mRNA), and non-coding RNAs is compulsory to achieve this important function.

A pivotal factor hereby are RNA secondary structures. One of the most stable and diverse representatives is the G-quadruplex structure (G4) implicated in many cellular mechanisms, such as mRNA processing and translation. In protein biosynthesis, G4s often act as obstacles but can also assist in this process. However, their presence has to be tightly regulated, a task which is often fulfilled by helicases.

One of the best characterized G4-resolving factors is the DEAH-box protein DHX36. The *in vitro* function of this helicase is extensively described and individual reports aimed to address diverse cellular functions as well. Nevertheless, a comprehensive and systems-wide study on the function of this specific helicase was missing, so far.

The here-presented doctoral thesis provides a detailed view on the global cellular function of DHX36. The binding sites of this helicase were defined in a transcriptome-wide manner, a consensus binding motif was deviated, and RNA targets as well as the effect this helicase exerts on them were examined. In human embryonic kidney cells, DHX36 is a mainly cytoplasmic protein preferentially binding to G-rich and G4-forming sequence motifs on more than 4,500 mRNAs. Loss of DHX36 leads to increased target mRNA levels whereas ribosome occupancy on and protein output of these transcripts are reduced. Furthermore, DHX36 knockout leads to higher RNA G4 levels and concomitant stress reactions in the cell. I hypothesize that, upon loss of this helicase, translationally-incompetent structured DHX36 target mRNAs, prone to localize in stress granules, accumulate in the cell. The cell reacts with basal stress to avoid cytotoxic effects produced by these mis-regulated and structured transcripts.

2. Zusammenfassung

Die Umsetzung genetischer Information in Proteine stellt einen Schlüsselaspekt des Lebens dar. Dabei ist die effiziente und exakte Regulierung dieses Prozesses für die Zelle essentiell, um die korrekte Menge dieser Effektormoleküle in einer gegebenen Situation zu produzieren. Zu diesem Zweck haben eukaryotische Zellen viele verschiedene Ebenen der transkriptionellen und posttranskriptionellen Genregulation entwickelt. Diese Mechanismen wiederum beruhen insbesondere auf den Interaktionen von Proteinen mit assoziierten Nukleinsäuren. Im Fall der posttranskriptionellen Genregulation ist ein abgestimmtes Wechselspiel zwischen RNA-bindenden Proteinen, Boten-RNAs und nicht-kodierenden RNAs zwingend erforderlich um diese wichtige Funktion zu erfüllen.

Ein zentrales Element hierbei bilden RNA-Sekundärstrukturen. Einer der stabilsten und variantenreichsten Vertreter dieser Strukturen ist die G-Quadruplexstruktur (G4), die in vielen zellulären Mechanismen, wie zum Beispiel Prozessierung und Translation der Boten-RNA, involviert ist. Während der Proteinbiosynthese agieren G4s häufig als Hindernisse, können diesen Prozess allerdings auch unterstützen. In beiden Fällen muss deren Präsenz genau reguliert werden, was häufig durch Helikasen erfolgt.

Einer der bestcharakterisiertesten, G4-entwindenden Faktoren ist das DEAH-Box Protein DHX36. Die *in vitro* Funktion dieser Helikase wurde bereits ausführlich beschrieben und einzelne Berichte haben darüber hinaus versucht, ihr verschiedene Funktionen in der Zelle zuzuweisen. Nichtsdestotrotz fehlt bislang eine umfassende und systemweite Studie zur Funktion dieser speziellen Helikase.

Die hier präsentierte Doktorarbeit liefert einen detaillierten Blick auf die globale Funktion von DHX36 in der Zelle. Bindestellen dieser Helikase im Transkriptom wurden definiert, ein allgemeines Bindemotiv abgeleitet und RNA-Bindeziele sowie der Effekt, den diese Helikase auf jene ausübt, untersucht. In humanen embryonalen Nierenzellen ist DHX36 ein vorwiegend zytoplasmatisches Protein, das bevorzugt G-reiche und G4-bildende Sequenzmotive auf über 4.500 Boten-RNAs bindet. Verlust von DHX36 führt zu einem erhöhten Level dieser Boten-RNAs in der Zelle, wobei deren Besetzung mit Ribosomen und die damit verbundene Proteinproduktion reduziert ist. Weiterhin führt der Verlust von DHX36 zu einem höheren RNA G4 Level und zu gleichzeitigen Stressreaktionen in der Zelle. Meine Vermutung ist, dass sich bei einem Verlust von DHX36 translationsinkompetente, strukturierte und leicht akkumulierende Ziel-Boten-RNAs in der Zelle anreichern. Die Zelle reagiert darauf mit basalem Stress um zytotoxische Effekte dieser miss-regulierten und strukturierten Transkripte zu vermeiden.

3. Introduction

3.1 Eukaryotic gene expression

The term "*Gene Expression"* describes the exploitation of genetically stored information for the synthesis of functional macromolecules. This key process is conserved in all forms of life and is essential for the growth and maintenance of cells. In multicellular organisms the diversity of cell types, their individual developmental stages, and their adaptation to environmental changes are determined by differences in the expression of identical genomic information[1,2].

Gene expression is often depicted as a series of sequential steps, however, many of them are interdependent and need exact orchestration as well as regulation[3]. Consequently, defects in or mis-regulation of involved processes can lead to cancer and other severe developmental and metabolic diseases[1]. Therefore, in-depth knowledge of these complex pathways and regulative mechanisms is required to understand the causes of these diseases and to find suitable therapies and cure.

Molecular basis of gene expression is the storage of information in *deoxyribonucleic acid* (DNA) by using a varying sequence of nucleic bases[4]. The most direct product of gene expression is *ribonucleic acid* (RNA) which itself can possess mechanistical or structural function. In these cases, the RNA belongs to the group of *non-coding RNAs* (ncRNAs) and represent the final effector molecule[5]. Contrary, coding RNAs exist known as *messenger RNAs* (mRNAs) which carry genetic information into the translational process. Here, they serve as templates for the generation of proteins, the eventual product of this kind of gene expression[6]. The entirety of the levels of all cellular proteins to a given time point determines cellular characteristics and reflects the balance between protein synthesis and degradation. Cellular production of proteins is regulated at the transcriptional and posttranscriptional level by the modulation of corresponding mRNAs at any stage of their existence[7].

3.1.1 The complex life of eukaryotic mRNAs

In eukaryotes, production and processing of mRNAs are tightly connected and occur in close spatial and temporal proximity[3,8]. Contrary, their translation into proteins is a separated process for which the mRNAs have to be intracellularly ex- and transported. Eventually, no longer required or defect mRNAs are needed to be degraded[9,10].

The following chapters provide an overview of the canonical life of eukaryotic mRNAs (Figure 3-1).

3.1.1.1 mRNA Transcription – Copying the information

Eukaryotic mRNA is produced as a precursor (pre-mRNA) during a process called transcription[11,12]. Most human mRNAs are transcribed in the nucleus by the *DNA-dependent RNA polymerase II* (RNAP II)[13] (Figure 3-1 I.). For transcription initiation, this multiprotein complex binds the promoter region of the protein-coding gene assisted by *general transcription factors* (GTFs)[14]. The so-formed *preinitiation complex* (PIC), to which often various other proteins contribute, opens the DNA double helix and positions the template strand at the active site of RNAP II resulting in the formation of a 'transcription bubble'[15]. RNAP II begins at the *transcription start site* (TSS) to catalyze the polymerization of ribonucleotides complement to the template DNA. After several aborted attempts, RNAP II breaks its interaction with the promoter and most initiation factors as soon as the nascent transcript has reached a critical length (>8 nucleotides (nt))[16]. Following this 'promoter escape', RNAP II acquires elongation factors to facilitate its proceeding through the DNA double strand and to overcome pausing sites[15,16]. Throughout transcription elongation, RNAP II continues to build up the new RNA molecule in 5´-3´ direction. Termination of mRNA transcription starts by the recognition of a so-called *polyadenylation signal* (poly-A signal) on the nascent RNA by the multimeric 3´-end processing complex[17,18]. The transcribed pre-mRNA is released after cleavage from the polymerase and subject of further processing (see 3.1.1.2). RNAP II, however, continues transcription after pausing and is later actively released from the DNA template strand[19].

3.1.1.2 mRNA Processing – Editing the information

The pre-mRNA has to undergo three essential processing steps, *capping*, *splicing*, and *polyadenylation*, in order to mature to a translationally competent mRNA.

As a result of the directionality of transcription, the 5´-end of the mRNA is the first part of the pre-mRNA which becomes accessible for further processing[8]. After 20 - 30 nts are polymerized, a cap structure, canonically consisting of a methylated guanine nucleotide (m7G) is linked in a multi-step process via an unusual 5´-5´ triphosphate bridge[20] (Figure 3-1 II.). This structure is important for protein synthesis by contributing to translation initiation[21] (see 3.1.1.4). Other functions of the cap are prevention of degradation by exonucleases, supporting mRNA nuclear export, and assisting in the splice process of 5´-terminal located introns[21,22].

Most eukaryotic protein genes consist of coding sequences named *exons* and large interspacing regions called *introns*[11,12,23]. During the splicing process, non-coding introns are excised from the transcript and the remaining exons are covalently joined. Consequently, the mature mRNA continuously consists only of exons. Nevertheless, not all exons are always needed and therefore removed resulting in expression of different

protein isoforms[24]. This process, known as *alternative splicing*, enables the cell to direct the expression of a dramatically higher number of proteins than expected from the number of protein-coding genes[25,26]. Splicing is performed at distinct sites located all over the pre-mRNA by a molecular machinery called *spliceosome* consisting of *uridine-rich small nuclear ribonucleoproteins* (UsnRNPs)[27] (Figure 3-1 III.). Correct splicing of the prospective mRNA is marked by the addition of the heteromeric *exon junction complex* (EJC) to each exon-exon-connection site. EJCs are important elements of the quality control system for mRNAs and their export in eukaryotic cells[28].

Finally, the 3´-end of the pre-mRNA is also processed, a mechanism which is tightly connected to the cleavage of the nascent RNA during the termination of transcription[18]. Being a part of the 3´-end processing complex (see 3.1.1.1) the enzyme *polyadenylate polymerase* (PAP) adds together with other factors a *poly-A-tail* consisting of hundreds of adenines (Figure 3-1 IV.). Like the cap structure, poly-A-tails contribute to translation, prevention of degradation, and nuclear export of mRNAs[17].

3.1.1.3 mRNA export – Trafficking the information

Whereas all these processes occur in the nucleus the place of translation is the cytoplasm[10]. Consequently, mature mRNAs have to be exported by different mechanisms through the *nuclear pore complex* (NPC)[29,30] (Figure 3-1 V.) whereas defective or incompletely processed transcripts are retained.

A subset of endogenous mRNAs uses the same pathway as proteins or tRNAs by being bound via adaptor molecules to the karyopherin *exportin 1* (XPO1) and transported along a RanGTP gradient. However, the bulk of cellular mRNAs is exported via a different mechanism. Here, the mRNA is bound by a heterodimer consisting of the non-karyopherins *nuclear RNA export factor 1* (NXF1) and *NTF2-related export protein 1* (NXT1). This transport does not rely on RanGTP[31]. Nevertheless, in both pathways the export factors (XPO1 or Nxf1/Nxt1) interact with the nucleoporins of the NPC and facilitate translocation[30]. Spatial controlled remodeling of the mRNA-bound proteins facilitates its release from the NPC to the cytoplasm.

In cases in which a defined cytosolic localization is of importance, mRNAs are further targeted intracellularly e.g. actively by motor proteins along the cytoskeleton or passively with a diffusion-and-trapping mechanism[32].

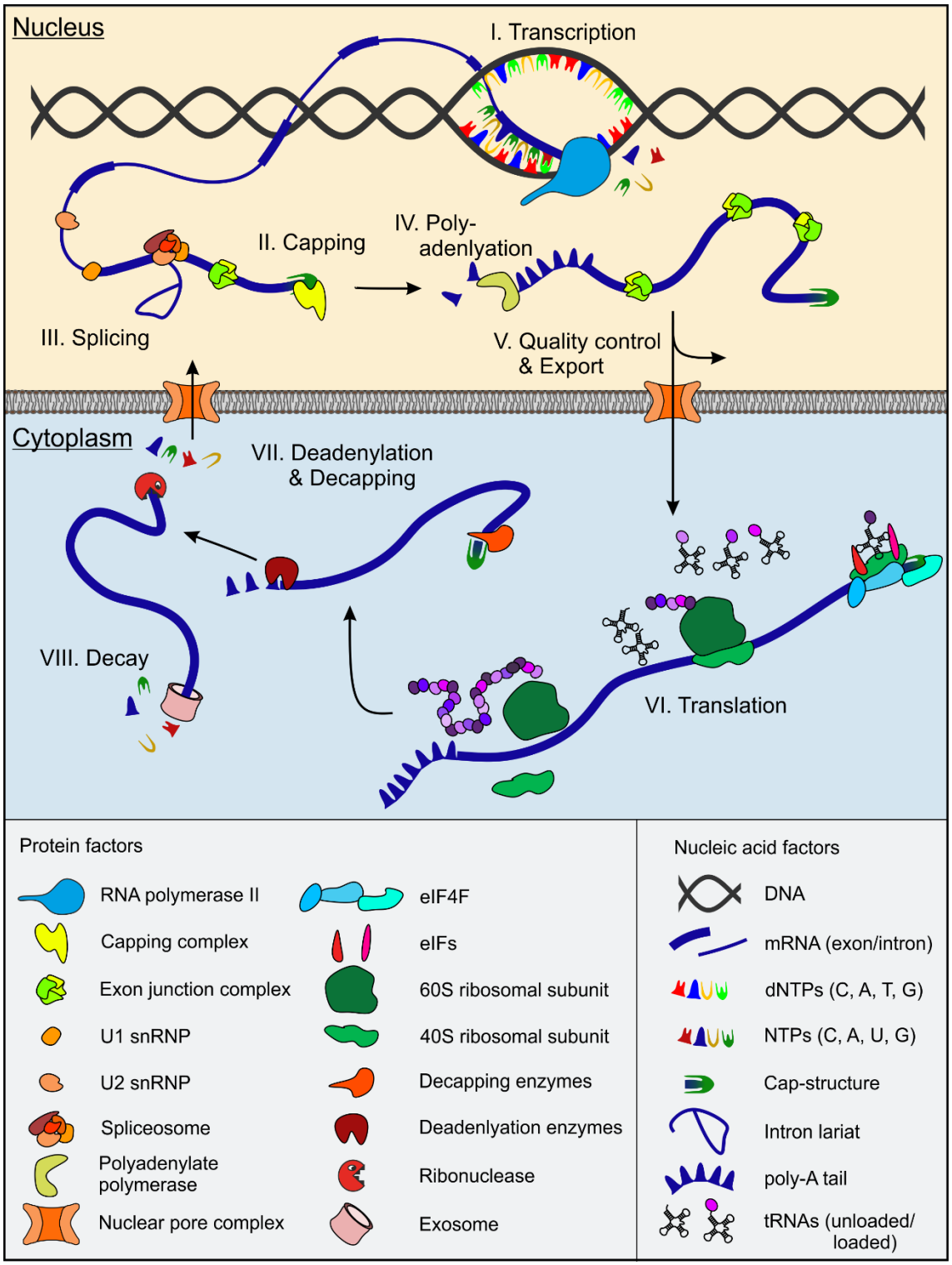

Figure 3-1: The life cycle of an eukaryotic mRNA.

Most eukaryotic mRNAs are transcribed by the DNA-dependent RNA polymerase II in the nucleus (I.). Co-transcriptionally, newly-synthesized pre-mRNAs are processed to mature mRNAs: 5´-end addition of a cap structure (II.), splicing (III.), and 3´-end polyadenylation (IV.). Quality of the mRNAs is assessed for efficient export through the nuclear pore complex into the cytosol (V.). Here, translation is initiated by translation factors and the 43S preinitiation complex. During mRNA translation by ribosomes, tRNAs deliver amino acids to the translational machinery (VI.). For decay, mRNAs are deadenylated, decapped (VII.), and finally degraded by ribonucleases or the exosome (VIII.).

3.1.1.4 mRNA Translation – Implementing the information

The main function of mRNAs is to provide the genetic information for proteins to the cellular protein synthesis machines, the ribosomes[6] (Figure 3-1 VI.).

The canonical eukaryotic translation starts with the recognition of the cap structure by the trimeric *eukaryotic initiation factor* (EIF) *EIF4F*[33]. This protein complex facilitates the recruitment of and provides a binding site for the *43S preinitiation complex* consisting of the *small ribosomal subunit* (40S), other EIFs, and the activated initiator $tRNA^{Met}$.[34,35] Then, the mRNA is scanned by the 43S complex in 5´-3´ direction until the start codon (AUG) is recognized and complement base-pairing with the initiator $tRNA^{Met}$ anticodon (UAC) takes place[36]. This interaction is the signal for a substantial reorganization of the complex which enables the binding of the *large ribosomal subunit* (60S) to form the translation-competent ribosome (80S)[37] bound to the mRNA[35,38].

After translation initiation, the polypeptide chain is built up in the elongation phase by a repetitive mechanism. Fundamental for this is the interaction of amino acid-loaded tRNAs (aminoacyl-tRNA) with the catalytic center of the ribosome which provides three consecutive tRNA binding sites. At the very beginning of this mechanism, the initiator $tRNA^{Met}$ is located in the *peptidyl site* (P-site). Next, different aminoacyl-tRNAs enter the ribosome´s *aminoacyl site* (A-site) and are checked for proper base-pairing with the codon triplet of the mRNA. If this is the case, the peptidyl-tRNA bond between the methionine and the initiator tRNA is attacked by the amino group of the new, tRNA-bound amino acid (aa). This results in the orchestrated transfer of the methionine from the initiator tRNA to this amino acid and the formation of a peptide chain. Additionally, by the 5´-3´ translocation of the ribosome on the mRNA, a new codon is placed to the active site, the polypeptide-bound tRNA moves to the P-site, and the initiator tRNA enters the *exit site* (E-site) from which it is finally released. Thereby, the A-site is free for new aminoacyl-tRNAs and the mechanism is repeated resulting in a growing polypeptide chain[39-41]

Elongation is assisted by several *eukaryotic elongation factors* (EEFs) until the ribosome encounters a stop codon (UAG, UGA, or UUA)[42]. These base triplets induce translation termination by the *eukaryotic release factor 1* (ERF1). The structure of ERF1 mimics a tRNA which enables this protein to enter the A-site and recognize the anticodon by a short peptide sequence. It also catalyzes the hydrolysis of the peptidyl-tRNA bond. This results in the release of the newly synthesized polypeptide whereas the ribosome is actively recycled[41,43].

3.1.1.4 mRNA Decay – Degrading the information

Intracellular mRNA levels are constantly adjusted by an equilibrium between mRNA synthesis and decay. To some extent, actively translated mRNAs are protected from degradation by proteins involved in translation shielding their ends. However, the major pathway for eukaryotic mRNA degradation starts with deadenylation of the 3´-end[44]. This is followed by either de-capping of the mRNA and 5´-3´ exoribonucleolytic cleavage facilitated by nucleases or by 3´-5´ degradation via the *exosome* complex[45]. In addition, also both deadenylation-independent and endoribonucleolytic mRNA decay pathways in parallel, are possible[44] (Figure 3-1 VII. and VIII.).

mRNA degradation is often connected to mRNA quality control because translating defect messages can have severe problems for the cell producing malfunctional proteins. Therefore, the cell must prevent translation of defective mRNAs and degrade them rapidly[44,46,47]. In addition, gene expression itself has to be tightly controlled and regulated to maintain physiologically optimal levels of each and every protein.

3.1.2 Regulation of eukaryotic gene expression

The integrity of mRNAs and proteins is of high importance, but also the correct cellular levels at a given time point are essential for the cell. The ability to regulate protein expression enables the cell to differentiate, proliferate, and adopt to environmental conditions. Consequently, various regulating and counteracting mechanisms during the above introduced steps of gene expression have been evolved in eukaryotic cells.

3.1.2.1 Major pathways of transcriptional gene regulation

The first step in modulating gene expression is changing the chromatin structure. In eukaryotes, DNA does not exist solitary but is wrapped around octameric *histone* complexes[48] and forms the so-called chromatin of which two types exist: lightly packed *euchromatin* is enriched in genes which are often transcribed whereas *heterochromatin* is more condensed and mostly transcriptionally silenced[49]. This is due to the decreased accessibility of the densely packed DNA for transcription factors and polymerases. By changing euchromatin to heterochromatin, large genomic regions or whole chromosomes, like the Barr body during X-chromosome inactivation in females[50,51], can be silenced and gene expression repressed. *Vice versa*, expression of heterochromatic regions can be upregulated by turning it into euchromatin[52]. The biochemical background for this regulative mechanism is methylation and demethylation of DNA and different *posttranslational modifications* (PTMs) on the histones[53-55]. Several families of histone-modifying proteins exist to write the correct histone code for a given situation and

developmental stage[56]. Being part of epigenetics, this mechanism is the only inheritable mechanism of regulation of gene expression[57].

In addition, unusual DNA secondary structures are predicted to have functions during transcriptional regulation in eukaryotes. To this kind of non-B-DNA structures belong Z-DNA, cruciform structures, triplex and quadruplex DNA[58,59]. Regions which can form into Z-DNA, for example, are enriched in 5´regions and TSSs of genes[60].

Another level of transcriptional regulation is provided by specific transcription factors (TFs)[61,62]. Contrary to the before-mentioned GTFs (see 3.1.1.1), TFs do not upregulate transcription in general but bind to specific *cis*-acting DNA sequences to manipulate expression of certain genes[61]. Depending on the resulting up- or downregulation, they are referred as *activator* or *repressor*, respectively[63,64]. Their binding to specific DNA sequences (*enhancer* or *silencer*, respectively)[65] can influence expression of associated genes over long distances. Activity of a transcription factor is determined by its cellular localization, its ability to bind the *cis*-acting element, and to transduce signals to the transcription machinery[66]. Many TFs are localized in the cytoplasm where they can be activated by a signal transduction cascade. Upon a distinct stimulus, suitable transcription factors are posttranslationally modified and translocate to the nucleus where they can bind to their DNA target sites. Posttranslational modifications are also able to control the binding strength of TFs to the DNA and the activity of their activation/silencing domain[66,67].

Via this domain transcription factors interact with transcriptional coregulators like the *mediator complex* which bridges transcription factors and RNAP II[68]. The mediator complex interacts with the *C-terminal domain* (CTD) of RNAP II which consists of up to 52 heptapeptide repeats (sequence: YSPTSPS)[19]. Regulation of RNAP II activity and processivity is determined by the phosphorylation status of the serine (S) residues in the CTD. To add or remove this form of PTMs, the CTD is targeted by different kinases and phosphatases, respectively, representing another level of transcriptional regulation of gene expression[69].

If and how much of a mRNA is produced is determined by the orchestration of these regulatory mechanisms. However, the fate of the individual mRNA molecule is under the control of processes summarized by posttranscriptional regulation of gene expression. Like transcriptional regulation it has a major impact on protein translation.

3.1.2.2 Major pathways of posttranscriptional gene regulation

Regulation of gene expression on the posttranscriptional level can occur at every stage of a mRNA´s existence (see 3.1.1, Figure 3-1) and is determined by different aspects (Figure 3-2). However, it mostly follows three basic principles: 1.) how abundant is the mRNA, 2.) how good is it exported, and 3.) how effective is it translated. Manipulating any of these parameters results in a change in the quantitative output of the encoded protein. In addition, some mechanisms can also alter the form of the expressed protein (e.g. isoforms).

As mentioned above (see 3.1.1.2) the first processing mechanism on newly-synthesized pre-mRNAs is capping[8]. The formation of the cap structure, additionally bound by the heterodimeric *nuclear cap-binding complex* (nCBC)[70,71], greatly increases mRNA stability, export, and translation efficiency[21,72]. Capping of pre-mRNAs depends, beside others, on the phosphorylation status of RNAP II. However, controlling this process results in a more general than specific regulation of gene expression[9].

Contrary, splicing can be extremely specific by the expression of different protein isoforms often in a tissue- or situation-dependent manner. Splicing is regulated by an interplay of *trans*-acting proteins (*splice activator* or *repressor*) and *cis*-acting sequences on the pre-mRNA. Depending on their function and location, these sequences are referred to *exonic* or *intronic splice enhancer* or *silencer* (ESE, ISE, ESS, ISS)[73,74], respectively. Consequences of alternative splicing can be changed function, localization, activity, or stability of the protein[75].

Like capping, 3´-end polyadenylation increases mRNA stability, export, and translation efficiency by recruiting the *nuclear poly-A-binding protein* (PABPN1)[76,77]. Polyadenylation can also be alternative: Usage of different poly-A signals, assisted by diverse auxiliary elements, results in altered lengths of 3´ *untranslated regions* (UTRs). This is of special interest if by this, for example, docking elements for other mRNA-binding proteins or RNAs (like miRNAs, see below) are excluded from the 3´UTR[78].

The next step, mRNA export, itself represents also a level of gene regulation. As part of the mRNA quality control improperly processed mRNAs are retained in the nucleus and defect transcripts are degraded during this step. Contrary, export of well-processed transcripts is achieved by interactions between export factors, e.g. Nxf1/Nxt1[31] or XPO1[79], with the NPC. Efficient translocation depends furthermore on the presence of nCBC (marks for efficient capping), EJCs (splicing), and PABPN1 (polyadenylation), on the mRNA. Thus, it is guaranteed that typically only processed mRNAs– capped, spliced, and polyadenylated – are exported[30,80].

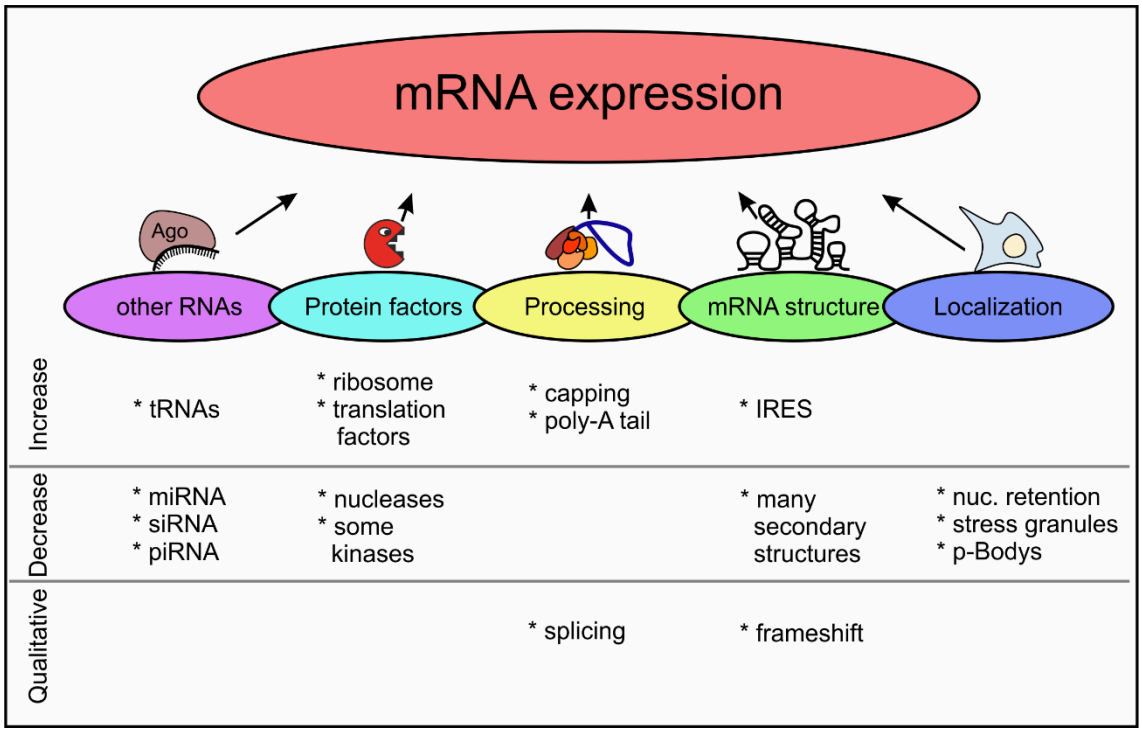

Figure 3-2: Overview of important determinants of mRNA expression.
Expression of mRNAs can be modulated by *trans*-acting RNA, protein factors, or complexes of both. Processing can have tremendous effects on mRNA expression, not only quantitatively but also regarding protein isoforms. Structures on the mRNA itself can also increase or decrease their expression. In some cases, a change of the form of the resulting protein is possible. Nuclear retention or sequestration in granules results in decreased mRNA expression.

Once the mRNA has reached the cytoplasm the second round of mRNA quality control occurs. For this, a dramatic reorganization of the mRNA-bound proteins takes place in the so-called *pioneer round of translation*[81]. The nCBC facilitates the initiation of a ribosome on the mRNA which then starts to translate the message. Rapid degradation of the mRNA can be triggered at this point, if 1.) a *premature termination codon* (PTC) is recognized by the existence of downstream EJCs on the mRNA leading to *nonsense-mediated decay* (NMD), or 2.) the lack of a stop codon results in stalling of the ribosome at the 3´end of the mRNA and *non-stop decay* (NSD), or 3.) strong secondary structures prevent the progression of the ribosomes and the mRNA is marked for *no-go decay* (NGD)[82]. If the mRNA is translational competent, nCBC is replaced by EIF4E, EJCs are removed by the pioneer ribosome, and PABPN1 is exchanged with its cytosolic counterpart PABPC1[72]. The mRNA is now ready for steady-state translation.
The translational process itself can be regulated in all its three phases (see 3.1.1.4). To achieve this, eukaryotic cells possess an arsenal of different active and passive factors and elements of which examples are presented in the following.

Initiation of translation can be actively regulated by phosphorylation of the alpha subunit of *eukaryotic initiation factor 2* (EIF2). This prevents formation of the active 43S complex and halts translation initiation[35]. This kind of reversible posttranslational modification is done by one of four kinases (EIF2AK1, EIF2AK2, EIF2AK3, EIF2AK4) which are sensing for and get activated by induction of certain kinds of cellular stress[83]. The rationale of this mechanism is to conserve resources while a new gene expression program is set up to react on the distinct stress stimuli[84].

Translation initiation is passively affected by structured RNA elements in 5´UTRs[85] which can decrease efficiency of the recruitment of the 43S complex, decelerate the scanning process, or completely abolish 43S progression. One of the best-studied examples is the *iron responsive element* (IRE) which regulates the translation of a subset of mRNAs that are involved in iron homeostasis[86]. Contrarily, specific structured RNA sequences can positively affect and promote translation by serving as an *internal ribosomal entry site* (IRES)[87]. Here, cap-independent initiation is possible which enables the cell, for example, to circumvent stress-induced blockage of canonical translation initiation and produce distinct mRNAs in response[88]. First discovered in viral RNA, more than 100 IRES are meanwhile reported in mammalians and it is expected that over ten percent of mammalian mRNAs dispose such a mechanism[89].

An example for active regulation of translation elongation is phosphorylation of a here-involved key factor: The sole function of the protein *elongation factor 2 kinase* (EF2K) is to phosphorylate the *eukaryotic elongation factor 2* (EEF2) which inhibits the binding of this essential factor to the ribosome. Without active EEF2, tRNAs cannot translocate from the A- to the P-site and the ribosome gets stuck on the mRNA. Also, the general availability of aminoacyl-tRNA can alter the speed of translational elongation[90].

In translation termination, structural elements and specific sequences can cause a read-through of the stop codon which results in an additional C-terminal domain, such as a localization signal, at the newly-synthesized protein[91,92].

Translation of a protein-coding sequence can also be downregulated by the presence of an *upstream open reading frame* (uORF). This sequence element, coding for a non-functional polypeptide, decreases translation of the downstream-located major coding sequence[93].

Another important level of posttranscriptional gene regulation is the sequestration of mRNAs. Sequestration means that the mRNA is removed from the sites of active translation in the cytosol and stored in granules like *stress granules* (SGs) or *processing bodies* (P-bodies)[94]. Stress granules are enriched in mRNAs, 40S ribosomal subunits, and translation initiation factors as well as RNP-remodeling enzymes[95]. However, their composition varies depending on different stress stimuli. P-bodies contain, beside

others, mRNAs and degradation factors[96]. The exact functions of both stress granules and P-bodies are still controversial. Suggestions range from degradation of messages to their remodeling, stabilization, and storage for translation as soon as they are needed again[95,96]. Importantly, mRNAs which are localized to SGs or P-bodies are translationally silenced.

Translational silencing is also the result of a molecular mechanism known as RNA interference (RNAi). This term comprises related pathways in which small non-coding RNAs associate with proteins of the Argonaute or the Piwi family. This results in the formation of an *RNA-induced silencing complex* (RISC), a gene-silencing ribonucleoprotein targeted against groups or individual mRNAs[97]. Specificity is achieved by base complementarity between the small guiding RNA and its target mRNA. The RNA classes acting in this mechanism are *micro RNAs* (miRNAs), *small interfering RNAs* (siRNAs), and *PIWI—interacting RNAs* (piRNAs), although they differ in their origin and biogenesis. First discovered in nematoda, RNA interference is widely spread in higher-ordered eukaryotes. For example, 5% of the human genome encodes for more than 1,000 miRNAs regulating ~30% of all protein-coding genes[98,99]. RNA interference fulfills roles in cellular development and immunity against parasitic nucleotide sequences like viruses and transposons.

Finally, mRNA decay is the ultimate posttranscriptional regulation mechanism. Destroying is the irreversible fate of every mRNA independent of whether this is induced by quality control (NMD, NGD, NSD)[82], RNA interference[97], or general instability of the transcript (short poly-A tail, de-capping)[44]. These destructive mechanisms give the cell the opportunity to react on its needs and on external stimuli by deleting old and producing new messages.

All the steps of posttranscriptional gene regulation are performed by an orchestrated interplay between RNA and proteins. At no time mRNA is without any interaction to a protein in the cell. The constant change of the mRNP code, the protein composition on the mRNA, determines the fate and the functional status of every individual mRNA molecule in the cell[7]. Basis for this are interactions of *RNA-binding proteins* (RBPs) with mRNAs via their RNA-binding domains. In addition to the canonical, binding motif-based recognition of target mRNAs secondary RNA structures can also serve as recruitment platforms for RBPs. In the next section some of these structural aspects will be presented before RBPs are discussed in greater detail.

3.2 Structural aspects of RNA

For decades, RNA has been considered to mainly have a passive function for example as carrier of information between DNA and protein or as a structural component. However, together with the identification of other RNA classes, it is today well-established that RNA is a versatile macromolecule that can also actively and functionally interact with biological processes[100]. An important feature for many RNA-mediated regulatory processes is the inherent ability of RNA to form stable inter- or intramolecular structures[101]. Consequently, the exact and tight regulation of the formation or resolution of RNA structures is essential to control their function and therewith associated biological mechanisms.

3.2.1 Classical RNA structures and their functions

Single-stranded (ss) RNA has a greater potential to form higher-ordered structures than double-stranded (ds) DNA due to the lack of a competing complementary strand. However, compared to proteins, RNA can adopt fewer structural variations[102]. The lower number of different building blocks, 20 amino acids versus 4 ribonucleotide bases, and their less intricate nature in terms of charge, polarity, size, and individual structure result in only few basic structural elements in RNA. Nevertheless, by combining these, RNA can adapt three-dimensional assemblies, often as complex as tertiary protein structures. Classic features of RNA structures are helices, loops, bulges, and junctions[103] (Figure 3-3 A) formed by canonical Watson-Crick base-pairing. Helices combined with a loop on top are called stem-loops or hairpin structures. A hairpin structure´s loop that forms the stem of another hairpin and *vice versa* is named pseudoknot[104]. Stability of these structures is determined by the length and GC-content of the stem, the length of the loop, and the number of mismatches (bulges). Complex tertiary RNA structures are the consequence of higher-ordered interactions resulting from a hierarchical and sequential folding of RNA[105]. The function of many RNAs greatly depends on their precise three-dimensional structure, thus some examples for this are presented here.

Most commonly RNA structures serve as interaction or recruitment platforms for proteins and manipulate or direct the function of the polypeptide. Some prominent examples are hairpin structures on *clustered regularly interspaced short palindromic repeats* (CRISPR) RNAs (crRNAs) recruiting *CRISPR-associated* (Cas) enzymes[106] (Figure 3-3 B), structured IRESs on mRNAs serving as alternative ribosome recruiting sites[87], or sequestration of the *positive transcription elongation factor b* (P-TEFb) in an inactive RNP by a 5´-located stem loop structure in the small nuclear RNA (snRNA) 7SK[107]. In translation, the tertiary structure of tRNAs has function not only in binding proteins but

also in fitting the RNA into protein complexes. In two-dimensional representations tRNAs resemble a cloverleaf (Figure 3-3 C) but coaxial stacking of the four stems ends in a L-shaped tertiary structure which perfectly fits in the A-, P-, and E-sites of the ribosomal catalytic center[108]. Recognition and binding of RNA structures by enzymes can also be important for their own maturation. An example is the exact processing of pri- and pre-miRNAs by Drosha and Dicer ribonucleases, respectively, facilitated by their hairpin structure[109].

RNAs that exhibit own catalytic activity (ribozymes) also greatly depend on correct folding. Here, structures form catalytic sites or keep the catalytically-active nucleotides in a single-stranded conformation. Examples are the eukaryotic 28S ribosomal RNA (rRNA), presumably consisting of 94 conserved stem-loop structures[110], self-splicing introns (Figure 3-3 D), RNase P, and hammerhead ribozymes[102].

RNA structures also act as functional obstacles. Many ribonucleases have a preference for structured (e.g. dsRNA in hairpins) or unstructured (ssRNA) stretches. For example, among the eight human canonical ribonucleases, RNase 1 shows the highest enzymatic activity on dsRNA whereas RNase 2 and 3 are completely inactive on them[111]. Thus, RNA structures influence, to some extent, transcript stability and decay.

These few examples show only an excerpt of possibilities why RNA structures are important for cellular mechanisms; possibilities which are further extended by another nucleic acid formation, the so-called G-quadruplex structure.

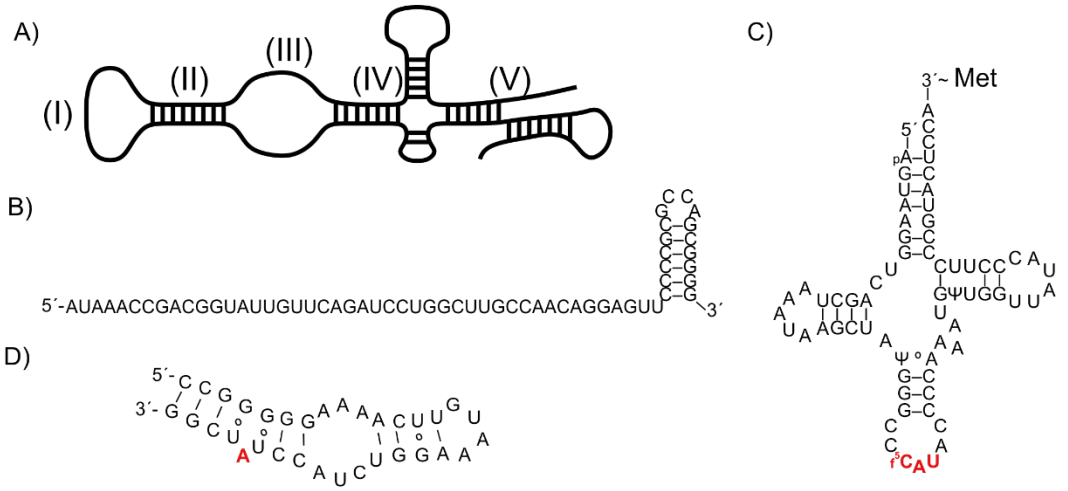

Figure 3-3: Characteristics and examples of RNA secondary structures.
A) Schematic of examples of general RNA secondary structures are loops (I), helices (II), bulges (III), junctions (IV), and pseudoknots (V). Combination of a loop and a helix results in the formation of a stem loop or hairpin structure. B) Example of a mature crRNA derived from the *E. coli* R44 CRISPR. The crRNA is bound by Cas nucleases via its 3´-terminal hairpin[112]. C) Cloverleaf-shaped human initiator tRNA^{Met}. The anticodon for the start codon is highlighted in red[113]. D) Secondary structure of the D6 domain of the self-splicing group II intron[114]. The 2´-OH group of the conserved adenosine (red) acts as the nucleophile in the first step of splicing (branch point of splicing). The RNA structure keeps the branch point unpaired and in an exposed position.

3.2.2 G-quadruplex structures

Ten years after Watson and Crick´s groundbreaking study[4], Karst Hoogsteen reported an alternative base-pairing in nucleic acids[115] in which pairing of guanine bases is possible (Figure 3-4 A, B, C). Recently, Hoogsteen base-pairing was demonstrated to occur in canonical duplex DNA in a thermal equilibrium with Watson-Crick base pairs[116]. Hoogsteen hydrogen bonding in nucleic acids enables the formation of higher-order arrangements like triple-stranded helices[117] and, of great importance for this thesis, G-quadruplex structures (G4s).

3.2.2.1 General characteristics of G-quadruplexes

G-quadruplexes are stable secondary structures of nucleic acids. The building block of a G4 is the G-quartet, a square and planar formation of four guanines arranged by eight hydrogen bonds (Figure 3-4 C). Stacking of two or more G-quartets results in the formation of the three-dimensionally G-quadruplex structure with single-stranded loop regions connecting the G-tracks. Under physiological conditions, monovalent cations (mostly K^+ and Na^+) in the center between two G-quartets stabilize this four-stranded structure by minimizing the electrostatic repulsion of the guanine O6 oxygens (Figure 3-4 A, C). Stability generally increases with the number of stacked G-quartets[59,118,119].

G4s can vary in the number of participating molecules (intra- or intermolecular), the directionality of the strands (parallel, antiparallel, hybrid), or the length and orientation of the bridging loops (propeller, lateral, diagonal) making them a topological highly diverse form of secondary structures[59,120] (Figure 3-4 D). The classic *putative quadruplex-forming sequence* (PQS) is $G_3L_xG_3L_xG_3L_xG_3$ (with x = 1-7) resulting in the formation of a three-layered G-quadruplex structure[121,122]. However, additional PQSs have been found forming G4s consisting of only two G-quartets or with a more variable loop length. Particularly in RNA more flexible parameters are tolerable because RNA G4s (rG4s) have generally a higher thermodynamic stability than DNA G4s (dG4s), are more prone to form stable secondary structures, and lack a competing complementary strand[123-125]. However, rG4s exhibit almost exclusively a parallel conformation due to steric constraints. So far, only one human telomeric RNA sequence was shown to form an antiparallel rG4[126].

In recent years much effort focused on the synthesis and characterization of G4-stabilizing ligands[127,128]. These small chemical compounds are always composed of several aromatic rings non-covalently interacting with the external (top) G-quartet and stabilizing by this the G4 structure. Over the past 15 years more than 800 G4 ligands with different specificities and binding abilities were discovered[129]. Among these is the rG4-specific ligand carboxypyridostatin[130] (cPDS) which was also used in this study[127,128].

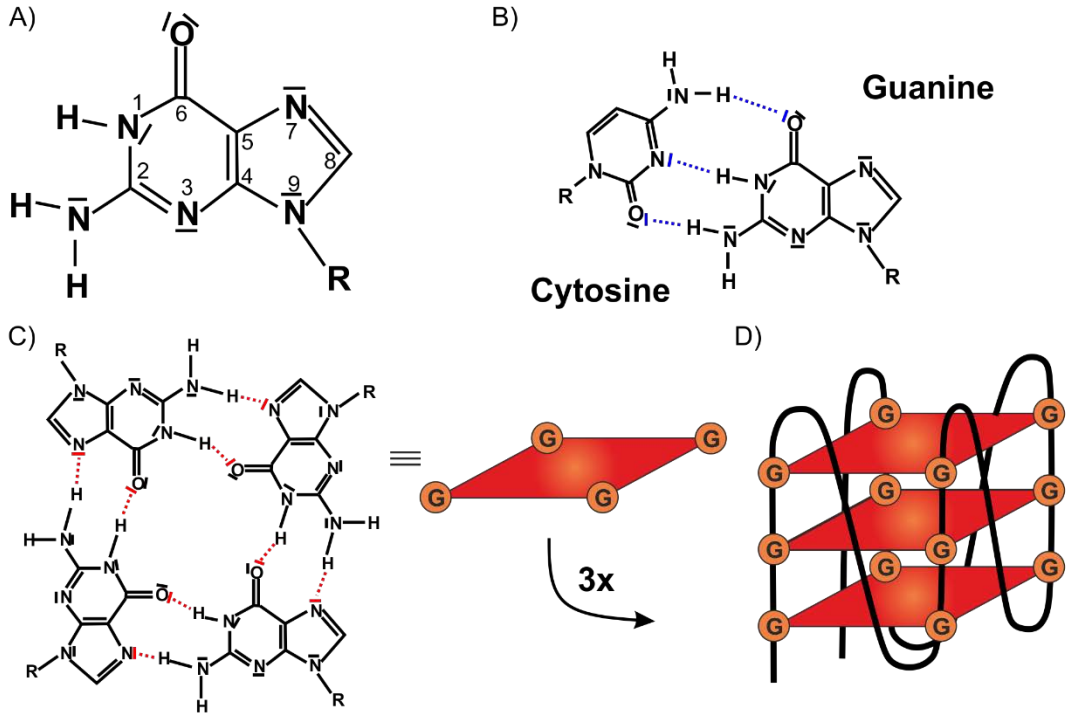

Figure 3-4: G-quadruplex structure formation
A) A guanine base with numbered positions of the purine ring. B) Canonical Watson-Crick base-pairing of cytosine with guanine. The intermolecular hydrogen bonds (HN_1, $HN\text{-}C_2$, and $O{=}C_6$) are highlighted in blue. C) Four guanine bases form a G-quartet by Hoogsteen hydrogen bonds highlighted in red (HN_1, $HN\text{-}C_2$, $O{=}C_6$, N_7). Stacking of G-quartets results in formation of a G-quadruplex structure. Here, a parallel G4 is depicted as an example[131].

Early reports in the G-quadruplex field focused on dG4s and still most work is concentrated on them. Consequently, less is known about rG4s and their biological implications. In the following chapters general information about G-quadruplexes will be provided together with specific information about rG4s, which are in the focus of this thesis.

3.2.2.2 G-quadruplex distribution in the human genome

Bioinformatic analyses have identified more than 375,000 dG4 motifs in the human genome following the rules for canonical PQSs ($G_3L_xG_3L_xG_3L_xG_3$, x = 1-7)[121,122]. By this, the total numbers of dG4 motifs within the human genome is higher than expected from its GC-content[121,122]. An experimental high-throughput analysis utilizing a polymerase stop assay coupled with next-generation sequencing detected ~525,000 observable dG4 sites. This assay allowed the formation of non-canonical dG4s which are not restricted by the canonical PQS motif[132]. This number grew to over 700,000 individual sites after the addition of the dG4-stabilizing ligand pyridostatin (PDS)[133,134]. Analysis of the location of those sites revealed that dG4 motifs are not randomly distributed over the genome but are enriched at functional positions. They are overrepresented in promoter regions,

transcriptional start sites, meiotic and mitotic double strand break sites, telomeres, and replication origins[59,121,135,136]. Additionally, they are enriched at the sites of proto-oncogenes while being underrepresented in tumor suppressor genes[135,137].

Further bioinformatic analysis showed that ~75,000 of the before-mentioned ~375,000 dG4 sites in the human genome (~20%) are located within genes and consequently present in the human transcriptome[122]. Among those, ~14,000 sites are located in exonic sequences.

In an experimental approach named rG4-seq Kwok and coworkers were able to identify 3,845 G4-forming sites in the transcriptome of HeLa cells by a combination of rG4-induced reverse transcriptase stalling (RTS) and next-generation sequencing[138]. After addition of cPDS[130] 13,423 rG4 sites could be detected by this approach[138]. In parallel, the group of David Bartel was able to identify 12,009 and 12,035 rG4s in the transcriptome of HeLa and HEK293 cells, respectively, using a RT-stop assay[139]. The overlap of the HeLa rG4s detected in these two independent approaches by two different research groups was notable ~65%[140]. Analyzing the distribution of rG4 motifs in the transcriptome revealed their enrichment in 5´- and 3´UTRs of mRNAs[140,141]. Also, many non-coding RNAs, especially TERC (*telomerase RNA component*) and (TERRA) (*telomeric repeat-containing RNA*), harbor rG4 motifs[142].

Although the formation of G-quadruplexes was considered to be an *in vitro* artifact and controversially discussed for a long time accumulating experimental evidence supports their formation *in vivo*.

3.2.2.3 Existence of G-quadruplex structures *in vivo*

The concept of *in vivo* formation and function of G-quadruplexes provoked a lot of skepticism in the scientific community. Critics raised concerns that formation of G4 structures is on the one hand too slow to regulate processes *in vivo* and on the other hand too stable impeding others, like replication[59]. Indeed, it is documented that intermolecular dG4s form and resolve slowly under physiological conditions[143]. However, intramolecular dG4s consisting of human telomeric DNA were shown to fold in a millisecond timescale[144]. Additionally, the existence of G4-specific chaperones and helicases allows the cell to actively and rapidly control their formation or disruption (see 3.2.2). Nevertheless, despite many years of research, experimental evidences for the presence of G4 structures in cells are limited or even controversial[131].

The discovery of G4-manipulating proteins is a first, albeit indirect, evidence for G4 structure formation *in vivo* (see 3.3.2). The necessity for the cell to develop a molecular toolbox to specifically regulate G4 structure formation and resolving shows that 1.) these structures are present *in vivo* and 2.) they can be regulated and used to achieve an evolutionary advantage[145]. Additionally, the before-mentioned enrichment of G4 motifs

at regulatory sites in the genome is evolutionary conserved from yeast[146] up to human[147] and a similar pattern is present in prokaryotes[148], too. This suggests that a selective pressure exists for organisms to retain and preserve specific G4 motifs that might exhibit a favorable biological role[59].

First experimental proof for the *in vivo* existence of dG4 structures was provided by the identification of target sites of the dG4-stabilizing ligand PDS[149] that overlap with binding sites of a DNA helicase that can resolve dG4s in human cells. A breakthrough was the development of telomeric G4-specific antibodies for the ciliate model organism *S. lemnae*[150] (Sty3 and Sty49) and of an antibody able to detect various G4 structures in human cells (BG4)[151]. Using these antibodies, it was for the first time possible to visualize G4s directly by immunofluorescence microscopy. Experiments in human cells with BG4 showed corresponding speckles in the nucleus which are sensitive to DNase 1 treatment. Reciprocally, increased BG4 signals were detected by both exogenic transfection of *in vitro* folded DNA G4s and treatment of the cells with PDS.

The development of G4 ChIP-Seq (**Ch**romatin-**I**mmuno**p**recipitation DNA-**Seq**uencing) using the BG4 antibody recovered ~10,000 dG4 sites *in vivo* in human chromatin representing ~1.5% of the G4 sites observed *in vitro* [134,152,153]. Those *in vivo* dG4s were enriched in regulatory as well as promoter regions of highly-transcribed genes and oncogenes.

Finally, many studies demonstrated that dG4s indeed affect cellular processes *in vivo*. On the DNA level G4 structures were shown to influence telomere homeostasis, replication[154,155], and transcription[145]. As only one example, a G4 structure in the c-MYC promoter controlled 90% of c-MYC transcription. Both a single base substitution that disrupted this G4 and the addition of the stabilizing ligand TMPyP4 to the cells greatly altered expression of the gene[156].

As already mentioned above rG4s have been less investigated compared to their DNA counterparts. Therefore, less evidence of their *in vivo* existence is published. rG4 structures can be detected in the cytoplasm of human cells by immunofluorescence microscopy using the BG4 antibody, albeit these signals are less intense compared to the nuclear speckles resulting from dG4s[157]. In these experiments BG4 signals were sensitive to RNase A treatment but not to DNase 1 treatment and increased upon specific stabilization with cPDS, thus proving their RNA identity[157].

In addition to the identification of rG4 sites in the human transcriptome (see 3.2.2.2) Guo and Bartel tried to couple chemical probing *in vivo* to selective stabilization of G4s by monovalent cations (Na^+ vs. K^+) and next-generation sequencing to detect folded rG4s *in vivo*[139]. Unexpectedly, they found rG4s to be globally unfolded in human cells. They proposed the existence of a robust rG4-unfolding machinery because addition of cPDS

had only minor effects whereas depletion of ATP or of the rG4-resolving helicase DHX36 showed almost no effect[139]. However, this approach had some critical limitations: First, a correct positive control was lacking, therefore it cannot be assured that this approach is generally suitable to detect rG4s *in vivo*. Second, this approach used differences in the stabilization of G4 structures by sodium and potassium cations. However, sodium cations do also stabilize G4 structures, even though less than potassium[124]. It would have been better to use non-stabilizing lithium cations in the control experiment[158]. Third, in a particular mRNA, a G-quadruplex may only be folded during a fraction of the mRNA´s lifetime (e.g. during splicing or transport), yet still be biologically relevant[131]. The Bartel approach provided most likely only a snapshot of a mixed cell population in which individually formed rG4s were probably lost in the background[140].

An important proof for rG4s are their addressed biological functions in many published reports. In the following, I will present the biological functions of rG4s in greater detail.

3.2.2.4 Functions of RNA G-quadruplexes

Several functions for rG4s have been predicted among some are experimentally proven[119]. rG4s can have functions in the nucleus by regulating telomere homeostasis and during the early steps of gene expression, including transcriptional termination and pre-mRNA processing (splicing and polyadenylation)[119], as well as in miRNA processing[159]. Possible cytoplasmic functions are mRNA targeting and translation[119] and miRNA targeting[160] (Figure 3-5). Additionally, viral rG4s are supposed to affect human cells[161]. In the following, these functions are constituted in greater detail.

Telomeres as well as TERC and TERRA[142] have a high G-content and are prone to fold into G-quadruplexes[59,162]. So far, only one mechanism is known how an rG4 can regulate telomere homeostasis. This concrete model comprises the presence of an rG4-motif at the 5´-end of human TERC. Formation of this rG4 impedes the folding of a helical motif named P1 which is crucial for accurate reverse transcription by the telomerase enzyme. Solely, when this G4 structure is actively resolved by a helicase P1 can form and telomerase is proficient for telomere extension[162].

In the genome G4 motifs are overrepresented directly after the 3'-end of genes especially when a downstream gene is in direct proximity allowing the conclusion that they fulfill biological functions at this position[141]. A proposed mechanism implies the contribution to transcriptional termination by acting as a polymerase pausing site. The existing theories about the mode of action comprise dG4s, rG4s, and/or hybrid structures[119]. While the exact nuclear mechanism for rG4 action here is still unclear, it was demonstrated for polycistronic gene expression in mitochondria that the formation of an rG4 structure stimulates transcriptional termination. Single G-to-A mutations reduced the amount of premature termination in a similar range like addition of the rG4-destabilizing compound

7-Deaza-GTP[163]. The proposed model resembles the prokaryotic intrinsic termination (Rho-independent termination) of transcription. This rG4 also acts as a switch between mitochondrial transcription and RNA primer synthesis for replication: Binding of the human mitochondrial transcription factor TEFM prevents formation of the rG4 and facilitates transcription[163].

Tightly connected to transcriptional termination is 3´-end processing of mRNAs. rG4s are described to act as a structured auxiliary element for alternative poly-A sites that promote interactions with polyadenylation regulators like hnRNP H/F[164]. Experimentally validated is a rG4 in the 3´UTR of the *Fragile X mental retardation syndrome-related protein 1* (FXR1) mRNA which facilitates the expression of a shorter 3´UTR version (short: 355 nt vs long: 972 nt) by enabling the usage of an alternative poly-A signal 60 nt upstream of the rG4. Disruption of this structure by a single nucleotide mutation prevents expression of the short mRNA[165]. Unfortunately, this report does not include the reciprocal experiment that stabilization of the rG4 leads to increased amounts of short mRNA. However, treatment with the small stabilizing ligand PhenDC3 resulted in increased reporter gene expression. This is most probably due to exclusion of miRNA target sites which stabilizes the mRNA. Another report showed the involvement of an rG4 in maintaining 3´-end processing of the pivotal DNA damage response protein p53 upon UV irradiation. To fulfill its function, p53 itself has to circumvent its own downregulation by recruitment of hnRNP H/F to the rG4 maintaining p53 expression[164].

Another pre-mRNA processing step can be affected by rG4 formation, too. G-quadruplexes can be part of *cis*-acting regions facilitating both up- and down regulation of splicing by acting as splice enhancer or silencer, respectively. rG4 structure formation enhances splicing efficiency of the PAX9 intron 1[166]. Chemical stabilization of an rG4 in intron 6 of human TERC changes its splicing pattern resulting in increased expression of an inactive form and downregulation of telomerase activity[167]. Similarly, a rG4 antagonizes the usage of a major 5´ splice site in the Bcl-X mRNA not resulting in expression of the anti-apoptotic protein form but in the pro-apoptotic version[168]. In contrast, an rG4 located in intron 3 of the p53 mRNA favors splicing of the adjacent intron 2 whereas its disruption by mutation leads to intron 2 retention[169]. Also, this seems to be disease-related because a *single nucleotide polymorphism* (SNP) probably changing the topology of the G-quadruplex is associated with increased risk for several cancers[170]. The functional mechanism behind the splicing regulation by rG4s is often the recruitment of effector proteins involved in splicing, like hnRNP H/F or FMRP.

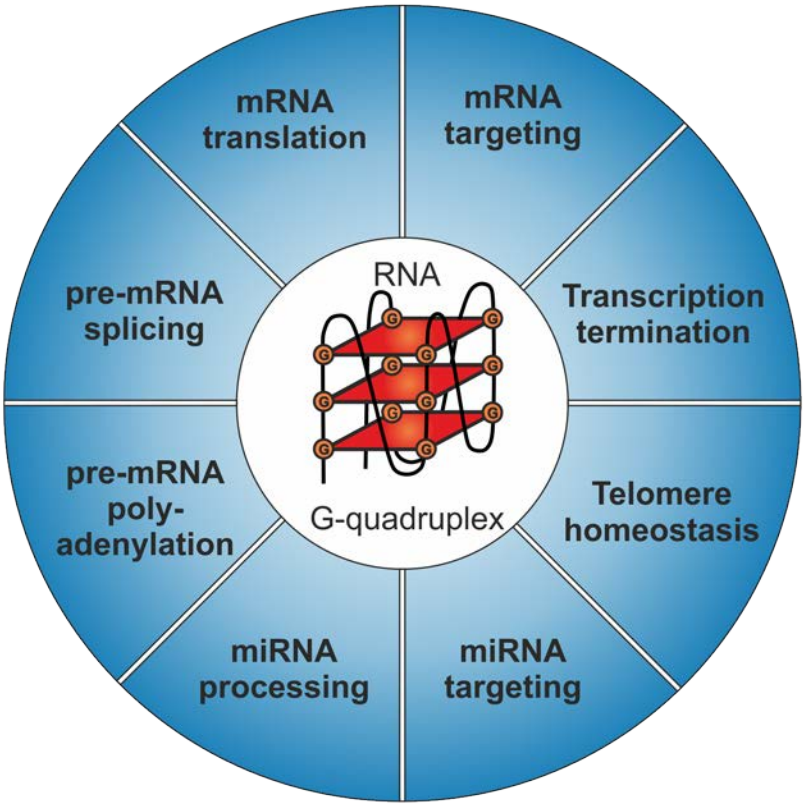

Figure 3-5: Functions of RNA G-quadruplex structures
RNA G-quadruplexes have functions in several cellular processes: In pre-mRNA processing rG4s can assist (alternative) poly-adenylation and splicing. In mature mRNAs rG4s can act as translational obstacles or initiation sites, or as docking sites for transport factors in mRNA targeting. Unclear is the exact mechanism of rG4-mediated transcriptional termination whereas rG4 formation in TERC results in downregulation of telomerase activity. A rather new field is the involvement of rG4s in miRNA biogenesis and function. General figure layout was adapted from Millevoi et al[119]. Figure was modified and updated according to recent insight in the RNA G-quadruplex field.

The largest wealth of experimental data for rG4 function is available for its regulatory role in translation. First hints on a potential regulative mechanism was published in 2001 when Schaeffer et al. determined an extremely G-rich minimal binding site for FMRP on its own mRNA which can form a ´purine quartet structure´. Cloning of this sequence in a reporter system resulted in a 1.5-fold decrease in protein expression, *in vitro*[171]. Many reporter gene assays with native G4-containing 5´UTRs have been followed *in vitro* and *in vivo* concluding that those structures mainly result in downregulation of protein synthesis[172-177]. This effect seems to be dependent on structure stability and position with the most prominent effect when rG4s are proximal to the 5´-cap structure or the start codon[178]. Furthermore, it was shown that translation of oncogenes in T-cell acute lymphoblastic leukemia is caused by 5´UTR-located G4s in an EIF4A-dependent manner[179]. Contrary, in rare cases ribosome binding can be assisted by rG4s by being integral parts of IRESs and facilitate cap-independent translation[180], for example in the VEGF mRNA.

Less investigated is the function of rG4s in the CDS and 3´UTR. For the CDS, only one report suggested a FMRP-mediated translation inhibitory effect on an individual mRNA[181] whereas another publication described a role of rG4s in ribosomal frameshifting[182]. Repression of translation was also reported for a rG4 located in the 3´UTR of the proto-oncogene PIM1[183]. However, such structures in PSD95 and CaMKIIa mRNAs have been identified as *cis*-acting element facilitating correct mRNA targeting in dendritic cells[184]. Here, they are supposed to act as binding sites for kinesin-associated proteins like FMRP[185]. Regarding cellular localization, another interesting concept comprises changes in stability of rG4s because of intracellular cation fluctuations due to neuronal activities. Thereby, both the localization and the translation of G4-bearing mRNAs can possibly be influenced[119].

Of great interest is also the interplay of rG4s and miRNA-mediated gene silencing. Recently, it was shown that rG4s often overlap with miRNA target sites. Consequently, a rG4 was shown to prevent binding of the miRNA 331-3p to the FADS2 mRNA[160]. But not only miRNA function can be regulated by G-quadruplex structures, also processing of the pri-mRNA can be affected in a negative or positive manner shown on the examples of miRNA200c and miRNA497 as well as miRNA451a, respectively[159]. Additionally, the involved research group identified over 700 pri-miRNA with the potential to form a rG4 structure.

Conclusively, it has to be mentioned here that potential G4 motifs and rG4 structures have also been identified to be highly abundant in the genomes of several RNA viruses. Among them are pathogenic viruses like the *human immunodeficiency virus* (HIV), the *human papillomavirus* (HPV), or the Epstein-Barr virus (EBV) which have rG4 structures at genomic sites critical for their replication[161]. In the *Flaviviridae* family, some notable members are the *Zika virus* (ZIKV), the *West Nile virus* (WNV), *dengue virus* (DENV), and the *yellow fever virus* (YFV), 7 highly conserved rG4 regions were identified suggesting a functional role for the viral life cycle[186]. However, so far little is known about viral rG4s, their exact role in the virus, and their function during infection of and propagation in human cells.

As presented above, various functions for rG4s are possible and their cellular existence has to be tightly regulated. Therefore, rG4 formation and especially their disruption is assisted by RNA-interacting proteins. Getting insight into this network helps us to understand, why, when, how rG4s form and which consequences this has for the cell.

3.3 Proteins interacting with RNA

RNAs form complexes with a variety of RNA-binding proteins (RBPs). A ´conventional` RBP participates in the formation of a ribonucleoprotein particle (RNP) to generate, process, edit, modify, transport, translate, or degrade the RNA or form a functional unit. However, this convention is broken by various novel RNA species. Many lncRNAs are assumed to recruit, organize, sequester, or inhibit proteins representing a mechanism in which the RNA and not the protein is controlling the RNP[187]. Nevertheless, proteins with the ability to bind or be bound by RNA play pivotal roles in fundamental cellular mechanisms. Consequently, proteins involved in RNA metabolism are among the proteins with the highest evolutionary conservation[188].

3.3.1 General aspects of RNA-binding proteins

RNA-binding protein is an umbrella term including all proteins interacting with RNA of any kind. They can be grouped according to different aspects like preferred RNA binding partner (e.g. mRNA, tRNA, ncRNA, rRNA), involvement in biological processes (e.g. RNA biogenesis, detection, splicing, translation, localization), or common molecular functionality (e.g. ribosomal proteins, translation factors, tRNA ligases, helicases, nuclease)[189]. Additionally, RBPs can be clustered in accordance to the presence of RNA-binding domains (RBDs) in the protein which mediate the RNA-protein interaction.

These structurally well-defined domains, among them the RNA recognition motif (RRM), RGG box, zinc-finger domain, K homology (KH) domain, or DEAD box domain, are present in many different RBPs and act often in combination to increase binding strength and specificity[189]. Unspecific binding is often mediated by interactions of positively charged protein side chains with the negatively charged sugar-phosphate backbone of the RNA whereas specific binding is achieved by interactions with the nucleobases. Here, an interplay between different preferences for the formation of aromatic interactions, electrostatic interactions, intermolecular hydrogen bonds, kinetics, and surface complementarity can lead to a high level of specificity[188,190]. Computational analyses revealed that many RBPs bind a relatively small, well-defined subset of RNA sequences of low-complexity motifs composed of just one or two major base types. However, linear sequence motifs are often insufficient to fully achieve specific RBP binding and contextual features, such as RNA secondary structures, are often needed in addition[191].

In a computational census of RNA-binding proteins, Gerstberger and coworkers came to a number of 1,542 manually curated RBPs in humans meaning that ~7.5% of all protein-coding genes interact with all known classes of RNA[189]. Experimental approaches to capture the mRNA interactome revealed 860 and 791 RBPs in HeLa and HEK293 cells,

respectively[192-194]. 543 RBPs are present in both data sets. Reliability and robustness of these methods were underlined by the detection of most well-established mRNA binding proteins as well as by the strong enrichment of classical RBDs in the analyzed proteins. Other RBDs are unique and characteristic for certain proteins (the 169 ribosomal proteins have distinct 119 RBDs exclusively present in them)[189]. Interestingly, about half of the proteins in each data set lacked known RBDs, results which are in line with a hypothesis suggesting that complex RNA-protein interaction do not require canonical RBDs and unconventional RNA binding is a broad phenomenon[187]. Consequently, novel types of RNA binding, like binding by *intrinsically disordered protein regions* (IDRs) or shape complementary-based interactions must be taken in account when searching for and analyzing RNA-binding proteins[86].

3.3.2 G-quadruplex-interacting proteins

In the G-quadruplex field, much effort is invested in the identification and characterization of specific G4-interacting proteins; however, this is a complex undertaking for several reasons.

G4 structures are topological highly diverse (see 3.2.2.1) and specificity for different structures can be determined through strand orientation, loop length, duplex-quadruplex junctions, groove of the G-quartet barrel, and/or a combination of these properties[195]. Even though rG4s are almost exclusively parallelly orientated they show an immense structural variety, too. A consequence of this is that every approach (e.g. pulldown) using an individual G4 structure will probably detect only proteins which can interact with this specific G4-bait but other proteins binding to different G4s might be missed. Also, in *in vivo* experiments, it is often not clear if a specific G4 is folded to the certain time point of analysis. Therefore, it has to be checked whether an identified candidate really binds G-quadruplex structures or only G-rich sequences. Despite this, many individual studies contributed to the identification of over 70 DNA and RNA G4-binding proteins in humans, recently listed in the *G4-interacting protein database* (G4IPDB)[196-198]. However, due to the large number of human cellular proteins and the few conditions tested so far, this database is very likely not complete and other G4-interacting factors will be identified in near future.

Recently, a computational comparison of the amino acid composition of the human G4-binding proteins from the G4IPDB revealed prominent enrichment for glycine (G) and arginine (R). Moreover, many of the analyzed proteins share a similar 20-aa long RG-rich motif[195]. Glycine and arginine are well known disorder-promoting amino acid residues[199], providing space for the speculation that G4-binding could be more due to

interaction with IDRs then to classical RBDs. In structural studies specific interactions between rG4s and the extended RGG motif of FMRP are demonstrated[200]. Search for such long RG-rich motifs in the human proteome ended in over 100 new potential G4-interacting proteins[195,201]. Nevertheless, further investigation of these candidates and validation of the predicted G4-binding is inevitable.

Examples for biological consequences of interactions between proteins and rG4s were already presented above (see 3.2.2.4) and their formation, resolution, and function in the cell must be tightly controlled[139,202,203]. This regulation can be achieved by either promoting the formation of G4s, preventing their folding, or actively resolve them, respectively. Stabilization of G4s was reported for the human protein nucleolin (NCL) and, especially on the DNA level, for TEBPβ in ciliates and Rap1 in yeast *in vitro*[204-207]. Additionally, several proteins have been identified facilitating the correct folding of RNA, like members of the hnRNP family[208], which are also found to be able to interact with rG4s.

In contrast, disruption of rG4s has been shown experimentally for several proteins *in vitro* and *in vivo*. The most prominent representatives and most likely further candidates belong to the enzyme class of helicases[202,203,209,210]. These ATP-dependent motor proteins bind and remodel nucleic acids and/or nucleic acid-protein complexes and have essential functions in virtually all aspects of nucleic acid metabolism[211]. Consequently, mutations in or functional deficiencies of helicases is often associated with severe diseases and cancer[203,212].

The human genome encodes 95 helicases of which 64 act preferentially on RNA and 31 on DNA, respectively[213]. Several of them have been shown to resolve or at least bind G-quadruplex structures *in vitro* and/or *in vivo* (see Table 1).

Additionally, the zinc-finger protein CNBP was recently identified to destabilize G4 structures *in vitro* in an ATP-independent fashion presumably by tightly binding to the unfolded motif and shifting the equilibrium towards the unfolded state[214]. This study further contributes to the hypothesis of the Bartel group suggesting a robust and redundant cellular network of rG4-disrupting proteins consisting of helicases and non-ATP-dependent factors[139]. One of the potential key players in this system is the helicase DHX36.

Table 1: Human G-quadruplex helicases[203,209]

Name	Family	Specificity	Involved in
DDX1	DEAD	DNA, RNA	Recombination, Processing
DDX11	DEAD	DNA	Replication, DNA repair
DDX21	DEAD	DNA, RNA	Transcription, Translation
ElF4A	DEAD	RNA	Translation
DHX9	DEAH	DNA, RNA	Transcription, Translation
DHX36	DEAH	DNA, RNA	Telomere, Transcription, Translation,
PIF1	Pif1-like	DNA	Telomere, Replication
RTEL	Rad3/XPD	DNA	Telomere, DNA repair
XPD/XPB	Rad3/XPD	DNA	Transcription, DNA repair
BLM	RecQ-like	DNA	Telomere, Replication, DNA repair
FANCJ	RecQ-like	DNA	Replication, DNA repair
WRN	RecQ-like	DNA	Telomere, Replication, DNA repair
ATRX	Swi/Snf	DNA	Telomere, Replication
DNA2	Upf1-like	DNA	Replication, DNA repair
DDX3X	DEAD	RNA, DNA	Only binding was shown so far
DDX5	DEAD	RNA	Only binding was shown so far
DDX17	DEAD	RNA	Only binding was shown so far

3.3.3 DEAH-box helicase DHX36

DHX36 is an ATP-dependent 3´-5´ helicase of the DEAH-box family which was described in many, sometimes conflicting reports resulting in a so far unclear cellular localization, *in vivo* target specificity, and function: In literature, this protein is also referred to as **R**NA **h**elicase associated with **AU**-rich element (RHAU)[215], due to its initial characterization as an AU-rich binding protein, or **G4 r**esolvase **1** (G4R1)[216], because of its ability to bind and unwind G4 structures *in vitro* and *in vivo*. In addition to these contradictory binding preferences it is not clear whether its preferred target nucleic acid species *in vivo* is DNA or RNA, or what its major cellular localization is. Two isoforms have been detected (Figure 3-6) in HeLa cells of which the full-length protein seems to localize predominantly in the nucleus excluding the nucleoli while the shorter version remains mostly in the cytoplasm[215].

3.3.3.1 Structure

DHX36 consists of two *RecA-like domains* in its core with highly conserved sequence elements; among them a variation (DEIH) of the name-giving DEAH box (= Walker B box), which is responsible for ATP-binding and hydrolysis[213] (Figure 3-6). At its N-terminal end a *glycine-rich element* is followed by a characteristic ´domain´ of this protein, the 13-aa evolutionary conserved *DHX36-specific motif* (DSM)[162,217,218]. This unique motif is necessary but not sufficient for high-affinity G-quadruplex binding by DHX36, as defined by dissociation constants of DHX36 G4-binding of below 10 pM for the full-length helicase and 310 nM for the isolated DSM, respectively[219-221]. Even though the glycine-rich element resembles the before-mentioned 20-aa long RGG motif, present in most G4-binding proteins (see 3.3.2), it seems not to be essential for DHX36-G4 binding[220] but for its recruitment to stress granules[217]. Recently, the first crystal structure of (bovine) DHX36 in complex with a parallel dG4 was published underlining the role of the helical DSM in G4-recognition[218]. Additionally, the previously observed preference of DHX36 for parallel G-quadruplexes[222,223] is likely a result from the steric interference of diagonal and lateral loops with DSM binding whereas other G4-helicases, which are missing this domain, show no such biased interaction behavior. Unfortunately, this crystal structure lacks the glycine-rich element, most probably due to its highly disordered structure and flexibility, resulting in incomplete information about the G4-association of this motif. Of the structural domains in the C-terminal region (*degenerate-winged-helix* (WH), *ratchetlike* (RL), and two *oligonucleotide and oligosaccharide-binding-fold-like* (OB) subdomains (OI and OII) followed by a *C-terminal extension* (CTE)) only the OI assists in G4 binding[218,222]. Parts of the other domains form a positively charged groove which is too narrow to accommodate double-stranded nucleic acid consistent with the requirements for a 3´ ss extension for DHX36 G4-resolving activity[216,223,224].

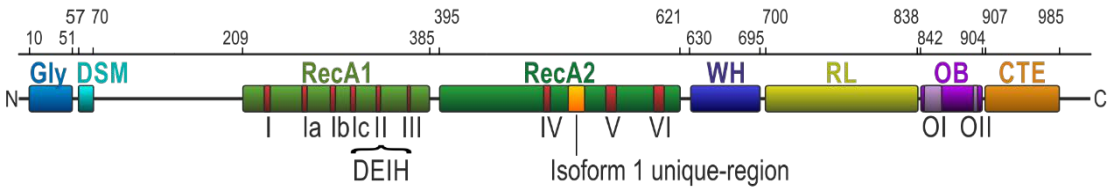

Figure 3-6: Schematic representation of the DHX36 domains.
Individual domains (Glycine-rich element (Gly), DHX36-specific motif (DSM), helicase domains RecA1 and RecA2, degenerate-winged-helix (WH), ratchetlike (RL), oligonucleotide and oligosaccharide-binding-fold-like (OB) subdomains (OI and OII), C-terminal extension (CTE)) are labeled and depicted in different colors. Conserved helicase motifs are indicated (dark red) and the sequence of the Walker B box (II) is indicated (DEIH). The 14 aa-long isoform 1 unique-region is only present in isoform 1. Numbers above show absolute position of domains in the polypeptide[218].

3.3.3.2 Mechanism

DHX36 is described as the major DNA and RNA G4 resolving factor in HeLa cells[221]. Disruption of G4s is presumably its major role as rates for unwinding of other DNA or RNA substrates (duplex, 5´overhang duplex, 3´overhang duplex, Y-form duplex) are up to 1000 times lower[216,225].

In the supposed G4-resolving mechanism, DHX36 initially binds to a 3´ss segment. This interaction induces conformational changes between the DHX36 core domains. Approaching a G4, DHX36 binds it with its DSM and translocates in the 5´direction resulting in disruption of the bottom G-quartet of the adjacent G4[224,226]. The destabilized G-quadruplex can either refold or adapt a resolved conformation. ATP-independent unfolding events occur repetitively by opening of the helicase core and reciprocal rotation of the C-terminal domain. ATP is likely to be only required for release of DNA from the helicase. This mechanism is similar to that of the double strand-specific DEAD-box helicase EIF4A or other members of the DEAH-box family suggesting that G4-quadruplex helicases, such as DHX36, have evolved in order to selectively recognize G4-substrates without modifying their mechanism of unwinding[218,226].

3.3.3.3 Biological relevance

Complete loss of DHX36 in mice is embryonically lethal[227,228]. Tissue-specific deletion of this helicase results in strong developmental defects: DHX36 loss in the hematopoietic system causes hemolytic anemia and differentiation defects at the proerythroblast stage[227]; Deletion in the cardiac system results in defective heart development and growth retardation[228]; Ablation of DHX36 in germ cells causes decreased spermatogonial differentiation[229].

DHX36 is upregulated in most breast cancer cell lines and in more than 30% of human lung cancers deposited on the cBioPortal[230,231]. DHX36 is also upregulated upon PRRSV (porcine reproductive and respirator syndrome virus) infection where it is involved in the DHX36-MyD88-p95 signaling cascade. DHX36 deficient mouse embryonic fibroblasts show reduced IFN-β production in response to long or short poly(I:C)[232] as well as influenza and reovirus infection, respectively. As mentioned before, the genomes of many human pathogenic viruses are enriched in sequences prone to fold G4s[161,233] and it was shown recently, that binding of human nucleolin to a viral rG4 suppresses HCV replication[234]. Therefore, also a connection between DHX36 and viral rG4s during infection is expected.

3.3.3.4 Function

Involvement of DHX36 in cellular mechanisms was demonstrated for telomere maintenance, transcription, pre-mRNA processing, translation, mRNA decay, and in response to stress and viral infection. In most cases, DHX36 was suggested to act by an G4-mediated mechanism. However, all these functions were only demonstrated for individual target molecules.

DHX36 was shown to act in telomere homeostasis by interacting with the telomerase holoenzyme[235]. Binding and resolving the 5´-located G4 of TERC[162,236] (see 3.2.2.4) enables the formation of the P1 helix critical for defining the template boundary for telomer elongation. Consequently, siRNA-knockdown of DHX36 in HEK293 cells resulted in shortened telomeres which will be eventually prolonged by alternative telomerase-independent mechanisms[162].

Furthermore, DHX36 is suggested to control gene expression. DHX36 is supposed to upregulate expression of YY1, a potential transcription factor and tumor suppressor, by binding to dG4s localized in the promoter of this gene[231]. Congruent to this, data obtained by gene arrays of different breast cancer samples indicated a significant positive correlation between DHX36 and YY1 expression.

DHX36 was also shown to be involved in pre-mRNA 3´-end processing which is often altered during cell differentiation, cancer, or viral infection[237]. As an example, DHX36 is necessary to maintain the correct 3´-processing of the p53 pre-mRNA upon UV-induced DNA damage. DHX36 binds and resolves an rG4[238] located within the 3´UTR to provide a binding site for hnRNP H/F which then ensures correct 3´-end processing.

Posttranscriptionally, DHX36 is involved in the regulation of gene expression. For example, DHX36 is required for facilitating translation of the transcription factor NKx2-5 in mice. DHX36 binds and resolves an rG4 located in the 5´UTR of this mRNA to allow the correct initiation of the 43S preinitiation complex. Conversely, DHX36 is described to depress expression of the transcription factor PITX1. This effect seems to be independent of 3´UTR-located RNA G4s in the mRNA although DHX36 is able to bind them *in vitro*. The exact mechanism of this regulation is still elusive.

An early report suggested the involvement of DHX36 in the destabilization of AU-rich element (ARE)-bearing mRNAs[215]. Also, interaction of DHX36 with components of the exosome were described suggesting a role in mRNA degradation.

Additionally, recruitment of DHX36 to stress granules via its N-terminal region has been observed even though the function there is unclear. Related to stress DHX36 is involved in sensing *pathogen-associated molecular patterns* (PAMPs) like dsRNA[239] together with other factors, such as DHX9, DDX1, DDX21, and PKR (EIF2AK2). DHX36 interacts also with signaling proteins like TRIF[239] or RIG-I[232] to support antiviral host defense.

Finally, DHX36 was shown to mediate dendritic localization of the neuronal pre-miRNA 134[240].

Despite all these individual findings, we have only an isolated view on the function and the relevance of the DEAH-box helicase DHX36. What is the preferred target of DHX36 – G-quadruplex structures or AU-rich elements? What are the overall biological consequences of DHX36 binding to and action on its nucleic acid targets? Do functional differences exist between both described isoforms? The biochemical characterization of this helicase and the corresponding resolving mechanism are published but a transcriptome-wide screening for RNA binding targets and its effect on them has not been done, yet. Consequently, little is known to date about the global DHX36 function in the cell.

3.4 Aim of this thesis

The overall aim of this thesis is to shed light on the cellular function of DHX36 *in vivo* with a special focus on its proposed main function, the regulation of G-quadruplex structures.

For this, a transcriptome-wide study in HEK293 cells will be performed together with various biochemical approaches to identify, categorize, and analyze the RNA interactome of DHX36. Based on the obtained results, the involvement of DHX36 in posttranscriptional gene regulation will be investigated as well as its *in vivo* influence on G-quadruplex biology. Possible links between these two research fields should be illustrated and confirmed with experimental data.

To accomplish this, a crosslink-assisted immunoprecipitation approach coupled to next-generation sequencing was performed. Obtained data were analyzed and grouped *in silico* and DHX36 binding preferences and determinants were characterized and new results were compared with published data. Effects of DHX36 binding and action on its targets was further investigated in loss-of-protein and loss-of-function analyses. A special focus was set on target RNAs harboring G-quadruplex forming sequences additionally accompanied with G4-related analyses. Potential models explaining the investigated effects and phenotypes mechanistically were developed and further addressed.

4. Results

Several previous studies have been focused on the *in vitro* characterization of DHX36[216,218,221,226]. It was demonstrated that this protein possesses helicase activity on both DNA and RNA[221] and depends on ATP and a 3´-ss-stretch at the substrate[218]. However, only a handful have also tried to elucidate the function of DHX36 *in vivo* by identifying and investigating distinct cellular target molecules[162,228,231,241] (see also 3.3.3.4). These studies resulted in many conflicting information about the *in vivo* target specificity, binding preferences, and intracellular localization, and consequently about the main function of DHX36 in the cell. Therefore, a systems-wide analysis of DHX36 function is indispensable to complement our knowledge about this special helicase and to understand its role in the cellular regulation of RNA G-quadruplex structures.

Therefore, my PhD project is built up on three consecutive elements: First, identifying the RNA binding targets of DHX36 on a global level. Second, knowing the identity of these target RNAs, the overall effect this helicase cause on them will be addressed. Third, assigning a biological function and mechanism to the observed effects. This study will be contained in *human embryonic kidney cells* (HEK293) with a special focus on RNA G-quadruplex biology.

4.1 DHX36 is a cytoplasmic mRNA-binding protein

As a starting point for my analysis, I used the Flp-In™ System (Thermo Fisher Scientific) to generate stable DHX36 expressing HEK293 cell lines with both isoforms being N-terminally FLAG/HA-tagged (FH-) (Figure 4-1 A).

Total RNA was isolated from HEK293 cells and reverse transcribed into complementary DNA (cDNA). This was used to clone the DHX36 isoform 1 (DHX36-1) gene into the multiple cloning site (MCS) of a recombination site-containing plasmid (pFRT-TO-FLAG-HA)[242]. The DHX36 isoform 2 (DHX36-2) expression plasmid was generated by site-directed mutagenesis through deletion of the 14 aa-long isoform 1-unique region (Figure 4-1 A). The transgenes, which were stably integrated into the genome of HEK293 cells, were under the control of a tetracycline-inducible promoter enabling the regulated expression of the tagged DHX36 isoforms. Quantitative Western blotting revealed that induction with tetracycline for 16 h resulted in approx. 4-fold FH-DHX36 levels compared to endogenous DHX36 levels (Figure 4-1 B). To see, if the increased amount of cellular DHX36 altered general cell characteristics, an analysis of the cellular growth was performed. FH-DHX36 overexpressing cells and the parental wildtype cell line showed no significant changes in growth upon tetracycline induction (Figure 4-1 C) indicating that moderately higher DHX36 levels do not harm the cell.

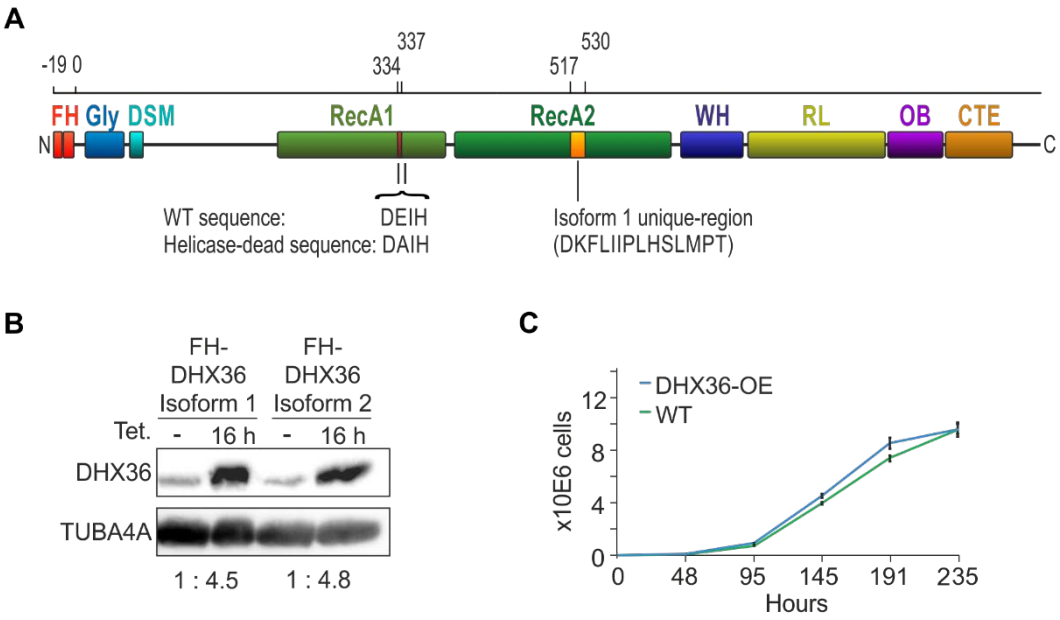

Figure 4-1 Generation of DHX36 overexpressing cells.
A, Schematic representation of DHX36 isoforms generated for expression in HEK293 cells. Individual domains are labeled as described in the legend of Figure 3-6. DHX36 isoforms are N-terminally tagged with FLAG/HA (FH, bright red) polypeptides. Isoform 2 lacks the isoform 1 unique-region (orange). **B,** Western blot analysis of the transgenic cell lines FH-DHX36-1 and FH-DHX36-2 with (16 h) and without (-) tetracycline (Tet.) induction. Normalized DHX36 levels are indicated below. **C,** DHX36-1 overexpression in cells has only marginal effects on cellular growth compared to wildtype (WT) HEK293 cells. Error bars represent standard deviations of three independent experiments.

The major source of nucleic acid in the cytoplasm is RNA, whereas in the nucleus DNA dominates but also different RNA species are present. DHX36 was shown to act on both DNA and RNA *in vitro*. To identify the major nucleic acid binding target of this protein it is therefore reasonable to define its subcellular localization. A previous report described a differential but not exclusive localization of both isoforms in HeLa cells with DHX36-1 being preferentially localized in the nucleus and DHX36-2 in the cytoplasm. Here, overexpression of both isoforms by transient transfection of DHX36-encoding plasmids was followed by immunofluorescence microscopic analyses.

For a more biochemical approach, the two newly-generated stable cell lines were used together with the parental wildtype (WT) cells to perform subcellular fractionation with subsequent Western blot analysis (Figure 4-2 A). In HEK293 cells, both moderately overexpressed DHX36 isoforms predominantly localized to the cytoplasm with only minor signals in the nuclear fractions. More importantly, endogenous DHX36 in WT cells was exclusively detectable in the cytoplasmic fraction. This leads to the conclusion that the major cellular localization of both DHX36 isoforms in HEK293 cells is the cytoplasm.

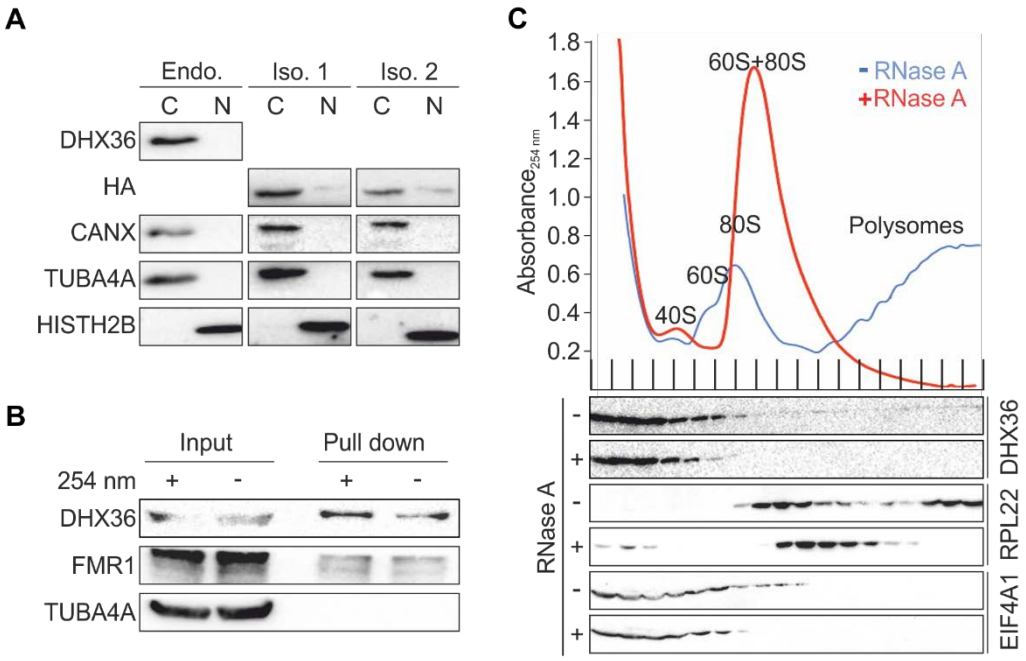

Figure 4-2 DHX36 localizes to the cytoplasm and binds to mRNAs.

A, Endogenous (Endo.) and transgenic FH-DHX36 isoforms 1 (Iso. 1) and 2 (Iso. 2) show mostly cytoplasmic localization in fractionation experiments from HEK293 cells. Cytoplasmic (C) and nuclear (N) fractions were probed with anti-HISTH2B (nuclear marker), anti-TUBA4A (cytoplasmic marker), and anti-CANX (endoplasmic reticulum marker) antibodies. Endogenous DHX36 and transgenic FH-DHX36 isoforms 1 and 2 were detected with anti-DHX36 antibody and anti-HA antibody, respectively. **B,** DHX36 can be detected in an oligo d(T) pulldown of polyadenylated RNA with and without UV crosslinking. The RBP FMR1 serves as a positive, TUBA4A as a negative control, respectively. **C,** UV absorbance at 254 nm of untreated (blue) HEK293 cell extract separated by sucrose gradient centrifugation shows peaks corresponding to 40S, 60S, 80S ribosomes, and polysomes. After RNase A-treatment (red) polysomal signals are missing. Western blots of the gradient fractions probed for DHX36, RPL22, and EIF4A1 are shown below.

Because RNA is the dominant nucleic acid species in the cytoplasm and DHX36 is suggested to act in posttranscriptional regulation it would be reasonable that DHX36 interacts with mRNAs. To analyze this, a pulldown of polyadenylated RNAs was performed to see if DHX36 can be co-purified. Even without formation of covalent bonds by crosslinking the cells with UV-light of a wavelength of 254 nm, a clear signal for DHX36 was detectable when analyzing the bound mRNP fraction by Western blotting (Figure 4-2 B). This result is consistent with two previous reports identifying DHX36 as an mRNA-binding protein in human cultured cells[192,194].

Next, it was consequent to investigate if DHX36 associates with actively translated mRNAs and whether it interacts directly with the translational machinery. In a polysome profiling experiment translation of mRNAs was blocked by the addition of cycloheximide, a potent inhibitor of eukaryotic translation elongation. After lysis, cells extracts were fractionated by ultracentrifugation through a sucrose gradient. UV-profile of the gradient was measured during harvesting, and individual fractions were analyzed by Western

blotting (Figure 4-2 C). This analysis revealed that the main portion of endogenous DHX36 in HEK293 cells migrated in the soluble fractions, however the remainder ~5-10% sedimented in the polysomal fractions, resembling the sedimentation behavior of the translation elongation factor EIF4A1. A similar result was obtained by using the FH-DHX36 overexpressing cells for this approach (see Appendix, Figure 7-1). A direct association to actively translating ribosomes is furthermore unlikely because treatment of extracts with RNase A before ultracentrifugation resulted in the collapse of polysomes and a complete loss of DHX36 from ribosomal fractions. These results indicated that DHX36 does not affect translation elongation but could possibly be involved in translation initiation.

4.2 DHX36 binds defined G-rich sites on thousands of mRNAs

Defining RNA as the major target nucleic acid of DHX36 opens several new questions: Is mRNA the main RNA species this helicase interacts with? Which individual RNAs are bound by DHX36? Is this interaction determined by general characteristics of RNA or by specific RNA recognition elements?

To answer these questions, a transcriptome-wide analysis of DHX36 binding to RNA at nucleotide resolution was accomplished by performing 4-thiouridine (4SU)-assisted **P**hoto**a**ctivatable-**R**ibonucleoside-Enhanced **C**ross**l**inking and **I**mmuno**p**recipitation (PAR-CLIP). In this approach, cultured cells were fed with the photoreactive uridine analog 4SU which was then incorporated into nascent transcripts. Irradiation of the cells with low-energy UV-light of 365 nm wavelength resulted in highly specific, short-ranged crosslinking events between the RNA and bound proteins. Isolation of the crosslinked RNPs by immunoprecipitation, directed against the Flag-tag of the transgenic protein, was followed by radiolabeling with ^{32}P and an additional purification step via gel electrophoresis. Protein components of gel-recovered RNPs were digested and RNA fragments were isolated. After adapter ligation, small RNA cDNA libraries were generated by reverse transcription and eventually sequenced. A characteristic T-to-C conversion at sites of crosslinked 4SU allowed the discrimination between specifically bound RNA fragments and background (Figure 4-3 A).

Variable parameters of this approach (e.g. RNase T1 concentrations) were optimized to and performed in the DHX36-overexpressing cell lines described above (see 4.1). Additionally, two new cell lines were created in which the ATPase activity of the helicase was disrupted by introducing an E-to-A mutation at position 335 in the RecA1 domain (DHX36-E335A)[241]. Consequently, these mutants are expected to bind its targets but remain stuck at sites of DHX36 action resulting in a deeper coverage of sequencing data.

Upon induction with tetracycline, the newly created cell lines were able to express the catalytically dead FH-DHX36 isoform 1 E335A (FH-DHX36-1-E335A) and FH-DHX36 isoform 2 E335A (FH-DHX36-2-E335A), respectively (Figure 4-3 B). The expression levels of the transgenes were comparable (~4-fold) to the before-mentioned transgenic wildtype versions (Figure 4-1 B).

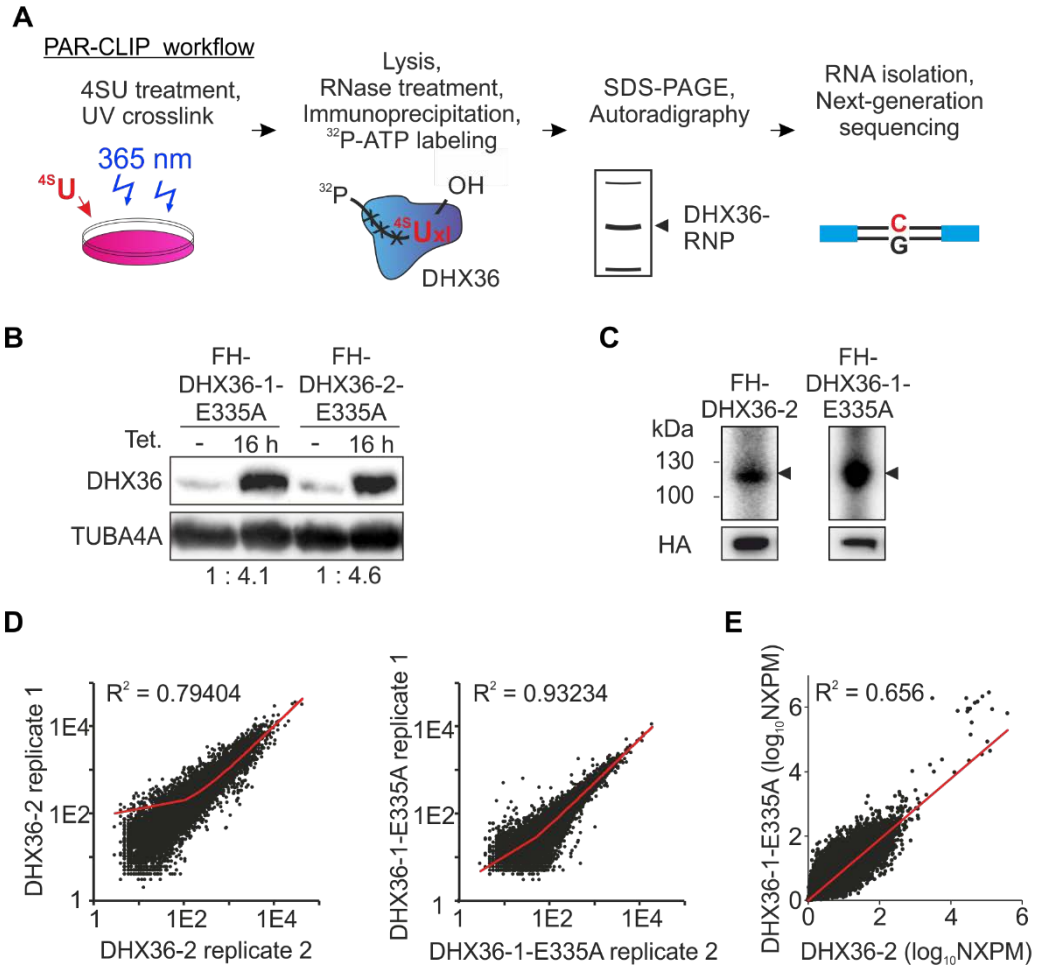

Figure 4-3 PAR-CLIPs of FH-DHX36 forms can be successfully reproduced.
A, Workflow of the performed PAR-CLIP experiments. **B,** Western blot analysis of the transgenic cell lines FH-DHX36-1-E335A (Iso1-E335) and FH-DHX36-2-E335A (Iso2-E335A) with (16 h) and without (-) tetracycline (Tet.) induction. **C,** Autoradiographs and Western blot identification (anti-HA antibody) of crosslinked and radiolabeled FH-DHX36-2 and FH-DHX36-1-E335A separated by SDS-PAGE (black arrowheads). **D,** Correlation of the biological replicates of FH-DHX36-2 (left panel) and FH-DHX36-1-E335A (right panel) PAR-CLIP experiments. **E,** Correlation of normalized crosslinked reads per million (NXPM) from FH-DHX36-2 and FH-DHX36-E335A-1 PAR-CLIP experiments revealed a high degree of correlation between binding targets of both (active and inactive) isoforms.

PAR-CLIP was performed with each of these four cell lines in duplicates. Initial analysis revealed comparable results in all these PAR-CLIP experiments. To cover the whole spectra of available DHX36 forms (different isoforms, helicase activity), FH-DHX36-2 wildtype and FH-DHX36-1-E335A were chosen for in-depth analysis due to their high-quality sequencing results and the great sequencing depth. In the following FH-DHX36 refers to the FH-DHX36-2 wildtype data set and FH-DHX36-E335A to the FH-DHX36-1-E335A data set, respectively.

Autoradiography of the crosslinked, ribonuclease-treated, and radiolabeled FLAG-immunoprecipitate showed the isolation of one main ribonucleoprotein particle (RNP) (Figure 4-3 C). Immunoblotting against HA confirmed the identity of the RNPs as FH-DHX36 and FH-DHX36-E335A, respectively. RNA fragments were isolated from the RNPs and used for generation of small RNA cDNA libraries which were subject of next-generation sequencing.

Using the PARalyzer software, clusters of overlapping reads were determined harboring characteristic T-to-C conversions diagnostic of 4SU-crosslinking events between the RNA and DHX36 at higher frequencies than expected by chance. The two FH-DHX36 replicates showed good correlation (R^2 = 0.79) and the 19,585 clusters which were present in both data sets were defined as high-confidence binding sites (Figure 4-3 D). The larger data sets of FH-DHX36-E335A replicates have an excellent correlation factor of R^2 = 0.93 (Figure 4-3 D) and 67,660 high-confidence binding sites. These results led to the conclusion that DHX36 binds thousands of different sites in the HEK293 transcriptome and DHX36 PAR-CLIPs are highly reproducible. Additionally, Figure 4-3 E showed high correlation (R^2 = 0.66) between DHX36 wildtype (isoform 2) and DHX36-E335 (isoform 1) indicating that no major difference in binding targets exist neither between both isoforms nor because of inactivating the helicase function.

The major RNA target species of DHX36 was mRNA which collected 70% - 75% of all high-confidence binding clusters (Figure 4-4 A). This is in line with the result of Figure 4-2 B verifying mRNA-binding of DHX36. Additionally, only a small fraction of the binding sites fell upon intronic sequences which is consistent with the mainly cytoplasmic localization of this helicase in HEK293 cells (Figure 4-2 A).

In total, DHX36 binding maps to exons of over 4,500 mRNAs (for top 100 targets see Table 3 in Appendix). No enrichment of targets involved in special biological pathways was detected as determined by GO (Gene Ontology Term Enrichment)[243] and KEGG (Kyoto Encyclopedia of Genes and Genomes)[244] analysis. Of the 26 previously reported DHX36 mRNA targets, identified in different cell types, 22 were recovered by my PAR-CLIP experiments in HEK293 cells, showing the validity of the performed approach (Table 2). The non-coding RNA TERC was included, BYCRN1 was not.

Table 2: Overlap of previously identified DHX36 targets with the PAR-CLIP data.

Name	RNA type	DHX36 crosslinked reads	Source
NKx2-5	mRNA	506	228
ELAVL1	mRNA	2301	228
PITX1	mRNA	1541	241
PLAU	mRNA	15	215
DHX9	mRNA	3851	241
NOTCH1	mRNA	68	241
SLC29A1	mRNA	554	241
USP9X	mRNA	2002	241
HTT	mRNA	242	241
PSMD1	mRNA	509	241
RPS3	mRNA	7214	241
PSMD4	mRNA	890	241
RPL23	mRNA	3030	241
P4HA2	mRNA	108	241
UBE2S	mRNA	3510	241
TCEAL8	mRNA	389	241
EI2AK4	mRNA	308	241
USP22	mRNA	3023	241
HAT1	mRNA	719	241
GAPDH	mRNA	480	241
YY1	mRNA	3498	231
TP53	mRNA	641	169
TERC	ncRNA	106	162,236,241
BCYRN1	lncRNA	0	241
IL4	mRNA	0	241
DDX39B	mRNA	0	241
IQSEC2	mRNA	0	241
SLC25A1	mRNA	0	241

In contrast to other proteins on which PAR-CLIP was performed[214,245], DHX36 did not preferentially bind to any of 5´UTR, CDS, or 3´UTR. The distribution of the binding clusters over these functional regions did not dramatically differ from what was mathematically expected by their average length (Figure 4-4 B). Again, no significant differences between DHX36 and DHX36-E335A was observable in this analysis. Despite this equally distributed interaction pattern to regions on mRNAs, DHX36 binding clusters were enriched at the start site of CDSs and directly after the stop codon in 3´UTRs as revealed by a metagene analysis (Figure 4-4 C). A more detailed analysis showed that this enrichment was most prominent within the first 100 nucleotides downstream of the start codon and within the 250 nucleotides following the stop codon (Figure 4-4 D).

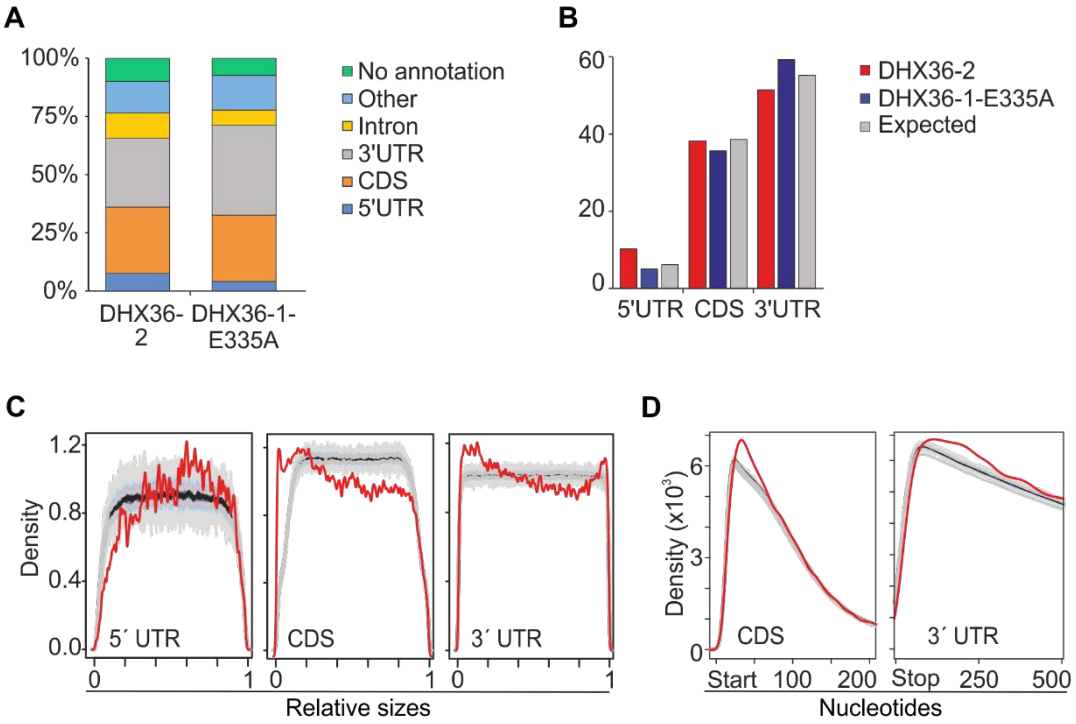

Figure 4-4 DHX36 interacts with mature mRNAs at thousands of sites.
A, Distribution of PAR-CLIP-derived binding sites across different RNA annotation categories. **B,** Distribution of FH-DHX36 (red) and FH-DHX36-E335A (blue) binding sites across 3'UTR, CDS, and 5'UTR matches the distribution expected based on the length of the annotation categories (gray). **C,** DHX36 binds preferentially close to the start and the stop codon as determined by a metagene analysis of the distribution of DHX36 binding clusters on mRNAs subdivided into 5'UTR, CDS, and 3'UTR (red lines). The distribution of 1,000 mismatched randomized controls is shown in gray lines. **D,** Metagene analysis of the distribution of DHX36 binding clusters 200 nt downstream of the start codon and 500 nt downstream of the stop codon, respectively (red). The distribution of 1,000 mismatched randomized controls is shown in gray. The black line indicates the mean of the gray distribution.

Nevertheless, for investigation of the function of DHX36 it is important to understand how DHX36 binding to its targets is determined. Based on the so far presented data, it cannot be excluded that these interactions are not determined at all, but occur randomly, meaning that DHX36 does not have any binding preferences. This explanation would even more fit to the observation, that DHX36 shows no preferences in binding to functional regions of mRNAs (Figure 4-4 B).

Random interactions of an RNP to target RNAs would result in a correlation of the binding to the expression levels or the transcript length of the targets. However, neither FH-DHX36 nor FH-DHX36-E335A PAR-CLIP data correlated with these parameters (Figure 4-5 A, B) suggesting sequence- or structure-specific determinates for DHX36 binding to its targets. Next, it was analyzed whether DHX36 preferentially bound to a specific set of nucleotides. For this, a 5-mer search of all possible nucleotide combinations was

performed in the PAR-CLIP data and their Z-scores over a background of shuffled sequences were calculated. The wildtype FH-DHX36 data set showed clear enrichment of G-rich 5-mers (defined as 5-mers containing at least 3 guanines) (Figure 4-5 C, upper panel). Interestingly, the deeper data set resulting from mutant FH-DHX36-E335A PAR-CLIPs provided an additional minor enrichment of A/U-rich 5-mers (Figure 4-5 C, lower panel).

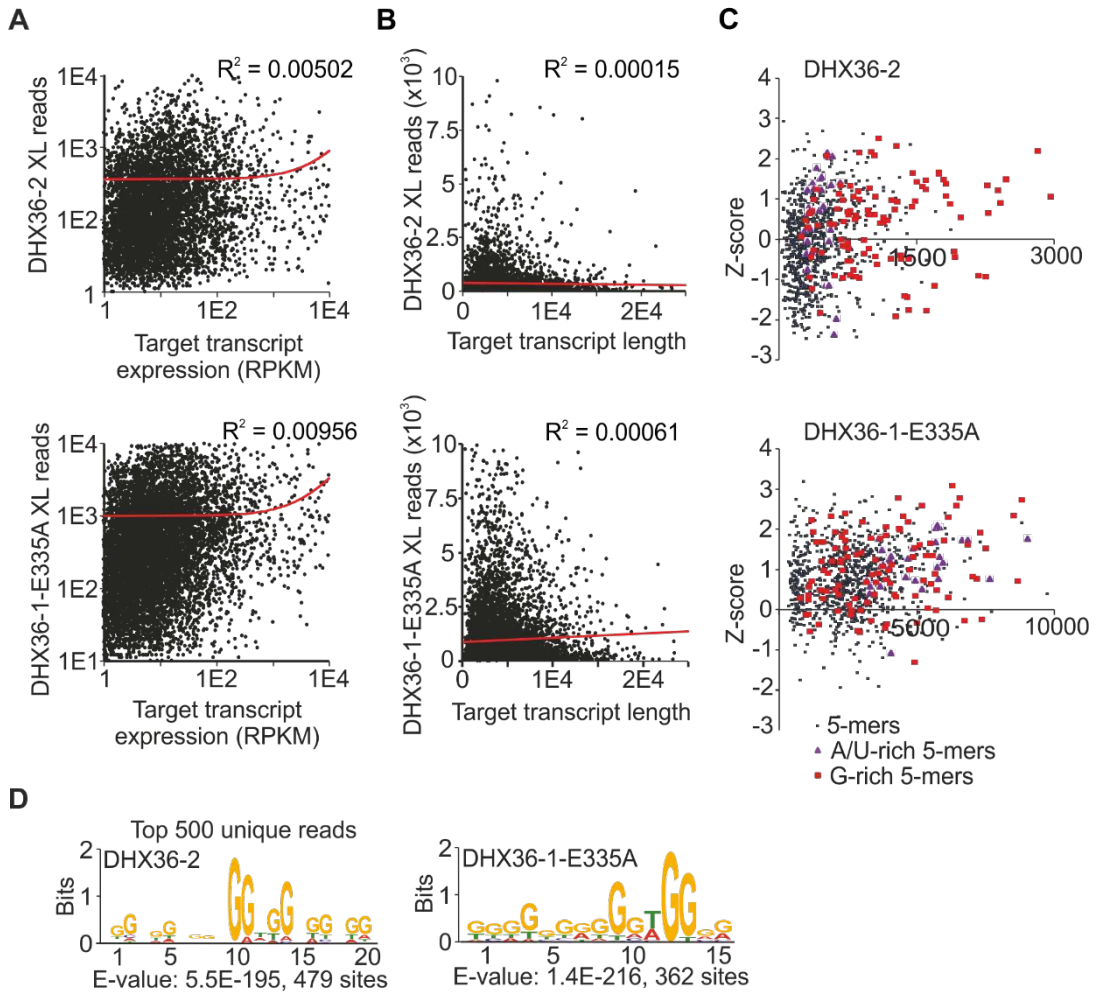

Figure 4-5 DHX36 recognizes G-rich sequence stretches on mRNA.

A, DHX36 binding on target mRNA is not determined by target transcript expression. Correlation between crosslinked reads from FH-DHX36 (top panel) or FH-DHX36-E335A (bottom panel) PAR-CLIPs and target transcript expression is indicated. **B,** DHX36 binding on target mRNA is also not determined by target transcript length. Correlation of DHX36 binding with target transcript length from FH-DHX36 (top panel) or FH-DHX36-E335A (bottom panel) PAR-CLIPs is indicated. **C,** Comparison of Z-scores and occurrence of all possible 5-mers shows an enrichment of G-rich sequences in FH-DHX36 PAR-CLIP binding sites (top panel). For FH-DHX36-E335A PAR-CLIP (bottom panel) an additional minor enrichment of A/U-rich 5-mers can be detected as well. 5-mers containing at least three Gs (red squares) or being A/U-rich (purple triangles) are highlighted. **D,** Representation of the RNA recognition element of FH-DHX36 (left panel) and FH-DHX36-E335A (right panel) PAR-CLIP binding sites generated by MEME using the top 500 unique reads (p-value less than 0.0001).

To accomplish the analysis of sequence-specific binding determinants, a search for a sequence logo, representing the DHX36 binding motif, was performed using the online-available MEME suite tool and the top 500 unique reads of both data sets. As expected from the 5-mer search, highly G-rich binding motifs with consecutive runs of guanines were identified for both DHX36 and DHX36-E335A (Figure 4-5 D) representing the consensus DHX36 RRE.

4.3 Binding sites of DHX36 can form G-quadruplexes *in vitro*

Binding of DHX36 to the consensus sequence motif, which was identified by the MEME analysis on basis of my PAR-CLIP data, was reciprocally tested by microscale thermophoresis. Sigmoidal curve progression obtained by the consensus binding motif indicated a binding event by DHX36 (Figure 4-6 A). Contrary, this progression is not observable for a mutated oligonucleotide.

The DHX36 RRE formally matched the criteria for flexible rG4 formation[246]. Therefore, it is possible that the DHX36 RNA recognition elements fold into a parallel G4 structure *in vitro*. To test this, an *in vitro* G4-folding protocol was performed on synthetic DNA oligonucleotides representing the DHX36 RRE. In addition, mutated versions of these motif (G to A or G to C mutations) which should be incapable of G4 formation were also treated in the same way. Potential G4 formation was explored by circular dichroism (CD) spectroscopy. Indeed, the typical shifts in the spectrum of minima and maxima towards wavelengths characteristic for parallel G4 structures (245 nm and 262 nm)[247] revealed G-quadruplex formation of the DHX36 RRE (Figure 4-6 B, red). The TP-G4 was used as positive control (Figure 4-6 B, gray). In contrast, mutated versions of this motif (shown in blue) do not shift. This allows the conclusion that the consensus DHX36 RRE can fold into a G4 structure *in vitro*.

However, this motif is an artificial product of an online software tool analyzing the PAR-CLIP reads. To see if native DHX36 binding sites can form G-quadruplex structures as well, the same analysis was performed using the sequences of the binding sites of four representative mRNAs, WAC, PURB; NAA50, and SLMO2 (Table 3). Again, variants of these sequences with mutations abolishing G4-formation served as negative control. In all cases, the CD spectra indicated G4 formation for the native DHX36 binding motif (Figure 4-6 C), whereas negative controls failed. These results showed that also *in vivo* DHX36 binding sites are capable to form G4s *in vitro*.

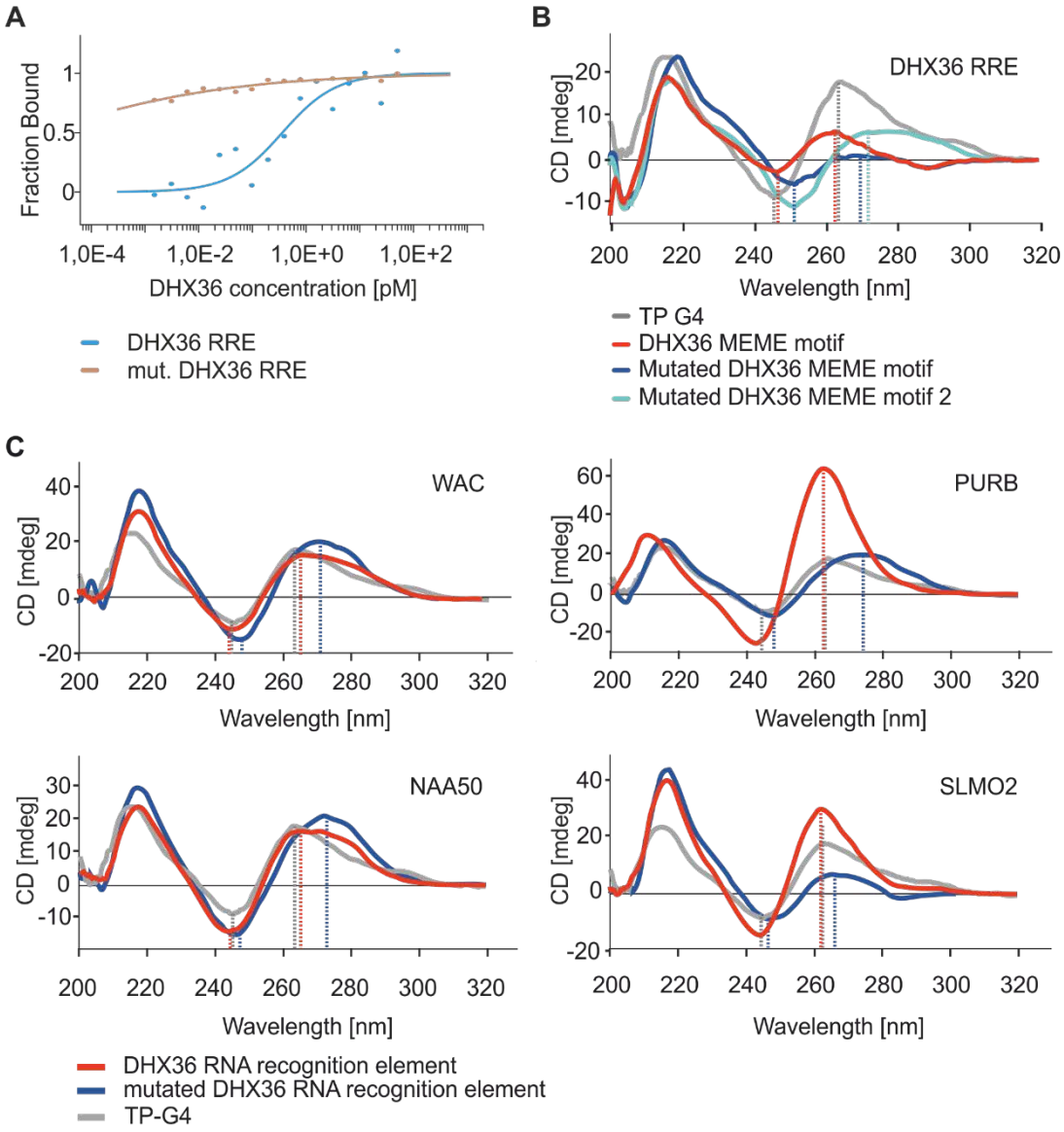

Figure 4-6 DHX36 RNA recognition elements can form G-quadruplexes *in vitro*.

A, Microscale thermophoresis analysis shows binding of DHX36 to the DHX36 RNA recognition element *in vitro* but not to a mutated form. **B,** Circular dichroism spectra of oligonucleotides after performing an *in vitro* G4-folding protocol. The FH-DHX36 PAR-CLIP-derived RRE (red) shifts towards peaks of the positive control TP-G4 (gray), whereas two mutated binding motifs (light and dark blue) do not shift. Lines represent mean of ten subsequent measurements. **C,** Same as B**,** but with representative native DHX36 RREs of the WAC, PURB, NAA50, and SLMO2 mRNAs (red) and mutated versions (blue). Sequences of used oligonucleotides are listed in 6.1.9.

Additional to this proof of G4 folding of DHX36 binding sites *in vitro,* I have investigated if DHX36 was preferentially bound to sites in the human transcriptome that reportedly form rG4s *in vitro*. These sites have been recently identified by rG4-seq, a transcriptome-wide *in vitro* approach on RNA isolated form HeLa cells[138] (also mentioned in 3.2.2.2). With the focus on mRNAs, the rG4-formings site were grouped based on their location in 5´UTRs, CDSs, or 3´UTRs. Next, the overlap of these sites with DHX36 binding clusters were calculated (Figure 4-7 A). In case of the 5´UTR-located rG4 sites, 59% were detected in the FH-DHX36-E335A PAR-CLIP data set whereas the overlap of rG4s in CDSs was 44%. Notably, exceptional 74% of rG4s located in 3´UTRs were recovered by DHX36 PAR-CLIP.

DHX36 binding to these sites is additionally demonstrated with screenshots generated with the Integrative Genomics Viewer software (Figure 4-7 B, C). Here, the gene structure of the two representative target mRNAs WAC and PURB is presented together with the mapping reads of an RNA sequencing experiment in HEK293 cells. Additionally, mapping reads obtained by the PAR-CLIP experiments of both FH-DHX36 and FH-DHX36-E335 are shown and the location of the recently identified rG4s[138] is indicated (orange bars). Sharp peaks in the PAR-CLIP reads indicate distinct binding clusters of DHX36.

In the close-up views, the sequences and the exact overlap of individual binding sites with rG4 sites can be seen. Blue highlighted are uridines in the sequencing reads which have undergone crosslinking events with FH-DHX36 or FH-DHX36-E335A, respectively, during PAR-CLIP. The different range of the scales for both PAR-CLIP sequencing tracks indicates the greater sequencing depth obtained by the mutant protein form. Here, also some extra peaks can be seen representing additional binding sites detected by the usage of FH-DHX36-E335A in PAR-CLIP.

At this point it can be concluded that DHX36 abundantly bound mature mRNAs in the cytoplasm of HEK293 cells with no preference to any of the subregions. However, DHX36 binding was not random but determined by G-rich binding elements able to fold into a G-quadruplex structure. No significant difference between isoform 1 and 2 were detected.

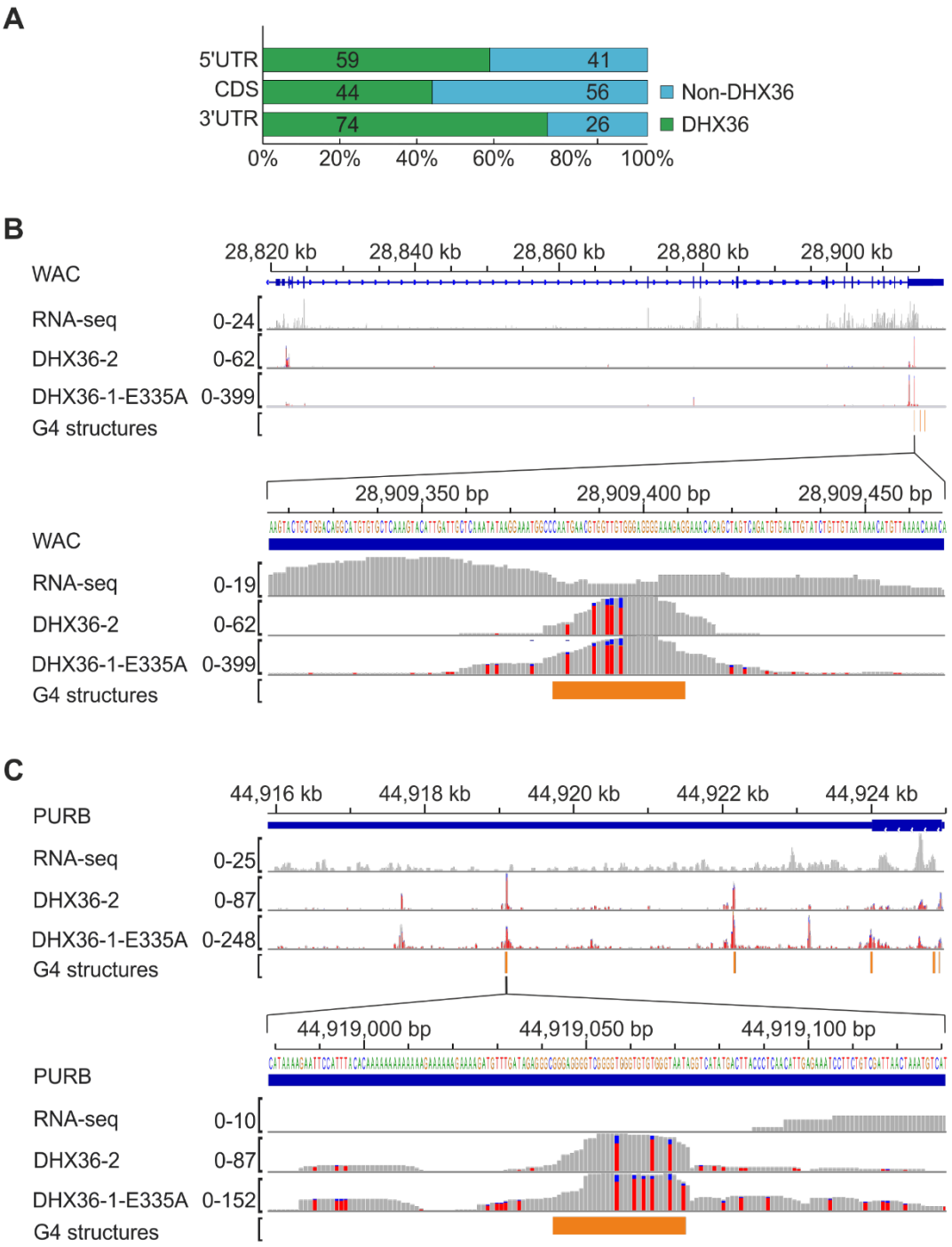

Figure 4-7 DHX36 binds to *in vitro* demonstrated G-quadruplexes.

A, Percentage of sites in the human transcriptome which were previously identified to form rG4s *in vitro*[138] categorized by 5'UTR, CDS, and 3'UTR found in DHX36-1-E335A PAR-CLIP binding sites (green). **B-C,** Top panels: Screenshot of the FH-DHX36 and FH-DHX36-E335A PAR-CLIP binding sites for representative target mRNAs **B,** WAC and **C,** PURB. The gene structure is presented, as well as coverage from a HEK293 RNA-seq experiment. The bottom two tracks show the alignment of sequence reads with characteristic T-to-C mutations from a FH-DHX36 and FH-DHX36-E335A PAR-CLIP experiment. Bottom panels: Close-up of the indicated regions in the 3'UTR of WAC and PURB, respectively. rG4s formed *in vitro*[138] are indicated in orange.

4.4 Loss of DHX36 results in increased target mRNA abundance

After successfully identifying target RNAs, binding preferences, and the RNA recognition element of DHX36, it will be addressed in the following part what function the binding of this helicase to its targets has. Several large-scale loss-of-function studies will be performed to address this question in a global way. Due to previous reports [228,231,241] and my data shown above (Figures 4-2, 4-4) a special focus will be set on the regulation of gene expression and the influence of DHX36 on mRNAs harboring G-quadruplex structures.

As an initial step, DHX36-knockout (KO) HEK293 cells were created using CRISPR/Cas9 technology. For this, a CRISPR RNA (crRNA) was designed to target the Cas9 nuclease from *S. pyogenes* to the DHX36 gene in order to disrupt it. After transfection of the crRNA, a fluorescent tracrRNA, and Cas9 into the parental HEK293 cells, fluorescence-activated single cell sorting by flow cytometry was used to select positively transfected single cells. Next, individual clones were screened for an effective disruption of the DHX36 gene.

Sequencing of the genomic DNA of one positive clone showed an extensive deletion in the DHX36 5´region (Figure 4-8 A). The result of this CRISPR-based gene disruption was the loss of detectable DHX36 protein in these cells confirmed by Western blotting (Figure 4-8 B).

In contrast to moderate DHX36 overexpression, complete knockout of this helicase resulted in obvious phenotypes. Growth rates of DHX36-KO cells were only half as high as those of the parental wildtype cells (Figure 4-8 C) with no concomitant decrease in cell viability (Figure 4-8 D). Additionally, loss of DHX36 caused a changed morphology of the cell population detected by light microscopy (Figure 4-8 E). DHX36-KO cells appeared to be incapable of spreading evenly in the culture dish resulting in the formation of dense cell piles, consisting of cells which grew over each other.

The establishment of those cells and the obvious phenotypes lead to the conclusion that DHX36 is not essential in HEK293 cells, but its deletion has global negative effects for the cell[227].

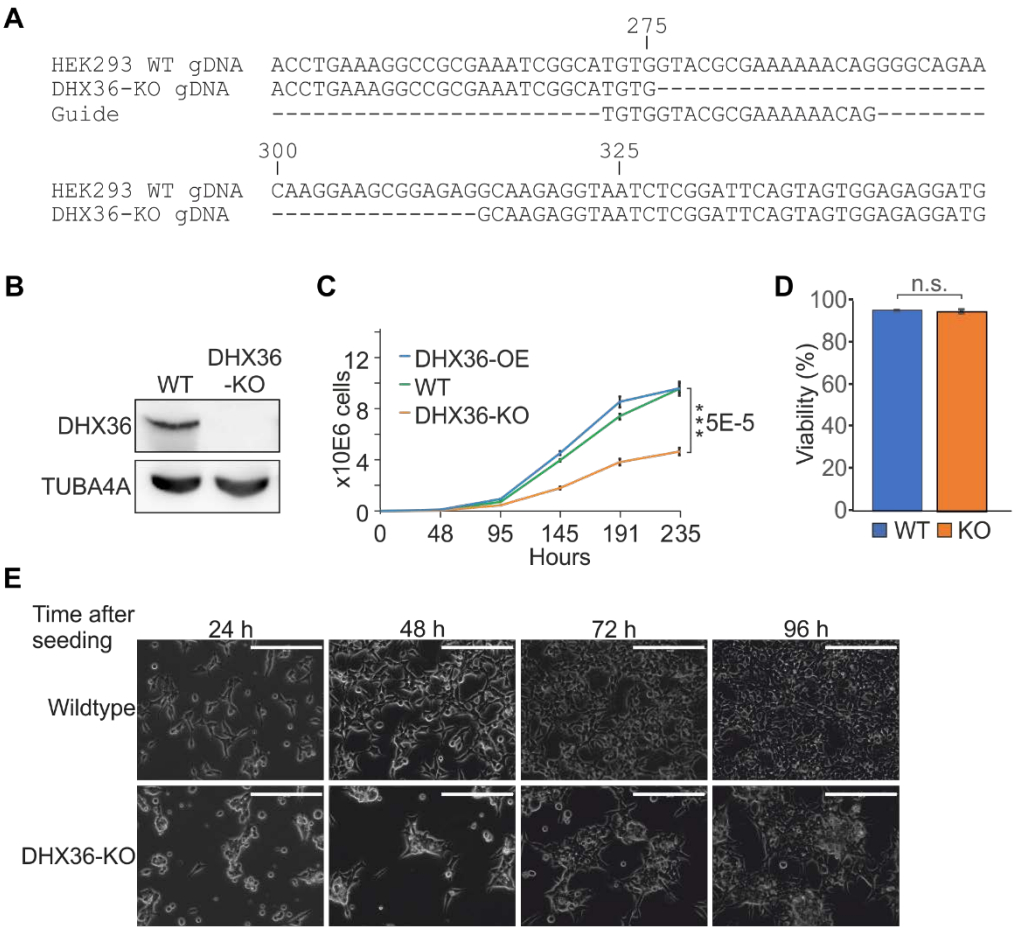

Figure 4-8 Generation of DHX36 knockout cells.
A, Western blot analysis of parental HEK293 cells and the DHX36-KO clone generated by Cas9-mediated gene editing. TUBA4A serves as loading control. **B,** Sequence of genomic DNA of the DHX36 knockout clone shows the disruption of the gene. Sequence of guide RNA (Guide) is indicated. **C,** DHX36-KO cells have a growth defect compared to wildtype and DHX36 overexpression cells. Error bars represent standard deviations of three independent experiments. **D,** DHX36-KO cells do not show a decreased cell viability as compared to wildtype cells. Error bars represent standard deviations of three independent experiments. **E,** DHX36-KO cells have a changed morphology compared to parental wildtype HEK293 cells. KO cells are not able to spread equally over the culture plate surface. Images were taken at indicate timepoints after seeding. Bar represents 200 µm.

RNA helicases can have function on target mRNA stability and decay[211]. To see whether loss of DHX36 resulted in changes of the target RNA abundance, I performed RNA sequencing of the transcriptomes of wildtype and DHX36-KO cells. For this, I isolated total RNA from triplicates of both cell types, depleted highly abundant ribosomal RNA, and generated cDNA libraries for sequencing (Figure 4-9 A).

For analyzing the effect of DHX36 binding to its targets, cellular mRNAs were grouped either by the number of DHX36 binding sites (clusters) or the number of crosslinked reads per target mRNA normalized by overall mRNA abundance (normalized crosslinked reads per million, NXPM) obtained by PAR-CLIP. Both metrics are described to correlate well with the occupancy of an RBP on its targets[245,248].

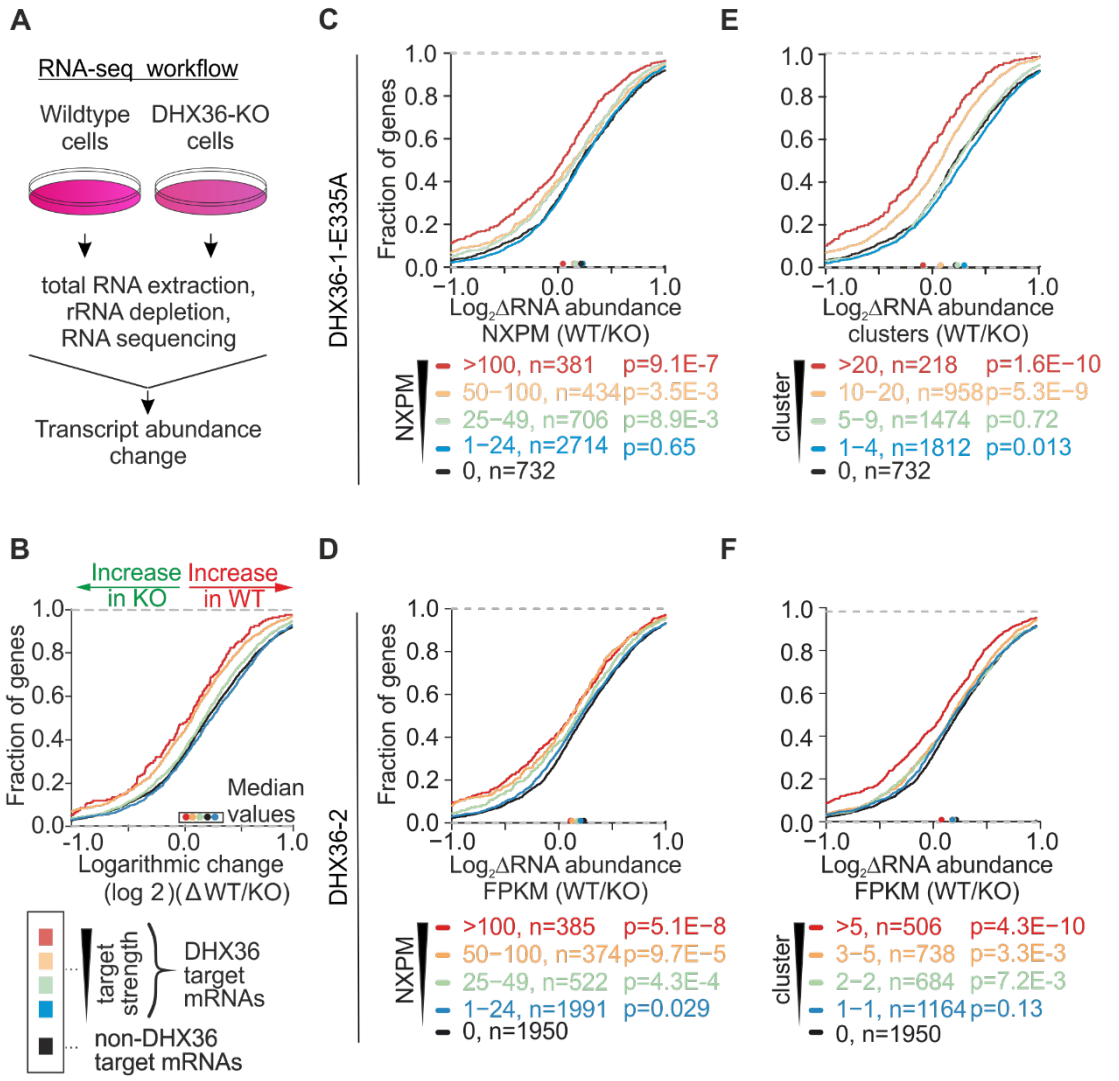

Figure 4-9 DHX36 knockout results in increased target mRNA abundance

A, Workflow of the RNA-seq experiments **B,** Schematic explanation of cumulative distribution functions (CDFs). mRNAs are grouped in “DHX36 non-targets” (black line) and according to the binding strength sorted by colors (red > yellow > green > blue) represented by normalized crosslinked reads per million (NXPM) or the number of binding sites (clusters). Conditions of binning, number (n), and significances are indicated. Fractions of genes are blotted against the logarithmic change of wildtype over DHX36-KO. Shifting of a colored line to the left compared to the non-targets (black) means higher amounts in wildtype compared to the DHX36-KO and shifting to the right means higher amounts in DHX36-KO compared to wildtype, respectively. Colored dots on the x-axis represent median values. **C - F,** DHX36-KO results in an increased target mRNA abundance shown by CDFs comparing changes in target mRNA abundance of DHX36-KO (n=3) and parental HEK293 cells (n=3). Target mRNAs are binned in accordance to the number of NXPM obtained by **C,** DHX36-E335A PAR-CLIP, **D,** DHX36 PAR-CLIP or to the number of clusters obtained by **E,** DHX36-E335A PAR-CLIP, **F,** DHX36 PAR-CLIP. Significance was determined using a two-sided Kolmogorov-Smirnov (KS) test.

Results are presented in cumulative distribution functions (CDFs) which depict the logarithmic change in the abundance of mRNAs in the defined gene groups between wildtype and DHX36-KO cells. Shift of the gene group to the left, relative to non-DHX36 targets, indicated an increase in the knockout compared to wildtype. Accordingly, a shift of the gene groups to the right shows a decrease of the mRNA abundance upon DHX36 deletion (Figure 4-9 B).
Loss of DHX36 led to significantly higher levels of both target mRNAs of wildtype FH-DHX36 and FH-DHX36-E335A. When gene groups were defined by the number of NXPM (Figure 4-9 C, D) the abundance of the top targets was ~15 % higher in KO cells. Moreover, grouping in accordance to the number of binding clusters (Figure 4-9 E, F) resulted in an ~25% increase. In general, the stronger the binding of DHX36 to its targets was (red lines in the CDFs represent the gene group collecting the most NXPM or binding clusters, respectively) the more significant was the effect on the target mRNA abundance.
These data were analyzed in greater detail by looking on the effect of DHX36 binding to the functional regions of mRNAs. For this, mRNAs with binding of DHX36 to the 5´UTR, CDS, or 3´UTR, respectively, were combined and analyzed as mentioned above (Figure 4-10). Interestingly, the general increase of target mRNA levels upon DHX36 knockout seemed to result almost exclusively on binding of the helicase at the UTRs whereas interaction with the CDS has a less strong effect on their abundance. This observation was again independent on whether grouping based on either the FH-DHX36-E335A (Figure 4-10 A-C) or FH-DHX36 wildtype (Figure 4-10 D-F) data and whether NXPM or cluster sites were used to determine binding strength.

Considering the high correlation of both PAR-CLIP data sets, the identification of similar RREs, and the comparable results of the differently grouped RNA-seq analysis, I will focus the functional investigation on mRNA targets obtained by the deeper FH-DHX36-E335A PAR-CLIP data, to maintain a clear representation in this thesis. Nevertheless, comparable results were obtained using the wildtype DHX36 PAR-CLIP data (see Figures 7-2, 7-3, 7-4 in the Appendix section).

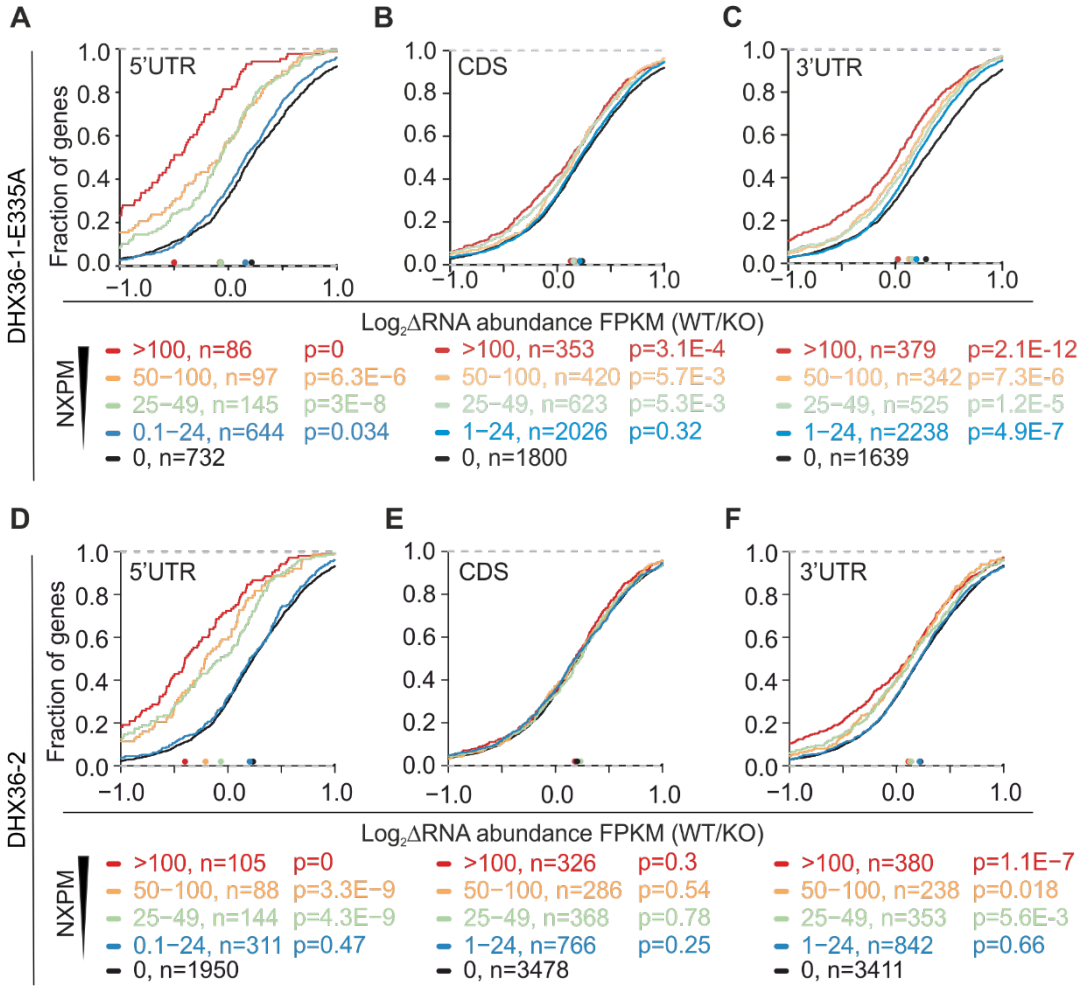

Figure 4-10 Target mRNA abundance is affected by DHX36 binding to UTRs.
A, A CDF comparing changes in target mRNA abundance of DHX36-KO (n=3) and parental HEK293 cells (n=3). Target mRNAs are binned in accordance to the number of NXPM in the 5´UTR obtained by DHX36-E335A PAR-CLIP. **B,** Same as in **A,** except mRNAs are binned based on the number of NXPM in the CDS. **C,** Same as in **A,** except mRNAs are binned based on the number of NXPM in the 3´UTR. **D - F,** Same as in **A - C,** except mRNAs are binned in accordance to the number of NXPM obtained by DHX36-PAR-CLIP. Significance was determined using a two-sided Kolmogorov-Smirnov (KS) test.

4.5 DHX36 decreases target mRNA stability

Two parameters determine the level of an mRNA in the cell: rates of transcription and rates of decay. One publication reported DHX36 to modulate expression of a gene (YY1, see 3.3.3.4) on the transcriptional level by resolving a dG4 in the promoter region[231]. Although, the mainly cytoplasmic localization of the helicase in HEK293 cells, which speaks against a major transcriptional role, it cannot be excluded that increased DHX36 target mRNA levels in knockout cells are, at least partially, due to elevated transcription rates.

To exclude this possibility, chromatin-associated nascent RNA from wildtype and DHX36-KO cells was isolated. As expected, sequencing of this RNA subpopulation and analyzing the obtained data revealed that no changes in transcription rates of target mRNA can be observed upon DHX36 loss (Figure 4-11 A).

Consequently, the only possible reason for the observed abundancy effect is that knocking out DHX36 leads to a concomitant stabilization of its target mRNAs. To test this, a transcriptional shut-off experiment was performed in wildtype and DHX36-KO cells. Total RNA was isolated at specific timepoints (0 h, 2 h, 4 h, and 8 h) after treating the cells with actinomycin D, a potent inhibitor of RNA synthesis. Levels of the four representative target mRNAs WAC, PURB, NAA50, and SLMO2 were quantified by qPCR (Figure 4-11 B). As expected, half-lives of these transcripts were increased upon DHX36 knockout leading to the conclusion that the loss of the helicase stabilizes its target mRNAs.

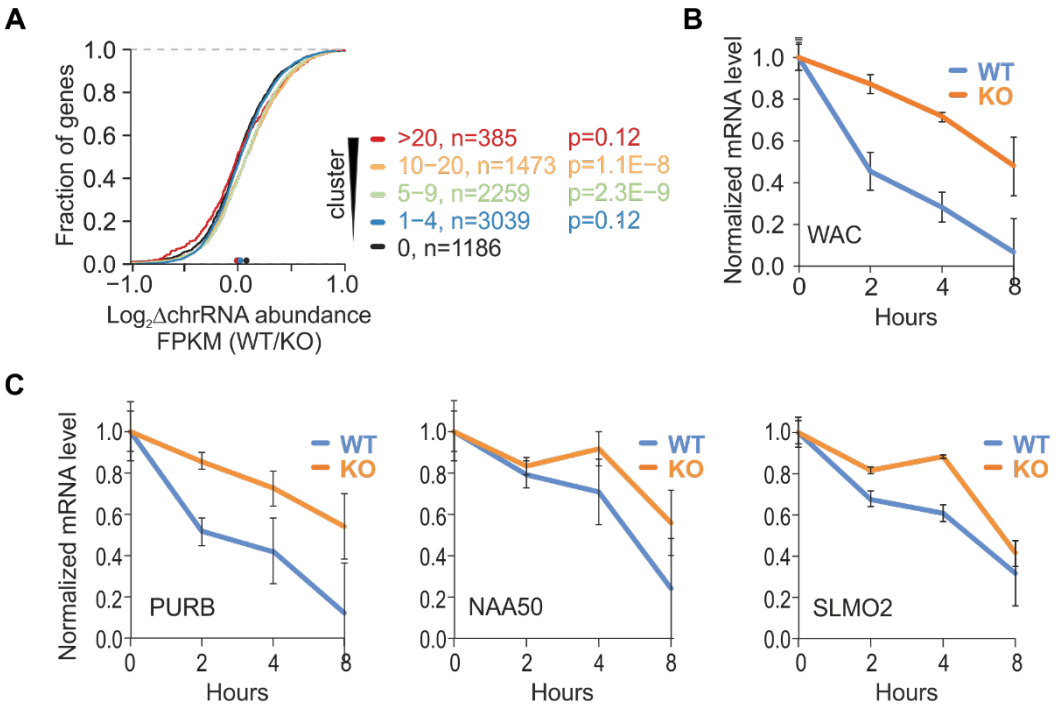

Figure 4-11 DHX36-KO results in increased target mRNA stability.

A, DHX36-KO does not result in higher transcription of target mRNAs compared to wildtype shown by a CDF comparing changes in nascent target mRNA abundance purified from the chromatin of DHX36 knockout cells (n=3) and parental HEK293 cells (n=3). Target mRNAs are binned in accordance to the number of binding clusters obtained by DHX36-E335A PAR-CLIP. Significance was determined using a two-sided KS test. **B,** DHX36 target mRNA WAC shows increased half-life upon DHX36-KO confirmed by qPCR after transcriptional shutoff with actinomycin D. **C,** same as in **B,** but with DHX36 target mRNAs PURB, NAA50, and SLMO2.

4.6 Target mRNA abundance is influenced by DHX36 helicase activity

Next, I investigated whether the helicase activity of DHX36 is necessary to produce the observed effect on target mRNA levels or whether sole binding of the helicase is sufficient. Therefore, a new set of cell lines was created. Using the DHX36-KO cells as paternal cell line, the loss of the helicase was rescued by reintroducing the DHX36 gene via the Flp-In system and the before-mentioned FH-DHX36-1 and FH-DHX36-1-E335A expression plasmids. After selection for positive transfection and construct integration, single clones were picked and analyzed by Western blotting. These rescue cells showed re-expression of FH-DHX36-1 and FH-DHX36-1-E335A, respectively, in the DHX36-KO background, upon induction with tetracycline (Figure 4-12 A).

First of all, I checked if re-introducing DHX36 in the DHX36-KO background is able to rescue the significant proliferation defect of this cell line. Analysis of the proliferation rates of these cells by a colorimetric assay (MTT) revealed also by this method that DHX36 knockout resulted in decreased proliferation rates compared to wildtype cells. However, re-introduction of wildtype DHX36-1 rescued this effect and even slightly exceeded wildtype proliferation rates. Contrarily, DHX36-1-E335A not only failed to rescue the proliferation defect but overexpression of this mutant helicase in DHX36 knockout cells resulted in even greater proliferative defects (Figure 4-12 B).

Next, I tested, if also the effect on target mRNA levels is depending on the helicase function of DHX36. Therefore, I isolated total RNA of both newly-created rescue cell lines and compared their transcriptome by RNA-seq. This analysis revealed that DHX36 target mRNAs were more abundant in the cells expressing the mutant DHX36 version compared to those KO cells rescued with wildtype DHX36 (Figure 4-12 C). Thus, the re-introduced transgenic DHX36 is capable to fulfill the proposed function of wildtype DHX36. Additionally, mRNA levels of the two representative target mRNAs WAC and PURB were analyzed by quantitative PCR in wildtype, DHX36-KO cells, and both rescue cell lines. As expected from the previous results, wildtype HEK293 cells and knockout cells expressing wildtype DHX36 had comparable low WAC and PURB mRNA levels (Figure 4-12 D, blue bars). In contrast, levels of those target mRNAs were elevated in DHX36-KO cells and mutant rescue cells (Figure 4-12 D, orange bars). Concluding from these results, DHX36 needs its helicase function to affect target mRNA levels.

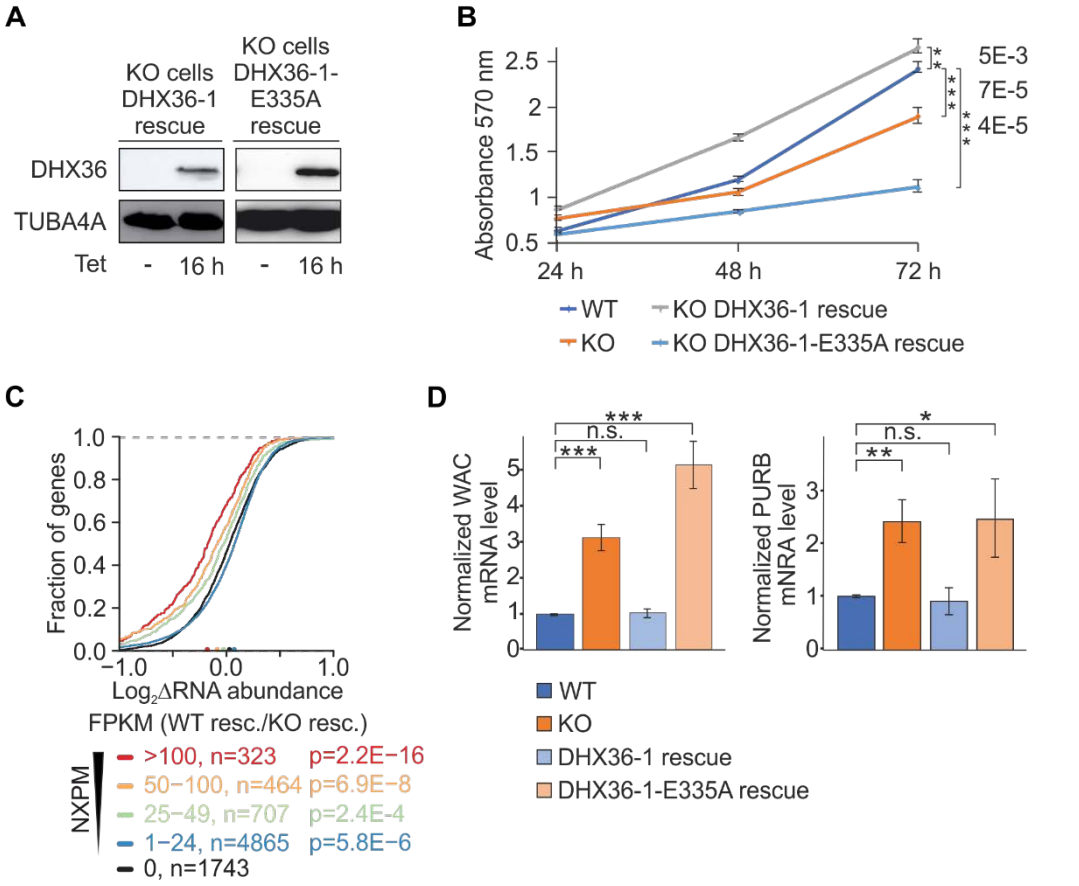

Figure 4-12 Target mRNA abundance depends on DHX36 helicase activity.

A, Western blot analysis of DHX36-KO cells stably expressing FH-DHX36-1 (DHX36-1 rescue) or FH-DHX36-1-E335A (DHX36-1-E335A rescue), respectively, with (16 h) and without (-) tetracycline (Tet.) induction. TUBA4A serves as loading control. **B,** MTT-proliferation assay depicted by blotting the normalized absorbance at 570 nm (maximum of the formazan product) over time (range of 72 h). Overexpression of FH-DHX36-1 in DHX36-KO cells leads to exceeded proliferation rates compared to DHX36-KO cells and wildtype cells whereas overexpression of FH-DHX36-1-E335A in DHX36-KO cells leads to less proliferation. Error bars represent standard deviations of four independent experiments. **C,** Changes in mRNA abundance shown by a cumulative distribution function comparing target mRNA abundance in DHX36-1 rescue cells (n=3) over target mRNA abundance in DHX36-1-E335A rescue cells (n=3). Target mRNAs are binned in accordance to the number of NXPM obtained by DHX36-1-E335A PAR-CLIP. Significance was determined using a two-sided Kolmogorov-Smirnov (KS) test. **D,** Quantification of WAC and PURB mRNA levels in wildtype (WT), DHX36-KO (KO), DHX36-1 rescue, and DHX36-1-E335A rescue by RT-qPCR. mRNA levels are normalized to the level of U6 snRNA. Median WT levels are scaled to 1. Significance was calculated using a Student´s T-Test (n=3). Significance levels: *P < 0.05, **P < 0.01, and ***P < 0.001 compared to normalized WT WAC and PURB level, respectively.

4.7 Loss of DHX36 does not result in more translated protein

An obvious conclusion from the data presented so far is provided by a mechanism in which DHX36 binds to G-rich elements and G-quadruplex-forming sequences on its target mRNAs to destabilize them. As a consequence, DHX36 should act as a negative regulator for the expression of its target mRNAs. If this theory is true DHX36 knockout cells should show, in addition of the higher abundance of target mRNAs, increased levels of proteins encoded by them.

To investigate this, I analyzed the differences in protein levels between wildtype and DHX36-KO cells by **s**table **i**sotope **l**abeling with **a**mino acids in **c**ell culture (SILAC) followed by quantitative proteomics. Both cell lines were cultured in reduced growth media supplemented with light arginine and light lysine. To distinguish between both cell lines during mass spectrometry, growth media were differentially changed to a special growth media containing medium heavy (control wildtype cells) or heavy (DHX36-KO cells) arginine and lysine, respectively. Equal protein amounts of both approaches were mixed and analyzed by mass spectrometry (Figure 4-13 A). Obtained results were combined with the determination of DHX36 binding targets by the DHX36-E335A PAR-CLIP experiment.

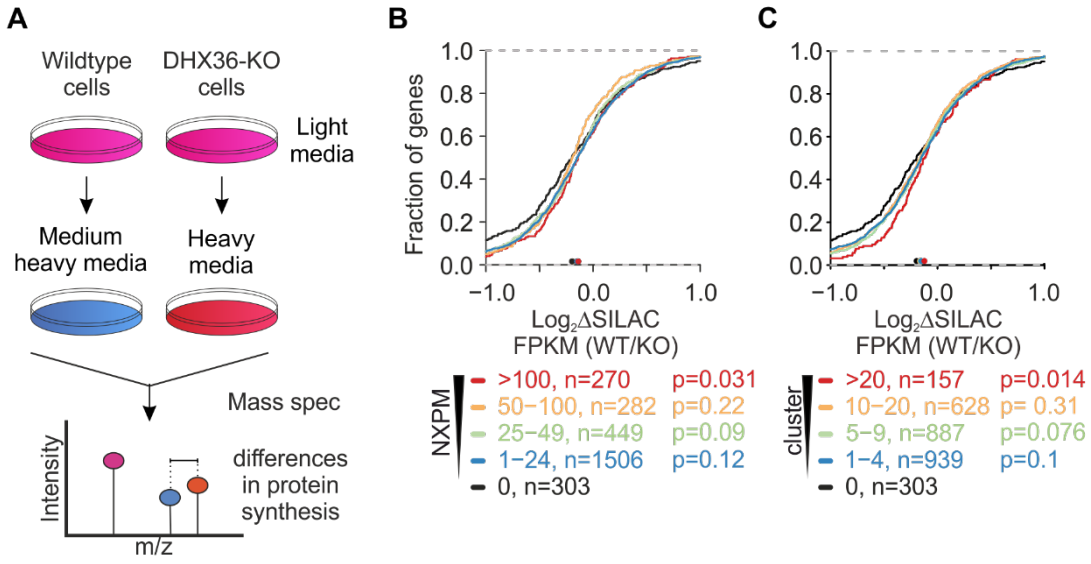

Figure 4-13 Higher target mRNA stability does not result in more translated protein.
A, Workflow of the SILAC experiments. **B,** A CDF comparing changes in protein abundance of DHX36-KO (n=3) and parental HEK293 cells (n=3) determined by SILAC and quantitative proteomics. Target mRNAs are binned in accordance to the number of NXPM obtained by DHX36-E335A PAR-CLIP. **C,** Same as in B, except target mRNAs are binned in accordance to the number of binding clusters obtained by DHX36-E335A PAR-CLIP. Significance was determined using a two-sided KS-test.

Interestingly, higher target mRNA levels in DHX36 knockout cells did not result in globally elevated protein levels from these genes (Figure 4-13 B, C). Quite the contrary, a modest but significant decrease in protein levels from DHX36 top targets upon DHX36 loss was detected (Figure 4-13 B, C; red lines). This leads to the conclusion that in the absence of DHX36 higher target mRNA levels does not lead to higher expression levels of corresponding proteins.

To dissect these paradox phenomena, I determined the coverage of ribosomes on cellular mRNAs in wildtype and DHX36 knockout cells by ribosome footprinting. For this, triplicates of wildtype HEK293 and DHX36-KO cells were treated with cycloheximide, which interferes with the translocation step in protein synthesis. After cell lysis and RNase treatment, stalled ribosomes were purified together with bound and protected mRNA fragments by ultracentrifugation through a sucrose cushion. mRNA fragments were isolated and used for the preparation of small RNA cDNA libraries for next-generation sequencing (Figure 4-14 A). Obtained results were analyzed with the focus on DHX36 target transcripts.

Comparing the recovered ribosome-protected fragments (RPFs) of wildtype and DHX36 knockout cells revealed only minor changes in RPF coverage: Nevertheless, wildtype cells showed overall significantly higher ribosome occupancy on target mRNAs (Figure 4-14 B, C) compared to knockout cells. This can mean that either the overall target mRNA population in DHX36-KO cells is less occupied by ribosomes or that a subpopulation (e.g. the extra mRNAs compared to WT) is by any reason not bound by ribosomes.

However, it has to be mentioned here that this view on the overall changes of RPFs can also be misleading since it is possible that ribosomes are stalling at specific (possibly structured) sites of mRNAs in the knockout cells. This would result in high RPF levels at these stalling sites but concomitantly to overall lower RPF coverage on downstream mRNA sites. This can be investigated by having a closer look on individual genes and their ribosome profiles on distinct sites on their mRNAs. As shown by analysis with the Integrative Genome Viewer for the two representative DHX36 targets WAC and PURB no significant enrichment of RPFs for specific sites on these mRNAs could be observed (Figure 4-14 D, E) whereas a general decrease in ribosome occupancy of these targets in DHX36 knockout cells is more obvious.

This means that, albeit their higher cellular abundance in DHX36 knockout cells, less ribosomes are translating the DHX36 target mRNAs compared to wildtype cells.

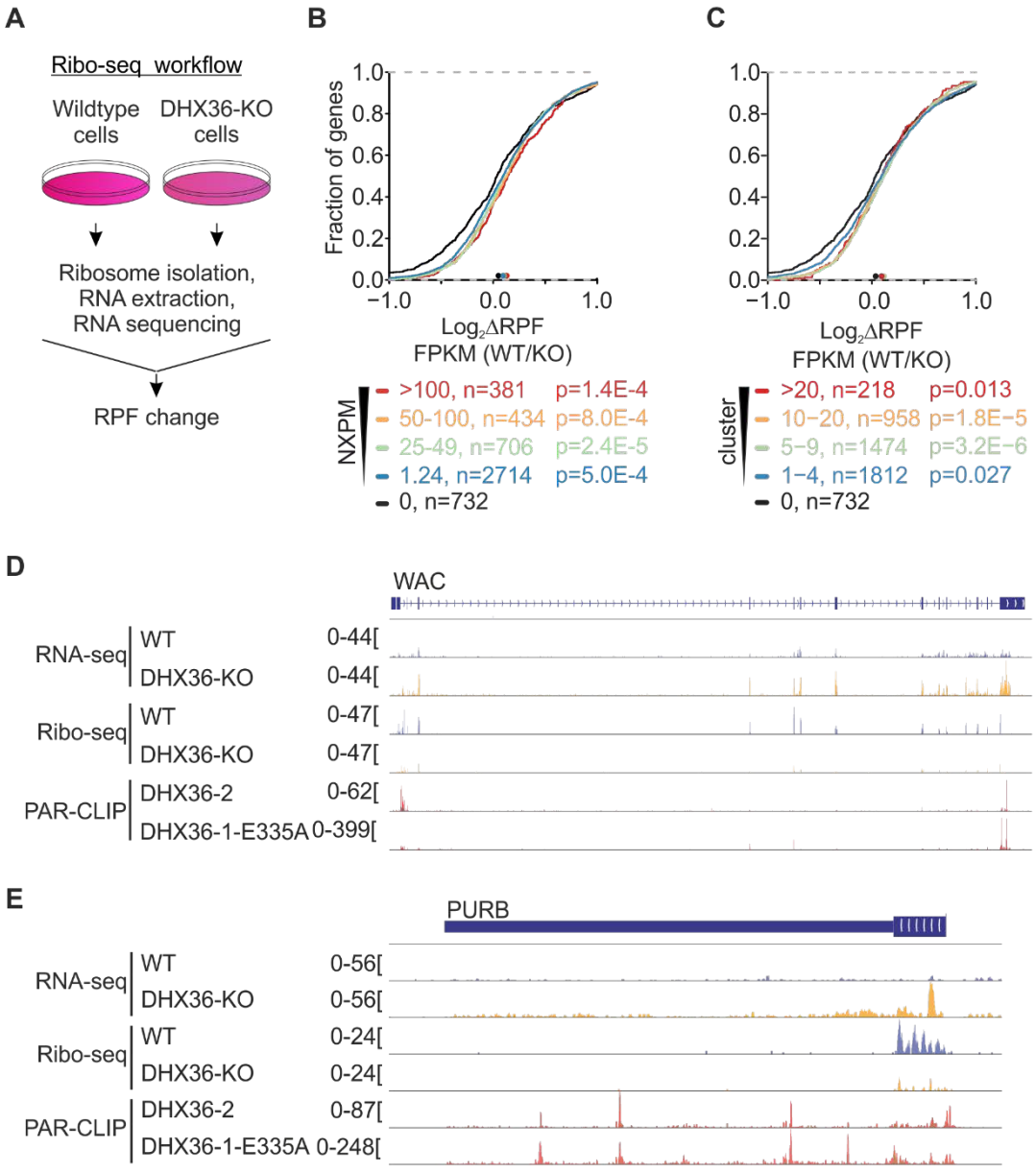

Figure 4-14 RPF levels on target mRNAs are not elevated in DHX36-KO cells.

A, Workflow of the ribosome footprinting experiments. **B,** A CDF comparing changes in ribosome-protected fragments (RPFs) of DHX36-KO (n=3) and parental HEK293 cells (n=3). Target mRNAs are binned in accordance to the number of NXPM obtained by DHX36-E335A PAR-CLIP. Significance was determined using a two-sided KS-test. **C,** Same as in B, except target mRNAs re binned in accordance to the number of binding clusters obtained by DHX36-E335A PAR-CLIP. **D – E,** Screenshots of RNA-seq and ribosome footprinting-seq coverage in wildtype and DHX36-KO HEK293 cells on the representative DHX36 targets **D,** WAC and **E,** PURB. Bottom two tracks show the coverage for the wildtype DHX36 and the DHX36-E335A PAR-CLIP.

As done with the RNA-seq data, these data were further analyzed in greater detail, too, by regrouping DHX36 target mRNAs in accordance to DHX36 binding to the functional regions of an mRNA. For representation of these data, cumulative distribution functions of mRNAs with binding of DHX36 to the 5´UTR, CDS, or 3´UTR, respectively, were generated.

No significant changes between WT and DHX36-KO cells in ribosome occupancy were detected for DHX36 top targets which were bound at 5´- or 3´ UTRs (Figure 4-15 A, C). However, in wildtype cells, a modest but significantly higher amount of RPFs on target mRNAs which were bound by DHX36 in the CDS is detectable.

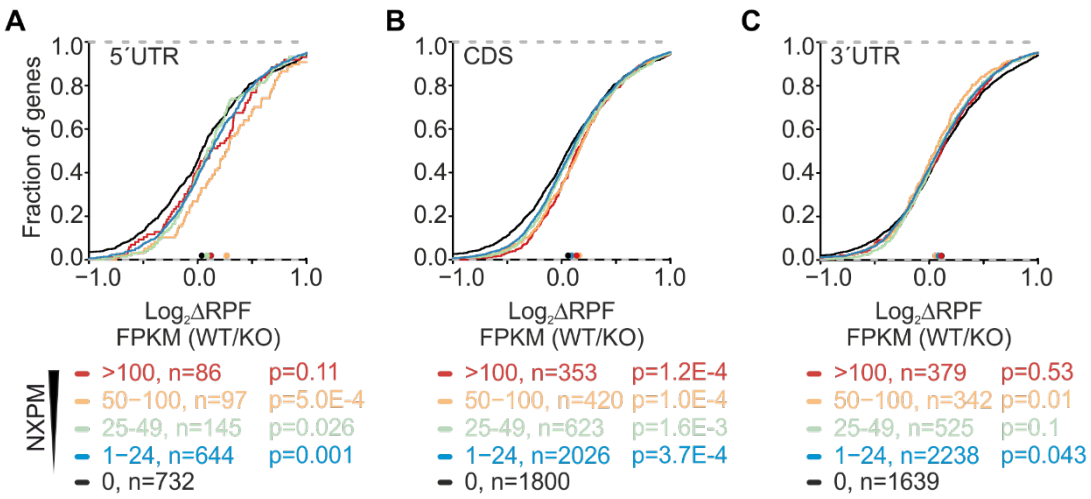

Figure 4-15 RPF levels in target CDSs are modestly lower in DHX36-KO cells.

A - C, A CDF comparing changes in ribosome-protected fragments (RPFs) of DHX36-KO (n=3) and parental HEK293 cells (n=3). Target mRNAs are binned in accordance to the number of NXPM in **A,** the 5´UTR, **B,** the CDS, or **C,** the 3´UTR obtained by DHX36-E335A PAR-CLIP. Significance was determined using a two-sided KS-test.

4.8 DHX36 increases translation efficiency of target mRNA

With the two previous experiments, RNA-seq and ribosome footprinting, it could be demonstrated that loss of DHX36 results 1.) in higher target mRNA abundance and 2.) less ribosome occupancy, at the same time. To combine these two data sets, the average density of ribosomes on each mRNA was calculated. For this, the number of RPFs was normalized with the mRNA abundance in both wildtype and DHX36-KO cells. This resulted in a score known as the translational efficiency (TE) which is a metric to approximate the translational output per mRNA molecule in the cell of a given gene[249].

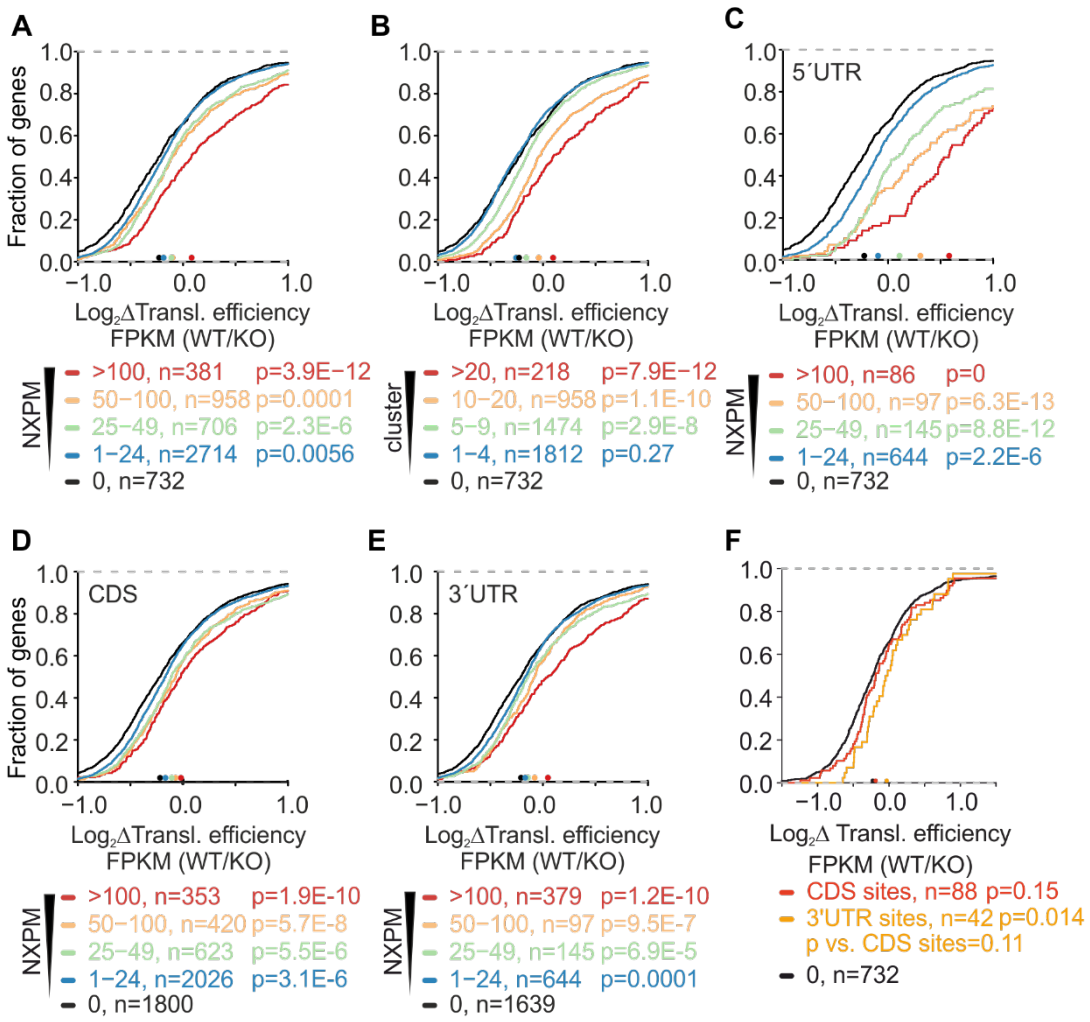

Figure 4-16 DHX36 knockout results in lower translation efficiency of target mRNAs.
A, A CDF comparing changes in translation efficiency (RPF/RNA abundance) of DHX36-KO (n=3) and parental HEK293 cells (n=3). Target mRNAs are binned in accordance to the number of NXPM obtained by DHX36-E335A PAR-CLIP. Significance was determined using a two-sided KS test. **B,** Same as in **A**, except mRNAs were binned based on the number of clusters. **C - E,** Same as in **A**, except mRNAs were binned based on the number of NXPM in **C,** the 5´UTR, **D,** the CDS, and **E,** the 3´UTR. **F,** A CDF comparing changes in TE of DHX36 knockout cells (n=3) and parental HEK293 cells (n=3). DHX36-target mRNAs (determined by binding clusters) are grouped by exclusively coding-sequence bound and 3' UTR bound.

This analysis revealed that loss of DHX36 strongly correlated with a decreased TE of DHX36 targets independent on which parameter the grouping based (Figure 4-16 A, B). For DHX36 top targets (NXPM > 100 with n= 381) a ~27% reduction was calculated. The decrease in target mRNA translation efficiency upon DHX36-KO was more pronounced in targets bound in the UTRs than in the CDS (Figure 4-16 C, D, E, F).

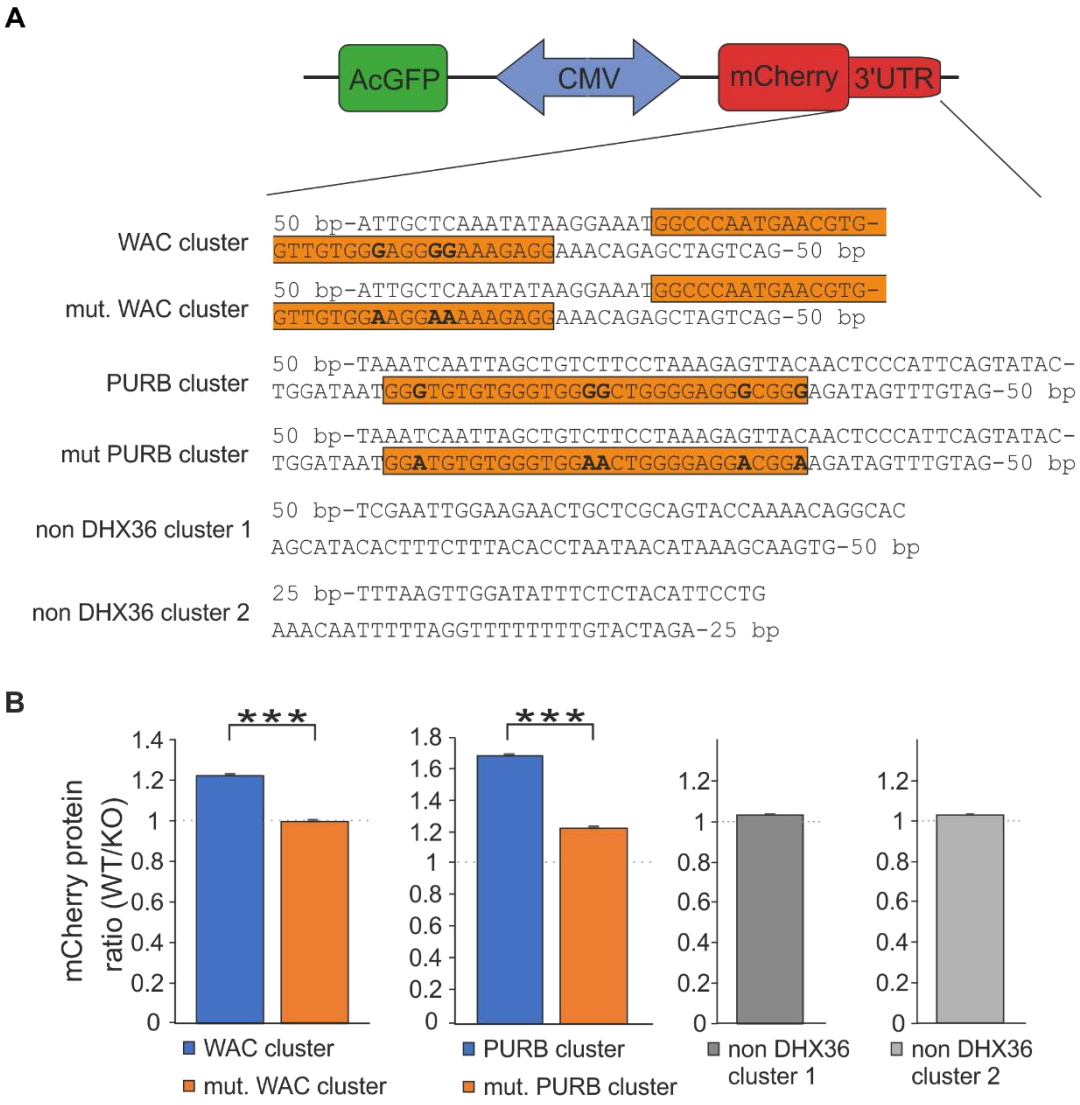

Figure 4-17 Reporter assays with DHX36 binding sites confirm translation effect.
A, Schematic representation of the reporter gene constructs stably integrated in HEK293 wildtype and DHX36-KO cells. Used native WAC and PURB binding clusters (rG4-forming sequence highlighted in orange), mutated forms (G-to-A mutations in bold), and non-DHX36 binding sequences in their natural sequence context (+ 25/50 bp upstream and downstream) are indicated. **B,** WT over KO ratios of the flow cytometry quantification of mCherry protein expressed from a reporter gene construct containing the WAC and the PURB DHX36 binding sites as well as two non-binding sites in the 3´UTR. Median fluorescence intensities of three times 100,000 double positive cells are presented. Significance was calculated using a Student´s T-Test (n=3. Significance levels: *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$.

This systems-wide analysis was finally complemented using a fluorescent reporter system. The rG4-forming DHX36 binding sites of the WAC and PURB mRNAs (see also Figure 4-7 C) were cloned into the 3´UTR of an mCherry gene encoded on the reporter plasmid. In addition, reporter constructs with mutated versions of these motifs, unable to form an rG4 (see also Figure 4-6 C), were generated as well as two negative controls with sequences of the DDX5 mRNA which are not bound by DHX36, as determined by PAR-CLIP) (Figure 4-17 A). These reporter constructs were stably integrated into wildtype HEK293 and DHX36-KO cells. As expected, more mCherry protein was expressed from wildtype clusters in wildtype cells, compared to DHX36-KO cells (ratio >1). This difference was strongly reduced for mutated binding clusters near to the values of the negative controls (ratio ~ 1). By this, the reporter assay with isolated DHX36 binding sites confirmed the systems-wide data that DHX36 assists in the expression of its target mRNAs.

4.9 DHX36 affects translation efficiency of rG4-bearing target mRNAs

DHX36 function was published to be strongly connected to RNA G4 biology[202,203], in particular to the regulation and resolving of these structures *in vivo*. In addition, the analysis of my DHX36 and DHX36-E335A PAR-CLIP data revealed that DHX36 binds to G-rich sequence elements (Figure 4-5) with the potential to form a G4 structure (Figure 4-6). Furthermore, a significant overlap of DHX36 binding sites with previously identified rG4 sites exists (Figure 4-7). Therefore, a detailed view of the influence of DHX36 function on the cellular rG4s is inevitable.

To start with this, the systems-wide sequencing data (PAR-CLIP, RNA-seq, and ribosome footprinting) were re-analyzed and combined with the rG4-seq data of the Balasubramanian group[138] to see which effect loss of DHX36 has on mRNAs harboring a rG4 structure. According to this data set, 2,731 mRNAs were combined to the group of rG4-bearing mRNAs. As a comparison, 2,493 mRNAs were selected harboring no such verified rG4 motif.

As expected, loss of DHX36 stabilized the subgroup of mRNAs, harboring an rG4 motif (Figure 4-18 A). In contrast, no remarkable change was detected in ribosome occupancy of those transcripts (Figure 4-18 B). Resulting from this, rG4-forming mRNAs show significantly reduced translation efficiencies (~17%) upon DHX36 knockout (Figure 4-18 C). These results further support the role of DHX36 in resolving cellular rG4s on mRNAs and, thereby, contributing to their translation.

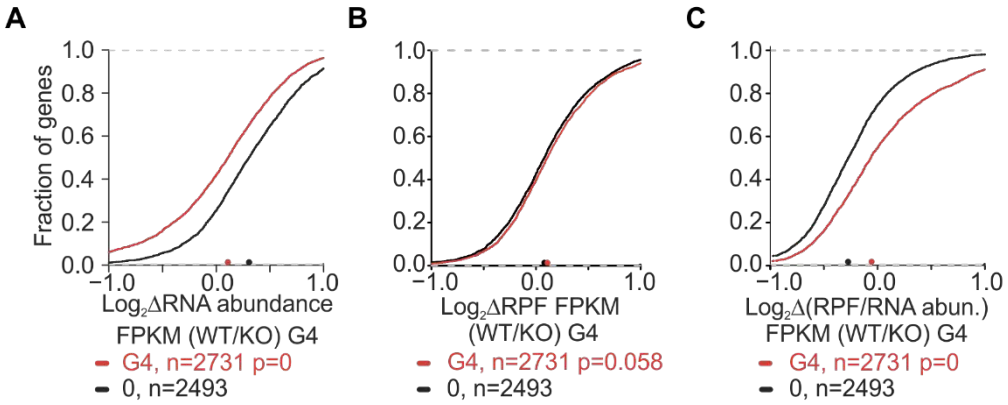

Figure 4-18 DHX36 affects rG4-bearing mRNAs.
A, DHX36-KO results in an increased abundance of rG4-mRNAs. A CDF is shown comparing changes in the levels of mRNAs harboring a previously identified rG4-site[138] between DHX36-KO (n=3) and parental HEK293 cells (n=3). **B,** Same as in **A,** except changes in ribosome-protected fragments were plotted. **C,** Same as in **A,** except changes in translation efficiency were plotted. Significance was determined using a two-sided KS test.

Since DHX36 is described as a G4-resolving helicase, those structures should be stabilized upon loss of this protein. [139]To analyze enrichment of rG4s in the DHX36-KO cells, immunofluorescence microscopy with the G4-binding antibody BG4 was performed, expanding recently published work[151,157]. As already mentioned, BG4 detects both dG4 and rG4 structures which makes it possible to see if DHX36-KO affects none, one, or both kinds of those structures.

A strong BG4 signal from the nucleus, representing dG4, was observable (Figure 4-19 A, B). This signal only marginally changed upon DHX36-KO (Figure 4-19 B). However, the overall weaker cytoplasmic BG4 signal increased by ~1.5 fold (Figure 4-19 A, C). To verify the identity of this cytoplasmic signal as derived from rG4s, wildtype HEK293 cells were treated with carboxypyridostatin (cPDS), a small chemical ligand specifically stabilizing rG4s. In addition, fixed DHX36-KO cells were incubated with RNase A. The staining signal of a protein, located in the cell membrane (ATP1A1), was used for correct identification of the cytoplasm of individual cells. cPDS treatment resulted in even elevated cytoplasmic BG4 signals compared to DHX36-KO cells, whereas increased signal in DHX36-KO cell was sensitive to RNase A. This indicated an accumulation of rG4s upon DHX36 loss corresponding in magnitude with the levels of DHX36 target mRNA stabilization. Concluding from these data, loss of DHX36 resulted in the stabilization of cellular rG4 structures, most likely on the mRNA level.

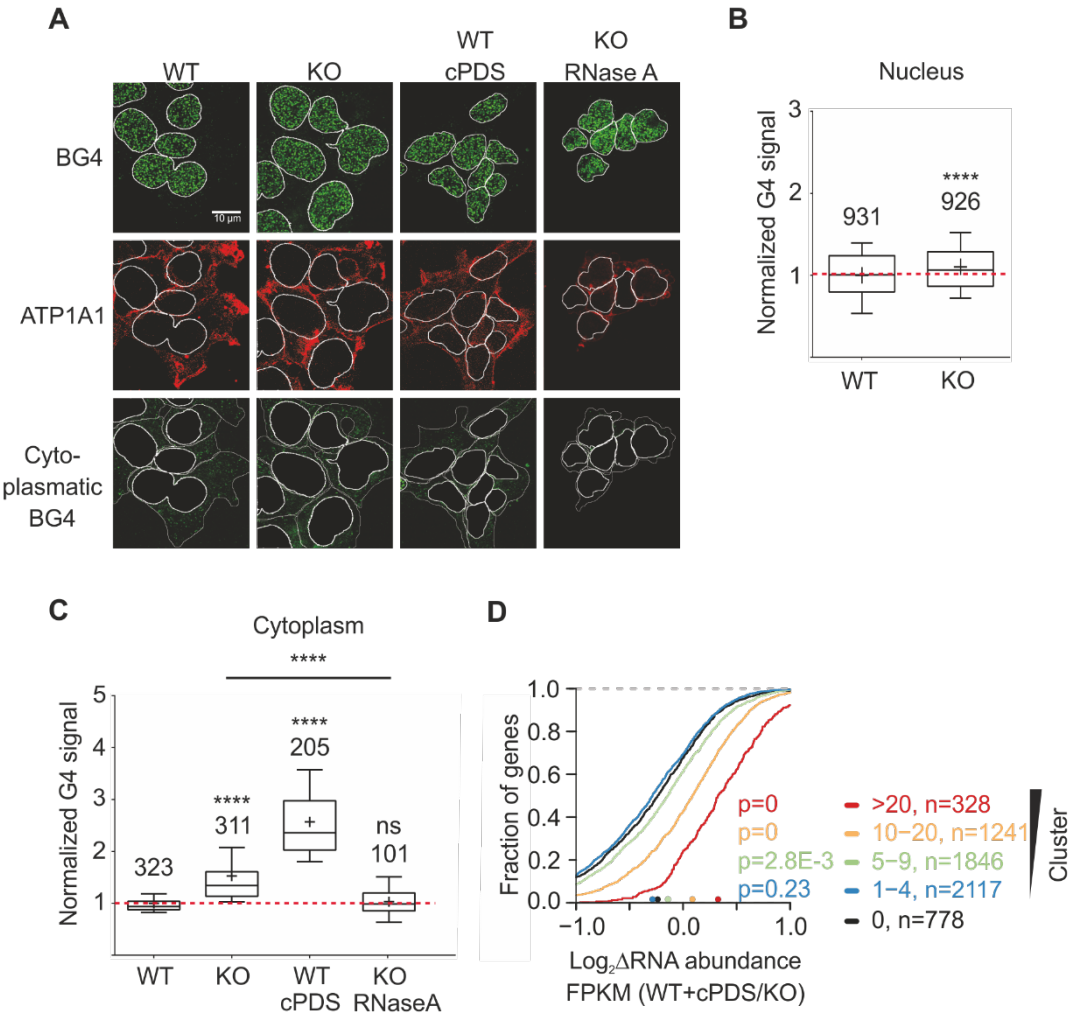

Figure 4-19 DHX36 is involved in the regulation of rG4s in living cells.

A, DHX36 loss results in increased cytoplasmic rG4 signals, revealed by labeling of wildtype and DHX36-KO HEK293 cells with BG4 antibody (green). Anti-alpha 1 Sodium Potassium ATPase (ATP1A1) marks the cytoplasm (red). Panels with masked nuclear signals allow visualization of the increased BG4 signal in the cytoplasm in DHX36-KO cells. Bar represents 10 µm. **B,** Nuclear G4 levels determined by fluorescence intensity of cells normalized over WT. Mean of two biological replicates. Numbers indicate analyzed nuclei. Horizontal lines and plus signs represent median and mean values, respectively. Error bars show the distribution from the 10th to the 90th percentile. Significance was determined using a two-sided KS test (*P<0.05; **P<0.01; ***P<0.001; ****P<0.0001). **C,** Same as in **B**, except cytoplasmic G4 levels are determined. Additional treatment of cell is indicated. Fluorescence intensity of the cytoplasmic compartment was calculated by subtracting the nuclear signal from total cell fluorescence. **D,** Stabilization of rG4 structures using carboxypyridostatin (cPDS) in HEK293 cells results in the accumulation of DHX36-E335A PAR-CLIP targets to a larger degree than DHX36-KO as shown by a CDF comparing changes in target mRNA abundance of DHX36 knockout cells (n=3) and parental HEK293 cells treated with cPDS (n=3). Significance was determined using a two-sided KS test.

Consequently, I speculated that the observed effect of DHX36 knockout on target mRNA abundance was the result of stabilized rG4 structures. To experimentally underline this assumption, RNA sequencing was performed using wildtype HEK293 cells which were treated with cPDS. The results were compared with the transcriptomic data obtained by RNA-seq of DHX36-KO cells (Figure 4-19 D). External stabilization of rG4s by cPDS even exceeded the stabilizing effect of DHX36-KO on its target mRNAs. This provides the conclusion that formed rG4s can contribute to the stabilization of mRNAs harboring those structures. However, this stabilization likely reduces translation efficiency or keep a subset of the mRNA translation-incompetent.

4.10 Connection of DHX36 and rG4s to cellular stress

To explain these observations, a closer look on the biological function of DHX36 was taken reconsidering the global effects of knocking out this helicase. DHX36-KO cells showed reduced proliferation rates and an abnormal morphology, both signs of cellular stress, but no change in cell viability (Figure 4-8). In addition, a direct role of DHX36 during translation elongation can be excluded due to its sedimentation behavior in the polysome gradients (Figure 4-2 C). Also, the in-depth analysis of the RNA-sequencing, ribosome profiling, and reporter gene data (Figures 4-10, 4-15, 4-16, 4-17) demonstrated that DHX36 binding in 3´UTRs is as efficient as binding in 5´UTRs of target mRNAs regarding regulation of translation. Consequently, this leads to the conclusion that reduced translation initiation upon loss of DHX36 is also unlikely to account for the decreased translation efficiency.

Thus, it can be speculated that the reduced translational competence of target mRNAs upon DHX36 knockout is, at least to some extent, due to a potential sequestration of these transcripts into translationally inactive stress granules. Interestingly, DHX36 was demonstrated to be recruited into SGs via interactions with RNA[217].

Consequently, I checked whether DHX36 target mRNAs are prone to localize to SG, too. In a recent study, the group of Roy Parker was able to isolate SGs and analyzed their RNA content by sequencing[250]. They grouped the cellular mRNAs whether they were enriched in SGs, depleted in SGs, or show no significant change compared to total RNA. Comparing the data of the Parker group with the DHX36 PAR-CLIP results revealed that DHX36 target mRNAs were significantly enriched in stress granules (Figure 4-20 A). Redefining this analysis by focusing on the rG4-bearing DHX36 targets showed that these mRNAs are also among those transcripts prone to localize in stress granules (Figure 4-20 B).

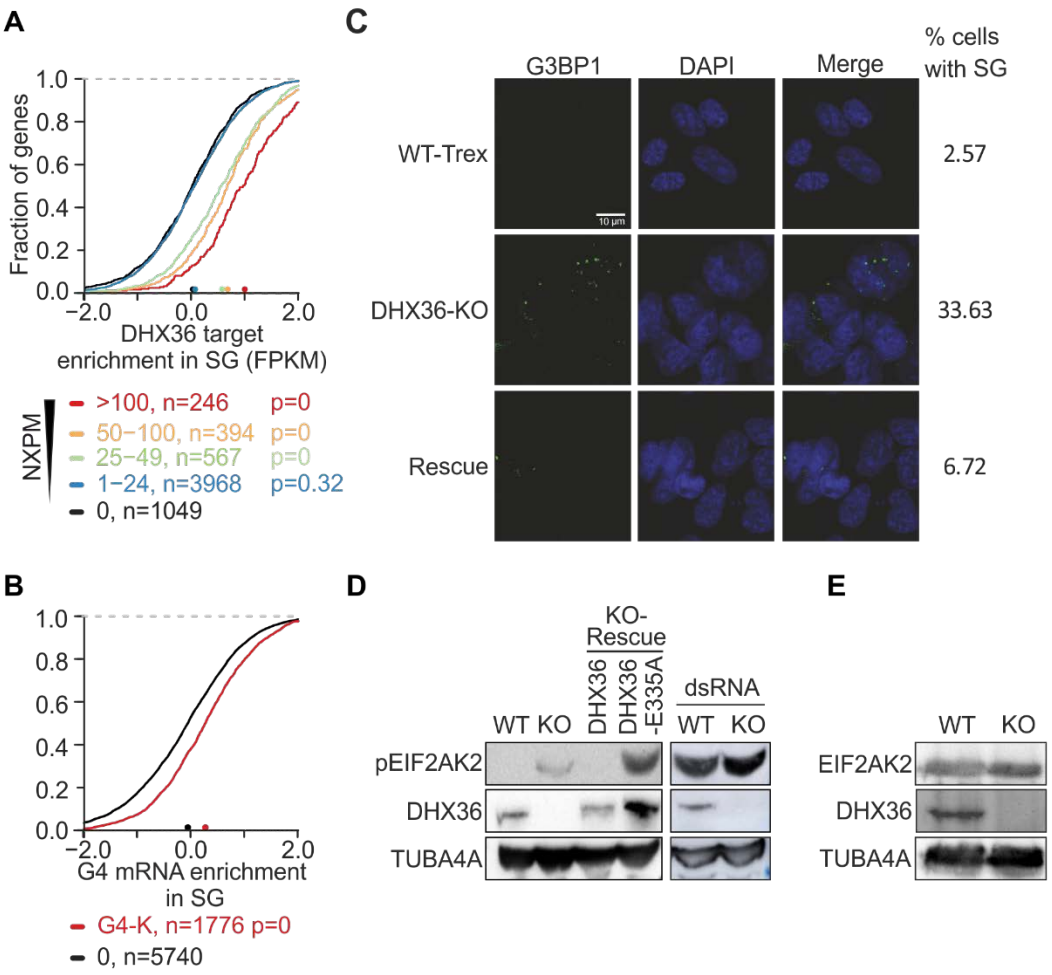

Figure 4-20 DHX36-KO leads to cellular stress.

A, A CDF showing enrichment of DHX36 mRNA targets in stress granules compared to non-targets. Target mRNAs are binned in accordance to the number of NXPM obtained by DHX36-E335A PAR-CLIP. Significance was determined using a two-sided KS-test. **B,** Same as in **A,** except DHX36 target mRNAs are grouped in accordance to whether they are harboring a previously identified rG4 site or not[138]. **C,** DHX36 loss results in increased SG formation, revealed by labeling of stress granules in wildtype, DHX36-KO, and DHX36-KO wildtype rescue cells with a G3BP1 antibody (green). Stress granules were counted and divided by the number of counted cells. Scalebar represents 10 µm. **D,** Western blot analysis of wildtype HEK293 cells (WT), DHX36-KO cells (KO), DHX36-KO cells with transgenic FH-DHX36 expression (DHX36-KO rescue), and DHX36-KO cells with FH-DHX36-E335A expression (DHX36-E335A-KO rescue). Positive control for PKR phosphorylation (pEIF2AK2) by dsRNA transfection is shown. **E,** Western blot analysis of wildtype HEK293 cells (WT) and DHX36-KO cells (KO) shows similar levels of unphosphorylated PKR (EIF2AK2).

To further investigate the connection of DHX36 function and cellular stress, fluorescent microscopy was performed in wildtype HEK293, DHX36-KO, and DHX36-KO wildtype rescue cells using an antibody targeted against the stress granule marker protein G3BP1 (Figure 4-20 C). Truly, more than 33% of the DHX36-KO cells showed clear stress granule formation detectable by microscopy compared to only ~2% wildtype cells. This effect was almost completely rescued by reintroducing the wildtype FH-DHX36 transgene into the DHX36-KO cells.

In addition, the activation of the stress marker *protein kinase A* (PKR, EIF2AK2) was analyzed by Western blotting with an antibody detecting the phosphorylated form of this protein. PKR is involved in detection of nucleic acid-based stress in mammalian cells and was shown to interact with DHX36[232]. Indeed, in contrast to WT HEK293 cells, DHX36-KO cells showed a detectable level of phosphorylated PKR (Figure 4-20 D). This effect was rescuable by expression of the wildtype FH-DHX36 transgene, but not by expression of the helicase mutant form. Quite the contrary, FH-DHX36-E335A expression resulted in even elevated levels of phospho-PKR.

To determine if this was really a sign of stress or just a secondary effect of the DHX36 knockout resulting in overall increased PKR levels and consequently unspecific autophosphorylation, the total levels of unphosphorylated PKR in WT and DHX36-KO cells were analyzed (Figure 4-20 E). Here, no dramatically higher amounts of PKR were observable arguing for a specific, stress-related induction of this pathway.

These data suggested that the increased rG4 formation resulted by DHX36-KO on one hand triggers the cellular stress response, and on the other hand leads to the sequestration and stabilization of rG4 transcripts in SGs.

5. Discussion

Gene expression determines the fate of the cell by influencing how the genotype gives rise to the phenotype. Over 20,000 protein genes are present in the human genome which can express a much higher number of proteins by alternative splicing. Nevertheless, only a fraction of those proteins is expressed to a certain timepoint in a cell dependent on its developmental stage, cell cycle phase, the growth conditions, and extra- or intracellular stimuli. Consequently, accurate and fast regulation of gene expression on the transcriptional and posttranscriptional level is inevitable for the cell and, in multicellular systems, for the whole organism. Its vitally importance becomes even more obvious considering that mis-regulation of involved mechanisms and other defects are the cause of various human diseases including cancer, autoimmunity, neurological disorders, diabetes, cardiovascular diseases, and obesity[251-253]. To fully understand the root cause of underlying diseases the characterization of their molecular mechanism is highly warranted.

Of utmost importance for correct posttranscriptional gene expression is the exact interplay of RNA-binding proteins with RNAs. In cells, RNAs exist as ribonucleoprotein complexes that are composed of one or more RNAs and typically numerous RNA-binding proteins[254]. These interactions serve in a variety of different purposes, including RNA biogenesis, processing, transport, stabilization, and degradation, but also coordination, stabilization, targeting, and control of protein complexes. Many RBP binding events are specific and require a mechanism for recognition of the appropriate binding target. Such a mechanism can be the recognition of the primary sequence of RNA, but also of secondary structure elements and three-dimensional arrangements[254]. In addition, the structure of RNA can be of immense importance, not only acting as a protein binding site but having function in biological processes by itself. Beside of the few classical elements of RNA structure, uncanonical formations exist which are less well studied. Among them is the highly stable RNA G-quadruplex structure which provides both great potential and challenges for biologically relevant functions. Thus, they are in need of exact and efficient regulation.

My thesis provides a systems-wide view on the function of the G4-resolving helicase DHX36 and its role in posttranscriptional gene regulation and RNA G-quadruplex biology. This protein is a mainly cytoplasmic helicase in HEK293 cells directly binding to polyadenylated mRNA whereas interactions with the translational machinery are mainly indirect. Transcriptome-wide determination of the binding behavior of this helicase revealed highly G-rich binding clusters which significantly overlapped with sites able to form rG4 structures *in vitro.* Loss of this helicase resulted in a significant decrease in

translation efficiency of accumulating target mRNAs, especially of those which are bound in the UTRs by the helicase. DHX36 knockout further leads to higher levels of cytoplasmic rG4 structures as well as clear signs of cellular stress.

In the following, the newly obtained data will be compared to state of the scientific knowledge, controversial results in the literature will be discussed, and the relevance of DHX36 in biological phenomena like development, cancer, and infection will be depicted.

5.1 Characterization of DHX36 in HEK293 cells

Several previous reports contributed to the *in vitro* characterization of DHX36 as a helicase able to bind and resolve parallel G-quadruplexes[218,221,222,226]. However, only few studies so far, tried to elucidate the *in vivo* role of DHX36 resulting in less and sometimes conflicting data. It has to be recapitulated that this study was conducted solely in HEK293 cells and it is possible that DHX36 has differing characteristics in other cell types, especially in cancer cells reconsidering its upregulation in certain cancer types.

The initial study about DHX36, here named RHAU, reported two isoforms with differential intracellular localization[215]. According to the authors, full-length DHX36 isoform 1 is predominantly present in the nucleus whereas isoform 2 remains mostly in the cytoplasm. In my study I tried to identify functional differences between both isoforms whereas others only cite this study in cases they mention this topic[225,255]. However, in none of my experiment (Figures 4-2, 4-3, 4-4, 4-5, 4-9, 4-10, 4-14, 4-15, 4-16, 7-2, 7-3, 7-4), I was able to detect a significant difference between both isoforms, thus I do not discriminate between them in respect of their function.

For determination of the cellular localization of the helicase, the initial report used fluorescence microscopy of transiently overexpressed, eGFP-tagged DHX36 isoforms in HeLa cells[215]. In my experiments, I used a more biochemical approach by carefully fractionating HEK293 wildtype and isoform 1 and -2 overexpressing HEK293 cells, respectively, followed by immunoblotting. This experiment revealed that DHX36 is almost exclusively localized in the cytoplasm (Figure 4-2 A) in HEK293 cells. The mainly cytoplasmic localization of this helicase in HEK293 cells is also congruent with its binding to mature mRNAs (Figure 4-4 A), which are mostly present in the cytoplasm. Only in the DHX36 overexpressing HEK293 cells, this helicase can also be slightly detected in the nuclear fraction. However, it cannot be excluded that endogenous DHX36 is present in the nucleus in low levels. It is possible that under certain circumstances, like high DHX36 levels or maybe also under specific stimuli, this protein might translocate partially to the nucleus. For this, previously reported nuclear functions, like telomere homeostasis or transcriptional regulation[162,231], of DHX36 are an argument. The differences in the

observations between the mentioned study[215] from Tran and colleagues and my work might be caused by the usage of different cell lines. Summing up, DHX36 is a mainly cytoplasmic DEAH-box helicase in HEK293 cells.

The major source of nucleic acid in cytoplasm is RNA and an oligo d(T) pull-down showed direct association of DHX36 with polyadenylated mRNAs (Figure 4-2 B). This result is in accordance with two recent mRNA interactome studies in HeLa and HEK293 cells, where DHX36 scored as an RBP binding polyadenylated RNA[192,194]. These data further suggested DHX36 to be involved in translation regulation.

To start the analysis of this potential main function of DHX36, I performed polysome profiling experiments. Here, only a slight sedimentation of DHX36 with polysomes was observable whereas the major portion was found in the lighter fractions of the gradients. This distribution was also recognized for the overexpression DHX36 cell line (Figure 7-1) and resembles the sedimentation behavior of the translation initiation factor EIF4A. However, these results are in contrast with a recently published report performed in HeLa cells[210]. Here, the authors see a stronger association of DHX36 with mono- and polysomes. A possible reason for this discrepancy could be the use of different cell types. In the cancer-derived cells, such as HeLa cells, global translation rates, especially of oncogenic mRNAs, are elevated[256]. The proposed main target site of DHX36, RNA G-quadruplexes, are linked to cancer and oncogenic translation[138,179], for example in T-cell acute lymphoblastic leukaemia. Here, EIF4A is essential for translation of oncogene mRNAs harboring rG4 in their 5´UTRs. Specific inhibition of EIF4A with silvestrol shows powerful therapeutic effects[179]. It could be possible that in some cancer cell types, e.g. HeLa cells, DHX36 is stronger associated to the translational machinery to facilitate fast ribosome progression and high levels of translation.

As presented here, some different results were obtained by analyzing DHX36 in other cell lines (e.g. localization) whereas others are comparable (e.g. mRNA binding protein). Therefore, I assume that also differences in the importance of DHX36 in fast proliferating cells and ´normal´ somatic cells exist. Two observations, made in the Paeschke group, underline this: 1.) although several attempts were made, it was not possible to create a HeLa DHX36-KO cells line, so far, because positively crRNA-transfected cells stop proliferating after few divisions. 2.) DHX36 (protein and mRNA) is upregulated in stem cells whereas these levels decrease during the differentiation to a basal level (Figure 7-5). In addition, conditional DHX36 knockout in mice showed severe developmental defects and early embryonic lethality[227,228].

It would be interesting to deeper analyze this hypothesis about the function of DHX36 in rapidly proliferating cells, to find possible unique functions of this helicase during cancer and development.

5.2 RNA binding of DHX36

One major aim of this thesis was the identification of the RNA binding targets of DHX36. For this, I adapted and performed 4-thiouridine (4SU) PAR-CLIP on DHX36 which allows to pinpoint the interaction sites between an RBP and RNA.

This method is based on crosslinking of the RNA-protein complexes to the photoactivatable uridine analog 4SU, incorporated in the cellular RNA. Therefore, it requires the presence of uridine within a few nucleotides of the binding site[248], otherwise this interaction cannot be captured by 4SU-crosslinking. However, only a very small fraction (~0.4%) of a 32-nt long sequence segment (average length of Solexa sequence reads) are U-less, corresponding to an occurrence of one such segment in every 8 kb of a transcript[248]. It would have been also possible to use 6-thioguanosine (6SG) as photoactivatable ribonucleoside, but two reasons spoke against this modification: 1.) 6SG has a lower crosslinking efficiency than 4SU[248] and 2.) concerns could be raised that the expected G-rich binding motif of DHX36 would be the result of a biased crosslinking with 6SG.

PAR-CLIP was performed not only in FH-DHX36 transgenic cells but also in cells expressing the helicase mutant version FH-DHX36-E335A. This was done to avoid the losing of short-term interactions between the progressive DHX36 helicase and its RNA targets, because the mutant version is not able to dissociate from its place of action. Indeed, the sequencing data obtained with FH-DHX36-E335A were deeper and resulted in the additional identification of the enrichment of A/U-rich 5-mers, beside the expected G-rich 5-mers (Figure 4-5 C). This allows the conclusion that the major site of action of DHX36 are G-rich sequence elements (and associated structures). However, transient and fast interaction with (less structured) AU-rich sites are also possible and may serve, for example, as a loading platform for the helicase.

Formally, it was investigated if DHX36 binding is determined by target transcript length and/or expression. This would argue for a more unspecific binding of DHX36 to its targets[257]. However, no correlation of DHX36 binding to one of these two parameters was observable (Figure 4-5 A, B) pointing to sequence- or structure-specific binding properties.

Consequently, PAR-CLIP data were used to determine a general binding motif of DHX36, which was G-rich, as expected. In addition, this consensus DHX36 RNA recognition element is able to form a parallel G-quadruplex structure *in vitro* (Figure 4-6 B). Native binding sites of four top target mRNAs were also tested by circular dichroism spectrometry and showed the same behavior (Figure 4-6 C). These experiments were performed with DNA oligonucleotides due to practical restrictions, however, this does not

derogate the main conclusion. The critical rationale of this analysis is that the sequence elements, which were identified in living cells, can form parallel G4 *in vitro*. It is well established that the G4 specificity of DHX36 is more determined by the differences between parallel and antiparallel G4s, rather than the identity of the target molecule (RNA or DNA)[218]. DHX36 was also shown to bind and resolve dG4s *in vitro*, too. The dissociation constants for DHX36 interactions with both nucleic acid types are in the low picomolar range (K_D rG4: 39 pM +/- 6 pM, K_D dG4: 77 pM +/- 6 pM)[221]. Furthermore, rG4s are even more stable than dG4s of the same sequence[125].

The rG4-specific binding of DHX36 is also underlined by a comparison of the PAR-CLIP results with data obtained by rG4-seq[138]. Here, a strong overlap of DHX36 binding clusters in HEK293 cells with a large proportion of the *in vitro* identified rG4-forming motifs in HeLa cells could be shown (Figure 4-6).

All this strengthens the conclusion that many of the sequences DHX36 binds *in vivo* can form parallel G4 structures.

5.3 The RNA interactome of DHX36

The main cellular RNA species DHX36 interacts with is mRNA (Figure 4-4 A), but this helicase shows no preferential binding to any of the functional regions of the target mRNAs (Figure 4-4 B). However, metagene analysis revealed an enrichment of DHX36 binding sites within the first 100 nts downstream of the start codon as well as within 250 nts downstream of the stop codon (Figure 4-4 C, D). Interestingly, the CDS binding of DHX36 proximal to the start codon resembles the interaction profile of another cytoplasmic G-rich binding protein, CNBP[214]. This zinc-finger protein increases the translation efficiency of its targets, too, although rG4 motifs are not enriched in CDSs[141].

PAR-CLIP showed DHX36 binding to over 4,500 different mRNAs in HEK293 cells. Among them are 22 of the 26 previously identified DHX36 mRNA targets showing the validity of the performed PAR-CLIP experiments (Table 2). Of the four mRNAs which were not detected in this transcriptome-wide study, one is not expressed in cultured HEK293 cells (IL4) as determined by RNA-seq. The other three transcripts (DDX39B, IQSEC2, SLC25A1) are expressed but do not collect any crosslinked reads suggesting that those are only indirect bound by DHX36 and PAR-CLIP was therefore not able to detect an interaction.

79% of the top 100 target mRNAs harbor an annotated rG4 structure[138] (Table 3). Interestingly, one of the absolute top binding targets is the DHX36 mRNA itself, with over 30,000 crosslinked reads. DHX36 targets its own RNA almost exclusively in the CDS, especially in the extremely G-rich sequence stretch near the start codon, which is coding

for the glycine-rich element (see 3.3.2, 3.3.3.1, and Figure 3-6). In theory, this could represent a negative feedback loop by which DHX36 represses its own translation to maintain the cellular DHX36 level in an optimal physiological range, like it is described for other RBPs[258]. A precondition for this would be that DHX36 binding results in impeded translation of the transcript. However, the presented data in this thesis strongly argue for a positive function of DHX36 in posttranscriptional regulation by increasing its target mRNAs translation efficiency, also by binding to CDSs. Thus, it can be speculated that DHX36 assists in the translation of its own mRNA. This would resemble more an expression cascade than a repressive feedback mechanism. Interestingly, DHX36 is described to have a function in antiviral immunity response[232] and its expression is upregulated upon viral infection[259]. In this case, this mechanism could have an important biological function for the cell in defending against viral infection. Nevertheless, it has to be stated that it is not possible to investigate such a mechanism with the DHX36-KO cells, but it should be manageable to continue research on this topic with the wildtype rescue in comparison to the E335A rescue cells. In addition, DHX36 targets two intronic sites in its own mRNA. This could point to a function in splicing or nuclear export. However, due to the significantly less crosslinked reads, compared to the binding sites in exons, and the preferred localization of DHX36 in the cytoplasm of HEK293 cells, this seems to be at least not the major function of DHX36 binding to its mRNA.

5.4 The function of DHX36 on its target mRNAs

With PAR-CLIP, DHX36 binding sites were identified in a transcriptome-wide manner, consensus binding motifs were delineated, and the global RNA interactome of this helicase was defined. Consequently, the function of DHX36 binding to its target mRNAs was investigated using these data and the DHX36-KO cells obtained by CRISPR/Cas9 technology (Figure 4-8).

5.4.1 Abundance of DHX36 target mRNAs

Standard RNA-seq experiments with parental wildtype and DHX36-KO HEK293 cells were performed to analyze potential changes in the transcriptomic composition of these cells. This analysis revealed an enrichment of DHX36 target mRNAs upon loss of the helicase (Figure 4-9). This effect is the result of an increase in target mRNA stability and not of higher transcription rates (Figure 4-11). Increased target mRNA stability is not caused by a general decrease in general RNA decay, because non-target mRNAs do not show this effect.

A similar effect, increased mRNAs stability without higher transcription rates, was reported for an individual DHX36 target, the NKx2-5 mRNA[228], in mice. Like the initial DHX36 report[215], the authors proposed a model in which DHX36 mediates mRNA decay by association to an AU-rich element on the target mRNA. My data point more to a rG4-mediated direct or indirect stabilization of target mRNAs upon DHX36 loss, since the main interaction site of DHX36 are G-rich elements. In addition, alternative stabilization of rG4s by cPDS shows the same, albeit stronger, effect (Figure 4-19 D) on DHX36 targets.

Detailed analysis further revealed that the RNA abundancy effect is almost exclusively mediated by DHX36 binding to UTRs than to CDSs of its target transcripts (Figure 4-10). This is interesting since rG4 motifs are approx. 4-5-fold enriched in UTRs compared to CDSs[138]. In addition, it was shown that target mRNA abundance depends on the helicase activity of DHX36 (Figure 4-12). Reintroducing wildtype FH-DHX36 in the KO cells rescued the target mRNA effect whereas mutant FH-DHX36-E335A failed and, in fact, intensified it to some extent. This could be due to unfunctional DHX36 bound to unresolved rG4s preventing other helicases and resolving factors to bind to these rG4 structures. Consequently, a compensation for the loss of DHX36 is not possible.

All this is pointing to a mechanism of DHX36 actively resolving G-rich structural elements in UTRs of its target mRNAs which results in a changed transcript abundance.

5.4.2 Translation of DHX36 target mRNA

Quantitative proteomics surprisingly revealed that the higher target mRNA levels did not lead to a concomitant increase in cellular proteins expressed by those transcripts (Figure 4-13). In contrast, levels of the corresponding proteins are modestly, but significantly lower in DHX36-KO cells compared to wildtype cells. This result was congruent with a study on a single target mRNA in mice[228].

To further investigate the molecular background for this, the ribosome density on mRNAs was determined in wildtype and DHX36-KO cells (Figure 4-14). This analysis revealed a slight decrease in ribosome-protected fragments upon DHX36 loss. Contrary to the effect on the mRNA abundance, this is primarily mediated by DHX36 binding to the coding sequences of its target transcripts (Figure 4-15). This could hint at a minor role of this helicase in translation elongation by directly resolving rG4s and G-rich structures in coding sequences to clear the way for the progressing ribosome. However, compared to the much stronger CDS-mediated regulation of translation by CNBP[214], another rG4-destabilizing factor which was characterized in a systems-wide manner, this effect seems to be negligible.

Hence, and because of the clearly observed mRNA abundancy effect (Figure 4-10), a special focus on DHX36 binding to UTRs was set. A recently published study by the Balasubramanian group reported that siRNA-mediated knockdown of DHX36 in HeLa cells resulted in a decreased translation efficiency for a subset of the cellular mRNAs[210]. They reported an increase of RPFs from 5´UTRs and a simultaneous decrease in CDSs, respectively. Interpreting their data, they argued with an enriched translation of uORFs upon DHX36 knockdown resulting in decreased translation of the main ORF. However, such an enrichment in 5´UTR ribosome occupancy is not detectable in my ribosome profiling data (Figures 4-14, 4-15) and this theory can therefore not be confirmed. Furthermore, they do not investigate at all, whether this mRNA population consists of targets of DHX36 or not, like it was done in this thesis (Figure 4-4). Nevertheless, the conflicting data between my work and this study, which is the same one observing a stronger DHX36 association to polysomes, might be reflected by the usage of different cell lines and, consequently, alternative target mRNAs in the HeLa cells (see also 5.1).

My experiments in HEK293 cells showed no prominent co-sedimentation of DHX36 with polysomes and complete loss of DHX36 did not affect the distribution of ribosome densities at either translation initiation, termination, or rG4-forming sites within the CDS (Figures 4-15, 4-18). This pointed to a mechanism in which DHX36 does not directly affect the translation machinery to facilitate translation of its mRNA targets. Nevertheless, loss of DHX36 resulted in a significant decrease in the translation efficiency of a subset of the cellular mRNA (Figure 4-16). This was demonstrated in my work presented here as well as in the conflicting report of the Balasubramanian group[210]. However, in contrast to their study, my data enabled a direct connection of DHX36 binding to its targets, G-rich structures, and translation efficiency. Additionally, my work allowed a close investigation not only of a potential 5´UTR-mediated function of DHX36 but also of binding of the helicase to the 3´UTR.

It is necessary to address this topic as well, because the 3´UTR-mediated effect on translation efficiency was as prominent as the 5´UTR-derived one (Figure 4-16). Moreover, the number of transcripts bound by DHX36 at the 3´UTR is higher than the number of mRNAs with DHX36 binding sites in the 5´UTR.

To further address this, reporter gene constructs (Figure 4-17) were used in which single native DHX36 binding sites were introduced in the 3´UTR of the reporter mRNA construct. DHX36-KO cells were not capable to produce reporter protein in the same amount as the parental wildtype cells did. In contrast to this, introduction of non-target sequence stretches, which show no DHX36 binding in PAR-CLIP at all, displayed similar amounts of mCherry protein in both cell lines. Interestingly, wildtype WAC and PURB

binding clusters were capable to form a rG4 structure (Figure 4-6). Reporter constructs with mutated WAC and PURB binding sites, in which rG4 formation was abolished by G-to-A mutations, showed similar reporter protein levels in WT and DHX36-KO cells. This directly delineated the importance of rG4 structures for this DHX36-mediated mechanism of translational regulation via the 3´UTR.

Searching for the main function of DHX36, a direct effect on translation was excluded, because 1.) affecting translation elongation was unlikely due to poor association with polysomes and 2.) translation initiation would perhaps fit for 5´UTR-bound targets but not for 3´UTR-bound mRNAs showing the same effects.

Considering this, it would be more likely, that loss of DHX36 activity led to the uncontrolled formation of rG4s and other structures in target mRNAs resulting in their sequestration into translationally inactive subcellular compartments, such as stress granules.

5.4.3 Sequestration of DHX36 target mRNAs

All in all, the here-presented data together with published reports in the literature, point to a mechanism in which DHX36 target mRNAs accumulate in SGs upon DHX36 loss where they render translationally incompetent: 1.) DHX36-KO cells exhibited basic levels of cellular stress (Figure 4-20 C, D), an effect which can be reverted by re-expression of transgenic wildtype DHX36. 2.) DHX36 target mRNAs in general and those that form rG4s *in vitro* in particular enrich in SGs (Figure 4-20 A, B). 3.) Upon stress, DHX36 localizes to stress granules[217].

The fact, that only ~33% of the DHX36-KO cells showed stress granules is not surprising, considering that SG formation is preceded by seeding of submicroscopic granules. Also, this might mean that loss of this helicase resulted in a more basic kind of stress than a complete outburst. An uninduced full stress response in DHX36-KO cells, with shutdown of translation and the localization of a large proportion of the cellular mRNA in SGs, was also not expected, as such a permanently highly-stressed cell line would hardly be viable.

The theory of the basic stress in DHX36-KO cells is also congruent with the moderate level of PKR phosphorylation which was detectable in these cells but not as extreme as in the dsRNA-transfected positive control (Figure 4-20). Interestingly, reintroducing mutant DHX36-E335A in the system resulted in elevated phospho-PKR levels compared to the parental DHX36-KO cells. An explanation for this might be that unfunctional DHX36 bound to unresolved rG4s was even more problematic for the cell than solely unregulated rG4s because other helicases and resolving factors could not target these rG4 anymore and were not able to compensate for the loss of DHX36.

It can be speculated that the molecular cause of the stress response in DHX36-KO cells was the accumulation of unfavorable, unresolved rG4s. The cell recognized these and sequestered the rG4-bearing mRNAs in SGs. Putting the uncontrolled presence of rG4s in the cell into a greater context, this mechanism might also be linked to a role in innate immunity. Three aspects allow this speculation: 1.) a number of DExD/H-box helicases act as sensors for foreign RNA triggering antiviral stress response[239], 2.) DHX36 enhances RIG-I signaling by facilitating PKR-mediated antiviral stress granule formation[232], 3.) viruses frequently contain G4s in their genomes[161]. This opens a room for the hypothesis that higher cellular rG4 levels might contribute to the recognition of these pathogens by the cell and DHX36 plays a pivotal role in this mechanism.

It would be interesting to see how DHX36-KO cells react to infection with different viruses as well as to different kinds of additional stresses. Also, it would be useful to see if *in vitro* transcribed and folded rG4s alone are able to induce a comparable stress or even immune reaction. It would be also worth to investigate if DHX36 has also function in clearance of SGs. Future experiments should further try to broaden our knowledge in this issue by investigating whether DHX36-KO cells have a defect in SG clearance and what effect external stabilization of rG4s has.

In addition to this, standard degradation of rG4-bearing mRNAs might also be problematic in a DHX36-negative background. To my knowledge, few is known about the activity of ribonucleases on rG4s and whether they are capable to cut these structures without the help of helicases. However, at least for FEN1, EXO1, and DNA2, nuclease activity on dG4s is described[260]. Other (d)G4 helicases, such as Pif1, are known to interact with nucleases[261], but so far little is known about nucleases as DHX36 interaction partner. Data of a preliminary co-immunoprecipitation experiment coupled to mass spectrometry indicated no such interaction whereas the initial DHX36 report suggested an association of this helicase with the exosome[215]. It would be interesting to deeper analyze this subject and investigate if DHX36 interacts under special conditions, such as stress, with ribonucleases.

All in all, it would be worth to further investigate the role of DHX36 in cellular stress in general, and in response to viral infection in particular.

5.5 DHX36 and its role in rG4 biology

As already mentioned in the introductory part, a special focus of this thesis was DHX36 and its role in the biology of RNA G-quadruplex structures. Albeit some aspects have already been discussed in the part about the function of this helicase on its target mRNAs (see 5.4), it is still worth and necessary to arrange the current results in the scientific field of rG4 structures.

It was recently proposed that rG4s are globally unfolded in eukaryotic cells[139], however some conceptual and technical limitations of this study have already been addressed in the introduction (see 3.2.2.3). In addition, others and myself have clearly demonstrated rG4-mediated effects in human cells[159,160,162-164,179,210]. To my knowledge this doctoral thesis represents the first study in which the enrichment of rG4 signal upon the knockout of a critical resolving factor was demonstrated in the most direct way (Figure 4-19). Considering that Guo and Bartel were not able to detect cellular rG4s or an increase of them after knocking out DHX36 by their DMS-seq highlights the need to improve this method or to develop others to chemically probe rG4 structures *in vivo* and better understand their impact on RNA biology.

Nevertheless, Guo and Bartel proposed the existence of a redundant network of rG4-resolving cellular factors which compensate for each other´s loss. In point of fact, at least nine different RNA helicases of the DEAH or DEAD-box family are known to interact with rG4s[209,221,262-265] (see Table 1) among them EIF4A[179]. In addition, other proteins destabilizing rG4s or preventing their formation without any classical helicase function seem to exist, as it was demonstrated for CNBP[214]. For these two mentioned candidates, genome-wide data are available, however, in the specific case of compensating loss of DHX36 both proteins can be excluded. They both show a direct effect on translation which is almost exclusively mediated by binding to 5´UTRs, in the case of EIF4A, and CDSs, for CNBP, respectively[179,214]. A new prime candidate for compensating DHX36 loss is the DEAH-box helicase DHX9. Balasubramanian and coworkers proposed a correlation in the change of the TE upon depletion of both helicases as well as a common mechanism for DHX36 and DHX9[210]. I personally consider other helicases to partially compensate for the loss of DHX36. Compensation among G4-resolving helicases is already documented[155] in yeast where loss of dG4-resolving activity by Pif1 deletion can be rescued by Rrm3. Another argument for the existence of a redundant rG4-resolving machinery in humans is the fact that DHX36-KO cells only exhibit an ~1.5-fold increase in rG4 detectable by immunofluorescence (Figure 4-19) and an ~25% accumulation of the best DHX36 target mRNAs (Figure 4-9). The amplitude of these effects is not as drastic as it would be expected if only DHX36 would function in those mechanisms.

Consequently, it is not unlikely that such a rG4-resolving machinery exists and other rG4 helicases and destabilizing factors will be identified in future.

Most studies in the current rG4 literature focus on the presence of these structures in 5´UTRs of mRNAs and their resulting impact on translation[172,176,210,228]. Here, they are assumed to have the greatest impact on this basic biological mechanism by manipulating translation initiation and the scanning process of the 43S complex[173,266]. However, the data of this thesis demonstrate that rG4s in the 3´UTR can be as efficient in translational regulation as rG4s located elsewhere on the mRNA. To my knowledge, this study is the first comprehensive link between regulation of 3´UTR-located rG4s and stability as well as translation efficiency of mRNAs.

In general rG4s are rather seen as a fine-tuning mechanism than as an “on/off-switch”[267]. This is congruent with the here-presented data (Figures 4-9, 4-11, 4-13, 4-17). Of course, a stronger or more prominent effect is always favorable for a scientist to address, but an important feature of translational control is that multiple regulatory mechanisms affect the rate of protein synthesis[94]. This functional redundancy complicates mechanistic studies, because interfering with one mechanism does not necessarily alter the overall protein output to a dramatic extent[94] due to compensation effects. Therefore, a given mechanism may itself cause only a minor change in protein synthesis rate. However, when multiple weak mechanisms act on the system together, significant translational control can result[94]. Therefore, I expect rG4s to act as one part in a system of several alternative mechanisms of regulation.

All in all, the presented study significantly contributes to our understanding of rG4 regulation itself and rG4-mediated regulation of gene expression in particular and opens the field for future research further determining the role of RNA G-quadruplex structures in living cells.

5.6 A model for DHX36 function

Considering all the data obtained and presented in this thesis and combining it with the current state of research, a model for DHX36 function can be proposed (Figure 5-1). Note, that I still expect DHX36 to be a multifunctional protein. The here-depicted model comprises only the function this thesis deals with, which is, however, most probably the main role of this helicase.

DHX36 interacts with a wide range of cellular mRNAs by structured G-rich elements and rG4s. This interaction, which is mostly independent of the translational machinery and leads to the simultaneous resolution of such G-rich structures, prevents mRNA aggregation in stress granules or in a related precursor form. Thereby, DHX36 maintains the steady-state levels of these mRNAs and keep them translationally active. In cases of additional cellular stress, DHX36 is passively recruited to SGs and assists in the extraction of its target mRNAs to either restore their translation competence or to facilitate their decay.

Loss of DHX36, however, leads to an increase in the cellular rG4 levels. The cell reacts with the formation of mRNA aggregates to protect the pool of active ribosomes and not get them trapped by these transcripts. By this, total levels of DHX36 targets and rG4-bearing mRNAs accumulate but render translationally inactive since degradation is impeded.

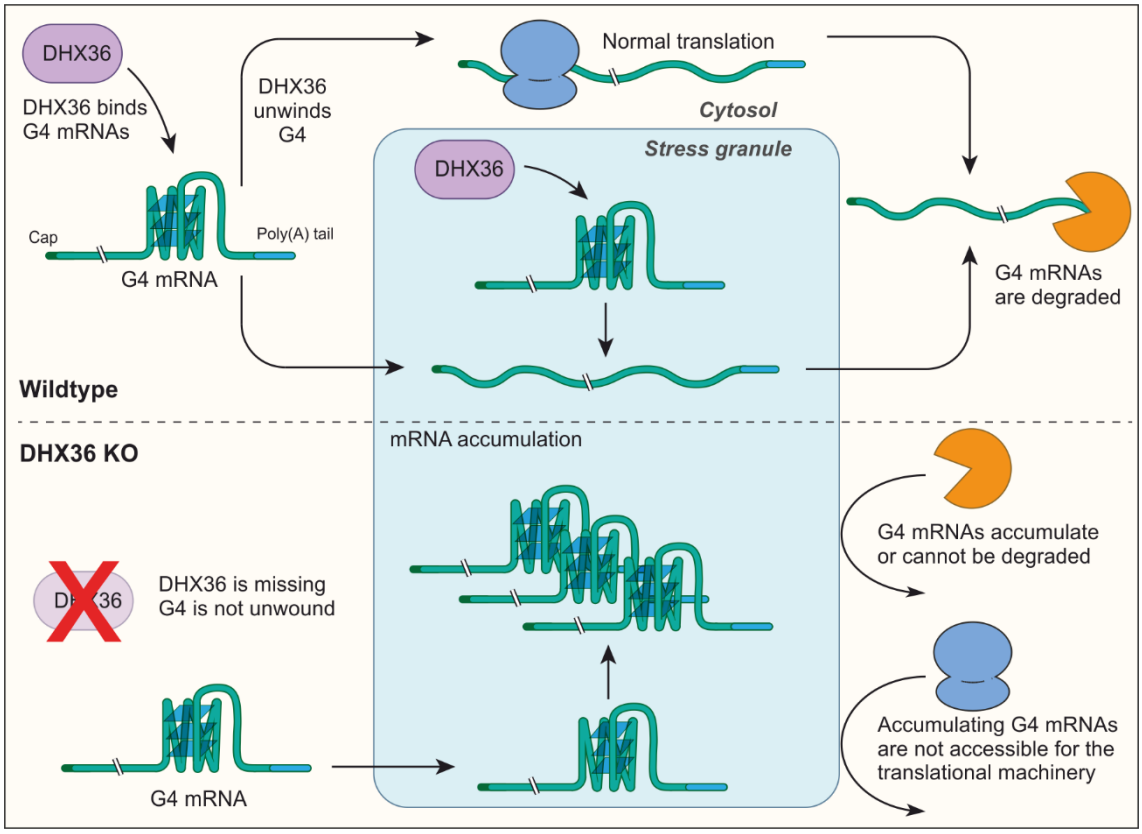

Figure 5-1 Model for DHX36 function.
DHX36 helps to unwind rG4 structures forming in mRNAs to keep them translational competent and prevents their accumulation in SG and/or helps releasing them from these structures. In DHX36-KO cells, rG4 containing mRNA cannot be translated and accumulate in SG.

5.7 DHX36 relevance for life and concluding remarks

To finally conclude this thesis, I want to shortly address the overall relevance of DHX36 function for living cells and multicellular organisms.

Albeit the obvious changes in the transcriptome of DHX36 cells, alteration in the protein output of DHX36 target mRNAs does significantly but not dramatically change (Figure 4-13). At the same time stress-related processes are activated and proliferation and morphology of those cells are reduced and altered, respectively (Figure 4-8). Nevertheless, cultured DHX36-KO cells are viable and thus, they have adapted to the loss of this helicase to maintain and regulate posttranscriptional gene regulation. Therefore, it can be assumed that loss of DHX36 changes the landscape of posttranscriptional regulation and other mechanisms of PTGR compensate for the functional loss of DHX36 activity, to some extent. All these observations were made in the isolated system ´cell culture´.

However, it has to be stated that in living organisms, like mice, 1.) complete knockout of DHX36 leads to early embryonic lethality[227], whereas 2.) tissue-specific knockout results in severe developmental defects in the affected organic systems[227-229]. To my knowledge no genetic disease is known in humans resulting from mutated DHX36. Expression levels of this helicase are high during early steps of differentiation but decrease eventually. All this suggests an inevitable function of DHX36 in humans during cell differentiation and development. In a broader context, this also underlines the importance to further investigate the known G-quadruplex regulating helicases and factors, like WRN, BLM, or EIF4A, and to identify the so far unknown potential members of the cellular G4-regulating machinery.

Contrary, DHX36 gene amplification is associated with cancer and many cancer cell types show elevated DHX36 levels. In addition, G-quadruplex structures are connected to oncogenes and oncogenic translation. This allows the hypothesis that mis-regulation of this interconnected system contributes to cancer formation and/or propagation. Illumination of such mis-controlled pathways can help to identify potential new targets for anticancer therapy against distinct cancer types and contribute to customized medicine.

In addition, more research is necessary to understand the impact of DHX36 and rG4s during viral infection. Deciphering this system can help to understand virulent strategies of viruses and can provide possible targets for therapy.

In summary, this study provides a detailed and systems-wide characterization of the targets and function of the major G-quadruplex-resolving helicase DHX36. The here presented doctoral thesis serves as a comprehensive resource for investigating target interactions of RNA helicases and offers a new link between rG4 formation and transcript stability on one hand and the translation efficiency and stress response on the other hand. This basic research data can provide valuable information for applied and clinical research on the way to better understand developmental diseases, cancer, and infection and to find suitable cure and therapy.

6. Material & Methodology

6.1 Material

6.1.1 General equipment

Standard laboratory devices used in this study were purchased from the following companies (alphabetical order):

BD Bioscience, Beckman Coulter GmbH, BioComp Instruments, Biorad, Eppendorf AG, GE Healthcare Life Science, Illumina Inc., Jasco Deutschland GmbH, Leica Microsystems, NanoTemper Technologies GmbH, Thermo Fisher Scientific

6.1.2 Consumables

All standard laboratory consumables used in this study were purchased from the following companies:

Sigma-Aldrich Co. LLC., Eppendorf AG, Applied Chemical Technology Inc., STARLAB INTERNATIONAL GmbH, or SARSTEDT AG & Co. KG.

6.1.3 Chemicals

All standard chemicals used in this study were purchased from one of the following companies:

Applied Chemical Technology Inc., Carl Roth GmbH + Co. KG, Merck KGaA, or Sigma-Aldrich Co. LLC.,

Distilled water was filtrated by a TKA-LAB HP filtration system and will be referred as ddH_20. Special chemicals and commercial kits are listed below in alphabetical order:

Chemical	Identifier	Supplier
Carboxypyridostatin trifluoroacetate salt	SML1176	Sigma-Aldrich
Dynabeads Protein G	10003D	Thermo Fisher Scientific
Ethidium bromide solution (1%)	2218.1	Carl Roth GmbH
Fluoroshield with DAPI	F6057-20ML	Sigma-Aldrich
γ-^{32}P-Adenosine 5´triphosphate (γ-ATP)	SRP-501	Hartmann Analytic
GeneRuler 1kb Plus DNA Ladder	SM1331	Thermo Fisher Scientific
GeneRuler Low Range DNA Ladder	SM1193	Thermo Fisher Scientific
L-Arginine:HCl (13C6, 99%)	CLM-2265-H-0.05	Cambridge Isotope Laboratories
L-Arginine:HCl (13C6, 99%, 15N4, 99%)	CNLM-539-H-0.1	Cambridge Isotope Laboratories

L-Lysine:2HCl (4,4,5,5-D4, 96-98%)	DLM-2640-0.25	Cambridge Isotope Laboratories
L-Lysine:2HCl (13C6, 99%, 15N2, 99%)	CNLM-291-H-0.05	Cambridge Isotope Laboratories
Oligo $d(T)_{25}$ Magnetic Beads	S14193	New England Biolabs
PageRuler Unstained Protein Ladder	26614	Thermo Fisher Scientific
Poly-D-Lysine	A-003-M	Sigma-Aldrich
Proteinmarker prestained (11 – 245 kDa)	1123YL500	Biofroxx
4-thiouridine (4SU)	n/a	Chemistry Department, University of Würzburg
SUPERaseIn RNase Inhibitor	AM2696	Thermo Fisher Scientific
Quantitas DNA Marker 100 bp – 1 kb	250212	Biozym

6.1.4 Commercial systems

Kit/System	Identifier	Supplier
CellTiter 96® Non-Radioactive Cell Proliferation Assay (MTT)	G4000	Promega
Genomic DNA Purification Kit	K0512	Thermo Fisher Scientific
innuPREP Gel Extraction Kit	845-KS-5030250	Analytik Jena
innuPREP PCRpure Kit	845-KS-5010250	Analytik Jena
innuPREP Plasmid Mini Kit	845-KS-5041250	Analytik Jena
IQ SYBR Green	170-8885	BioRad
Lipofectamine 2000 Transfection Reagent	11668027	Thermo Fisher Scientific
Lipofectamine CRISPRMAX Cas9 Transfection Reagent	CMAX00015	Thermo Fisher Scientific
NEBNext Multiplex Oligos for Illumina	E7335	New England Biolabs
NEBNext rRNA Depletion Kit	E6310	New England Biolabs
NEBNext Ultra Directional RNA Library Prep Kit for Illumina	E7420	New England Biolabs
NuPAGE 4-12% Bis-Tris Protein Gels	NP0321BOX	Thermo Fisher Scientific
QuantiTect Reverse Transcription Kit	205313	Qiagen

6.1.5 Strains and cell lines

6.1.5.1 Bacterial strains

Strain	Genotype	Supplier
DH5α	F- ϕ80*lac*ZΔM15 Δ(*lac*ZYA- *arg*F)U169 *deo*R *rec*A1 *end*A1 *hsd*R17(rk$^-$, mk$^+$) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ$^-$	Thermo Fisher Scientific
One shot *ccdB* Survival 2T1	F$^-$*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*Δ139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G *fhu*A::*IS2*	Thermo Fisher Scientific

Name	Strain	Plasmid	Origin
MSB4	DH5α	pFRT FlagHA-DHX36-1	This study
MSB8	DH5α	pOG44	This study
MSB35	DH5α	pFRT FlagHA-DHX36-2	This study
MSB36	DH5α	pFRT FlagHA-DHX36-1 E335A	This study
MSB37	DH5α	pFRT FlagHA-DHX36-2 E335A	This study
MSB95	One shot *ccdB* Survival 2T1	pDONR-221	This study
MSB97	DH5α	pDONR-221-PURB site	This study
MSB98	DH5α	pDONR-221-WAC site	This study
MSB100	DH5α	pcDNA5-FRT-GFP-mCherry-3pGW	This study
MSB102	DH5α	pcDNA5-FRT-GFP-mCherry-3pGW-PURB site	This study
MSB103	DH5α	pcDNA5-FRT-GFP-mCherry-3pGW-WAC site	This study
MSB107	DH5α	pcDNA5-FRT-GFP-mCherry-3pGW-mutPURB site	This study
MSB108	DH5α	pcDNA5-FRT-GFP-mCherry-3pGW-mutWAC site	This study
MSB112	DH5α	pDONR-221-non-target site1	This study
MSB113	DH5α	pDONR-221-non-target site 2	This study
MSB114	DH5α	pcDNA5-FRT-GFP-mCherry-3pGW-non-target site 1	This study
MSB115	DH5α	pcDNA5-FRT-GFP-mCherry-3pGW-non-target site 2	This study

6.1.5.2 Eukaryotic cell lines

Cell line	Identifier	Supplier
T-Rex™-HEK 293	R71007	Thermo Fisher Scientific

Cell line	Paternal cell line	Description	Origin
MSC5	T-Rex™-HEK 293	Inducible expression of FH-DHX36-1	This study
MSC6	T-Rex™-HEK 293	Inducible expression of FH-DHX36-1-E335A	This study
MSC7	T-Rex™-HEK 293	Inducible expression of FH-DHX36-2	This study
MSC8	T-Rex™-HEK 293	Inducible expression of FH-DHX36-2-E335A	This study
MSC9	T-Rex™-HEK 293	DHX36-gene knock out by CRISPR/Cas9	This study
MSC10	MSC9	Inducible overexpression of FH-DHX36-1 in DHX36-KO cells	This study
MSC11	MSC9	Inducible overexpression of FH-DHX36-1 E335A in DHX36-KO cells	This study
MSC13	T-Rex™-HEK 293	pcDNA5-FRT-GFP-mCherry-3pGW-PURB site	This study
MSC14	T-Rex™-HEK 293	pcDNA5-FRT-GFP-mCherry-3pGW-WAC site	This study
MSC17	MSC9	pcDNA5-FRT-GFP-mCherry-3pGW-PURB site	This study
MSC18	MSC9	pcDNA5-FRT-GFP-mCherry-3pGW-WAC site	This study
MSC21	T-Rex™-HEK 293	pcDNA5-FRT-GFP-mCherry-3pGW-mutPURB site	This study
MSC22	T-Rex™-HEK 293	pcDNA5-FRT-GFP-mCherry-3pGW-mutWAC site	This study
MSC25	MSC9	pcDNA5-FRT-GFP-mCherry-3pGW-mutPURB site	This study
MSC26	MSC9	pcDNA5-FRT-GFP-mCherry-3pGW-mutWAC site	This study
MSC28	T-Rex™-HEK 293	pcDNA5-FRT-GFP-mCherry-3pGW-non-target site 1	This study
MSC29	MSC9	pcDNA5-FRT-GFP-mCherry-3pGW-non-target site 1	This study
MSC30	T-Rex™-HEK 293	pcDNA5-FRT-GFP-mCherry-3pGW-non-target site 2	This study
MSC31	MSC9	pcDNA5-FRT-GFP-mCherry-3pGW-non-target site 2	This study

6.1.6 Enzymes

Enzyme	Identifier	Supplier
BamHI	ER0051	Thermo Fisher Scientific
Calf intestine phosphatase	M0290S	New England Biolabs
Cas9 Nuclease	1074181	Integrated DNA Technologies
DpnI	ER1701	Thermo Fisher Scientific
Elongase	10422322	Thermo Fisher Scientific
ExTaq DNA polymerase	RR001A	Clontech Laboratories
Gateway BP Clonase II Enzyme mix	11789020	Thermo Fisher Scientific
Gateway LR Clonase II Enzyme mix	11791100	Thermo Fisher Scientific
Homemade Taq DNA polymerase	n/a	Homemade
Phusion High-Fidelity DNA polymerase	F-531	Thermo Fisher Scientific
Proteinase K	7528.3	Carl Roth GmbH + Co. KG
RNase T1	EN0541	Thermo Fisher Scientific
Shrimp Alkaline Phosphatase	783901000UN	Thermo Fisher Scientific
Superscript III Reverse Transcriptase	18080044	Thermo Fisher Scientific
T4 RNA Ligase	EL0021	Thermo Fisher Scientific
T4 RNA Ligase 2, truncated K227Q	M0351S	New England Biolabs
T4 Polynucleotide Kinase	EK0032	Thermo Fisher Scientific
Turbo DNase I	AM2238	Thermo Fisher Scientific
XhoI	ER0691	Thermo Fisher Scientific

6.1.7 Antibodies

Western blotting	Dilution	Identifier	Supplier
Anti-CANX	1:1000	ab31290	Abcam
Anti-DHX36	1:500	sc-377485	Santa Cruz Biotechnology
Anti-EIF4A1	1:1000	SAB2700953	Sigma-Aldrich
Anti-FLAG	1:2000	F1804	Sigma-Aldrich
Anti-FMR1	1:2000	n/a	Linder et al., 2008
Anti-HA	1:2000	MMS-101R	Covance
Anti-HISTH2B	1:2000	ab1790	Abcam
Anti-PKR/EIF2AK2	1:1000	Ab32052	Abcam
Anti-PKR/EIF2AK2 (phospho T446)	1:1000	ab32036	Abcam
Anti-RPL22	1:2000	sc-136413	Santa Cruz Biotechnology
Anti-TUBA4A	1:5000	T5168	Sigma-Aldrich
HRP-conjugated goat anti-mouse	1:5000	sc-2031	Santa Cruz Biotechnology
HRP-conjugated goat anti-rabbit	1:5000	sc-2030	Santa Cruz Biotechnology

IF-microscopy	Identifier	Supplier
Anti-ATPA1A	ab7671	Abcam
Anti-G3BP1	611126	BD Bioscience
Anti-FLAG	2368	Cell signaling
BG4	na	Expressed from the pSANG10-3F-BG4 plasmid[151]
Alexa Fluor 488 goat anti-mouse IgG	A11001	Thermo Fisher Scientific
Cyanine 3 goat anti-rabbit IgG	A10520	Thermo Fisher Scientific

6.1.8 Plasmids

Plasmid name	Generated by	Marker	Origin
pFRT-Flag-HA		Amp	Spitzer et al. 2013
pFRT-FlagHA-DHX36-iso1	Restriction cloning	Amp	This study
pFRT-FlagHA-DHX36-iso1 E335A	Site-directed mutagenesis	Amp	This study
pFRT-FlagHA-DHX36-iso2	Site-directed mutagenesis	Amp	This study
pFRT-FlagHA-DHX36-iso2 E335A	Site-directed mutagenesis	Amp	This study
pDONR-221		Kan	Thermo Fisher Scientific
pDONR-221-PURB site	Gateway cloning	Kan	This study
pDONR-221-WAC site	Gateway cloning	Kan	This study
pDONR-221-mutPURB site	Gateway cloning	Kan	This study
pDONR-221-mutWAC site	Gateway cloning	Kan	This study
pDONR-221-non-target site 1	Gateway cloning	Kan	This study
pDONR-221- non-target site 2	Gateway cloning	Kan	This study
pcDNA5-FRT-GFP-mCherry-3pGW		Amp	Addgene
pcDNA5-FRT-GFP-mCherry-3pGW-PURB site	Gateway cloning	Amp	This study
pcDNA5-FRT-GFP-mCherry-3pGW-WAC site	Gateway cloning	Amp	This study
pcDNA5-FRT-GFP-mCherry-3pGW-mutPURB site	Restriction-free cloning	Amp	This study
pcDNA5-FRT-GFP-mCherry-3pGW-mutWAC site	Restriction-free cloning	Amp	This study
pcDNA5-FRT-GFP-mCherry-3pGW-non-target site 1	Gateway cloning	Amp	This study
pcDNA5-FRT-GFP-mCherry-3pGW-non-target site 2	Gateway cloning	Amp	This study

6.1.9 Synthetic oligonucleotides

All DNA oligonucleotides used in this study were purchased from Sigma-Aldrich Co LLC in reverse-phase cartridge purification grade.

RNA was purchased from Integrated DNA Technologies.

Name	Sequence (5´ - 3´)
Cloning	
For DHX36	GGATCCGGATCCATGAGTTATGACTACCATCAG
Rev DHX36	CTCGAGCTCGAGTCAGCTGTAATATCCATCCTG
For GWC PURB	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATAGTAAGTGAATGAGATTATCAG
Rev GWC PURB	GGGGACCACTTTGTACAAGAAAGCTGGGTGCTGTATTTTCTTAAGGTAAATGTG
For GWC WAC	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAATGGACTTAAAAGTACTGCTGGATCGCTCAATGGACTTAAAAGTACTGCTG
Rev GWC WAC	GGGGACCACTTTGTACAAGAAAGCTGGGTGAACAATTTTTGTTTGTTTTAACATG
For PURB mut RF	CAATTAGCTGTCTTCCTAAAGAGTTACAACTCCCATTCAGTATACTGGATAATGAGTGTGTGGGTGAAGCTG
Rev PURB mut RF	CTTAAGGTAAATGTGTTTTTTTTTTTTTCTTTTTTCTTTTCTACAAACTATCTCTCGCTCTCCCCAGCTTC
For WAC mut RF	GACAGGCATGTGTGCTCAAAGTACATTGATTGCTCAAATATAAGGAAATGGCCCAATGAACGTGGTTG
Rev WAC mut RF	GTTTATTACAACAGATACAATTCACATCTGACTAGCTCTGTTTCCTCTTTCTTCTCTCACAAC
For non-target site 1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAAATTTGTCATCAATTATGACTACCC
Rev non-target site 1	GGGGACCACTTTGTACAAGAAAGCTGGGTGTTGGGATTAATTGCTTGATTAGC
For non-target site 2	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGTAATTATGGTGCACTTTTTCG
Rev non-target site 2	GGGGACCACTTTGTACAAGAAAGCTGGGTGTTTTGTGAAAACACTGCCTGC
Mutagenesis	
For DHX36-E335A	GTACTTGATGCAATCCATGAAAGAAATCTGCAGTCAG
Rev DHX36-E335A	CATGGATTGCATCAAGTACGATATGACTAACACTGGAC

For DHX36-Del14	GATCTCTTGATGTCACAAGTAATGTTTAAATCAGTTAACCAGACACAGGTGTTTAAAAGAACCCCTCCTGGTGTTCGGAAAATAGTAATTGC
Rev DHX36-Del14	GGGGTTCTTTTAAACACCTGTGTCTGGTTAACTGATTTAAACATTACTTGTGACATCAAGAGATCATGTAAAGTGCTGATATTGTCCCAGCCTGG

Sequencing

For Seq 618 DHX36	GCTGCCTTCGTATGGAATGC
For Seq 841 DHX36	GCAGAATCTTGTGGCAGTGG
For Seq 1350 DHX36	GCGAAGAAGGTATTCTGCAAG
For Seq 1995 DHX36	GGCAGTGTTACTCTCCATAAG
For Seq 2580 DHX36	CACAAAAACCGATGGCCTG
Rev Seq 490 DHX36	GTCAATATATGATCTGTTCCTG
Rev Seq 950 DHX36	GTCTGACTGGAGCCACTG
Rev Seq 2328 DHX36	GTATCTGAAACCACGTCGCC
For m13 seq	CCCAGTCACGACGTTGTAAAACG
Rev m13 seq	GGAAACAGCTATGACCATG
For check 3pGW	CATCGATAAGCTTGTCCAC
Rev check 3pGW	GGAGATATCGTCGACAAGC

Circular dichroism

TP-G4	AAAAAAAAAAGGGGGAGCTGGGGTAGATGGGAATGTGAGGG
DHX36 bind motif	AAAAAAAAAAGGAGGAGGAGGAGGA
DHX36 mut bind motif	AAAAAAAAAAGAAGAAGGAGAAGAA
DHX36 mut bind motif 2	AAAAAAAAAAGCAGCAGGAGCAGCA
NAA50 G4 motif	AAAAAGGGGACTGCACAAGGATGTGAATACTGGGAGGTGG
PURB G4 motif	AAAAAGGGTGTGTGGGTGGGGCTGGGGAGGGCGGG
WAC G4 motif	AAAAAGGCCCAATGAACGTGGTTGTGGGAGGGGAAAGAGG
SLMO2 G4 motif	AAAAAGGTGGGGAAGAACAAGCATAATGGTAGGGGGAGG
NAA50 mut G4 motif	AAAAAGGAGACTGCACAAGGATGTGAATACTGAGAGGTGG
PURB mut G4 motif	AAAAAGAGTGTGTGGGTGAAGCTGGGGAGAGCGAG
WAC mut G4 motif	AAAAAGGCCCAATGAACGTGGTTGTGAGAGAAGAAAGAGG

SLMO2 mut G4 motif	AAAAAGGTGGAGAAGAACAAGCATAATGGTAGAGAGAGG
Microscale thermophoresis	
DHX36 motif 5´ link Cy5	AAAAAGGAGGAGGAGGAGGA
DHX36 motif mut2 5´ link Cy5	AAAAAGCAGCAGGAGCAGCA
qPCR	
For RNU6 RT	GCTTCGGCAGCACATATACTAAAAT
Rev RNU6 RT	CGCTTCACGAATTTGCGTGTCAT
For NAA50 RT	TGGCACCTTACCGAAGGCTA
Rev NAA50 RT	TTGCCGACTCATTGCTGATCT
For PURB RT	GCCATCACCGTACCCTTCAAA
Rev PURB RT	CCCTCTGTCGTTCCTGGATTT
For WAC RT	GCCGGAGATCCTTCACCAC
Rev WAC RT	TTTGGCCTTACTGTGACCTGT
For SLMO2 RT	CCCAAACCCTATGAACCCAAG
Rev SLMO2 RT	TGTGCAACTTTCCAGAGGG
For mCherry RT	CCCGCCGACATCCCCGACTA
Rev mCherry RT	GGGTCACGGTCACCACGCC
For AcGFP RT 1	GCAACATCCTGGGCAATAAG
Rev AcGFP RT 1	GAAGTTCACCTTGATGCCATTC
crRNA	
DHX36 crRNA	UGUGGUACGCGAAAAAACAGGUUUUAGAGCUAU
tracrRNA-ATTO 550	GTTGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTT
PAR-CLIP	
3´Adapter	rApp-Barcode-TCGTATGCCGTCTTCTGCTTG
5´Adapter	GUUCAGAGUUCUACAGUCCGACGAUC
19 nt size marker	CGUACGCGGGUUUAAACGA
35 nt size marker	CUCAUCUUGGUCGUACGCGGAAUAGUUUAAACUGU
3´Primer	CAAGCAGAAGACGGCATACGA
5´Primer	AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA

6.1.10 Buffers & solutions

Buffer	Composition
DNA loading dye (6x)	48% (v/v) glycerol, 0.1% (v/v) Bromphenol blue, 0.1% (v/v) Xylene cyanol
TBE (5x)	445 mM Tris base, 445 mM Boric acid, 10 mM Na_2EDTA
SDS loading dye (6x)	300 mM Tris-HCl pH 6.8, 120 mM DTT, 9% (w/v) SDS, 48% (v/v) glycerol, 0.1% (v/v) Bromphenol blue
SDS running buffer (10x)	0.25 M Tris base, 1.92 M glycine, 1% (w/v) SDS
Separating gel buffer (4x)	1.5 M Tris-HCl pH 8.8, 0.4% (w/v) SDS
Stacking gel buffer (4x)	0.5 M Tris-HCl pH 6.8, 0.4% (w/v) SDS
TBS-T	10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% (v/v) Tween20
Western blotting buffer	50 mM Tris base, 50 mM glycine, 20% (v/v) methanol
Coumaric acid solution	90 mM coumaric acid in DMSO
Luminol solution	250 mM luminol in DMSO
NET-Gelatine	50 mM Tris base, 150 mM NaCl, 0.05% (v/v) Triton-X-100, 5 mM Na_2EDTA, 2.5% gelatine
RNA loading dye (2x)	50 mM Na_2EDTA, 0.05% Bromphenol blue, 88% (v/v) formamide
Proteinase K buffer	50 mM Tris-HCl pH 7.5, 75 mM NaCl, 6.25 mM Na_2EDTA, 1% (w/v) SDS
Luria-Bertani (LB) medium	1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract
SOC	2% (w/v) tryptone, 10 mM NaCl, 10 mM $MgSO_4$, 10 mM $MgCl_2$, 2.5 mM KCl, 0.5% (w/v) yeast extract, 2% (w/v) glucose
NP40 Lysis Buffer	50 mM HEPES pH 7.5, 150 mM KCl, 2 mM Na_2EDTA, 0.5% (v/v) NP-40 substitute, 1 mM NaF, freshly added before use: 0.5 mM DTT, 0.2 mM PMSF, 1 mM LP, 1 mM AP, 0.1 mM AEBSF
IP-Wash Buffer	50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5 mM Na_2EDTA, 0.1% (v/v) NP-40 substitute

Elution Buffer	10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 mg ml^{-1} FLAG peptide
Cytoplasmic Fraction Buffer	20 mM HEPES pH 7.6, 0.1% (v/v) NP-40 substitute, 10% (v/v) glycerol, 2 mM $MgCl_2$, freshly added before use: 0.5 mM DTT, 0.2 mM PMSF, 1 mM LP, 0.1 mM AEBSF, 1 mM AP, 0.1% (v/v) SUPERaseIn
Sucrose Buffer	20 mM HEPES pH 7.6, 10 mM NaCl, 1.5 mM $MgCl_2$, 10% (v/v) glycerol, 37% (w/v) sucrose, 0.5 mM Na_2EDTA, freshly added: 0.5 mM DTT, 0.2 mM PMSF, 1 mM LP, 1 mM AP, 0.1 mM AEBSF, 0.1% (v/v) SUPERaseIn
Nuclear Lysis Buffer	20 mM HEPES pH 7.6, 100 mM NaCl, 0.5 mM Na_2EDTA, freshly added: 0.5 mM DTT, 0.2 mM PMSF, 1 mM LP, 1 mM AP, 0.1 mM AEBSF, 0.1% (v/v) SUPERaseIn
NUN-Buffer (2x)	50 mM HEPES pH 7.6, 600 mM NaCl, 7.5 mM $MgCl_2$, 2 M urea, 0.2 mM Na_2EDTA, 2% (v/v) NP-40 substitute, 0.5 mM DTT
Ribosome footprinting buffer	20 mM Tris, pH7.4, 150 mM NaCl, 4 mM $MgCl_2$, 1 mM DTT, 100 µg ml^{-1} cycloheximide, 25 U ml^{-1} Turbo DNase I, 1% (v/v) NP-40
Sucrose cushion buffer	20 mM Tris, pH7.4, 150 mM NaCl, 4 mM $MgCl_2$, 1 mM DTT, 100 µg ml^{-1} cycloheximide, 20 U ml^{-1} SUPERaseIn, 1 M sucrose
LiDS Lysis Buffer	20 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.5% (w/v) LiDS, 1 mM Na_2EDTA, 5 mM DTT
Wash buffer 1 oligo d(T) pulldown	20 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.1% (w/v) LiDS, 1 mM Na_2EDTA, 5 mM DTT
Wash buffer 2 oligo d(T) pulldown	20 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM Na_2EDTA, 5 mM DTT
Wash buffer 3 oligo d(T) pulldown	20 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM Na_2EDTA
Elution buffer oligo d(T) pulldown	20 mM Tris-HCl pH 7.5, 1 mM Na_2EDTA
PCR Buffer for homemade Taq (10x)	750 mM Tris-HCl pH 8.8, 200 mM $(NH_4)_2SO_4$, 25 mM $MgCl_2$, 0.1% (v/v) Tween20
Polysome Buffer (10x)	200 mM Tris-HCl pH 7.5, 1 M KCl, 50 mM $MgCl_2$
Polysome Lysis Buffer	20 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 0.1 mg/ml Cycloheximide, 0.5% (v/v) NP-40 substitute, 0.5 mM Na_2EDTA, freshly added: 0.5 mM DTT, 0.2 mM PMSF, 1 mM LP, 1 mM AP, 0.1 mM AEBSF, 0.1% (v/v) SUPERaseIn

Hypotonic Lysis Buffer	10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 10% (v/v) glycerol, 0.3% (v/v) NP-40 substitute
Nuclear Lysis Buffer (NLB)	20 mM Tris-HCl pH 7.5, 150 mM KCl, 3 mM $MgCl_2$, 10% (v/v) glycerol, 0.3% (v/v) NP-40 substitute
HBS Buffer (2x)	270 mM NaCl, 10 mM KCl, 40 mM HEPES pH 7.05, 1.5 mM Na_2HPO_4
TB buffer	10 mM Pipes pH 6.7, 15 mM $CaCl_2$, 250 mM KCl, 3 mM $MgCl_2$, 55 mM $MnCl_2$
G4 folding buffer	10 mM Tris-HCl pH7.5, 100 mM KCl, 1 mM EDTA

6.1.11 Growth media & antibiotics

Solution	Identifier	Supplier
DPBS, no calcium, no magnesium	Cat#: 14190094	Thermo Fisher Scientific
DMEM, high glucose	Cat#: 41965039	Thermo Fisher Scientific
DMEM, high glucose, no glutamine, no lysine, no arginine	Cat#: A1443101	Thermo Fisher Scientific
Fetal Bovine Serum (FBS)	Cat#: 10270106	Thermo Fisher Scientific
Fetal Bovine Serum, dialyzed	Cat#: 26400036	Thermo Fisher Scientific
Trypsin-EDTA (0.25%), phenol red	Cat#: 25200056	Thermo Fisher Scientific
L-Glutamine (200 mM)	Cat#: 25030024	Thermo Fisher Scientific
Penicillin-Streptomycin (10,000 U/ml) (PenStrep)	Cat#: 15140122	Thermo Fisher Scientific
Opti-MEM I Reduced Serum Medium	Cat#: 31985070	Thermo Fisher Scientific

SILAC Media

Name	Composition
Light medium	DMEM, 10% dialyzed FBS, 4 mM glutamine, 1.74 mM L-proline, 0.8 mM L-lysine, 0.4 mM L-arginine
Medium-heavy medium	DMEM, 10% dialyzed FBS, 4 mM glutamine, 1.74 mM L-proline, 0.8 mM L-lysine [4,4,5,5-D4], 0.4 mM L-arginine [U-13C6]
Heavy medium	DMEM, 10% dialyzed FBS, 4 mM glutamine, 1.74 mM L-proline, 0.8 mM L-lysine [U-13C6, 15N2], 0.4 mM L-arginine [U-13C6, 15N4]

Antibiotic	Identifier	Supplier
Ampicillin	Cat#: K029.5	Carl Roth GmbH
Blasticidin	Cat#: ant-bl-1	Invivogen
Chloramphenicol	Cat#: C0378	Sigma-Aldrich
Cycloheximide	Cat#: C1988	Sigma-Aldrich
Hygromycin B	Cat#: ant-hg-5	Invivogen
Kanamycin	Cat#: T832.3	Carl Roth GmbH
Tetracycline	Cat#: 87128	Sigma-Aldrich
Zeocin	Cat#: ant-zn-1	Invivogen

6.2 Methodology

6.2.1 Bacteria cell culture

6.2.1.1 Cultivation of bacteria

E. coli strains were grown on LB agar plates at 37°C. For liquid growth cultures, 5 ml LB media were inoculated with one singly grown colony for overnight growth at 37°C and 200 rpm agitation.

All growth media were supplemented with appropriate amounts of antibiotics compatible to resistances of each bacteria strain (100 µg ml^{-1} ampicillin, 50 µg ml^{-1} kanamycin, 30 µg ml^{-1} chloramphenicol). OD600 was measured using a spectrophotometer.

6.2.1.2 Cryo-conservation of bacteria

Bacterial permanent stocks were prepared by mixing 600 µl bacterial overnight culture with 50% sterile glycerol for cryo-conservation. After snap-freezing in liquid nitrogen, stocks were stored at -80°C.

6.2.1.3 Preparation of chemical-competent E. coli

5 ml LB media without any antibiotics were inoculated with *E. coli* DH5α and incubated for overnight growth at 37°C and 200 rpm agitation. 250 ml SOB media were inoculated with 2 ml of this pre-culture and incubated at 18°C and 180 rpm agitation until the culture reaches OD600 = 0.6. The culture was chilled for 10 min on ice, cells were collected by centrifugation (4°C, 2,500 g, 10 min), and resuspended in 80 ml ice-cold TB buffer. After 10 min incubation on ice, cells were again pelleted and carefully resuspended in 20 ml freshly prepared, ice-cold TB buffer by gentle agitation. DMSO was added to a final concentration of 7% and the suspension was incubated on ice for 10 min. Finally, the cells were aliquoted to 100 µl, snap-frozen in liquid nitrogen, and stored at – 80°C.

6.2.1.4 Transformation of plasmid DNA in bacteria

A 100 µl aliquot chemical competent *E. coli* DH5α was thawed on ice for 5 min and carefully supplemented with 1 - 10 ng pre-chilled plasmid DNA. After incubation on ice for 30 min, cells were heat shocked (42°C, 1 min) and placed back on ice for 5 min. 450 µl of pre-warmed SOC media were added and cells were recovered for 1 h at 37°C and 500 rpm agitation. Cells were collected by centrifugation (1 min, 8,000 g, RT) and re-suspended in 200 µl fresh SOC media. The transformation mixture was plated on selective LB agar plates (100 µg ml^{-1} ampicillin, 50 µg ml^{-1} kanamycin, or 30 µg ml^{-1} chloramphenicol) and cultivated at 37°C.

6.2.2 Eukaryotic cell culture

6.2.2.1 Cultivation of adherent eukaryotic cells

Adherent HEK293 cells were cultured in DMEM, supplemented with 10% (v/v) FBS and 1% (v/v) PenStrep. For HEK293 wildtype cells, 100 µg ml^{-1} Zeocin and 10 µg ml^{-1} Blasticidin were added. Selective media for stable transfected cells contained 100 µg ml^{-1} Hygromycin B instead of Zeocin. Cells were grown in sterile cell culture dishes at 37°C and 5% CO_2.

100% confluent cells were regularly passaged on new plates. For this, cells were washed with sterile PBS and detached from the plate by incubating for 5 min with suitable amounts of Trypsin-EDTA solution. Trypsinization was stopped by adding fresh media and cells were separated by careful pipetting. If needed, cell concentration in this suspension was counted with a hemocytometer using 0.2% trypan blue to differentiate between living and dead cells. Desired amounts of cells were seeded in new cell culture dishes.

6.2.2.2 Cryo-conservation of eukaryotic cells

Eukaryotic cell cryo stocks were made by preparing a cell suspension and collecting the cells by centrifugation (5 min, 200 g, RT). The cell pellet was resuspended in freezing media (standard media + 10% FBS + 10% DMSO) to get a cell density of approx. 2-3 $x10^6$ cells per cryo-vial. Stocks were subsequently frozen for 2 – 3 h at -20°C, 2 – 3 days at -80°C, and finally stored in liquid nitrogen (-196°C).

6.2.2.3 Thawing of cryo-conserved eukaryotic cells

Cryo-conserved cells were rapidly thawed in a water bath and the cell suspension was transferred to a 15-ml tube containing 5 ml pre-warmed growth media. To eliminate residual DMSO, cells were carefully pelleted by centrifugation (5 min, 200 g) and the supernatant was aspirated. Cells were resuspended in fresh media and distributed on a 10-cm cell culture plate.

6.2.2.4 Generation of stable eukaryotic cell lines

Stable DHX36 expressing HEK293 cells were generated using the Flp-In™ T-REx™ system. For this, wildtype Flip-In™ T-REx™ 293 were seeded on a 10-cm plate in media containing no selective antibiotics. At ~40-50% confluency transfection was performed using the calcium chloride protocol. 9 µg of the desired integrative pFRT-plasmid, containing the desired gene construct, and 1 µg of the plasmid pOG44, expressing the Flp recombinase, were diluted in 450 µl sterile ddH_2O. 50 µl of 2.5 M $CaCl_2$ were added to the DNA solution and cells were treated with 40 µM chloroquine for 5 min. Next, 500 µl of 2x HBS buffer were dropwise added to the plasmid solution, mixed by flicking, and incubated for 3 min at RT. The entire transfection mix was carefully added to the cells

and distributed by swaying. 5 h post transfection, growth media was changed to media containing no selective antibiotics.

32 h post transfection, successfully transfected cells were selected for positive integration of the transgene construct using DMEM medium supplemented with Hygromycin and Blasticidin. Selective media was changed every 1 - 2 days until single clones could be identified. Those were picked by trypsinization and transferred to individual plates for further expansion.

6.2.2.5 Transient transfection of eukaryotic cells

For transient transfection of plasmid DNA, Lipofectamine 2000 was used according to the supplier´s instruction.

6.2.2.6 CRISPR/Cas9-mediated gene knockout in eukaryotic cells

4x 10^5 wildtype Flp-In T-REx HEK293 cells were seeded in a 6-well plate compartment. Next day, 100 pmol DHX36 crRNA and 100 pmol tracrRNA-ATTO 550 complementary to the crRNA were solved together in Nuclease-Free Duplex Buffer and boiled for 5 min at 95°C (total volume :100 µl). The RNA mixture was incubated for 15 min at RT to achieve strand annealing. 15 pmol annealed RNA was mixed with 15 pmol Cas9 protein and 5 µl Cas9+reagent (IDT) in Opti-MEM (total volume: 150 µl). In parallel, 125 µl Opti-MEM were mixed with 7.5 µl CRISPRMAX Transfection Reagent. After 5 min incubation at RT both solutions were combined, vortexed, and added to the cells. 48 h post transfection, ATTO 550 positive cells were sorted by flow cytometry and 1 cell per well were seeded in a 96-well plate in standard DMEM, supplemented with 10% (v/v) FBS and 1% (v/v) PenStrep. Single clones were expanded and successful loss of DHX36 was detected by Western blotting. Changes in the genomic region of DHX36 in DHX36-negative clones were analyzed by isolation of genomic DNA using the commercial Genomic DNA Purification Kit according to the supplier´s instructions followed by sequencing.

6.2.2.7 Light microscopy imaging

1.2 x10^6 wildtype HEK293 T-REx Flp-In cells and HEK293 T-Rex Flp-In DHX36-KO cells were seeded in 60-mm cell culture dishes. Microscopic images were taken 24 h, 48 h, 72 h, and 96 h after seeding using an EVOS FL cell imaging system.

6.2.2.8 Growth analysis of adherent eukaryotic cells

The growth of HEK293 T-REx Flp-In wildtype cells and HEK293 T-REx Flp-In DHX36 knockout cells was analyzed by counting the cells over a period of 10 days. 0.1 x10^6 of regular grown cells were seeded in 35-mm cell culture dishes. At different time points, cell numbers were determined in triplicates using a Fuchs-Rosenthal-hemocytometer.

6.2.2.9 MTT proliferation assay

Cell proliferation of different cell lines was analyzed using a commercial, MTT-based proliferation assay (Promega). Absorbance at 570 nm was measured on a Tecan Infinite®200 plate reader against a reference line at 700 nm.

6.2.2.10 Flow cytometry

Expression of mCherry protein in cell lines with stably integrated pcDNA5-FRT-GFP-mCherry-3pGW reporter constructs was induced with 500 ng ml^{-1} tetracycline for 16 h. Cells were trpysinized, washed twice with PBS, and filtered through a 35 µm nylon net. Fluorescent signals of AcGFP and mCherry were detected on a LSRFortessa™ system (BD Biosciences). Triplicates of 100,000 double positive cells were analyzed using the commercial FlowJo software (FlowJo LLC).

6.2.3 DNA biochemistry

6.2.3.1 Preparation of plasmid DNA

Plasmid DNA preparation from bacteria was performed with the innuPREP Plasmid Mini Kit (Analytik Jena). All steps were made according to the supplier´s instructions.

6.2.3.2 Polymerase Chain Reaction (PCR)

Standard PCRs were performed with homemade Taq polymerase, for preparative PCRs commercial ExTaq was used, respectively. Oligonucleotides which were used in this study are listed above (see 6.1.9).

Reagent	Homemade Taq	ExTaq
10x PCR buffer for homemade Taq	1x	- / -
10x ExTaq Buffer	- / -	1x
dNTP mix	200 µM	200 µM
Each primer	0.5 µM	0.5 µM
Template DNA	10 – 50 ng	10 – 50 ng
Homemade Taq	1 µl	- / -
ExTaq	- / -	1 U
ddH_2O	To 50 µl	To 50 µl

Step	Time	Temperature	Cycle
Initial denaturation	5 min	95°C	1x
Denaturation	20 s	95°C	
Annealing	15 s	53 - 58°C*	30x
Elongation	1 min/kb	72°C	
Final elongation	5 min	72°C	1x

* depending on melting temperature of primer sets

6.2.3.3 Agarose gel electrophoresis

Standard agarose gels were produced by melting 1 - 2.5% (w/v) agarose in 1x TBE. After 10 min mixing on a magnetic stirrer, 0.01% (v/v) ethidium bromide was added and the gel was poured. DNA samples were mixed with 6x DNA loading buffer (5:1 ratio) and loaded on the freshly poured gel. Standard conditions for agarose gel electrophoresis were 120 V for 30 – 120 min at RT. Fluorescent signals of ethidium bromide stained DNA bands were detected using the Molecular Imager ChemiDOC XRS+ Imaging System.

6.2.3.4 Purification of DNA from agarose gels and PCR reactions

DNA extraction from agarose gels was performed with the innuPREP Gel Extraction Kit (Analytik Jena). All steps were made according to the supplier´s instructions.
DNA from PCR reactions was purified and concentrated using the innuPREP PCRpure Kit (Analytik Jena). All steps were made according to the supplier´s instructions.

6.2.3.5 Restriction digest of DNA

Plasmid DNA and DNA fragments were restriction digested using enzymes purchased from Thermo Fisher Scientific (see 6.1.6) according to the supplier´s recommendations for buffer conditions. Standard restriction digestion was carried out for 2 h at 37°C. Restriction enzymes were afterwards inactivated by incubation for 20 min at 80°C.

6.2.3.6 Dephosphorylation of digested DNA

To avoid re-ligation of digested DNA plasmids, 5´phosphate groups were removed by Shrimp Alkaline Phosphatase (SAP). For this, 1 µl of SAP was added directly to inactivated restriction digest preparation and incubated for 1 h min at 37°C.

6.2.3.7 DNA ligation

Restriction digested DNA fragments and digested and dephosphorylated DNA plasmids were ligated in a 10 µl reaction by T4 DNA ligase in a molar ratio of 5:1. Total mass of DNA in the ligation reaction was 200 ng and incubation took place overnight at RT.

6.2.3.8 Site-directed mutagenesis

Site-directed mutagenesis was used to introduce the helicase-dead mutation E335A in the gene. In addition, isoform 1 was converted to isoform 2 by this method. Forward and reverse primer containing the desired mutations were used for amplification in separated reactions:

Reagent	Forward reaction	Reverse reaction
Elongase 5x Buffer A	0.5x	0.5x
Elongase 5x Buffer B	0.5x	0.5x
dNTP mix	200 µM	200 µM
Forward primer	0.5 µM	- / -
Reverse primer	- / -	0.5 µM
Template plasmid DNA	20 ng	20 ng
Elongase	0.5 µl	0.5 µl
ddH_2O	To 25 µl	To 25 µl

Oligonucleotides which were used in this study are listed above (see 6.1.9).

Step	Time	Temperature	Cycle
Initial denaturation	5 min	95°C	1x
Denaturation	30 s	95°C	
Annealing	1 min	55°C	9x
Elongation	1 min/kb	68°C	

Both reactions were mixed, 0.5 µl of Elongase was added, and another 20 cycles were performed. Methylated paternal plasmid DNA was digested with DpnI (see 6.2.3.5). Finally, 10 µl of this mixture were transformed in chemical competent *E. coli* DH5α.

6.2.3.9 Restriction-free cloning

Restriction-site independent introduction of small sequences (<500 bp) into plasmid DNA was achieved by restriction-free cloning. For the generation of native sequence megaprimers, HEK293 cDNA was used as template for adding desired flaking regions by PCR. Artificial motif megaprimer were produced with purchased oligonucleotides and PCR.based buildup of DNA stretches.

Megaprimer were introduced into the plasmids in a 20 µl reaction with Phusion polymerase using a 2-step PCR amplification (Template:megaprimer ratio ~ 1:20):

Recipe			
Phusion High-Fidelity PCR Master Mix (2x)		1x	
Megaprimer		100 ng	
Template DNA		100 ng	
ddH_2O		To 20 µl	
Step	**Time**	**Temperature**	**Cycle**
Initial denaturation	30 s	98°C	1x
Denaturation	8 s	98°C	20x
Elongation	1 min/kb	72°C	
Final elongation	5 min	72°C	1x
* depending on melting temperature of primer sets			

Paternal plasmid was digested with DpnI for 2 h at 37˚C. 10 µl of the reaction mix was transformed into *E. coli* DH5α cells.

6.2.3.10 Gateway cloning

Gateway cloning of reporter gene constructs was performed with the commercial BP an LR clonase kits according to the manufacturer´s instructions.

6.2.3.11 Single-colony Polymerase Chain Reaction (scPCR)

PCR-based screening for positive colonies was performed with homemade Taq polymerase. Results were analyzed by agarose gel electrophoresis. Oligonucleotides which were used in this study are listed above (see 6.1.9).

Recipe	
10x PCR buffer for homemade Taq	1x
dNTP mix	200 µM
Each primer	0.5 µM
Template	Single colony
Homemade Taq	0.5 µl
ddH_2O	To 20 µl

Step	Time	Temperature	Cycle
Initial denaturation	10 min	95°C	1x
Denaturation	20 s	95°C	
Annealing	15 s	53 - 58°C*	30x
Elongation	1 min/kb	72°C	
Final elongation	5 min	72°C	1x

* depends on primer set

6.2.3.12 DNA sequencing

Sequencing of DNA fragments or plasmids was performed by the company Eurofins Genomics. Premixed samples were prepared according to the provider´s instructions. Sequencing result were analyzed with the BioEdit software.

6.2.4 Protein biochemistry

6.2.4.1 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Discontinuous and denaturating sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out to separate proteins based on their molecular weight. By default, 10% PAA separating and 4% PAA stacking gels were used each containing 0.1% SDS. Polymerization of acrylamide monomers was achieved by addition of 1:100 APS and 1:1000 TEMED. Samples were mixed with 6x SDS loading dye to get a 1x solution and denatured for 5 min at 95°C. SDS-PAGEs were carried out for ~2-3 h at 80-120 V in 1x SDS running buffer.

6.2.4.2 Western blot

For detection of specific proteins after separation by SDS-PAGE (see 6.2.4.1), Western blotting was performed. Two Whatman paper of the size and shape of the gel were wetted in Western blotting buffer and placed onto the anode of a semi-dry blotting system. Then, the pre-wetted Protran BA83 Membrane was stacked on top followed by the SDS gel and another two wetted Whatman paper. Semi-dry electro blotting was performed for 1 h and 15 min. Applied amount of mA was calculated by multiplying length of the gel x width of the gel x 2. Afterwards, the membrane was blocked in TBS-T + 5% milk powder for 1 h to avoid unspecific binding of the antibodies.

After one wash with TBS-T for 5 min, the membrane was incubated overnight with the primary antibody solution at 4°C under constant agitation. Next morning, the membrane was washed again 3x with TBS-T and incubated with the secondary antibody solution

for 1.5 h at RT under constant agitation. After the final 3 washing steps, a solution consisting of 225 mM luminol, 9 mM cumaric acid, and 0.2% H_2O_2 was applied to the membrane. Chemiluminescent signals were detected using the Molecular Imager ChemiDOC XRS+ Imaging System.

6.2.4.3 Bradford assay

To determine the total protein concentration of a protein solution, 998 µl of 1x Bradford solution (Roth) was mixed with 2 µl of the sample. Bradford standard curve was determined by known concentration of bovine serum albumin (BSA). Comparison of the measured absorbance at 595 nm determined the protein concentration of the sample.

6.2.4.4 Small-scale purification of DHX36

To purify DHX36 on a small scale for microscale thermophoresis experiments (see 6.2.6.7), DHX36 overexpressing cells were grown on 5 150-mm dishes to 70% confluency. Transgene expression was induced by addition of 500 ng ml^1 tetracycline for 36 h. For harvesting, growth media was decanted, and cell were washed 3x with ice-cold PBS. Next, cells were scraped off the dishes in 1 ml ice-cold PBS per plate using a rubber policeman. After pelleting the cells by centrifugation (5 min, 300 g, 4°C), lysis was performed by resuspension in 10 ml NP-40 lysis buffer and incubation on ice for 30 min. Crude lysates were sheared 10x through a 26-gauge needle clarified by 15 min centrifugation at 20,000 g and 4°C. Afterwards, 200 µl ml^{-1} FLAG-M2 antibody conjugated to magnetic Dynabeads^{TM}Protein G were added. Antigen capture was performed for 5 h under constant agitation on a rotating wheel at 4°C. Magnetic beads were collected on a magnetic rack and washed 3x with NP-40 lysis buffer, 3x with IP-wash buffer. FH-DHX36 was eluted in 600 µl elution buffer for 10 min at 37°C.

6.2.5 RNA biochemistry

6.2.5.1 RNA purification from human cells

For the purification of total RNA from human cells commercial TRIzol solution was used according to the supplier's protocol. In detail, cells were taken up in suitable amounts of TRIzol directly from the plate. After 5 min incubation on RT 200 µl chloroform per ml TRIzol were added and vortexed for 15 s followed by 3 min incubation at RT. Phase separation was achieved by centrifugation (4°C, 15 min, 12,000 g). The RNA-containing upper aqueous phase was transferred to a new tube and mixed with 0.5 ml isopropanol per ml TRIzol. RNA precipitation was completed by incubation for 20 min on ice. RNA was pelletized (4°C, 10 min, 20,000 g), washed with 1 ml cold 75% ethanol, and air-dried for 5 min. The resulting pellet was resolved in a desired volume of ddH_2O. RNA concentration was determined on a Nanodrop device.

6.2.5.2 Reverse transcription

To carry out reverse transcription, the SuperScript™ III Reverse Transcriptase protocol was used according to the supplier´s instructions. Oligo d(T) primer were used for the generation of cDNA of polyadenylated mRNAs. 4 µg of total RNA, isolated with Trizol, were used (see 6.2.5.1).

6.2.5.3 5´end-labeling of RNA

Purchased RNA was 5´end labeled with radioactivity to use it as size markers for urea gel electrophoresis. For this, a mix of 20 pmol of the oligonucleotide and 1 U of polynucleotide kinase (PNK) in commercial 1x PNK buffer was prepared. 25 µCi radioactive γ -ATP were added and the preparation was incubated for 30 min at 37°C. Labeled size markers were then separated on a urea PAGE, excised, and eluted from the gel pieces.

6.2.5.4 Urea-Polyacrylamide gel electrophoresis

Denaturing urea polyacrylamide gel electrophoresis (urea-PAGE) was performed to separate RNA based on their sizes. For small RNA fragments (20 – 80 nt), 12 - 15% PAA separating gels with 8 M urea in 1x TBE were poured. Polymerization of acrylamide monomers was achieved by addition of 1:100 APS and 1:1000 TEMED. Gels were pre-run for 30 min at 30 W. Samples were mixed with 2x formamide loading dye to get a 1x solution and denatured for 5 min at 95°C. Urea-PAGEs were carried out for 60 – 90 min at 30 W in 1x TBE buffer. After separating labeled RNA on a urea PAGE, a blanked phosphor storage screen was exposed to the gel for 1 – 16 h. Signals were detected using a Typhoon FLA 7000 device.

6.2.5.5 Purification of RNA from urea gels

Labeled RNA of desired size was localized on the gel by aligning a printout of the image derived from an exposed phosphor storage screen (see 6.2.5.4). Gel pieces containing the RNA were excised, transferred to a microfuge tube, and smashed using a pestle. 400 µl 0.3% of NaCl were added and incubated overnight at 4°C under constant agitation. Remaining gel pieces were removed by filtration. RNA was precipitated in 70% EtOH.

6.2.5.6 Reverse transcription and quantitative Polymerase Chain Reaction (RT-qRT-PCR)

To analyze cellular RNA levels total RNA was isolated (see 6.2.5.1). 500 ng of this RNA was transcribed into cDNA using the commercial Quantitect Reverse Transcription Kit (Qiagen) according to the manufacturer´s instructions. Obtained cDNA was diluted 1:20 with ddH_2O.PCR was carried out in triplicates on a CFX96 RealTime System with primer against U6 snRNA as normalization.

Recipe	
IQ SYBR Green 2x Mastermix	6 µl
Diluted cDNA	3 µl
Primer Mix (1 µM Forward & Reverse Primer)	3 µl

Step	Time	Temperature	Cycle
Initial denaturation	3 min	95°C	1x
Denaturation	15 s	95°C	
Annealing	15 s	57.5°C	40x
Elongation	15 s	72°C	

6.2.6 Molecular biological experiments

6.2.6.1 Polysome profiling

HEK293 wildtype cells were grown on a 150-mm cell culture dish to 90% confluency. Media was change to growth media containing 25 µg ml^{-1} cycloheximide and cells were incubated for 10 min to stall ribosomes. Cells were washed once with ice-cold PBS and 100 µl of polysome lysis buffer were added (note: for samples used for RNase-treated lysates, no SUPERaseIn was added). Cells were scraped of the dish and transferred to a pre-chilled 1.5 microcentrifuge tube followed by 10 min incubation on ice. Lysate was cleared by 10 min centrifugation at 20,000 g and 4°C. For gradient preparation, 5% and 45% sucrose solutions, both in 1x polysome buffer, were prepared and sterile filtered. SV60 rotor tubes were filled half with the 5% solution. 45% solution was carefully layered beneath the 5% solution. Gradual mixing was facilitated by rotating (56 s, angle: 85°) on a Beckman´s Gradient Maker. Clarified lysate was loaded onto the 5 - 45% linear sucrose gradient and centrifuged for 60 min in a SW60Ti rotor (Beckman) at 38,000 rpm and 4°C. Gradients were fractionated using a Gradient fractionator (Biocomp) and UV profile (254 nm) was measured in parallel. Obtained fractions were further analyzed by standard Western blotting (see 6.2.4.1. and 6.2.4.2).

6.2.6.2 Oligo d(T) pulldown

To verify mRNA association of DHX36, polyadenylated cellular RNA was isolated and bound proteins were analyzed. For this, HEK293 wildtype cells were grown on two 150-mm cell culture dishes to 90% confluency. Growth medium was aspirated, and cells were washed with ice-cold PBS. One dish was kept untreated whereas the other one was crosslinked by irradiation with 0.15 J cm^{-2} 254 nm UV-light. Cells were scraped off the dishes using a rubber policeman and collected by centrifugation. Cell pellets were resuspended in 1.5 ml LiDS lysis buffer. For cell shearing, lysates were passed 3x through a 26-gauge needle, followed by 10 min incubation on ice. Input samples were taken and oligo d(T) magnetic beads, equilibrated in lysis buffer, were added to the lysates. At 4°C, capture of polyadenylated RNA was performed for 1 h under constant agitation. Beads were collected on a magnetic rack and washed twice with wash buffer 1, wash buffer 2, and wash buffer 3, respectively. Elution was achieved by incubation with 100 µl elution buffer for 3 min at 55°C. Eluate was concentrated using a Speedvac Concentrator (Eppendorf) and mRNA binding of proteins was analyzed by standard Western blotting (see 6.2.4.1. and 6.2.4.2).

6.2.6.3 Subcellular fractionation of human cells

Human HEK293 cells were fractionated as previously described[268] with minor changes to check for localization of DHX36 isoforms. Stable cell lines as well as wildtype cells were grown on 150-mm cell culture dishes to 80% confluency. Growth medium was aspirated, and cells were washed with ice-cold PBS. Expression of transgenic FLAG/HA-DHX36-Iso1 and -Iso2 was induced by addition of 500 ng ml^{-1} tetracycline for 16 h. Growth medium was aspirated, and cells were washed with ice-cold PBS. Cells were scraped off the 150-mm cell culture dish and collected by centrifugation. Pelleted cells were resuspended in 1 ml of hypotonic lysis buffer (HLB) per 75 mg cell pellet. After 10 min incubation on ice, cell lysates were briefly vortexed followed by 8 min centrifugation at 800 g and 4°C. The cytoplasmic fractions (supernatant) were carefully transferred to new tubes and 5 M NaCl were added to a final concentration of 150 mM. The remaining nuclear fractions (pellet) were gently washed 4x with HLB to remove all remaining cytoplasmic impurities. Then, pellets were resuspended in nuclear lysis buffer (NLB) and sonicated for 2 cycles (40% power, 30 s ON, 2 min OFF). Both, the cytoplasmic and the nuclear fractions were 15 min centrifuged at 18,000 g and 4°C to remove all insoluble debris. Supernatants were analyzed by standard Western blotting (see 6.2.4.1. and 6.2.4.2). Used markers for subcellular compartments: nuclear = anti-Histone 2B, cytosolic = anti-α-Tubulin, endoplasmic reticulum membrane = anti-Calnexin.

6.2.6.4 Transcriptional shut-off mRNA stability assay

HEK293 wildtype cells and HEK293 DHX36-KO cells were grown on 35-mm dishes to 80% confluency. At timepoint T=0, media was changed to media containing 3 µg ml^{-1} actinomycin D. Cells were harvested and lysed at timepoints 0 h, 2 h, 4 h, and 8 h after addition of actinomycin D by adding TRIzol directly to the plate. RNA was isolated (see 6.2.5.1) and mRNA levels were analyzed by reverse transcription followed by standard quantitative PCRs (see 6.2.5.6).

6.2.6.5 Reporter assay

Cell lines with stably integrated pcDNA5-FRT-GFP-mCherry-3pGW reporter constructs were treated with 500 ng ml^{-1} tetracycline (Merck) for 16 h to induce expression of mCherry protein. Cells were trypsinized, washed twice with PBS, and filtered through a 35 µm nylon net. Detection and analysis of the fluorescent signals of mCherry and AcGFP were performed on a LSRFortessa™ system (BD Biosciences). Triplicates of 100,000 double-positive cells were analyzed.

6.2.6.6 Stress induction with dsRNA

Wildtype and DHX36-KO cells were grown on 6-well dishes until 90% confluency. For positive control of phosphorylation of EIF2AK2, both cell lines were transfected with 1.5 µg of dsRNA using Lipofectamine 2000. 6 h post transfection, cells were harvested and lysed in NP40 lysis buffer on ice for 10 min. Lysates were cleared by centrifugation (15 min, 20,000 g, 4°C) and analyzed by standard Western blotting (6.2.4.2).

6.2.6.7 Microscale thermophoresis (MST)

For microscale thermophoresis, following binding reactions were prepared in 1x MST buffer, supplemented with 0.5% BSA and 5 mM DTT in a total volume of 40 µl: A constant concentration of 25 nM 5´Cy5-labeled and folded oligonucleotides were used. A 1:1 serial dilution of DHX36 (see 6.2.4.4), with 50 nM as highest concentration was prepared and added. In standard treated capillaries microscale thermophoresis analysis was performed with 80% LED, 20% MST power, on the Monollith NT.115. Figures were made with Hill fit using the MO Affinity analysis software.

6.2.7 Systems-wide experiments

6.2.7.1 PAR-CLIP

For determination of DHX36 RNA binding targets Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) was performed with minor modifications as described previously[269]. Important details are described in the following. HEK293 DHX36-Iso1, -Iso2, -Iso1-E335A, and Iso2-E335A were cultured on 15 150-mm cell culture dishes to 80% confluency. Transgene expression was induced by adding 500 ng ml^{-1} tetracycline for 16 h. For the same time range, cells were fed with 100 µM of 4-thiouridin (4SU). For harvesting, growth media was carefully decanted, and the cells were rinsed with ice-cold PBS. Next, RNPs were crosslinked by irradiation with 365 nm UV light for 5 min and scraped off the dishes in 1 ml ice-cold PBS per plate using a rubber policeman. After pelleting the cells by centrifugation, lysis was performed by resuspension in 7 ml NP-40 lysis buffer and incubation on ice for 12 min. Crude lysates were clarified by 15 min centrifugation at 20,000 g and 4°C. Lysates were treated with 1 U $µl^{-1}$ RNase T1 for 15 min at 22°C. Afterwards, 75 µl ml^{-1} FLAG-M2 antibody conjugated to magnetic Dynabeads™Protein G were added. Antigen capture was performed for 105 min under constant agitation on a rotating wheel at 4°C. Magnetic beads were collected on a magnetic rack and washed 3x with NP-40 lysis buffer. A second RNase T1 digestion with 10 U $µl^{-1}$ was performed for 15 min at 22°C to trim the unprotected parts of the co-captured RNA. 3´ends of the RNA fragments were dephosphorylated using 0.5 U $µl^{-1}$ calf intestinal phosphatase for 10 min at 37°C and 400 rpm. Co-immunoprecipitated RNA was radioactively 5´end-labeled using 1 U $µl^{-1}$ T4 polynucleotide kinase and 0.5 µCi $µl^{-1}$ ^{32}P-γ-ATP for 30 min at 37°C. Almost complete 5´end phosphorylation of RNA was accomplished by adding an excess of non-radioactive ATP up to 100 µM for 5 min. So-treated RNA-protein complexes were separated by SDS-PAGE on a commercial, precasted NuPAGE™ 4-12% Bis-Tris Protein Gel. For detection of immunoprecipitated and radiolabeled DHX36-RNA complexes, a blanked storage phosphor screen was exposed to the gel for 1 h. An exact print out of this was used to localize and excise DHX36-RNA complexes of expected size. The resulting gel pieces were crushed and RNPs were extracted by passive elution in 1x proteinase K buffer overnight at 4°C. Protein components of the complexes were digested with 5 mg ml^{-1} proteinase K for 2 h at 55°C. Released RNA fragments were isolated by phenol-chloroform-isoamyl alcohol (25:24:1) extraction and precipitated in ethanol using GlycoBlue as carrier. RNA fragments were resolved in 10 µl ddH_2O and used to generate a small RNA cDNA library preparation protocol as previously described[270]. In detail, preadenylated, barcoded 3´adapter oligonucleotides were ligated to isolated and labeled RNA as well as 19 nt and

35 nt radiolabeled RNA size marker (see 6.2.5.3). For this, 1 µl of 100 µM 3´adapter, 6 µl of 50% PEG8000, and 2 µl of 10x RNA ligase buffer were added. After denaturation for 45 s at 95°C, 1 µl of Rnl2 (1-249) K227Q ligase was added to the chilled reaction mixture. Ligation was performed overnight on ice. Next day, ligation reactions were separated on a 15% Urea PAGE (see 6.2.5.4), ligated RNA fragments detected with a phosphor imager screen, and excised (see 6.2.5.5). Elution of RNA fragments in 0.3 M NaCl was followed by ethanol precipitation. 5´adapter were ligated using RNA ligase 1 for 1 h at 37°C, accordingly. 5´- and 3´ligated RNA fragments were reverse transcribed using the commercial Superscript III system. Optimal PCR amplification cycles for the library were determined by pilot PCR.

So-obtained PAR-CLIP cDNA libraries were sequenced on an Illumina HiSeq 2500 platform. Sequencing reads were mapped against the human genome version hg19. Clusters of overlapping sequences were generated using the PARalyzer software[271] incorporated into a pipeline (PARpipe; https://ohlerlab.mdc-berlin.de/software/PARpipe_119/) with default settings. Binding sites were categorized using the Gencode GRCh37.p13 GTF annotation (gencode.v19.chr_patch_hapl_scaff.annotation.gtf, http://www.gencodegenes.org/releases/19.html).

6.2.7.2 Sample preparation for RNA sequencing

In HEK293 DHX36-KO DHX36 rescue cells and HEK293 DHX36-KO DHX36-E335A rescue cells, transgene expression was induced by addition of 500 ng ml^{-1} tetracycline. RNA from the induced before-mentioned cell lines, wildtype HEK293 cells, and HEK293 DHX36-KO cells was isolated with TRIzol (see 6.2.5.1). Triplicates were used for each cell line. Depletion of ribosomal RNA was accomplished by using the commercial NEBNext rRNA Depletion Kit following the supplier´s protocol.

RNA sequencing was performed as described below (see 6.2.7.5).

6.2.7.3 Isolation and preparation of chromatin-associated RNA for sequencing

HEK293 wildtype cells and HEK293 DHX36-KO cells were grown in triplicates on 150-mm cell culture dishes. After reaching 80 - 90% confluency, cells were trypsinized and collected by centrifugation. Obtained pellets were washed once with ice-cold PBS and transferred to pre-chilled 1.5 ml microcentrifuge tube. For lysis of the cell membrane, cells were resuspended in 400 µl ice-cold Cytoplasmic Fraction Buffer and incubated on ice for 5 min. Crude lysates were layered on a 400 µl sucrose buffer cushion and centrifuged for 20 min (20,000 g, 4°C). The supernatant, consisting of the cytoplasmic fraction, was collected and the pellet was resuspended in 200 µl ice-cold Nuclear Lysis

Buffer. 200 µl ice-cold 2x NUN-Buffer were added dropwise, followed by a pulsed vortexing and incubation on ice for 20 min to lyse the nuclear membrane. After 30 min centrifugation at 20,000 g and 4°C, resulting supernatant (= nucleoplasmic fraction) was collected. The pellet, consisting of the chromatin, was washed with 1 ml Nuclear Lysis Buffer and resuspended in 1 ml TRIzol. These chromatin solutions were heated to 65°C and run through a 26-gauge needle until no insoluble objects were visible. Chromatin-associated RNA was isolated according to the manufacturer´s instructions for Trizol-based RNA isolation (see also 6.2.5.1).
RNA sequencing preparation and analysis was performed as described below (see 6.2.7.5).

6.2.7.4 Sample preparation for cPDS-treated RNA sequencing

In a 6-well plate, HEK293 wildtype cells were grown to approximately 50% confluency. In triplicates, growth media was changed to media containing 2 µM carboxypyridostatin (cPDS). After 42 h incubation, cells were collected, and total RNA was isolated using TRIzol according to the manufacturer´s instructions (see 6.2.5.1). Depletion of ribosomal RNA was accomplished by using the NEBNext rRNA Depletion Kit according to the supplier´s protocol.
RNA sequencing preparation and analysis was performed as described below (see 6.2.7.5).

6.2.7.5 RNA-sequencing

For RNA-sequencing, cDNA preparation of the resulting rRNA-depleted RNA was done with the NEBnext Ultra Directional RNA Library Prep Kit for Illumina. Enrichment of the cDNA was performed using indexed primers of the NEBNext Multiplex Oligos for Illumina and sequencing was completed on a HiSeq 2500 platform. Sequencing reads were aligned to the hg19 human genome using Tophat 2[272]. Cufflinks was used to quantify reads on the UCSC hg19 annotation set differential expression was determined by Cuffdiff[272].

6.2.7.6 Ribosome profiling

Ribosome footprinting was performed as previously published[273]. In triplicates, HEK293 wildtype cells and HEK293 DHX36-KO cells were grown to ~80% confluency on 100-mm cell culture dishes. Growth media was changed to growth media containing 100 µg ml^{-1} cycloheximide to stall ribosomes and stop translation. After 30 s media was aspirated and dishes were immediately chilled on ice. After washing with ice-cold PBS, 400 µl ribosome footprinting buffer were added. Cells were scraped off the dish using a rubber policeman and collected in pre-chilled 1.5 ml microcentrifuge tubes. After 10 min incubation on ice, lysates were triturated by passing 10x through a 26-gauge needle and

cleared by 10 min centrifugation at 20,000 g and 4°C. 300 µl of cell extract was treated with 2.5 U μl^{-1} RNase I for 45 min at RT to digest unbound RNA. RNase digestion was stopped by adding 0.65 U μl^{-1} SUPERaseIn. Extracts were centrifuged through a 900 µl sucrose cushion for 4 h at 70,000 rpm and 4°C in a TLA 100.3 rotor. The supernatant was carefully aspirated, and ribosome-containing pellets were resuspended in 150 µl of ribosome footprinting buffer supplemented with 20 U ml^{-1} SUPERaseIn. RNA fragments were purified by phenol-chloroform extraction and precipitated with ethanol. After washing twice with 75% ethanol, RNA pellets were air-dried and resolved in 15 µl DEPC-treated water. Small RNA cDNA libraries for next-generation sequencing were prepared as previously described[270] with following modifications: 500 ng RNA were dephosphorylated using 0.7 U μl^{-1} calf intestinal phosphatase for 30 min at 37°C. Afterwards, samples were size-selected on a 15% urea-PAGE for 20-35 ribonucleotide long fragments (see also 6.2.5.4). Elution of RNA fragments in 0.3 M NaCl was followed by ethanol precipitation (see 6.2.5.5). RNA fragments were ligated with preadenylated, barcoded 3´adapters. Next, RNA was precipitated, washed and phosphorylated at the 5´end with 5 U μl^{-1} T4 polynucleotide kinase in 1x T4 DNA ligase buffer (20 µl reaction volume) for 30 min at 37°C. Samples were separated on a 15% urea-PAGE for selection of 3´-ligated fragments. Extracted, precipitated, and washed RNA was ligated with 5´adapters and samples were separated on a 12% urea-PAGE for selection of 5´- and 3´ligated fragments. After reverse transcription and cDNA library enrichment, samples were sequenced on an HiSeq 2500 platform.

After sequencing the reads were aligned to the human genome version hg19 using TopHat[272] and quantified on RNA defined in the UCSC hg19 annotation database using Cufflinks[272]. Overlaps of DHX36 cluster and different genomic regions were calculated with BEDTools[274].

6.2.7.7 Stable isotope labeling by amino acids in cell culture (SILAC)

HEK293 wildtype cells and HEK293 DHX36-KO cells were grown in duplicates to approximately 50% confluency on 35-mm cell culture dishes. Then, cells were fed with light media. 24 h later, cells were washed twice with pre-warmed PBS and wildtype cells were changed to medium-heavy media. KO cells were changed to heavy media. 24 h later, cells were washed twice with pre-warmed PBS and collected in 100 µl NP-40 lysis buffer. After incubation for 10 min on ice, lysates were cleared by 15 min centrifugation at 20,000 g and 4°C. Protein concentration was assessed by standard Bradford assay.

For in-gel digestion proteins were reduced and alkylated prior to SDS-PAGE by heating the cell lysates for 10 min at 70°C in NuPAGE LDS sample buffer supplemented with 50 mM DTT. Samples were alkylated by adding 120 mM iodoacetamide Simply Blue. Whole

lanes were cut into 15 bands. The bands were destained with 30% acetonitrile, shrunk with 100% acetonitrile, and dried in a vacuum concentrator. Digests with 0.1 µg trypsin per gel band were performed overnight at 37°C in 50 mM ammonium bicarbonate buffer. Peptides were extracted from the gel slices with 5% formic acid. NanoLC-MS/MS analyses were performed on an Orbitrap Fusion equipped with an EASY-Spray Ion Source or a PicoView Ion Source (New Objective) and coupled to an EASY-nLC 1000. Using the Easy-Spray Ion Source the peptides were loaded on a trapping column (2 cm × 75 µm ID, PepMap C18, 3 µm particles, 100 Å pore size) and separated on an EASY-Spray column (25 cm × 75 µm ID, PepMap C18, 2 µm particles, 100 Å pore size). Using the PicoView Ion Source, the peptides were loaded on capillary columns (PicoFrit, 30 cm x 150 µm ID, New Objective) self-packed with ReproSil-Pur 120 C18-AQ, 1.9 µm (Dr. Maisch). A 60-minute linear gradient from 3% to 40% acetonitrile and 0.1% formic acid was used. Both MS and MS/MS scans were acquired in the Orbitrap analyzer with a resolution of 60,000 for MS scans and 15,000 for MS/MS scans. HCD fragmentation with 35% normalized collision energy was applied a Top Speed data-dependent MS/MS method with fixed cycle time of 3 s was used. Dynamic exclusion was applied with a repeat count of 1 and exclusion duration of 120 s; singly charged precursors were excluded from selection. Minimum signal threshold for precursor selection was set to 50,000. Predictive AGC was used with an AGC target value of 5e5 for MS scans and 5e4 for MS/MS scans. EASY-IC was used for internal calibration.

For MS raw data file processing, database searches and quantification, MaxQuant version 1.5.7.4[275] was used. UniProt human reference proteome database was used in combination with a database containing common contaminants as a reverse concatenated target-decoy database. Protein identification was under control of the false-discovery rate (<1% FDR on protein and peptide level). In addition to MaxQuant default settings, the search was performed with tryptic cleavage specificity with three allowed missed cleaves. The search was performed with the following variable modifications: Gln to pyro-Glu formation and oxidation (on Met). Normalized H/M ratios were used for protein quantitation (at least two peptides per protein).

6.2.8 Fluorescence microscopy

6.2.8.1 Fluorescent staining of cellular G4 structures

BG4 antibody was expressed and purified as described previously[151]. Sterile coverslips were coated with 100 µg ml^{-1} poly-D lysine solution for 1 h at 37°C. Cells were seeded on these coverslips in 6-well plates with low confluency. If applied, treatment with 2 µM cPDS was performed 24 h before fixation. Cells were pre-fixed with a solution of 50% DMEM and 50% methanol/acetic acid (3:1) at RT for 10 minutes. After a brief wash with methanol/acetic acid (3:1), cells were fixed with methanol/acetic acid (3:1) at RT for 10 min. Next, cells were permeabilized with in PBS supplemented with 0.1% Triton X-100 at RT for 3 minutes. If RNase treatment was applied, coverslips were incubated with 100 µg ml^{-1} RNase A for 1 h at 37°C. Cells were blocked in blocking solution (2% milk in PBS, pH 7.4) for 1 h at RT and then incubated with 2 µg per slide of BG4 and Anti-alpha 1 Sodium Potassium ATPase antibodies in blocking solution for 2 h at RT. Cells were then incubated with 1:800 of a rabbit antibody against the FLAG-tag of the BG4 antibody in blocking solution for 1 h. Next, cells were incubated at RT with 1:1000 Alexa Fluor 488 goat anti-mouse IgG and Cyanine 3 goat anti-rabbit IgG in blocking solution for 1 h. After each step, cells were washed three times for 5 minutes with 0.1% Tween-20 in PBS under gentle rocking. The cover glasses were mounted with Fluoroshield mounting media containing DAPI for nuclear staining. Slides were visualized at RT by using a confocal microscope (Leica SP5 AOBS).

6.2.8.2 Detection of stress granules

Sterile coverslips were coated with 100 µg ml^{-1} poly-D lysine solution for 1 h at 37°C. Cells were seeded on these coverslips in 6-well plates with low confluency. 48 hours post seeding cells were pre-fixed with a solution of 50% DMEM and 50% methanol/acetic acid (3:1) at RT for 10 minutes. After washing the cells briefly with methanol/acetic acid (3:1), cells were fixed with methanol/acetic acid (3:1) at RT for 10 min. Cells were then permeabilized with 0.1% Triton X-100 in PBS at RT for 3 minutes. Blocking of the cells was performed in blocking solution (2% milk in PBS, pH 7.4) for 1 h at RT. Incubation with G3BP1 antibody (5 µg ml^{-1} antibody) in blocking solution was done for 2 h at RT. Next, cells were incubated at RT with 1:1000 Alexa Fluor 488 goat anti-mouse IgG. After each antibody stain, cells were washed three times for 5 minutes with 0.1% Tween-20 in PBS under gentle rocking. The cover glasses were mounted with Fluoroshield mounting media containing DAPI for nuclear staining. Slides were visualized at RT by using a confocal microscope (Leica SP5 AOBS).

6.2.9 Others

6.2.9.1 Circular dichroism

Synthetic oligonucleotides were diluted in G4 folding buffer to a molar concentration of 10 µM. Samples were boiled for 10 min at 95°C, followed by subsequent passive cooling to RT. Measurement of the CD spectra was performed using a Jasco J-810 spectropolarimeter in 0.2 ml quartz cuvettes. Measurements in the range of 200-350 nm were recorded. and averaged between ten accumulations with an instrument scanning speed of 10 nm sec^{-1}.

6.2.9.2 5mer Z-scoring

The BEDTools[274] utility *getfasta* was used to recover the genomic nucleotide sequences corresponding to the remaining intervals in these BED files. To produce background, target read sequences were scrambled while preserving GC content. The 5mers for targets and for background, respectively, were counted and calculated Z-score enrichment by proportion, $\mathrm{Z} = \frac{\mathrm{X} - \mu}{\sigma}$. X is the proportion $\frac{count}{total\ counts}$ for a given 5mer. Respectively, μ and σ are the average and standard deviation of 5mer proportions in background.

6.2.9.3 MEME motif search

Analysis of the DHX36 consensus binding motif was carried out using the MEME suite[276] using the top 500 unique cluster sequences defined by PARalyzer[271].

6.2.9.4 Statistical analyses

Statistical parameters are shown in the figures and listed in the figure legends. Statistical significance is claimed when $P < 0.05$. In the figures, asterisks mark statistical significance as follows: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, or ****$P < 0.0001$.

6.2.9.5 Cell image analysis

Fluorescence signal was determined using ImageJ software with the following formula: Corrected Total Cell Fluorescence (CTCF) = Integrate Density - (Area of selected cell x Mean Fluorescence of Background Readings). The cytoplasmic compartment was calculated by subtracting the nuclear signal, obtained by DAPI staining, from the total cell signal, obtained by labeling the cell membrane with the Anti-alpha 1 Sodium Potassium ATPase antibody. For graphical representation of signal distribution, box-and-whisker plots were generated using GraphPad Prism 6 software with the following settings: boxes: 25–75 percentile range; whiskers: 10–90 percentile range. Statistical significance was determined by “Kolmogorov-Smirnov” and “Mann-Whitney” non-parametric tests. Data analyses were performed with Excel and GraphPad, and all figures were prepared with Adobe Illustrator.

7. Appendix

7.1 Supplementary data

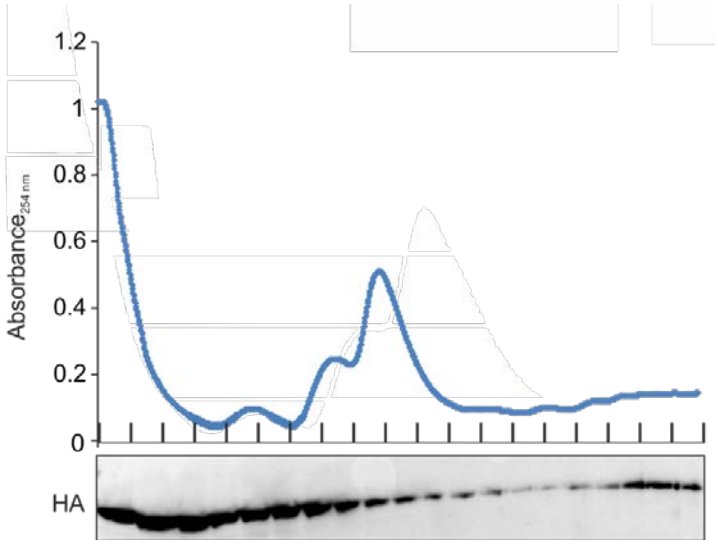

Figure 7-1 Polysome gradient of DHX36 overexpression cells.
UV absorbance at 254 nm of FH-DHX36-1 HEK293 cell extract separated by sucrose gradient centrifugation shows peaks corresponding to 40S, 60S, 80S ribosomes, and polysomes. Western blots of the gradient fractions probed for HA, is shown below.

Table 3: Top 100 DHX36 target mRNAs determined by PAR-CLIP.

Gene ID	Annotated G4s[138]	DHX36 cluster	DHX36 crosslinked reads
ZNF146	Yes	19	74361
ERC1	Yes	53	62956
GLO1	Yes	14	58038
EEF2	Yes	19	39253
G3BP1	Yes	44	34036
DHX36	Yes	27	30909
ZNF711	No	30	27897
HNRNPA2B1	Yes	21	26851
RAD23B	Yes	23	25379
KLHL15	Yes	35	23916
HIST1H4C	No	2	23235
PPP1CC	Yes	12	23010
UQCRH	No	4	22043
NUCKS1	Yes	35	21383
DDIT4	Yes	8	20527
PMAIP1	No	10	20416
RAN	Yes	18	19782
HNRNPA0	Yes	11	18906
DEK	Yes	24	18654
EIF2S3	No	11	18524
JUN	Yes	20	18335
IRS4	No	16	17911
HNRNPH3	Yes	19	16873
CANX	Yes	31	16056
HIST1H4H	Yes	1	16016
KMT2D	No	50	15973
SRSF1	Yes	17	15725
NFKBIB	Yes	8	15383
HECA	Yes	22	15350
CCND2	No	47	15341
SLMO2	Yes	17	14995
FEM1B	No	42	14573
LARP1	Yes	36	14494
MTPN	No	20	14484
PSMA4	Yes	9	14285
PURB	Yes	49	14278
SF3B3	Yes	30	14276
WAC	Yes	29	14177
HIST1H4B	No	3	14120
CD164	No	13	13519
EIF2S2	No	9	13385
MAZ	Yes	15	13381
SYNCRIP	Yes	49	13080
SF1	Yes	15	13028
DYNLL2	Yes	18	12875
ETV3	Yes	24	12863
CCT3	No	14	12752
NAA50	Yes	34	12740
TSC22D3	No	10	12605
ARPC5L	Yes	9	12533
PPP2CA	Yes	23	12450

ZNF664	Yes	20	12436
ANP32A	Yes	15	12407
HNRNPA1	Yes	13	12290
MORF4L2	Yes	18	12262
PTP4A1	Yes	27	12109
MCL1	Yes	24	12063
LUC7L2	Yes	29	12036
SPEN	Yes	54	11641
HNRNPR	Yes	24	11412
LDOC1L	Yes	15	11395
LMNB1	Yes	12	11369
SERBP1	Yes	40	11283
U2AF1	Yes	14	11234
HNRNPH1	Yes	23	11055
POU3F2	No	22	10990
BTG1	Yes	7	10970
CEBPG	Yes	23	10959
MAPK1IP1L	No	22	10934
ARL5B	Yes	50	10836
CSDE1	Yes	31	10827
STC2	Yes	24	10809
PTGES3	No	15	10803
SP1	Yes	44	10769
DDX17	Yes	30	10674
YKT6	Yes	8	10643
ZBTB7A	Yes	18	10576
LAMTOR5	No	7	10573
FAM168B	Yes	28	10554
DNAJC2	Yes	8	10552
FEN1	Yes	14	10543
GTF2H3	No	11	10485
ATL2	Yes	19	10476
PCNP	Yes	15	10340
RFC3	Yes	23	10283
UBE2E1	Yes	14	10138
CNBP	Yes	10	9925
HIST1H4A	Yes	2	9759
ERH	Yes	6	9733
HNRNPD	Yes	15	9700
TMEM123	Yes	21	9658
HNRNPU	Yes	35	9597
DBNDD2	Yes	7	9575
CUTC	Yes	11	9418
CBX6	Yes	20	9320
LDHA	No	12	9228
OAZ1	Yes	9	9225
HNRNPL	Yes	15	9181
PSMD11	Yes	15	9158

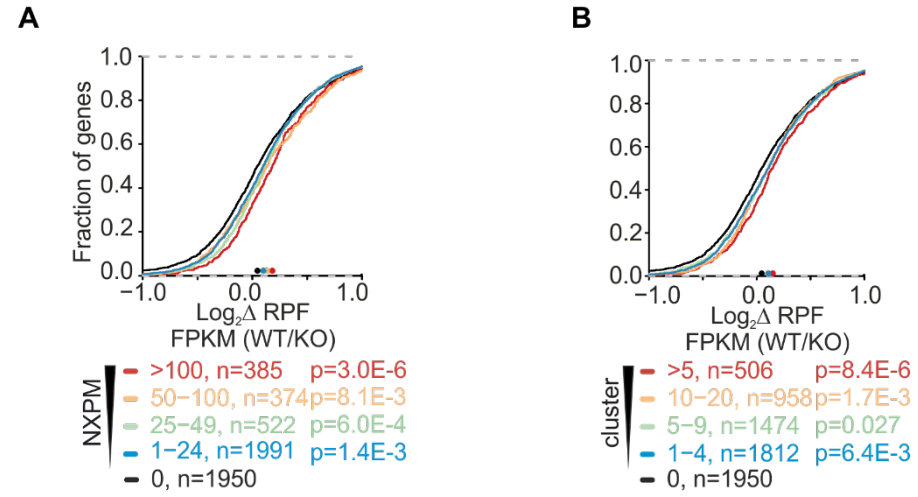

Figure 7-2 Supplementary data to Figure 4-14.

A, A CDF comparing changes in ribosome-protected fragments (RPFs) of DHX36-KO (n=3) and parental HEK293 cells (n=3). Target mRNAs are binned in accordance to the number of NXPM obtained by DHX36 PAR-CLIP. Significance was determined using a two-sided KS-test. **B,** Same as in A, except target mRNAs are binned in accordance to the number of binding clusters obtained by DHX36-E335A PAR-CLIP.

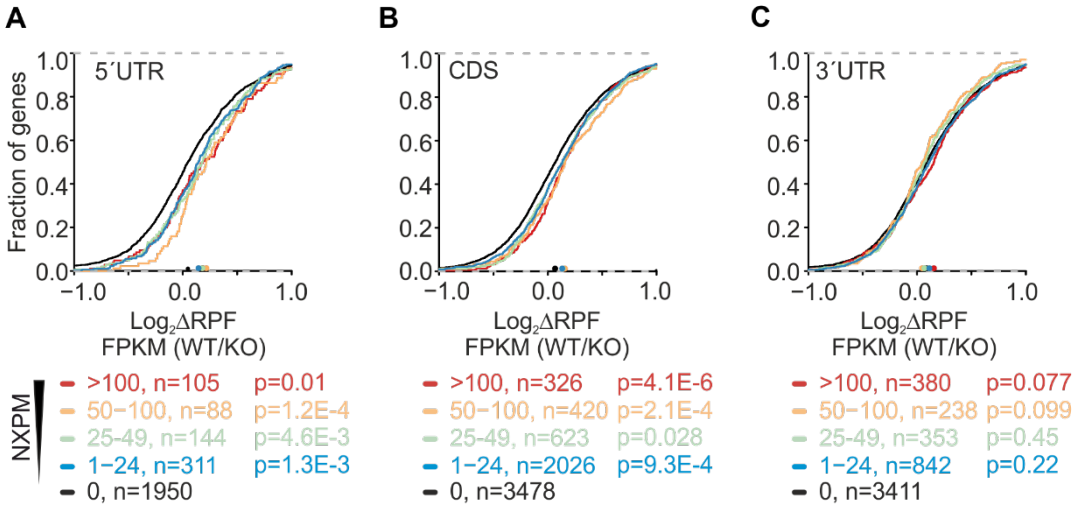

Figure 7-3 Supplementary data to Figure 4-15.

A - C, A CDF comparing changes in ribosome-protected fragments (RPFs) of DHX36-KO (n=3) and parental HEK293 cells (n=3). Target mRNAs are binned in accordance to the number of NXPM in **A,** the 5´UTR, **B,** the CDS, or **C,** the 3´UTR obtained by DHX36 PAR-CLIP. Significance was determined using a two-sided KS-test.

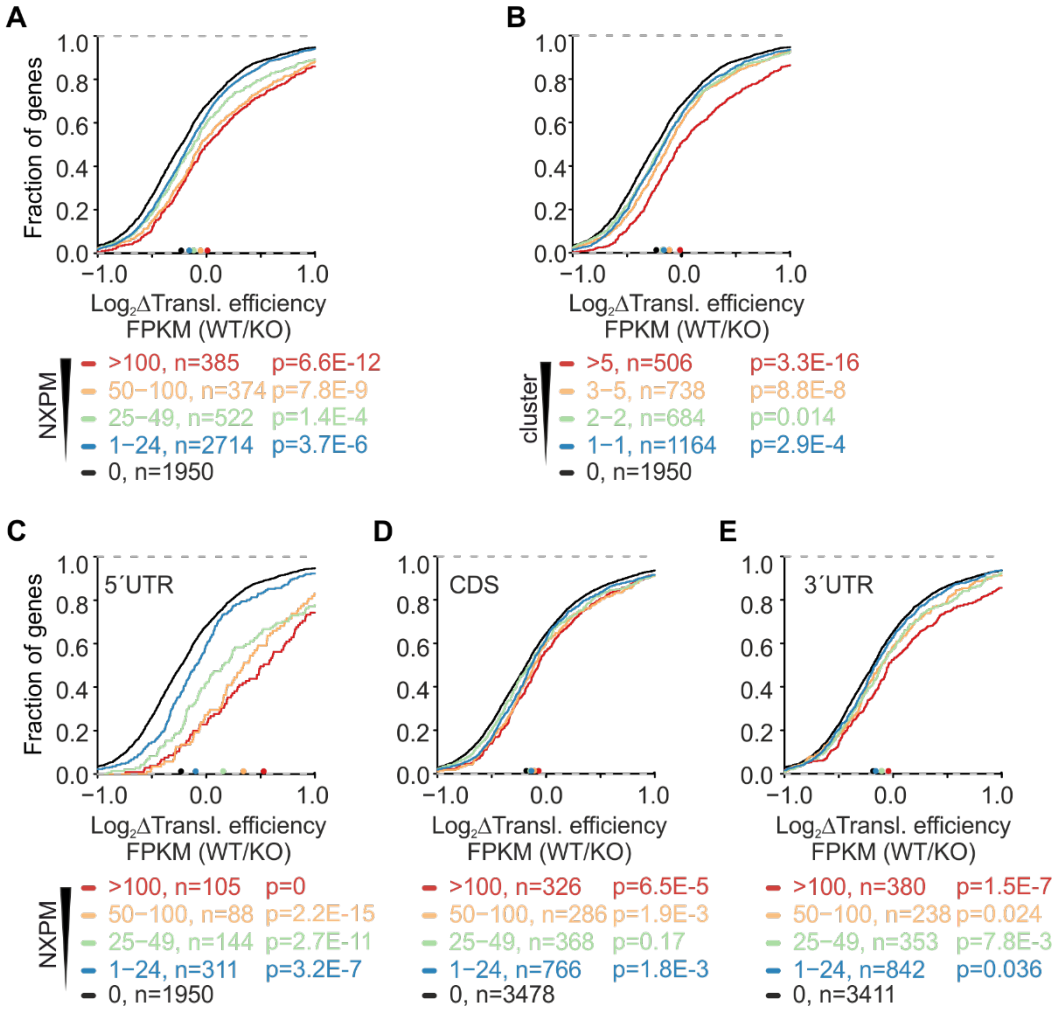

Figure 7-4 Supplementary data to Figure 4-16.

A, A CDF comparing changes in translation efficiency of DHX36-KO (n=3) and parental HEK293 cells (n=3). Target mRNAs are binned in accordance to the number of NXPM obtained by DHX36 PAR-CLIP. Significance was determined using a two-sided KS test. **B,** Same as in **A**, except mRNAs are binned based on the number of clusters. **C - E,** Same as in **A**, except mRNAs are binned based on the number of NXPM in **C,** the 5´UTR, **D,** the CDS, and **E,** the 3´UTR.

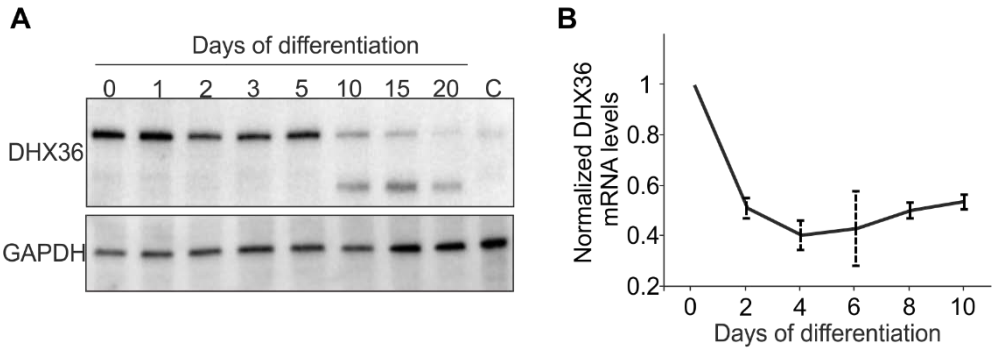

Figure 7-5 Expression of DHX36 during cell differentiation.

A, Western blot analysis of DHX36 protein level during cell differentiation of human stem cells. GAPDH serves as loading control. B, Corresponding DHX36 mRNA levels during cell differentiation.

7.2 Abbreviations

7.2.1 Names

E. coli	*Escherichia coli*
DENV	Dengue virus
G4IPDB	Guanine quadruplex structure interacting protein database
EPV	Epstein-Barr virus
HCV	Hepatitis C virus
HEK293	Human embryonic kidney 293 cell line
HeLa	Cervical cancer cell line
HIV	Human immunodeficiency virus
HPV	Human papillomavirus
PRRSV	porcine reproductive and respirator syndrome virus
S. lemnae	*Stylonychia lemnae*
WNV	West Nile virus
YFV	Yellow fever virus
ZIKV	Zika virus

7.2.2 Scientific terms

A	adenine
A	alanine
aa	amino acid
ARE	AU-rich element
A-site	aminoacyl site
ATP	adenosine triphosphate
bp	base pair
C	carbon
C	cytosine
CDF	cumulative distribution function
cDNA	complementary DNA
ChIP	Chromatin-Immuno-precipitation DNA-sequencing
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CTD	carboxy terminal domain
CTE	carboxy terminal extension
C-terminus	carboxyl terminus
D	aspartic acid
dG4	DNA G-quadruplex structure
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleotides
ds	double-stranded
DSM	DHX36 specific motif
E	glutamic acid
EEF	eukaryotic elongation factor
EIF	eukaryotic initiation factor
ESE	exonic splice enhancer
ESS	exonic splice silencer
E-site	exit site
G	glycine
G	guanine
G4	G-quadruplex structure
Gly	glycine-rich element
GO	Gene Ontology Term Enrichment
GTF	general transcription factor
H	hydrogen
H	histidine
I	isoleucine
IDR	intrinsically disordered protein region
IFN	interferon
IP	immunoprecipitation
IRE	iron responsive element
IRES	internal ribosomal entry site
ISE	intronic splice enhancer
ISS	intronic splice silencer
K	potassium
kb	kilobase

KEGG	Kyoto Encyclopedia of Genes and Genomes
KH	K homology
L	looping nucleotide
lncRNA	long non-coding RNA
m7G	methylated guanine cap
MCS	multiple cloning site
miRNA	micro RNA
mRNA	messenger RNA
mRNP	messenger ribonucleoprotein particle
N	nitrogen
Na	sodium
ncRNA	non-coding RNA
NGD	no-go decay
NMD	nonsense-mediated decay
NPC	nuclear pore complex
NSD	non-stop decay
nt	nucleotide
N-terminus	amino terminus
O	oxygen
OB	oligonucleotide and oligosaccharide-binding-fold-like
OD	optical density
ORF	open reading frame
P	proline
PAGE	polyacrylamide gel electrophoresis
PAMP	pathogen-associated molecular pattern
PAR-CLIP	photoactivatable-ribonucleoside-enhanced crosslinking immunoprecipitation
P-body	processing body
PCR	polymerase chain reaction
pre-miRNA	precursor micro RNA
pre-mRNA	precursor messenger RNA
pri-miRNA	primary micro RNA
PNK	polynucleotide kinase
poly-A signal	poly-adenylation signal
poly(I:C)	polyinosinic:polycytidylic acid
PQS	putative quadruplex-forming sequence
P-site	peptidyl site
PTC	premature termination codon
PTM	posttranslational modification
R	arginine
RBD	RNA binding domain
RBP	RNA binding protein
rG4	RNA G-quadruplex structure
RL	ratchetlike domain
RNA	ribonucleic acid
RNase	ribonuclease
RNP	ribonucleoprotein particle
RPF	ribosome-protected fragment
RRM	RNA recognition motif
rRNA	ribosomal RNA
RT	room temperature
S	serine
scPCR	single colony polymerase chain reaction
SILAC	stable isotope labeling with amino acids in cell culture
siRNA	small interfering RNA
SG	stress granule
SNP	single nucleotide polymorphism
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
ss	single-stranded
T	thymine
T	threonine
TF	transcription factor
tRNA	transfer RNA
TSS	transcription start site
U	uracil
uORF	upstream ORF
UsnRNP	uridine-rich snRNP
UV	ultraviolet light
UTR	untranslated region
WH	winged-helix domain
WT	wildtype
Y	tyrosine

7.2.3 Proteins, nucleic acids, and complexes

40S	Small ribosomal subunit
43S	Translational preinitiation complex
60S	Large ribosomal subunit
80S	Ribosome
ATP1A1	Sodium/Potassium-Transporting ATPase Subunit Alpha-1
ATRX	Transcriptional regulator ATRX
BG4	G4-specific antibody BG4
BLM	Bloom syndrome protein
CaMKIIa	Ca2+/calmodulin-dependent protein kinase II
Cas	CRISPR associated
CNBP	Cellular nucleic acid binding protein
CRE	Cre recombinase
DDX1	ATP-dependent RNA helicase DDX1
DDX3X	ATP-dependent RNA helicase DDX3X
DDX5	ATP-dependent RNA helicase DDX5
DDX11	ATP-dependent RNA helicase DDX11
DDX17	ATP-dependent RNA helicase DDX17
DDX21	ATP-dependent RNA helicase DDX21
Dicer	Dicer endoribonuclease
DHX9	ATP-dependent RNA helicase A
DHX36	ATP-dependent DNA/RNA helicase DHX36, also referred to as RHAU or G4R1
DNA2	DNA replication ATP-dependent helicase/nuclease DNA2
Drosha	Drosha ribonuclease
EE2K	Eukaryotic elongation factor 2 kinase
EEF2	Eukaryotic translation elongation factor 2
EF2K	Elongation factor 2 kinase
EIF2	Eukaryotic translation initiation factor 2
EIF2AK1	Eukaryotic translation initiation factor 2 alpha kinase 1, also known as HRI
EIF2AK2	Eukaryotic translation initiation factor 2 alpha kinase 2, also known as PKR
EIF2AK3	Eukaryotic translation initiation factor 2 alpha kinase 3, also known as PERK
EIF2AK4	Eukaryotic translation initiation factor 2 alpha kinase 4, also known as GCN2
EIF4A	Eukaryotic translation initiation factor 4A; DEAD-box helicase DDX48
EIF4E	Eukaryotic translation initiation factor 4E; Cap-binding protein
EIF4F	Eukaryotic translation initiation factor 4F; complex of EIF4E, EIF4G, and EIF4A
EIF4G	Eukaryotic translation initiation factor 4G; scaffolding protein
EJC	Exon Junction Complex
ERF1	Eukaryotic translation release factor 1
FMRP	Fragile X mental retardation protein 1
FXR1	Fragile X mental retardation syndrome-related protein 1
IFN	Interferon
hnRNP H/F	heterogeneous nuclear ribonucleoprotein H/F
nCBC	Nuclear cap-binding complex
NCL	Nucleolin
NKx2-5	Homeobox protein NKx2-5
NPC	Nuclear Pore Complex

NXF1	Nuclear RNA export factor 1
NXT1	NTF2-related export protein 1
PABPC1	Polyadenylate-binding protein cytoplasmic 1
PABPN1	Polyadenylate-binding protein nuclear 1
PAP	Polyadenylate polymerase
PAX9	Paired box gene 9, transcription factor
PIC	Preinitiation complex
PIF	ATP-dependent helicase PIF1
PIM1	Proto-oncogene serine/threonine-protein kinase
PITX	Paired-like homeodomain 1, transcription factor
PKR	see →EIF2AK2
Pol II	DNA-dependent RNA polymerase II
PSD95	Postsynaptic density protein 95. Encoded by the gene DLG4
P-TEFb	Positive transcription elongation factor b
RAP1	Repressor/Activator protein 1, part of the Shelterin complex on telomeres
RanGTP	GTP-bound Ras-related nuclear protein
RHAU	See → DHX36
RIG-I	Probable ATP-dependent RNA helicase DDX58
RISC	RNA-induced silencing complex
RNAP II	DNA-dependent RNA polymerase II
RTEL	Regulator of telomere elongation helicase 1 homolog
SAP	Shrimp alkaline phosphatase
TEBPβ	Telomere end-binding protein beta; (ciliate)
TEFM	Transcription Elongation Factor, Mitochondrial
TRIF	TIR-domain-containing adapter-inducing interferon-β
VEGF	Vascular endothelial growth factor
WRN	Werner syndrome ATP-dependent helicase
XPB	General transcription and DNA repair factor IIH helicase subunit XPB
XPD	General transcription and DNA repair factor IIH helicase subunit XPD
XPO1	Exportin 1
YY1	YY1 transcription factor

miRNA200c	Human micro RNA 200c
miRNA 331-3p	Human micro RNA 331-3p
miRNA451a	Human micro RNA 451a
miRNA497	Human micro RNA 497
pre-miRNA134	Human premature micro RNA 134

28S	28S ribosomal RNA
7SK	7SK snRNA
TERC	Telomerase RNA component
TERRA	Telomerase repeat containing RNA
$tRNA^{Met}$	Eukaryotic initiator tRNA

7.2.4 Reagents

4SU	4-thiouridine
6SG	6-thioguanosine
AEBSF	4-(2-Aminoethyl)benzensulfonylfluorid
Amp	Ampicillin
AP	Aprotinin
ATP	Adenosine triphosphate
APS	Ammonium persulfate
cPDS	carboxypyridostatin
DAPI	4',6-diamidino-2-phenylindole
7-Deaza-GTP	7-Deaza-2´-deoxy-guanosine-5´triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetate
FBS	Fetal bovine serum
[γ-P32]-ATP	[gamma-P32] Adenosine triphosphate
Kan	Kanamycin
LP	Leupeptin/Pepstatin
PAA	Polyacrylamide
PBS	Phosphate buffered saline
PDS	pyridostatin
PenStrep	Penicillin Streptomycin
$PhenDC_3$	3,3'-((1,10-Phenanthroline-2,9-dicarbonyl)bis(azanediyl))bis (1-methyl- quinolin-1-ium)bis-triflate
PMSF	Phenylmethylsulfonylfluorid
SDS	Sodium dodecyl sulfate
TBE	Tris/Borate/EDTA
TEMED	Tetramethylethylendiamine
TMPyP4	tetra-(N-methyl-4-pyridyl)porphyrin
TRIS	Tris(hydroxymethyl)aminomethane)

7.2.5 Units

A	Ampere	l	litre
Å	Angstrom	M	molarity
°C	degree Celsius	min	minute
Ci	Curie	rpm	revolutions per minute
Da	Dalton	S	Svedberg
g	gram	s	second
g	g-force (acceleration)	U	unit
h	hour	V	Volt
J	joule	W	Watt

7.2.6 Dimensions

M	Mega	(10^6)	µ	micro	(10^{-6})
k	kilo	(10^3)	n	nano	(10^{-9})
m	milli	(10^{-3})	p	pico	(10^{-12})

8. Bibliography

1. Singh, A.J., Ramsey, S.A., Filtz, T.M. & Kioussi, C. Differential gene regulatory networks in development and disease. *Cell Mol Life Sci* **75**, 1013-1025 (2018).
2. Rockman, M.V. & Kruglyak, L. Genetics of global gene expression. *Nat Rev Genet* **7**, 862-72 (2006).
3. Proudfoot, N.J., Furger, A. & Dye, M.J. Integrating mRNA processing with transcription. *Cell* **108**, 501-12 (2002).
4. Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-8 (1953).
5. Eddy, S.R. Computational genomics of noncoding RNA genes. *Cell* **109**, 137-40 (2002).
6. Brenner, S., Jacob, F. & Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190**, 576-581 (1961).
7. Gehring, N.H., Wahle, E. & Fischer, U. Deciphering the mRNP Code: RNA-Bound Determinants of Post-Transcriptional Gene Regulation. *Trends Biochem Sci* **42**, 369-382 (2017).
8. Bentley, D.L. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* **15**, 163-75 (2014).
9. Keene, J.D. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8**, 533-43 (2007).
10. Kohler, A. & Hurt, E. Exporting RNA from the nucleus to the cytoplasm. *Nat Rev Mol Cell Biol* **8**, 761-73 (2007).
11. Berget, S.M., Moore, C. & Sharp, P.A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* **74**, 3171-5 (1977).
12. Chow, L.T., Gelinas, R.E., Broker, T.R. & Roberts, R.J. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1-8 (1977).
13. Cramer, P. et al. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**, 640-9 (2000).
14. Roeder, R.G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* **21**, 327-35 (1996).
15. Kornberg, R.D. The molecular basis of eukaryotic transcription. *Proc Natl Acad Sci U S A* **104**, 12955-61 (2007).
16. Goldman, S.R., Ebright, R.H. & Nickels, B.E. Direct detection of abortive RNA transcripts in vivo. *Science* **324**, 927-8 (2009).
17. Mandel, C.R., Bai, Y. & Tong, L. Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* **65**, 1099-122 (2008).
18. Porrua, O. & Libri, D. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* **16**, 190-202 (2015).
19. Kuehner, J.N., Pearson, E.L. & Moore, C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol* **12**, 283-94 (2011).
20. Shatkin, A.J. Capping of eucaryotic mRNAs. *Cell* **9**, 645-53 (1976).
21. Lewis, J.D. & Izaurralde, E. The role of the cap structure in RNA processing and nuclear export. *Eur J Biochem* **247**, 461-9 (1997).

22. Mazza, C., Ohno, M., Segref, A., Mattaj, I.W. & Cusack, S. Crystal structure of the human nuclear cap binding complex. *Mol Cell* **8**, 383-96 (2001).

23. Gilbert, W. Why genes in pieces? *Nature* **271**, 501 (1978).

24. Black, D.L. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**, 291-336 (2003).

25. Matlin, A.J., Clark, F. & Smith, C.W. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* **6**, 386-98 (2005).

26. UniProt, C. UniProt: a hub for protein information. *Nucleic Acids Res* **43**, D204-12 (2015).

27. Meister, G. & Fischer, U. Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs. *EMBO J* **21**, 5853-63 (2002).

28. Le Hir, H., Gatfield, D., Izaurralde, E. & Moore, M.J. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J* **20**, 4987-97 (2001).

29. Nakielny, S. & Dreyfuss, G. Transport of proteins and RNAs in and out of the nucleus. *Cell* **99**, 677-90 (1999).

30. Carmody, S.R. & Wente, S.R. mRNA nuclear export at a glance. *J Cell Sci* **122**, 1933-7 (2009).

31. Herold, A. et al. TAP (NXF1) belongs to a multigene family of putative RNA export factors with a conserved modular architecture. *Mol Cell Biol* **20**, 8996-9008 (2000).

32. Palacios, I.M. How does an mRNA find its way? Intracellular localisation of transcripts. *Semin Cell Dev Biol* **18**, 163-70 (2007).

33. Grifo, J.A., Tahara, S.M., Morgan, M.A., Shatkin, A.J. & Merrick, W.C. New initiation factor activity required for globin mRNA translation. *J Biol Chem* **258**, 5804-10 (1983).

34. Aitken, C.E. & Lorsch, J.R. A mechanistic overview of translation initiation in eukaryotes. *Nat Struct Mol Biol* **19**, 568-76 (2012).

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

36. Hinnebusch, A.G. Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. *Trends Biochem Sci* **42**, 589-611 (2017).

37. Ben-Shem, A., Jenner, L., Yusupova, G. & Yusupov, M. Crystal structure of the eukaryotic ribosome. *Science* **330**, 1203-9 (2010).

38. Unbehaun, A., Borukhov, S.I., Hellen, C.U. & Pestova, T.V. Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes Dev* **18**, 3078-93 (2004).

39. Kapp, L.D. & Lorsch, J.R. The molecular mechanics of eukaryotic translation. *Annu Rev Biochem* **73**, 657-704 (2004).

40. Rodnina, M.V. & Wintermeyer, W. Recent mechanistic insights into eukaryotic ribosomes. *Curr Opin Cell Biol* **21**, 435-43 (2009).

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

42. Alkalaeva, E.Z., Pisarev, A.V., Frolova, L.Y., Kisselev, L.L. & Pestova, T.V. In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3. *Cell* **125**, 1125-36 (2006).

43. Song, H. et al. The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* **100**, 311-21 (2000).

44. Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763-76 (2009).

45. Makino, D.L., Baumgartner, M. & Conti, E. Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. *Nature* **495**, 70-5 (2013).

46. Maquat, L.E. & Carmichael, G.G. Quality control of mRNA function. *Cell* **104**, 173-6 (2001).

47. van Hoof, A. & Wagner, E.J. A brief survey of mRNA surveillance. *Trends Biochem Sci* **36**, 585-92 (2011).

48. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-60 (1997).

49. Grewal, S.I. & Jia, S. Heterochromatin revisited. *Nat Rev Genet* **8**, 35-46 (2007).

50. Heard, E. et al. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell* **107**, 727-38 (2001).

51. Chadwick, B.P. & Willard, H.F. Barring gene expression after XIST: maintaining facultative heterochromatin on the inactive X. *Semin Cell Dev Biol* **14**, 359-67 (2003).

52. Katan-Khaykovich, Y. & Struhl, K. Heterochromatin formation involves changes in histone modifications over multiple cell generations. *EMBO J* **24**, 2138-49 (2005).

53. Jeltsch, A. & Jurkowska, R.Z. New concepts in DNA methylation. *Trends Biochem Sci* **39**, 310-8 (2014).

54. Jiao, L. & Liu, X. Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science* **350**, aac4383 (2015).

55. Kalb, R. et al. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol* **21**, 569-71 (2014).

56. Soshnev, A.A., Josefowicz, S.Z. & Allis, C.D. Greater Than the Sum of Parts: Complexity of the Dynamic Epigenome. *Mol Cell* **62**, 681-94 (2016).

57. Dupont, C., Armant, D.R. & Brenner, C.A. Epigenetics: definition, mechanisms and clinical perspective. *Semin Reprod Med* **27**, 351-7 (2009).

58. van Holde, K. & Zlatanova, J. Unusual DNA structures, chromatin and transcription. *Bioessays* **16**, 59-68 (1994).

59. Bochman, M.L., Paeschke, K. & Zakian, V.A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet* **13**, 770-80 (2012).

60. Schroth, G.P., Chou, P.J. & Ho, P.S. Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. *J Biol Chem* **267**, 11846-55 (1992).

61. Mitchell, P.J. & Tjian, R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371-8 (1989).

62. Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147-51 (2003).

63. Ma, J. Transcriptional activators and activation mechanisms. *Protein Cell* **2**, 879-88 (2011).

64. Reynolds, N., O'Shaughnessy, A. & Hendrich, B. Transcriptional repressors: multifaceted regulators of gene expression. *Development* **140**, 505-12 (2013).

65. Kolovos, P., Knoch, T.A., Grosveld, F.G., Cook, P.R. & Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* **5**, 1 (2012).

66. Filtz, T.M., Vogel, W.K. & Leid, M. Regulation of transcription factor activity by interconnected post-translational modifications. *Trends Pharmacol Sci* **35**, 76-85 (2014).

67. Tootle, T.L. & Rebay, I. Post-translational modifications influence transcription factor activity: a view from the ETS superfamily. *Bioessays* **27**, 285-98 (2005).

68. Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H. & Kornberg, R.D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599-608 (1994).

69. Hsin, J.P. & Manley, J.L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26**, 2119-37 (2012).

70. Furuichi, Y., LaFiandra, A. & Shatkin, A.J. 5'-Terminal structure and mRNA stability. *Nature* **266**, 235-9 (1977).

71. Izaurralde, E. et al. A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78**, 657-68 (1994).

72. Wigington, C.P., Williams, K.R., Meers, M.P., Bassell, G.J. & Corbett, A.H. Poly(A) RNA-binding proteins and polyadenosine RNA: new members and novel functions. *Wiley Interdiscip Rev RNA* **5**, 601-22 (2014).

73. Goren, A. et al. Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Mol Cell* **22**, 769-81 (2006).

74. Wang, Z. & Burge, C.B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802-13 (2008).

75. Kelemen, O. et al. Function of alternative splicing. *Gene* **514**, 1-30 (2013).

76. Kuhn, U., Nemeth, A., Meyer, S. & Wahle, E. The RNA binding domains of the nuclear poly(A)-binding protein. *J Biol Chem* **278**, 16916-25 (2003).

77. Wu, Y. Unwinding and rewinding: double faces of helicase? *J Nucleic Acids* **2012**, 140601 (2012).

78. Tian, B. & Manley, J.L. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* **18**, 18-30 (2017).

79. Buhlmann, M. et al. NMD3 regulates both mRNA and rRNA nuclear export in African trypanosomes via an XPOI-linked pathway. *Nucleic Acids Res* **43**, 4491-504 (2015).

80. Katahira, J. Nuclear export of messenger RNA. *Genes (Basel)* **6**, 163-84 (2015).

81. Maquat, L.E., Tarn, W.Y. & Isken, O. The pioneer round of translation: features and functions. *Cell* **142**, 368-74 (2010).

82. Atkinson, G.C., Baldauf, S.L. & Hauryliuk, V. Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components. *BMC Evol Biol* **8**, 290 (2008).

83. Donnelly, N., Gorman, A.M., Gupta, S. & Samali, A. The eIF2alpha kinases: their structures and functions. *Cell Mol Life Sci* **70**, 3493-511 (2013).

84. Baird, T.D. & Wek, R.C. Eukaryotic initiation factor 2 phosphorylation and translational control in metabolism. *Adv Nutr* **3**, 307-21 (2012).

85. Leppek, K., Das, R. & Barna, M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol* **19**, 158-174 (2018).

86. Hentze, M.W. et al. Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science* **238**, 1570-3 (1987).

87. Gebauer, F. & Hentze, M.W. Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* **5**, 827-35 (2004).

88. Spriggs, K.A., Stoneley, M., Bushell, M. & Willis, A.E. Re-programming of translation following cell stress allows IRES-mediated translation to predominate. *Biol Cell* **100**, 27-38 (2008).

89. Mokrejs, M. et al. IRESite--a tool for the examination of viral and cellular internal ribosome entry sites. *Nucleic Acids Res* **38**, D131-6 (2010).

90. Varenne, S., Buc, J., Lloubes, R. & Lazdunski, C. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol* **180**, 549-76 (1984).

91. Dabrowski, M., Bukowy-Bieryllo, Z. & Zietkiewicz, E. Translational readthrough potential of natural termination codons in eucaryotes--The impact of RNA sequence. *RNA Biol* **12**, 950-8 (2015).

92. Schueren, F. et al. Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals. *Elife* **3**, e03640 (2014).

93. Sonenberg, N. & Hinnebusch, A.G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731-45 (2009).

94. Hershey, J.W., Sonenberg, N. & Mathews, M.B. Principles of translational control: an overview. *Cold Spring Harb Perspect Biol* **4**(2012).

95. Buchan, J.R. & Parker, R. Eukaryotic stress granules: the ins and outs of translation. *Mol Cell* **36**, 932-41 (2009).

96. Balagopal, V. & Parker, R. Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs. *Curr Opin Cell Biol* **21**, 403-8 (2009).

97. Wilson, R.C. & Doudna, J.A. Molecular mechanisms of RNA interference. *Annu Rev Biophys* **42**, 217-39 (2013).

98. Jinek, M. & Doudna, J.A. A three-dimensional view of the molecular machinery of RNA interference. *Nature* **457**, 405-12 (2009).

99. Macfarlane, L.A. & Murphy, P.R. MicroRNA: Biogenesis, Function and Role in Cancer. *Curr Genomics* **11**, 537-61 (2010).

100. Cech, T.R. & Steitz, J.A. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77-94 (2014).

101. Freier, S.M. & Altmann, K.H. The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA:RNA duplexes. *Nucleic Acids Res* **25**, 4429-43 (1997).

102. Doherty, E.A. & Doudna, J.A. Ribozyme structures and mechanisms. *Annu Rev Biophys Biomol Struct* **30**, 457-75 (2001).

103. Tinoco, I., Jr. & Bustamante, C. How RNA folds. *J Mol Biol* **293**, 271-81 (1999).

104. Cao, S., Xu, X. & Chen, S.J. Predicting structure and stability for RNA complexes with intermolecular loop-loop base-pairing. *RNA* **20**, 835-45 (2014).

105. Mahen, E.M., Watson, P.Y., Cottrell, J.W. & Fedor, M.J. mRNA secondary structures fold sequentially but exchange rapidly in vivo. *PLoS Biol* **8**, e1000307 (2010).

106. Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J.A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**, 1355-8 (2010).

107. Markert, A. et al. The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. *EMBO Rep* **9**, 569-75 (2008).

108. Agirrezabala, X. & Valle, M. Structural Insights into tRNA Dynamics on the Ribosome. *Int J Mol Sci* **16**, 9866-95 (2015).

109. Ha, M. & Kim, V.N. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* **15**, 509-24 (2014).

110. Gorski, J.L., Gonzalez, I.L. & Schmickel, R.D. The secondary structure of human 28S rRNA: the structure and evolution of a mosaic rRNA gene. *J Mol Evol* **24**, 236-51 (1987).

111. Sorrentino, S. The eight human "canonical" ribonucleases: molecular diversity, catalytic properties, and special biological actions of the enzyme proteins. *FEBS Lett* **584**, 2194-200 (2010).

112. Jore, M.M. et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* **18**, 529-36 (2011).

113. Sinha, A. et al. Biochemical characterization of pathogenic mutations in human mitochondrial methionyl-tRNA formyltransferase. *J Biol Chem* **289**, 32729-41 (2014).

114. Erat, M.C., Zerbe, O., Fox, T. & Sigel, R.K. Solution structure of domain 6 from a self-splicing group II intron ribozyme: a Mg(2+) binding site is located close to the stacked branch adenosine. *Chembiochem* **8**, 306-14 (2007).

115. Honig, B. & Rohs, R. Biophysics: Flipping Watson and Crick. *Nature* **470**, 472-3 (2011).

116. Novoa, I., Zeng, H., Harding, H.P. & Ron, D. Feedback inhibition of the unfolded protein response by GADD34-mediated dephosphorylation of eIF2alpha. *J Cell Biol* **153**, 1011-22 (2001).

117. Mergny, J.L. et al. Sequence specificity in triple-helix formation: experimental and theoretical studies of the effect of mismatches on triplex stability. *Biochemistry* **30**, 9791-8 (1991).

118. Chowdhury, S. & Bansal, M. Effect of coordinated ions on structure and flexiblity of parallel G-quandruplexes: a molecular dynamics study. *J Biomol Struct Dyn* **18**, 11-28 (2000).

119. Millevoi, S., Moine, H. & Vagner, S. G-quadruplexes in RNA biology. *Wiley Interdiscip Rev RNA* **3**, 495-507 (2012).

120. Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* **34**, 5402-15 (2006).

121. Huppert, J.L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* **33**, 2908-16 (2005).

122. Todd, A.K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res* **33**, 2901-7 (2005).

123. Arora, A. & Maiti, S. Differential biophysical behavior of human telomeric RNA and DNA quadruplex. *J Phys Chem B* **113**, 10515-20 (2009).

124. Joachimi, A., Benz, A. & Hartig, J.S. A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg Med Chem* **17**, 6811-5 (2009).

125. Zhang, D.H. et al. Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. *Biochemistry* **49**, 4554-63 (2010).

126. Xiao, C.D., Shibata, T., Yamamoto, Y. & Xu, Y. An intramolecular antiparallel G-quadruplex formed by human telomere RNA. *Chem Commun (Camb)* **54**, 3944-3946 (2018).

127. Duarte, A.R., Cadoni, E., Ressurreicao, A.S., Moreira, R. & Paulo, A. Design of Modular G-quadruplex Ligands. *ChemMedChem* **13**, 869-893 (2018).

128. Ruggiero, E. & Richter, S.N. G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Res* **46**, 3270-3283 (2018).

129. Li, Q. et al. G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res* **41**, D1115-23 (2013).

130. Di Antonio, M. et al. Selective RNA versus DNA G-quadruplex targeting by in situ click chemistry. *Angew Chem Int Ed Engl* **51**, 11073-8 (2012).

131. Fay, M.M., Lyons, S.M. & Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J Mol Biol* **429**, 2127-2147 (2017).

132. Bhattacharyya, D. et al. Engineered domain swapping indicates context dependent functional role of RNA G-quadruplexes. *Biochimie* **137**, 147-150 (2017).

133. Rodriguez, R. et al. A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J Am Chem Soc* **130**, 15758-9 (2008).

134. Chambers, V.S. et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol* **33**, 877-81 (2015).

135. Huppert, J.L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* **35**, 406-13 (2007).

136. Prioleau, M.N. G-Quadruplexes and DNA Replication Origins. *Adv Exp Med Biol* **1042**, 273-286 (2017).

137. Eddy, J. & Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res* **34**, 3887-96 (2006).

138. Kwok, C.K., Marsico, G., Sahakyan, A.B., Chambers, V.S. & Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat Methods* **13**, 841-4 (2016).

139. Guo, J.U. & Bartel, D.P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353**(2016).

140. Kwok, C.K., Marsico, G. & Balasubramanian, S. Detecting RNA G-Quadruplexes (rG4s) in the Transcriptome. *Cold Spring Harb Perspect Biol* **10**(2018).

141. Huppert, J.L., Bugaut, A., Kumari, S. & Balasubramanian, S. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res* **36**, 6260-8 (2008).

142. Martadinata, H. & Phan, A.T. Structure of human telomeric RNA (TERRA): stacking of two G-quadruplex blocks in K(+) solution. *Biochemistry* **52**, 2176-83 (2013).

143. Mergny, J.L., De Cian, A., Ghelab, A., Sacca, B. & Lacroix, L. Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res* **33**, 81-94 (2005).

144. Gray, R.D. & Chaires, J.B. Kinetics and mechanism of K+- and Na+-induced folding of models of human telomeric DNA into G-quadruplex structures. *Nucleic Acids Res* **36**, 4191-203 (2008).

145. Rhodes, D. & Lipps, H.J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res* **43**, 8627-37 (2015).

146. Capra, J.A., Paeschke, K., Singh, M. & Zakian, V.A. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. *PLoS Comput Biol* **6**, e1000861 (2010).

147. Nakken, S., Rognes, T. & Hovig, E. The disruptive positions in human G-quadruplex motifs are less polymorphic and more conserved than their neutral counterparts. *Nucleic Acids Res* **37**, 5749-56 (2009).

148. Rawal, P. et al. Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation. *Genome Res* **16**, 644-55 (2006).

149. Rodriguez, R. et al. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat Chem Biol* **8**, 301-10 (2012).

150. Schaffitzel, C. et al. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc Natl Acad Sci U S A* **98**, 8572-7 (2001).

151. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* **5**, 182-6 (2013).

152. Hansel-Hertsch, R. et al. G-quadruplex structures mark human regulatory chromatin. *Nat Genet* **48**, 1267-72 (2016).

153. Mao, S.Q. et al. DNA G-quadruplex structures mold the DNA methylome. *Nat Struct Mol Biol* **25**, 951-957 (2018).

154. Paeschke, K., Capra, J.A. & Zakian, V.A. DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell* **145**, 678-91 (2011).

155. Paeschke, K. et al. Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* **497**, 458-62 (2013).

156. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. & Hurley, L.H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A* **99**, 11593-8 (2002).

157. Biffi, G., Di Antonio, M., Tannahill, D. & Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat Chem* **6**, 75-80 (2014).

158. Bhattacharyya, D., Mirihana Arachchilage, G. & Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front Chem* **4**, 38 (2016).

159. Rouleau, S.G., Garant, J.M., Bolduc, F., Bisaillon, M. & Perreault, J.P. G-Quadruplexes influence pri-microRNA processing. *RNA Biol* **15**, 198-206 (2018).

160. Rouleau, S., Glouzon, J.S., Brumwell, A., Bisaillon, M. & Perreault, J.P. 3' UTR G-quadruplexes regulate miRNA binding. *RNA* **23**, 1172-1179 (2017).

161. Metifiot, M., Amrane, S., Litvak, S. & Andreola, M.L. G-quadruplexes in viruses: function and potential therapeutic applications. *Nucleic Acids Res* **42**, 12352-66 (2014).

162. Booy, E.P. et al. The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. *Nucleic Acids Res* **40**, 4110-24 (2012).

163. Wanrooij, P.H., Uhler, J.P., Simonsson, T., Falkenberg, M. & Gustafsson, C.M. G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. *Proc Natl Acad Sci U S A* **107**, 16072-7 (2010).

164. Decorsiere, A., Cayrel, A., Vagner, S. & Millevoi, S. Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage. *Genes Dev* **25**, 220-5 (2011).

165. Beaudoin, J.D. & Perreault, J.P. Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res* **41**, 5898-911 (2013).

166. Ribeiro, M.M. et al. G-quadruplex formation enhances splicing efficiency of PAX9 intron 1. *Hum Genet* **134**, 37-44 (2015).

167. Gomez, D. et al. Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res* **32**, 371-9 (2004).

168. Weldon, C. et al. Specific G-quadruplex ligands modulate the alternative splicing of Bcl-X. *Nucleic Acids Res* **46**, 886-896 (2018).

169. Marcel, V. et al. G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinogenesis* **32**, 271-8 (2011).

170. Gemignani, F. et al. A TP53 polymorphism is associated with increased risk of colorectal cancer and with reduced levels of TP53 mRNA. *Oncogene* **23**, 1954-6 (2004).

171. Schaeffer, C. et al. The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J* **20**, 4803-13 (2001).

172. Arora, A. et al. Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA* **14**, 1290-6 (2008).

173. Beaudoin, J.D. & Perreault, J.P. 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res* **38**, 7022-36 (2010).

174. Dai, J. et al. Discovery of Small Molecules for Up-Regulating the Translation of Antiamyloidogenic Secretase, a Disintegrin and Metalloproteinase 10 (ADAM10), by Binding to the G-Quadruplex-Forming Sequence in the 5' Untranslated Region (UTR) of Its mRNA. *J Med Chem* **58**, 3875-91 (2015).

175. Gomez, D. et al. A G-quadruplex structure within the 5'-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Research* **38**, 7187-7198 (2010).

176. Kumari, S., Bugaut, A., Huppert, J.L. & Balasubramanian, S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol* **3**, 218-21 (2007).

177. Shahid, R., Bugaut, A. & Balasubramanian, S. The BCL-2 5' untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry* **49**, 8300-6 (2010).

178. Kumari, S., Bugaut, A. & Balasubramanian, S. Position and stability are determining factors for translation repression by an RNA G-quadruplex-forming sequence within the 5' UTR of the NRAS proto-oncogene. *Biochemistry* **47**, 12664-9 (2008).

179. Wolfe, A.L. et al. RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513**, 65-70 (2014).

180. Bonnal, S. et al. A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J Biol Chem* **278**, 39330-6 (2003).

181. Westmark, C.J. & Malter, J.S. FMRP mediates mGluR5-dependent translation of amyloid precursor protein. *PLoS Biol* **5**, e52 (2007).

182. Endoh, T. & Sugimoto, N. Unusual -1 ribosomal frameshift caused by stable RNA G-quadruplex in open reading frame. *Anal Chem* **85**, 11435-9 (2013).

183. Arora, A. & Suess, B. An RNA G-quadruplex in the 3' UTR of the proto-oncogene PIM1 represses translation. *RNA Biol* **8**, 802-5 (2011).

184. Subramanian, M. et al. G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep* **12**, 697-704 (2011).

185. Davidovic, L. et al. The fragile X mental retardation protein is a molecular adaptor between the neurospecific KIF3C kinesin and dendritic RNA granules. *Hum Mol Genet* **16**, 3047-58 (2007).

186. Fleming, A.M., Ding, Y., Alenko, A. & Burrows, C.J. Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. *ACS Infect Dis* **2**, 674-681 (2016).

187. Hentze, M.W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* **19**, 327-341 (2018).

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

189. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat Rev Genet* (2014).

190. Auweter, S.D., Oberstrass, F.C. & Allain, F.H. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res* **34**, 4943-59 (2006).

191. Dominguez, D. et al. Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol Cell* **70**, 854-867 e9 (2018).

192. Baltz, A.G. et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* **46**, 674-90 (2012).

193. Castello, A. et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393-406 (2012).

194. Castello, A. et al. System-wide identification of RNA-binding proteins by interactome capture. *Nat Protoc* **8**, 491-500 (2013).

195. Brazda, V. et al. The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors. *Molecules* **23**(2018).

196. von Hacht, A. et al. Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res* **42**, 6630-44 (2014).

197. Brazda, V., Haronikova, L., Liao, J.C. & Fojta, M. DNA and RNA quadruplex-binding proteins. *Int J Mol Sci* **15**, 17493-517 (2014).

198. Mishra, S.K., Tawani, A., Mishra, A. & Kumar, A. G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. *Sci Rep* **6**, 38144 (2016).

199. DeForte, S. & Uversky, V.N. Order, Disorder, and Everything in Between. *Molecules* **21**(2016).

200. Vasilyev, N. et al. Crystal structure reveals specific recognition of a G-quadruplex RNA by a beta-turn in the RGG motif of FMRP. *Proc Natl Acad Sci U S A* **112**, E5391-400 (2015).

201. Huang, Z.L. et al. Identification of G-Quadruplex-Binding Protein from the Exploration of RGG Motif/G-Quadruplex Interactions. *J Am Chem Soc* (2018).

202. Mendoza, O., Bourdoncle, A., Boule, J.B., Brosh, R.M., Jr. & Mergny, J.L. G-quadruplexes and helicases. *Nucleic Acids Res* **44**, 1989-2006 (2016).

203. Sauer, M. & Paeschke, K. G-quadruplex unwinding helicases and their function in vivo. *Biochem Soc Trans* **45**, 1173-1182 (2017).

204. Tosoni, E. et al. Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res* **43**, 8884-97 (2015).

205. Fang, G. & Cech, T.R. Characterization of a G-quartet formation reaction promoted by the beta-subunit of the Oxytricha telomere-binding protein. *Biochemistry* **32**, 11646-57 (1993).

206. Giraldo, R. & Rhodes, D. The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J* **13**, 2411-20 (1994).

207. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. & Lipps, H.J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat Struct Mol Biol* **12**, 847-54 (2005).

208. Rajkowitsch, L. et al. RNA chaperones, RNA annealers and RNA helicases. *RNA Biol* **4**, 118-30 (2007).

209. Herdy, B. et al. Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Res* **46**, 11592-11604 (2018).

210. Murat, P. et al. RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol* **19**, 229 (2018).

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

212. Abdelhaleem, M. Over-expression of RNA helicases in cancer. *Anticancer Res* **24**, 3951-3 (2004).

213. Umate, P., Tuteja, N. & Tuteja, R. Genome-wide comprehensive analysis of human helicases. *Commun Integr Biol* **4**, 118-37 (2011).

214. Benhalevy, D. et al. The Human CCHC-type Zinc Finger Nucleic Acid-Binding Protein Binds G-Rich Elements in Target mRNA Coding Sequences and Promotes Translation. *Cell Rep* **18**, 2979-2990 (2017).

215. Tran, H., Schilling, M., Wirbelauer, C., Hess, D. & Nagamine, Y. Facilitation of mRNA deadenylation and decay by the exosome-bound, DExH protein RHAU. *Mol Cell* **13**, 101-11 (2004).

216. Vaughn, J.P. et al. The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J Biol Chem* **280**, 38117-20 (2005).

217. Chalupnikova, K. et al. Recruitment of the RNA helicase RHAU to stress granules via a unique RNA-binding domain. *J Biol Chem* **283**, 35186-98 (2008).

218. Chen, M.C. et al. Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* **558**, 465-469 (2018).

219. Giri, B. et al. G4 resolvase 1 tightly binds and unwinds unimolecular G4-DNA. *Nucleic Acids Res* **39**, 7161-78 (2011).

220. Lattmann, S., Giri, B., Vaughn, J.P., Akman, S.A. & Nagamine, Y. Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res* **38**, 6219-33 (2010).

221. Creacy, S.D. et al. G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J Biol Chem* **283**, 34626-34 (2008).

222. Heddi, B., Cheong, V.V., Martadinata, H. & Phan, A.T. Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex. *Proc Natl Acad Sci U S A* **112**, 9608-13 (2015).

223. Tippana, R., Hwang, H., Opresko, P.L., Bohr, V.A. & Myong, S. Single-molecule imaging reveals a common mechanism shared by G-quadruplex-resolving helicases. *Proc Natl Acad Sci U S A* **113**, 8448-53 (2016).

224. Yangyuoru, P.M., Bradburn, D.A., Liu, Z., Xiao, T.S. & Russell, R. The G-quadruplex (G4) resolvase DHX36 efficiently and specifically disrupts DNA G4s via a translocation-based helicase mechanism. *J Biol Chem* **293**, 1924-1932 (2018).

225. Gueddouda, N.M., Mendoza, O., Gomez, D., Bourdoncle, A. & Mergny, J.L. G-quadruplexes unfolding by RHAU helicase. *Biochim Biophys Acta* (2017).

226. Chen, M.C., Murat, P., Abecassis, K., Ferre-D'Amare, A.R. & Balasubramanian, S. Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase. *Nucleic Acids Res* (2015).

227. Lai, J.C. et al. The DEAH-box helicase RHAU is an essential gene and critical for mouse hematopoiesis. *Blood* **119**, 4291-300 (2012).

228. Nie, J. et al. Post-transcriptional Regulation of Nkx2-5 by RHAU in Heart Development. *Cell Rep* **13**, 723-32 (2015).

229. Gao, X. et al. A G-quadruplex DNA structure resolvase, RHAU, is essential for spermatogonia differentiation. *Cell Death Dis* **6**, e1610 (2015).

230. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).

231. Huang, W. et al. Yin Yang 1 contains G-quadruplex structures in its promoter and 5'-UTR and its expression is modulated by G4 resolvase 1. *Nucleic Acids Res* **40**, 1033-49 (2012).

232. Yoo, J.S. et al. DHX36 enhances RIG-I signaling by facilitating PKR-mediated antiviral stress granule formation. *PLoS Pathog* **10**, e1004012 (2014).

233. Lavezzo, E. et al. G-quadruplex forming sequences in the genome of all known human viruses: A comprehensive guide. *PLoS Comput Biol* **14**, e1006675 (2018).

234. Bian, W.X. et al. Binding of cellular nucleolin with the viral core RNA G-quadruplex structure suppresses HCV replication. *Nucleic Acids Res* **47**, 56-68 (2019).

235. Lattmann, S., Stadler, M.B., Vaughn, J.P., Akman, S.A. & Nagamine, Y. The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. *Nucleic Acids Res* **39**, 9390-404 (2011).

236. Sexton, A.N. & Collins, K. The 5' guanosine tracts of human telomerase RNA are recognized by the G-quadruplex binding domain of the RNA helicase DHX36 and function to increase RNA accumulation. *Mol Cell Biol* **31**, 736-43 (2011).

237. Hollerer, I., Grund, K., Hentze, M.W. & Kulozik, A.E. mRNA 3'end processing: A tale of the tail reaches the clinic. *EMBO Mol Med* **6**, 16-26 (2014).

238. Newman, M. et al. The G-Quadruplex-Specific RNA Helicase DHX36 Regulates p53 Pre-mRNA 3'-End Processing Following UV-Induced DNA Damage. *J Mol Biol* (2016).

239. Zhang, Z. et al. DDX1, DDX21, and DHX36 helicases form a complex with the adaptor molecule TRIF to sense dsRNA in dendritic cells. *Immunity* **34**, 866-78 (2011).

240. Bicker, S. et al. The DEAH-box helicase DHX36 mediates dendritic localization of the neuronal precursor-microRNA-134. *Genes Dev* **27**, 991-6 (2013).

241. Booy, E.P. et al. The RNA helicase RHAU (DHX36) suppresses expression of the transcription factor PITX1. *Nucleic Acids Res* **42**, 3346-61 (2014).

242. Spitzer, J., Landthaler, M. & Tuschl, T. Rapid creation of stable mammalian cell lines for regulated expression of proteins using the Gateway(R) recombination cloning technology and Flp-In T-REx(R) lines. *Methods Enzymol* **529**, 99-124 (2013).

243. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).

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

245. Yamaji, M. et al. DND1 maintains germline stem cells via recruitment of the CCR4-NOT complex to target mRNAs. *Nature* **543**, 568-572 (2017).

246. Jodoin, R. et al. The folding of 5'-UTR human G-quadruplexes possessing a long central loop. *RNA* **20**, 1129-41 (2014).

247. Gray, D.M. et al. Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality* **20**, 431-40 (2008).

248. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-41 (2010).

249. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. & Weissman, J.S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218-23 (2009).

250. Khong, A. et al. The Stress Granule Transcriptome Reveals Principles of mRNA Accumulation in Stress Granules. *Mol Cell* **68**, 808-820 e5 (2017).

251. Lee, T.I. & Young, R.A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237-51 (2013).

252. Corbett, A.H. Post-transcriptional regulation of gene expression and human disease. *Curr Opin Cell Biol* **52**, 96-104 (2018).

253. Cooper, T.A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777-93 (2009).

254. Re, A., Joshi, T., Kulberkyte, E., Morris, Q. & Workman, C.T. RNA-protein interactions: an overview. *Methods Mol Biol* **1097**, 491-521 (2014).

255. Booy, E.P. et al. RNA Helicase Associated with AU-rich Element (RHAU/DHX36) Interacts with the 3'-Tail of the Long Non-coding RNA BC200 (BCYRN1). *J Biol Chem* **291**, 5355-72 (2016).

256. Truitt, M.L. & Ruggero, D. New frontiers in translational control of the cancer genome. *Nat Rev Cancer* **17**, 332 (2017).

257. Jankowsky, E. & Harris, M.E. Specificity and nonspecificity in RNA-protein interactions. *Nat Rev Mol Cell Biol* **16**, 533-44 (2015).

258. Stapleton, J.A. et al. Feedback control of protein expression in mammalian cells by tunable synthetic translational inhibition. *ACS Synth Biol* **1**, 83-8 (2012).

259. Jing, H. et al. DExD/H-Box Helicase 36 Signaling via Myeloid Differentiation Primary Response Gene 88 Contributes to NF-kappaB Activation to Type 2 Porcine Reproductive and Respiratory Syndrome Virus Infection. *Front Immunol* **8**, 1365 (2017).

260. Vallur, A.C. & Maizels, N. Activities of human exonuclease 1 that promote cleavage of transcribed immunoglobulin switch regions. *Proc Natl Acad Sci U S A* **105**, 16508-12 (2008).

261. Budd, M.E., Reis, C.C., Smith, S., Myung, K. & Campbell, J.L. Evidence suggesting that Pif1 helicase functions in DNA replication with the Dna2 helicase/nuclease and DNA polymerase delta. *Mol Cell Biol* **26**, 2490-500 (2006).

262. Chakraborty, P. & Grosse, F. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst)* **10**, 654-65 (2011).

263. McRae, E.K.S. et al. Human DDX21 binds and unwinds RNA guanine quadruplexes. *Nucleic Acids Res* (2017).

264. Guo, M. et al. A distinct triplex DNA unwinding activity of ChlR1 helicase. *J Biol Chem* **290**, 5174-89 (2015).

265. Ribeiro de Almeida, C. et al. RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination. *Mol Cell* **70**, 650-662 e8 (2018).

266. Bugaut, A. & Balasubramanian, S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res* **40**, 4727-41 (2012).

267. Al-Zeer, M.A. & Kurreck, J. Deciphering the Enigmatic Biological Functions of RNA Guanine-Quadruplex Motifs in Human Cells. *Biochemistry* **58**, 305-311 (2019).

268. Gagnon, K.T., Li, L., Janowski, B.A. & Corey, D.R. Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nat Protoc* **9**, 2045-60 (2014).

269. Hafner, M., Lianoglou, S., Tuschl, T. & Betel, D. Genome-wide identification of miRNA targets by PAR-CLIP. *Methods* **58**, 94-105 (2012).

270. Hafner, M. et al. Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. *Methods* **58**, 164-70 (2012).

271. Corcoran, D.L. et al. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol* **12**, R79 (2011).

272. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-78 (2012).

273. Ingolia, N.T. Genome-wide translational profiling by ribosome footprinting. *Methods Enzymol* **470**, 119-42 (2010).

274. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).

275. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-72 (2008).

276. Bailey, T.L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-8 (2009).

9. Table of Figures

Figure 3-1: The life cycle of an eukaryotic mRNA. 6
Figure 3-2: Overview of important determinants of mRNA expression. 11
Figure 3-3: Characteristics and examples of RNA secondary structures. 15
Figure 3-4: G-quadruplex structure formation 17
Figure 3-5: Functions of RNA G-quadruplex structures 22
Figure 3-6: Schematic representation of the DHX36 domains. 28
Figure 4-1 Generation of DHX36 overexpressing cells. 33
Figure 4-2 DHX36 localizes to the cytoplasm and binds to mRNAs. 34
Figure 4-3 PAR-CLIPs of FH-DHX36 forms can be successfully reproduced. 36
Figure 4-4 DHX36 interacts with mature mRNAs at thousands of sites. 39
Figure 4-5 DHX36 recognizes G-rich sequence stretches on mRNA. 40
Figure 4-6 DHX36 RNA recognition elements can form G-quadruplexes *in vitro*. 42
Figure 4-7 DHX36 binds to *in vitro* demonstrated G-quadruplexes. 44
Figure 4-8 Generation of DHX36 knockout cells. 46
Figure 4-9 DHX36 knockout results in increased target mRNA abundance 47
Figure 4-10 Target mRNA abundance is affected by DHX36 binding to UTRs. 49
Figure 4-11 DHX36-KO results in increased target mRNA stability. 50
Figure 4-12 Target mRNA abundance depends on DHX36 helicase activity. 52
Figure 4-13 Higher target mRNA stability does not result in more translated protein. . 53
Figure 4-14 RPF levels on target mRNAs are not elevated in DHX36-KO cells. 55
Figure 4-15 RPF levels in target CDSs are modestly lower in DHX36-KO cells. 56
Figure 4-16 DHX36 knockout results in lower translation efficiency of target mRNAs. 57
Figure 4-17 Reporter assays with DHX36 binding sites confirm translation effect. 58
Figure 4-18 DHX36 affects rG4-bearing mRNAs. 60
Figure 4-19 DHX36 is involved in the regulation of rG4s in living cells. 61
Figure 4-20 DHX36-KO leads to cellular stress. 63
Figure 5-1 Model for DHX36 function. 77
Figure 7-1 Polysome gradient of DHX36 overexpression cells. 114
Figure 7-2 Supplementary data to Figure 4-14. 117
Figure 7-3 Supplementary data to Figure 4-15. 117
Figure 7-4 Supplementary data to Figure 4-16. 118
Figure 7-5 Expression of DHX36 during cell differentiation. 118

10. Publications

Substantial parts of this thesis were published in the following article:

Markus Sauer, Stefan A. Juranek, James Marks, Alessio De Magis, Hinke G. Kazemier, Daniel Hilbig, Daniel Benhalevy, Xiantao Wang, Markus Hafner, Katrin Paeschke
DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions.

Accepted for publication in Nature Communications (2019)

Related Publications:

Markus Sauer, Katrin Paeschke. (2017) *G-quadruplex unwinding helicases and their function in vivo.* Biochem. Soc. Trans. 45(5):1173-1182

Previous Publications:

Publications listed here are unrelated to this PhD thesis.

Stefanie Bock, Fabian Gauch, Yannik Giernat, Frank Hillebrand, Daria Kozlova, Lisa Linck, Rebecca Moschall, **Markus Sauer**, Christian Schenk, Kristina Ulrich, Jochen Bodem. (2013) *HIV-1: Lehrbuch von Studenten für Studenten*. OPUS Würzburg

Kemal Marc Akat, D´Vesharronne Moore-McGriff, Pavel Morozov, Miguel Brown, Tassos Gogakos, Joel Correa Da Rosa, Aleksandra Mihailovic, **Markus Sauer**, Ruiping Ji, Aarthi Ramarathnam, Hana Totary-Jain, Zev Williams, Thomas Tuschl and P. Christian Schulze: (2014) *Comparative RNA-sequencing analysis of myocardial and circulating small RNAs in human heart failure and their utility as biomarkers.* Proc. Natl. Acad. Sci. USA 111(30):11151-6.

11. Curriculum vitae

12. Acknowledgements

At the end of this thesis I would like to express my deepest gratitude to all people who helped me during this project and in preparation of this doctoral thesis:

- First and foremost, I would like to thank my primary supervisor Prof. Katrin Paeschke for giving me the opportunity to work in her lab, for always having an open door, for every precious advice and helpful discussion.

- I am grateful to the members of my thesis committee, my second supervisor Prof. Jörg Vogel and my third supervisor Prof. Utz Fischer, for the critical and insightful discussions and feedback on my work.

- I want to thank Prof. Manfred Gessler for chairing the oral defense committee.

- A special thanks to Dr. Markus Hafner for bioinformatical, practical, and conceptual help and advice and for the chance to perform some experiments in his laboratory at the National Institute of Health. Thanks to Dr. Daniel Benhalevy for technical training in ribosome footprinting during this time

- I am grateful to Prof. Alexander Buchberger for any precious advice and the opportunity to conduct parts of my research in his laboratory in 2017 and.

- I want to thank Dr. Stefan Juranek for countless answers on questions about cell culture and RNA biology as well as insightful discussions. Thanks for your help with the DHX36 knockout cells.

- At this point, I want to thank all the current and former members of the Paeschke group and the Juranek group for the good times. Especially, I want to thank Dr. Sanjay Gupta for many times providing advice and technical help at the begin of my thesis. Thanks to Dr. Alessio De Magis for providing his know-how and excellent skills in fluorescence microscopy. My fellow PhD students in the group, especially Eike Schwindt, Daniel Hilbig, and Mona Hajikazemi for help, the harmonic working atmosphere, and every scientific or non-scientific discussion. Thanks to Nadine Stoll, Theresa Schmitz, and especially Maria Gallant for the great technical support.

- Also, I want to thank all people at the Department of Biochemistry at the University of Würzburg. Especially, I want to thank Dr. Thomas Ziegenhals, Jürgen Ohmer, Dr. Cornelia Brosi, Dr. Michael Grimm, Dr. Georg Polleichtner, and Lissy Kunkel for many funny discussions during lunch breaks.

- Thanks to the people at the European Research Institute for the Biology of Ageing (ERIBA) in the Netherlands for their help and the uncomplicated way to work with them.

- Additionally, I want to say thanks to the people of the Medical Clinic III at the Biomedical Center I at the University Hospital Bonn, especially Pia Sauerborn and Andre Gauchel.

- I want to thank my friends and especially my family for their encouragement and help during the whole time of my studies.

- Finally, I want to thank Marlén for her help, encouragement, and love all over the time.

13. Affidavit

I hereby confirm that my thesis entitled "**DHX36 function in RNA G-quadruplex-mediated posttranscriptional gene regulation**" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Hiermit erkläre ich an Eides statt, die Dissertation **“Funktion von DHX36 in RNA G-Quadruplex-vermittelter posttranskriptioneller Genregulierung”** eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

……

Aschaffenburg, 30.04.2019 Signature