Functional characterisation of DHX proteins in the regulation of RNA metabolism and genome stability

Dissertation

for the award of the degree

“Doctor of Philosophy”

Division of Mathematics and Natural Sciences

of the Georg-August-Universität Göttingen

within the doctoral program Molecular Biology of Cells

of the Georg-August University School of Sciences (GAUSS)

Submitted by

Rebecca Rossen Falk

From Kolding, Denmark

Göttingen 2023

Thesis Committee

Prof. Dr. Markus T. Bohnsack	Department of Molecular Biology University Medical Center Göttingen
Dr. Ricarda Richter-Dennerlein	Department of Cellular Biochemistry University Medical Center Göttingen
Prof. Dr. Lutz Walter	Department of Primate Genetics German Primate Center

Members of the Examination Board

Referee: Prof. Dr. Markus T. Bohnsack	Department of Molecular Biology University Medical Center Göttingen
2nd Referee: Dr. Ricarda Richter-Dennerlein	Department of Cellular Biochemistry University Medical Center Göttingen

Further members of the Examination Board

Prof. Dr. Henning Urlaub	Department of Clinical Chemistry University Medical Center Göttingen
Dr. Sonja Lorenz	Research Group Ubiquitin Signaling Specificity Max Planck Institute for Multidisciplinary Sciences
Prof. Dr. Jörg Stülke	Department of General Microbiology Institute for Microbiology and Genetics

Date of oral examination: 01.03.2023

Acknowledgements

First and foremost, I would like to express my gratitude to Prof. Dr. Markus T. Bohnsack for giving me the opportunity to conduct my PhD research in his lab and for his supervision and support during my study.

I am very grateful to my thesis advisory committee members, Dr. Ricarda Richter-Dennerlein, and Prof. Dr. Lutz Walter, for their advice and support. In addition, I would like to thank my extended examination board, Prof. Dr. Henning Urlaub, Dr. Sonja Lorenz, and Prof. Dr. Jörg Stülke, for lending me their time and expertise.

I would like to express my deepest appreciation to Dr. Katherine Bohnsack for her support and scientific discussions, which were really influential to this project.

I am very grateful to all current and former colleagues at the Department of Molecular Biology for the cherished time we spent together in the lab and social settings. I would especially like to thank all the members of the Bohnsack lab for their kind help and support during my study, and I wish to extend my thanks to Philipp Hackert for technical support and Nicolás Lemus, MD PhD for bioinformatic analyses and scientific discussions.

I would like to thank my roommate Dora Radnai for providing me with endless support and continuous encouragement throughout my years of study.

Finally, I would like to express my appreciation to my family and friends who have supported me. This accomplishment would not have been possible without them.

Table of contents

Summary

Cells contain a multitude of different RNAs, which are highly versatile biomolecules that have crucial functions in almost all cellular processes. Furthermore, the inter-relationship between RNA and proteins is essential, as ribonucleoprotein complexes are responsible for the execution and control of fundamental cellular processes. RNA helicases are nucleoside triphosphate-dependent RNA-binding proteins capable of remodelling RNAs and ribonucleoprotein complexes and are, consequently, a key driving force in almost all aspects of RNA metabolism and are expressed not only in all three domains of life but also in some viruses.

DHX proteins are a diverse group of Superfamily 2 RNA helicases with shared structural features that display a wide array of RNA remodelling activities in a multitude of cellular processes. Consequently, strict spatial, temporal, and catalytic regulation of DHX proteins is essential in cells, and the dysregulation of helicase activity is often associated with tumorigenesis and disease. However, the cellular function and mode of regulation of several DHX proteins remain elusive.

The objective of this study was to broaden the understanding of the functional and regulatory repertoires of DHX proteins. To this end, a toolbox was established for exploring DHX protein functions in the cellular context. The toolbox included verified siRNAs, stably transfected cell lines for inducible expression of tagged DHX proteins, and tested antibodies. The usefulness of this toolbox was exemplified by the functional characterisation of DHX40. Analysis of DHX40 protein interactions by immunoprecipitation and mass spectrometry revealed that DHX40 interacts with several proteins implicated in regulating LINE-1 retrotransposition. Biochemical analysis using recombinantly expressed and purified proteins showed that DHX40 is an RNA-dependent ATPase that is stimulated by the putative prolyl isomerase PPIL4, a novel protein cofactor identified in this study. DHX40 was confirmed to be a substrate of the deubiquitinase USP7, and mutagenesis analysis was used to investigate the interaction of DHX40 with the E3 ligase TRIM27. Comprehensive analysis of the RNA interactome by the crosslinking and analysis of cDNA (CRAC) approach revealed that DHX40 interacts with LINE-1 RNAs in HEK293, and complementary differential expression analysis showed that LINE-1 RNAs are upregulated upon depletion of DHX40. Finally, a cell-based LINE-1 retrotransposition assay showed that DHX40 and PPIL4 are suppressors of LINE-1 retrotransposition in HEK293 cells.

1. Introduction

1.1 RNA helicase families, architecture, and activities

1.1.1 RNA helicase classifications

Ribonucleic acids (RNAs) are highly versatile biomolecules that have crucial functions in almost all cellular processes. The cellular environment hosts a multitude of different RNAs, which can generally be classified as either coding RNAs, i.e., messenger RNAs (mRNA) or non-coding RNAs (ncRNAs). ncRNAs encompass numerous sub-classifications, including micro RNAs (miRNAs), small nuclear RNAs (snRNAs), transfer RNAs (tRNAs), small interfering RNAs (siRNAs), small nucleolar RNAs (snoRNAs), long non-coding RNAs (lncRNAs), and ribosomal RNA (rRNA) (Brosius & Raabe, 2016). RNAs readily fold into simple secondary and tertiary structures and even complex structures which exhibit biochemical activities. However, a fundamental requirement for complex life is the inter-relationship between RNA and proteins, as ribonucleoproteins (RNPs) are responsible for the execution and control of fundamental biological processes (Ganser et al., 2019; Herschlag, 1995). Decades ago, it became apparent that RNAs constantly undergo dynamic changes on timescales ranging from picoseconds to seconds (Ganser et al., 2019). The correct folding of RNAs is paramount to their catalytic function and interactions, as RNP formation requires protein binding to a correctly folded RNA. However, RNAs can readily get caught in aberrant conformations due to their high plasticity and propensity for spontaneously folding into low-energy structures. RNA chaperones, such as RNA helicases, can aid RNA in achieving a functional conformation (Herschlag, 1995). The protein-RNA inter-relationship, therefore, serves not only a functional but also a structural purpose. RNA helicases are nucleoside triphosphate (NTP) dependent RNA binding proteins (RBPs) capable of remodelling RNAs and RNPs. Consequently, RNA helicases are a key driving force in almost all aspects of RNA metabolism and are expressed not only in all three domains of life but also in some viruses (Jankowsky & Fairman, 2008; Singleton et al., 2007; Sloan & Bohnsack, 2018; Tanner & Linder, 2001).

1.1.2 Characteristic features of helicases

Comparative structural and functional analyses have classified helicases into six superfamilies (SF1-6), in which ring-forming helicases comprise SF3-6, while the non-ring-forming helicases comprise SF1 and SF2 (Jankowsky, 2011). The greatest number of helicases implicated in manipulating nucleic acids belong to SF1 and SF2, which further divides into three and ten families, respectively (Fairman-Williams et al., 2010; Gorbalenya & Koonin, 1993; Singleton et al., 2007) (Figure 1). Of these families, five SF2 families include RNA helicases, namely, Ski2-

like, RIG-I-like, DEAD-box, DEAH/RHA, and NS3/NPH-11 helicases, while only one SF1 family, Upf1-like helicases, remodel RNA (Fairman-Williams et al., 2010; Jankowsky, 2011; Sloan & Bohnsack, 2018).

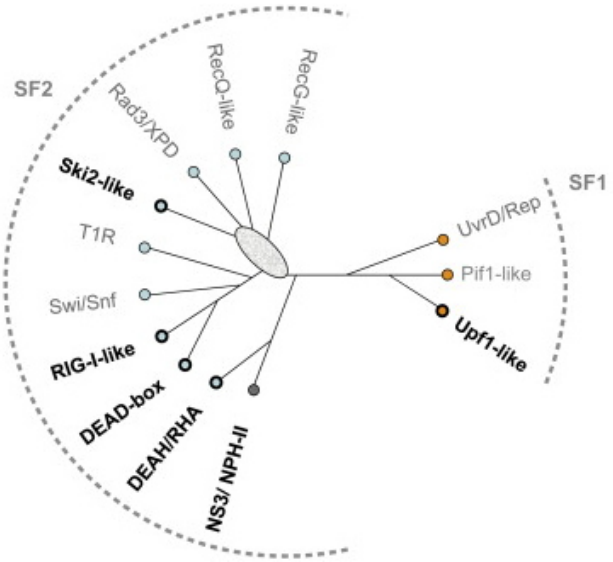


Figure 1 Superfamily 1 and superfamily 2 helicase families. *Uprooted cladogram showing the families of Superfamily 1 (SF1) and superfamily 2 (SF2) without branch length scaling. Family names in bold indicate families containing RNA helicase members. The oval indicates uncertainty in the cladogram topology in this region. From Jankowsky, 2011.*

The hallmark of SF1 and SF2 helicases is conserved catalytic cores that are structurally almost identical. Moreover, within each superfamily, the level of structural similarity is still higher (Fairman-Williams et al., 2010). The conserved catalytic core consists of two similar protein domains resembling the recombination protein, RecA; these RecA-like domains, termed RecA1 and RecA2, are arranged in tandem and connected by a flexible linker (Fairman-Williams et al., 2010; Jankowsky & Fairman, 2007). A notable signature of the conserved helicase core, shared by the SF1 and SF2 helicases, is the up to 12 characteristic sequence motifs implicated in substrate and NTP binding, energetic coupling, and translocation. The highest level of sequence conservation between the SFs is within motifs I, II, and VI, which are responsible for the coordination between binding and hydrolysis of NTPs (Fairman-Williams et al., 2010; Sloan & Bohnsack, 2018). The catalytic residues in motif II furthermore lend their names to two major families within SF2, i.e. DEAH/RHA and DEAD-box helicases (Jankowsky & Fairman, 2008). The Q motif, upstream of motif I, spatially arranges the adenine base through a conserved glutamine residue and confers adenosine triphosphate (ATP) selectivity. This motif is absent in DEAH/RHA and NS3/NPH-II helicases but present in other families within SF2 (Fairman-Williams et al., 2010; Singleton et al., 2007; Tanner & Linder, 2001). Motifs Ia-c contribute to nucleic acid substrate binding through interactions with the sugar-

phosphate backbone (Fairman-Williams et al., 2010; Sloan & Bohnsack, 2018), while motifs III and Va are responsible for coordinating nucleic acid and NTP binding (Fairman-Williams et al., 2010). Motifs IV and IVa contribute to RNA substrate binding, and motif V is important for translocation in DEAH/RHA helicases, where it senses the catalytic state and positions the RecA2 domain accordingly (Hamann et al., 2019; Sloan & Bohnsack, 2018). Notably, the RecA2 domain contains an antiparallel β-hairpin protruding from the domain between motifs V and VI. This β-hairpin element seems to be a conserved feature of the DEAH/RHA helicases (Hamann et al., 2019; Walbott et al., 2010). Additionally, a hook-turn and hook-loop, in the RecA1 and RecA2 domains, respectively, are essential to the function of some DEAH/RHA helicases (Hamann et al., 2021; Tauchert et al., 2017).

In most cases, amino- and carboxyl-terminal (N- and C-terminal) regions flank the conserved helicase core; these can play roles in the biochemical function and localisation of the helicase (Tauchert et al., 2017). However, the terminal regions vary significantly between, but also within, superfamilies. One exception to this is the C-terminal regions of DEAH/RHA helicases. While the N-terminal region of DEAH/RHA helicases is protein specific, this family of helicases is characterised by a conserved C-terminal domain (CTD). The CTD consists of three subdomains; a degenerate winged-helix (WH), a helix bundle (HB), and an oligosaccharide-binding fold (OB) (Chen & Ferré-D'Amaré, 2017; Hamann et al., 2019; Walbott et al., 2010). Another exception is RIG-I-like receptors (RLRs) which have a CTD responsible for binding to their double-stranded RNA (dsRNA) substrate and, except for DHX58, have two N-terminal caspase recruitment domains (CARD) essential for anti-viral signal transduction (Duic et al., 2020; Stok et al., 2022).

1.1.3 Biochemical activities

The term helicases originally meant proteins that move directionally across a phosphodiester backbone to separate two hybridised nucleic acid strands. However, with time the term has come to encompass a group of proteins with shared structural features and a wide array of functions involved in all aspects of nucleic acid metabolism. Moreover, as RNAs have a vast repertoire of structural conformations, so must the proteins that remodel them have a vast repertoire of molecular functions. Therefore, although the name "helicases" implies displacing hybridised strands, proteins classified as RNA helicases can also remodel their substrates by promoting RNA annealing, displacement of RBPs or by acting as RNA clamps in RNPs (Sloan & Bohnsack, 2018). For example, DEAD-box proteins unwind RNA strands locally in a translocation-independent fashion by inserting a conserved alpha-helix into an RNA duplex, thus destabilising it (Sloan & Bohnsack, 2018). DEAH/RHA helicases, in contrast, remodel their substrates by translocating along a single-stranded RNA (ssRNA), thus disrupting RNA structures and displacing RNAs and RBPs. Recently, a series of cryogenic electron

microscopy (cryo-EM) structures of DEAH/RHA helicases in different conformational states unravelled the mechanism of translocation, providing valuable mechanistic insight into this group of proteins (Hamann et al., 2019).

1.1.3.1 DEAH/RHA translocation mechanism

Within DEAH/RHA helicases, the spatial arrangement of the RecA1 and RecA2 domains forms a cleft in which the conserved motifs required for NTP binding and hydrolysis cluster on either side (Walbott et al., 2010). Interactions between the CTD and the helicase core form an RNA-binding tunnel that is paramount for DEAH/RHA helicase function as it stabilises RNA interactions, thereby enabling the helicase to translocate long substrate RNAs (de Bortoli et al., 2021; Hamann et al., 2019; Tauchert et al., 2017). Before translocation, the DEAH/RHA helicase loads onto its single-stranded substrate; this binding is influenced by the conformation of the RNA binding groove, which is affected by the catalytic state.

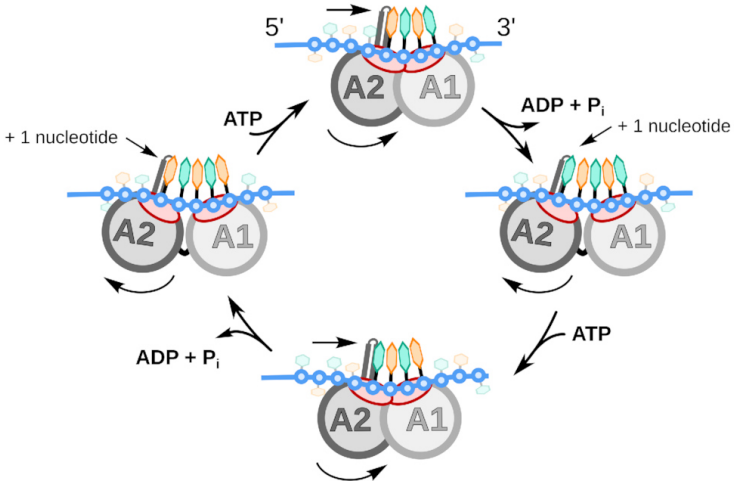


Figure 2 Molecular model of the translocation mechanism of DEAH/RHA helicases *DEAH/RHA helicases translocate at a step-size of one RNA nucleotide per hydrolysed ATP. The DEAH/RHA helicases exhibit a closed conformation of the helicase core with a stack of four RNA nucleotides bound to the RecA-like domains when bound to ATP (Top and bottom). The helicase core adopts an open conformation in the absence of an adenosine nucleotide allowing for the accommodation of an additional RNA nucleotide in the binding tunnel, resulting in a bound five-nucleotide stack (Left and right). In both conformations, the anti-parallel β-hairpin protruding from the RecA2 domain serves as a physical barrier for the nucleotide stack. Upon helicase core opening, a shift of the β-hairpin allows a fifth RNA nucleotide to be incorporated between the 5' RNA nucleotide in the stack and this structural motif. Helicase core closure, brought on by ATP binding, forces the RNA through the binding tunnel. The DEAH/RHA helicases can thus translocate along an ssRNA in the 3' to 5' direction by continuously cycling between these conformational states. The model is based on the crystallographic structures of Prp43 and Prp22 from Chaetomium thermophilum. The figure is adapted from Hamann et al., 2019.*

In an NTP-bound conformation, the RNA binding groove has a higher affinity towards its substrate than in a nucleoside diphosphate (NDP) bound state (Hamann et al., 2019; Tauchert et al., 2017). Cycling between the NTP- and NDP-bound conformational states allows the helicase to translocate in a 3′ to 5′ direction and couples NTP hydrolysis to RNA translocation (de Bortoli et al., 2021; Hamann et al., 2019; Pyle, 2008; Tauchert et al., 2017) (Figure 2). When NTP is bound, the two RecA domains are in a closed conformation that accommodates a stack of four RNA nucleotides (de Bortoli et al., 2021; Hamann et al., 2019; Tauchert et al., 2017). This closed conformation allows the conserved sequence motifs Ia, Ib, IV, and V, as well as the hook-loop, the hook-turn and the β -hairpin, to interact with the 3′ regions of the ssRNA (Hamann et al., 2019). NTP hydrolysis and the transition to an NDP-bound state lead to a shift of the RecA2 domain, which weakens the interaction between the RNA substrate and the RecA2 domain (de Bortoli et al., 2021; Hamann et al., 2019). Release of the NDP leads to an open conformation and the loss of all nucleotide-mediated interactions between the RecA-like domains. During the transition from the closed to the open conformation, the β-hairpin shifts one RNA nucleotide toward the 5′end, allowing the incorporation of a fifth RNA nucleotide between the existing four-nucleotide stack and this structural motif (de Bortoli et al., 2021; Hamann et al., 2019; Tauchert et al., 2017). Toggling between the conformational states thereby enables the DEAH/RHA helicase to translocate in a 3’-5‘ direction with a step size of one nucleotide per hydrolysed NTP (Figure 2) (Hamann et al., 2019).

1.2 DHX proteins in diverse cellular processes

DHX proteins are a group of 16 SF2 RNA helicases containing all 15 DEAH/RHA helicases and DHX58, a RIG-I-like receptor (RLR). DHX proteins are important for a diverse array of cellular processes where they contribute by translocating along or binding to deoxyribonucleic acids (DNAs), RNAs and RNPs (Jarmoskaite & Russell, 2014). DEAH/RHA helicases generally perform their function by binding to ssRNA and translocating in a 3’ to 5’ direction, allowing them to displace RNAs and proteins from their substrate through unwinding or winching (Semlow et al., 2016). In addition, some DEAH/RHA helicases bind to and manipulate DNA, dsRNA or RNA secondary structures (Fuller-Pace, 2006; Murat et al., 2018; Sweeney et al., 2021). Table 1 provides an overview of DHX proteins and their cellular functions, and the sections below explain the role of DXH proteins in diverse cellular pathways in more detail.

Table 1 Cellular pathways of DHX proteins.

DHX protein	Functions	References
DHX8	Pre-mRNA splicing	(Bertram et al., 2017; Semlow et al., 2016; Strittmatter et al., 2021)
DHX9	Transcription, translation	(Aratani et al., 2001; Fidaleo et al., 2016; Fuller-Pace, 2006; Murat et al., 2018; Nakajima et al., 1997)
DHX15	Pre-mRNA splicing, ribosome biogenesis	(K. E. Bohnsack et al., 2022; Memet et al., 2017; Tauchert et al., 2017; Walbott et al., 2010)
DHX16	Pre-mRNA splicing	(Bao et al., 2017; H.-L. Liu & Cheng, 2012)
DHX29	Translation, innate immune response	(Hashem et al., 2013; Sweeney et al., 2021; Zhu et al., 2018)
DHX30	Mitochondrial RNA metabolism, translation	(Bosco et al., 2021; Cruz-Zaragoza et al., 2021)
DHX32	Unknown	
DHX33	Translation	(Y. Zhang et al., 2015)
DHX34	RNA decay, pre-mRNA splicing	(Hug et al., 2022; Hug & Cáceres, 2014)
DHX35	Pre-mRNA splicing	(Ilagan et al., 2013; Sales-Lee et al., 2021)
DHX36	Translation	(Lyu et al., 2022; Murat et al., 2018; Sauer et al., 2019)
DHX37	Ribosome biogenesis	(Boneberg et al., 2019; Choudhury et al., 2019)
DHX38	Pre-mRNA splicing, chromosome segregation	(Nishimura et al., 2019; Zhan et al., 2018; Zhou & Reed, 1998)
DHX40	Unknown	
DHX57	Unknown	
DHX58	Innate immune response	(Duic et al., 2020)

1.2.1 In transcription

Among DHX proteins, DHX9 is uniquely able to translocate on both DNA and RNA, and interact with RNA:DNA hybrids, DNA forks, and guanine quadruplexes (Fidaleo et al., 2016; Fuller-Pace, 2006). The interaction of DHX9 with DNA substrates is a requirement for its role in transcriptional regulation, where it activates transcription by acting as a bridging factor between the CREB-binding protein (CBP) and RNA polymerase II (Pol II) (Aratani et al., 2001; Fidaleo et al., 2016; Fuller-Pace, 2006; Nakajima et al., 1997). As CBP is a general transcription coactivator and plays an important role in nuclear signalling, DHX9 is essential for various transcriptional pathways (Aratani et al., 2001).

A specialised NTD of DHX9 is responsible for binding to CBP, while a minimal transactivation domain between the NTD and the helicase core interacts with Pol II (Fuller-Pace, 2006). Interestingly, the helicase activity of DHX9 is required for the activation of CBP-dependent transcription but not for Pol II recruitment, indicating that DHX9 enhances the transcriptional machinery at promoter sites by influencing chromatin structure through local DNA unwinding (Aratani et al., 2001; Fuller-Pace, 2006). Moreover, DHX9 has been found to bridge the interaction between Pol II and the breast cancer-specific tumour suppressor BRCA1, which is involved in the transcriptional regulation of several genes involved in DNA repair and apoptosis, thus suggesting a general role for DHX9 in transcriptional regulation (Fuller-Pace, 2006).

1.2.2 In RNP biogenesis

1.2.2.1 In spliceosomal remodelling

Excision of introns from pre-mRNAs by splicing is a critical step in the maturation of mRNAs. In eukaryotic cells, pre-mRNA splicing is catalysed by the spliceosome, a highly dynamic RNP machinery involving five snRNAs and approximately 100 proteins (de Bortoli et al., 2021; Wilkinson et al., 2020). To initiate splicing, the U1 and U2 small nuclear ribonucleoproteins (snRNPs) mark the intron for splicing through recognition of the 5’ splice doner (5’SD) and the branch point (BP), respectively and form the pre-spliceosome (A complex; Figure 3)(Wilkinson et al., 2020). The U1 and U2 snRNPs then recruit the pre-assembled U4/U6.U5 tri-snRNP to form the fully assembled spliceosome called the pre-B complex. In the B complex, the transfer of the 5’SD from U1 snRNP to U6 snRNA triggers the unwinding of U6 snRNA from U4 snRNA, forming the B^{act} complex, thus allowing the U6 snRNA to fold and associate with U2 snRNA and form an active site harbouring two catalytic metal ions in (B* complex) (Wilkinson et al., 2020). Remodelling of the B* complex allows docking of the BP adenosine in the active site forming the C complex. In the C complex, the BP adenosine attacks the 5’SD positioned at the M1 catalytic metal ion, thus producing a cleaved 5’ exon. In the C* complex, the BP adenosine is replaced by the 3’ splice acceptor (3’SA) in the active site, allowing the 5’ exon to attack the

3'SA to produce mRNA and an excised lariat intron. After exon ligation, the complex is referred to as the P complex (Wilkinson et al., 2020). The release of the spliced mRNA leaves behind the intron-lariat spliceosome (ILS), which is disassembled to allow decay of the intron-lariat and recycling of the snRNAs and associated factors (Wilkinson et al., 2020).

The timely assembly, rearrangement, and disassembly of the spliceosome is partially contributed to eight helicases, including four DEAH/RHA helicases (Jarmoskaite & Russell, 2014). The DEAH/RHA helicases all function in the late stages of splicing, where they sequentially act to remodel and disassemble the spliceosome (de Bortoli et al., 2021). The first DEAH/RHA helicase, DHX16, stimulated by G-patch protein GPKOW (see section 1.3), promotes the transition from B^{act} to B* complex by dislodging proteins from the BP, freeing it for nucleophilic attack by the 5'SD site (Bao et al., 2017; K. E. Bohnsack et al., 2021; de Bortoli et al., 2021; Jarmoskaite & Russell, 2014). Subsequently, DHX38 rearranges the spliceosome for the second transesterification at the 3'SA site by destabilising the U2/U6 helix and displacing several proteins (Jarmoskaite & Russell, 2014; Zhan et al., 2018). After exon-exon ligation, DHX8 destabilise interactions between the mRNA and the U5 snRNP, thus releasing the mRNA from the spliceosome (Jarmoskaite & Russell, 2014; Strittmatter et al., 2021). Finally, DHX15, stimulated by G-patch protein TFIP11 (see section 1.3), mediates the disassembly of the spliceosome, thus allowing the release of the lariat and recycling of the spliceosome constituents (K. E. Bohnsack et al., 2021; Jarmoskaite & Russell, 2014; Tauchert et al., 2017).

DHX15 additionally plays an important role in kinetic proofreading in which suboptimal substrates are discarded (K. E. Bohnsack et al., 2022). Proofreading by DHX15 is performed in concert with DHX38 and DHX8 and ensures high splicing fidelity by aborting splicing if a suboptimal substrate is associated with the spliceosome. In cooperation with DHX38, DHX15 proofreads the 5'SD and rejects inappropriate substrates, and together with DHX8, DHX15 can detect exon ligation of suboptimal pre-mRNA substrates. Upon detecting an aberrant intermediate DHX15 abort splicing by prematurely disassembling the spliceosome, releasing the non-spliced mRNA and recycling the spliceosome constituents (K. E. Bohnsack et al., 2022). Two additional DEAH/RHA helicases, DHX34 and DHX35, have been identified in association with the spliceosomal C complex, but their functions have yet to be elucidated (Hug et al., 2022; Sales-Lee et al., 2021). Remarkably, all DEAH/RHA helicases implicated in spliceosomal transition localise to the periphery of the complexes. The helicases are, therefore, not in direct contact with the substrates that they remodel. This would indicate that these helicases do not remodel their substrate through traditional unwinding but rather pull at their substrate and remodel RNPs through a winching mechanism (Semlow et al., 2016; Wilkinson et al., 2020).

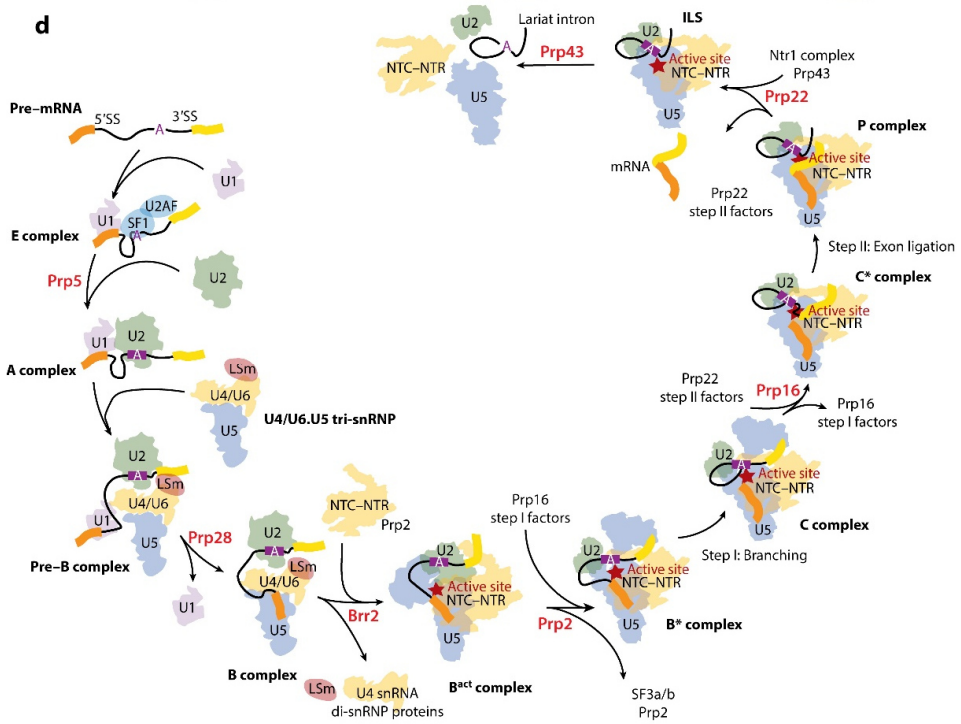


Figure 3 Schematic representation of the human splicing cycle *The eukaryotic pre-mRNA spliceosome is assembled in a highly ordered step-wise manner on its pre-mRNA substrate, characterised by a 5' splice site (5'SS or 5'SD), a BP adenosine, and a 3' splice site (3'SS or 3'SA). The early spliceosome (E complex) is initiated by binding U1 snRNP to the 5'SS. Next, the U2 snRNP is recruited to the BP sequence, yielding the pre-spliceosome (A complex). The fully assembled pre–B complex is formed by joining the pre-spliceosome to the pre-assembled tri-snRNP. In the B complex, 5'SS release from the U1 snRNP by Prp28/DDX23 induces a dramatic remodelling of the spliceosome. Once U4 snRNA is unwound from U6 snRNA, the U2 and U6 snRNAs can form the active site of the B^{act} complex cradled by the Prp19-associated complex (NTC) and the Prp19-related complex (NTR). Prp2/DHX16 stimulate remodelling resulting in the B* complex in which the active site is fully competent to catalyse the branching reaction, thus leading to the C complex. Prp16/DHX38 promotes remodelling of the spliceosome to form the C* complex, which adopts the conformation necessary for exon-ligation. After exon ligation, the complex is termed the post-catalytic spliceosome (P complex). Prp22/DHX8 catalyses the translocation-dependent release of the ligated exons (mRNA), thus separating it from the intron-lariat spliceosome (ILS). Disassembly of the ILS to release the intron lariat and recycle the snRNAs and associated factors is facilitated by Prp43/DHX15. SF2 helicases from the DEAD-box, DEAH/RHA, and Ski-2 families are marked in red. The figure is from Wilkinson et al., 2020.*

1.2.2.2 In ribosome biogenesis

Ribosome assembly is a complex, multistep process spanning several cellular compartments, and involves the assembly of four rRNAs (18S, 5.8S, 28S, and 5S), approximately 80 ribosomal proteins (RPs), and is mediated by a vast array of ribosome biogenesis factors (RBFs) (K. E. Bohnsack & Bohnsack, 2019). Biogenesis of the small subunit (SSU; 40S), and the large subunit (LSU; 60S) requires precise coordination of numerous transient and stable

protein-RNA and RNA-RNA interactions to ensure progression through a series of conformational states (Figure 2) (Duss et al., 2019). Eukaryotic ribosome assembly starts in the nucleolus with RNA polymerase I (POL I) transcription of the rDNA to produce the 35S pre-rRNA, which undergoes several processing steps to give rise to 18S, 5.8S and 28S rRNA (Peña et al., 2017) (Figure 4). Cotranscriptional cleavage of the 35S pre-rRNA release the 20S pre-rRNA and the 27S pre-rRNA, which, together with subunit-specific RPs and RBFs form the early 40S and 60S pre-ribosomes, respectively (Peña et al., 2017). The pre-60 and pre-40 subunits undergo independent maturation pathways, aided by various transiently associating assembly factors, as they are shuttled to the cytoplasm (Peña et al., 2017). Before the pre-ribosomes achieve translational competence, they undergo functional proofreading linked to the release of remaining RBFs (Peña et al., 2017).

Most of our current knowledge on ribosome biogenesis in eukaryotes comes from studies in the yeast *Saccharomyces cerevisiae*. Ribosome biogenesis is highly conserved amongst eukaryotes, and various human proteins have been identified as homologues of yeast proteins and are generally expected to fulfil the same function (Martin et al., 2013). Human ribosome biogenesis involves 23 RNA helicases, but only a few are functionally characterised in relation to ribosome biogenesis. Five of the 23 helicases belong to the DEAH/RHA helicase family, namely DHX9, DHX15, DHX16, DHX33, and DHX37 (Martin et al., 2013). RNA helicases, implicated in ribosomal maturation, are responsible for several rearrangement events, including the release of snoRNAs, recruitment and dissociation of RPs and RBF, and manipulation of rRNA secondary and tertiary structures (K. E. Bohnsack & Bohnsack, 2019). A crucial event in SSU maturation is the base pairing between U3 snoRNA and the 18S rRNA. The U3 snoRNA bridges key interactions in the pre-rRNA and is vital for subunit maturation. Subsequently, DHX37 promotes dissociation of the U3 snoRNA, thus allowing the pre-rRNA to progress to the next conformational state (Boneberg et al., 2019; Choudhury et al., 2019). Additionally, the DEAH/RHA helicase DHX15, stimulated by NKRF, is required for efficient cleavage of the precursor transcript by XRN2 (Memet et al., 2017).

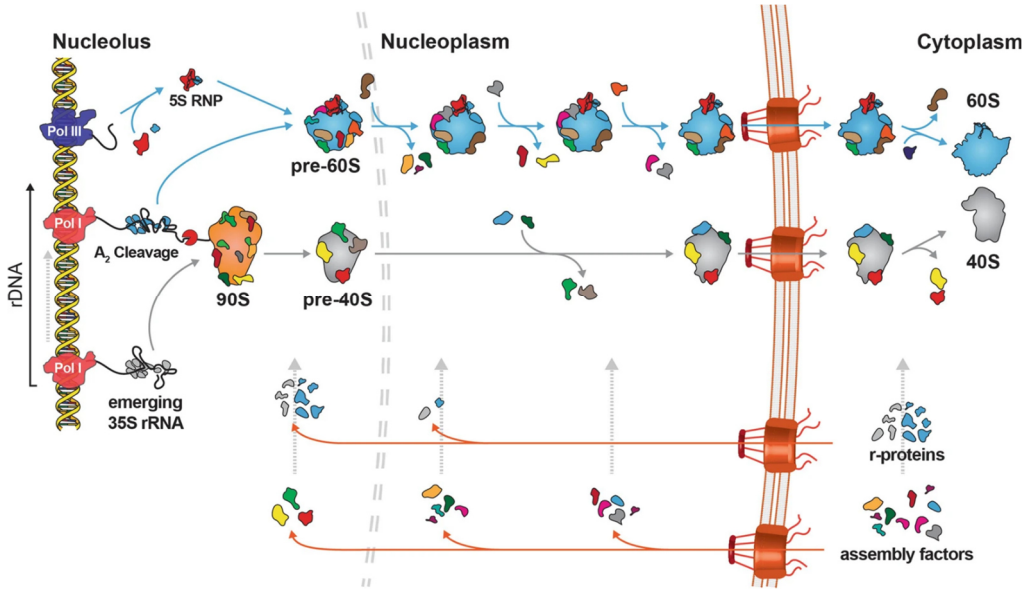


Figure 4: Schematic model for the eukaryotic ribosome assembly *Eukaryotic ribosome assembly starts in the nucleolus with RNA polymerase I (POL I) transcription of the rDNA to produce the 35S pre-rRNA, the precursor to 18S, 5.8S and 25S rRNAs, and Pre-5S rRNA transcription by RNA Pol III. A large RNP complex termed the 90S pre-ribosome, consisting of snoRNAs and 40S-specific RPs, assembles on the nascent 35S transcript. Cotranscriptional cleavage of the 35S separates the 40S pre-ribosome from the 27S pre-rRNA, which then recruits 60S-specific RPs and RBFs to form the pre-60S ribosome. The pre-60 and pre-40 subunits undergo independent maturation aided by multiple transiently associating RBPs and transport receptors that facilitate the translocation of pre-ribosomes through nuclear pore complexes into the cytoplasm (Oborská-Oplová et al., 2022; Peña et al., 2017). The Figure is from Oborská-Oplová et al., 2022.*

1.2.3 In mitochondrial RNA metabolism

Mitochondria shape their proteome through the import of nucleus-encoded proteins and the expression of their own genome. Human mitochondrial DNA (mtDNA), a circular molecule of ~16.5 kb, encodes thirteen core subunits required for oxidative phosphorylation, two mitochondrial ribosomal RNAs (mt-rRNAs) and 22 mt-tRNAs (Cruz-Zaragoza et al., 2021; Pearce et al., 2017). The mitochondrial mRNAs are transcribed, translated, and processed within the mitochondrial matrix. These steps, as well as ribosome maturation and RNA turnover, are thought to occur at defined foci termed mitochondrial RNA granules (RMG) (Cruz-Zaragoza et al., 2021; Pearce et al., 2017). In addition to the mt-DNA encoded proteins, an estimated 250-300 nucleus-encoded proteins are dedicated to serving mitochondrial gene expression (Pearce et al., 2017).

DEAH/RHA helicases are present not only in the nucleoplasm and cytoplasm but also in the mitochondria, as exemplified by DHX30. DHX30 localise to both cytoplasm and mitochondria, but it is more abundant in the latter (Bosco et al., 2021). Within mitochondria, DHX30 associates with mitochondrial ribosomes, where it non-specifically primes mRNAs for translation (Antonicka & Shoubridge, 2015; Bosco et al., 2021; Cruz-Zaragoza et al., 2021).

Although it associates with mitoribosomes, the role of DHX30 in mitochondrial ribosome assembly is currently under debate. Acute DHX30 depletion does not affect mitochondrial ribosome assembly, but translation defects might indirectly affect ribosome formation due to reduced levels of mitochondrial proteins (Antonicka & Shoubridge, 2015; Cruz-Zaragoza et al., 2021).

1.2.4 In translation

Translation initiation of most eukaryotic mRNAs occurs by scanning the 5' untranslated region (UTR) for an initiation codon starting from the cap-proximal region and progressing in the 3' direction (Sweeney et al., 2021). The scanning mechanism involves the 43S preinitiation complex (43S PIC), a complex functional interplay between the 40S ribosomal subunit, initiator methionine transfer RNA ($tRNA_i^{Met}$), guanosine triphosphate (GTP) and the eukaryotic initiation factors (eIF); eIF1, eIF1A, eIF2, and eIF3. The 43S PIC attaches to the mRNA, a process that requires the cooperative action of eIF4A, eIF4B and eIF4F, and then scans until initiation codon recognition, whereafter it forms the 48S initiation complex (48S IC) with established codon-anticodon base-pairing (Hashem et al., 2013; Sweeney et al., 2021). Scanning through stable secondary structures furthermore requires the recruitment of DHX29 to the mRNA entrance of the 43S PIC. In this context, DHX29 resolves secondary structures before the mRNA is threaded through the 43S PIC, thus ensuring a linear nucleotide by nucleotide inspection (Hashem et al., 2013; Sweeney et al., 2021).

Translation of mRNAs is negatively affected by the formation of G-quadruplexes (G4s), one of the most stable nucleotide secondary structures (Murat et al., 2018; Sauer et al., 2019). The DEAH/RHA helicase DHX36 is recruited to G4s on mRNAs by its N-terminal domain and resolves these structures by an unclear mechanism, thus improving translation efficiency (Lyu et al., 2022). DHX9 can similarly resolve these G4 structures in an NTP-dependent manner, and recruitment of DHX9 to mRNAs improves translation efficiency and prevents the accumulation of translationally incompetent mRNAs (Caterino & Paeschke, 2022; Murat et al., 2018; Sauer et al., 2019). Moreover, DHX33 and DHX30 also regulate translation by promoting translation initiation and reducing global translation, respectively (Bosco et al., 2021; Y. Zhang et al., 2015). However, the exact mechanisms of translational regulation by DHX30 and DHX33 are not yet fully understood (Bosco et al., 2021; Y. Zhang et al., 2015).

1.2.5 In RNA decay

Processing and degradation of RNAs is a crucial cellular function that allows the cell to recycle essential biomolecules, remove unwanted ncRNAs, and halt the translation of mRNAs where the products are unnecessary or even detrimental to the cell (Khemici & Linder, 2018). Eukaryotic cells degrade RNA in a multitude of ways accentuating the cells' demand for RNA degradation; these degradation pathways include deadenylation-dependent mRNA decay,

nonsense-mediated RNA decay (NMD), non-stop decay, and no-go decay (Tuck et al., 2020). NMD is a translation-coupled quality control mechanism that eliminates mRNAs that contain premature termination codons (PTCs) (Hug & Cáceres, 2014). In vertebrates, NMD is coupled to pre-mRNA splicing via the exon junction complex (EJC) to ensure efficient degradation of RNAs containing PTCs and reduce errors in gene expression (Hug & Cáceres, 2014). When the ribosome encounters a PTC, the surveillance complex (SURF) is on the mRNA. The SURF complex gives way to the decay-inducing complex (DECID), and ribosomal subunits, release factors, and the nascent peptide dissociate from the RNA. Finally, the decay of the mRNA is triggered (Nogueira et al., 2021).

RNA helicases have a central role in NMD as their ability to remodel RNPs aids NMD progression, as seen in the transition from the SURF complex to the DECID complex promoted by DHX34 (Hug et al., 2016). In this context, DHX34 associates with the SURF complex through interaction with hypophosphorylated UFP1 and subsequently promotes the recruitment of UFP2 and the dissociation of the eukaryotic release factors (eRF), eRF1 and eRF3 in an ATP-dependent manner (Hug et al., 2016; Hug & Cáceres, 2014).

1.2.6 In chromosomal segregation

In addition to its role in pre-mRNA splicing, DHX38 is involved in chromosomal segregation (Nishimura et al., 2019). Chromosome segregation is a tightly regulated process in which genetic information is divided equally between daughter cells. The human centromeres, which are paramount for segregation, consist of repetitive sequences from which the ncRNAs Satellite I, II, and III, are transcribed (Nishimura et al., 2019). Depletion of these ncRNAs causes chromosome segregation defects. DHX38 interacts with Satellite I, and the knockdown of DHX38 results in chromosome segregation defects independent of pre-mRNA splicing (Nishimura et al., 2019). However, the role of DHX38 in chromosome segregation requires additional elucidation.

1.2.7 In innate immunity

The innate immune system continuously surveys the cellular environment for any sign of infection. RNA helicases can act as cellular sensors to detect viral nucleic acids in the cytoplasm of infected cells and activate the appropriate immune response (Ranji & Boris-Lawrie, 2010; Stok et al., 2022). The RLRs, RIG-I and MDA5, are recruited to foreign dsRNA through their CTD and expose their N-terminal CARDs upon forming an RNA-supported oligomeric filament on their substrate. The exposed CARDs activate MAVS, facilitating the activation of transcriptional regulators responsible for initiating type I interferon (IFN) response (Duic et al., 2020; Stok et al., 2022). DHX58, the least understood member of the RLR family, lacks CARDs and cannot activate MAVS independently. Instead, it acts in concert with MDA5 to modulate signal transduction (Stok et al., 2022). DXH58 regulates the formation rate and

length of MDA5-RNA filaments by enhancing MDA5-RNA interactions while simultaneously restricting the length by dissociating MDA5 in an ATP-dependent manner. This adjustment of filaments leads to an enhanced antiviral response as short and dense filaments have greater antagonistic activity than longer filaments due to the increased number of contacts between the CARD domains in dense filaments (Stok et al., 2022).

The RLR MDA5 furthermore works in cooperation with the helicase DHX29 to sense infection by Picornaviridae family viruses, including encephalomyocarditis virus (EMCV) (Zhu et al., 2018). In this context, DHX29 recognise the structured 5'-UTRs typical for this family of viruses and promotes MDA5 binding and RNA filament formation (Sweeney et al., 2021; Zhu et al., 2018). This function of DHX29 is similar to its function in translation initiation; therefore, it is speculated that DHX29 recognises the viral RNA during the scanning of the mRNA by the 43S PIC (Zhu et al., 2018).

1.3 Modes of regulation of DHX proteins

Strict spatial, temporal, and catalytic regulation of RNA helicases is essential for cells, as dysregulation of helicase activity is associated with tumourigenesis and disease (K. E. Bohnsack et al., 2021; Steimer & Klostermeier, 2012). However, since the helicase core of SF2 helicases interacts primarily with the sugar-phosphate backbone of substrate RNAs, it generally lacks intrinsic specificity. This non-specificity necessitates dedicated strategies to limit the promiscuity of SF2 helicases and to ensure their target specificity (K. E. Bohnsack et al., 2021; Donsbach & Klostermeier, 2021). Proteins are regulated in a plethora of ways, including control of protein amount by altered transcription, translation, and turnover rate, spatial regulation by recruitment to cellular compartments and target RNAs/RNPs, and changes in conformation to modify the catalytic activity or substrate binding. Cells often require rapid and transient regulation of existing proteins, either by post-translational modification or by interaction with biomolecules (Cooper, 2019).

Regulation of DEAH/RHA helicases and RLRs is generally conferred either in *cis* by auxiliary domains or in *trans* by interacting partners. Regulatory auxiliary domains include auto-inhibitory domains that induce a conformational state incompatible with catalysis, domains that recognise specific RNA features, and domains that ensure appropriate localisation (K. E. Bohnsack et al., 2021; Donsbach & Klostermeier, 2021; Kato et al., 2021; Sloan & Bohnsack, 2018). In addition, post-translation modifications can regulate the catalytic activity, turnover rate, and localisation of RNA helicases (Barber & Rinehart, 2018; Donsbach & Klostermeier, 2021; Hage et al., 2022; Kato et al., 2021; Z. Zhang et al., 2013). However, the predominant mechanism of RNA helicase regulation is interactions with cofactor proteins (K. E. Bohnsack et al., 2021).

1.3.1 Regulation of RNA helicases by protein cofactors

Protein cofactors generally recruit their cognate RNA helicases to specific RNA substrates and often, with a few exceptions, double as enhancers of otherwise poor helicase activity (Hamann et al., 2020; Studer et al., 2020). Several families of protein cofactors have evolved alongside RNA helicases to ensure the precise and flexible regulation of RNA helicase. Examples of protein cofactor families include LOTUS domain cofactors which regulate Vasa, a DEAD-box helicase in *Drosophila melanogaster*, MIF4G domain cofactors that regulate several DEAD-box proteins in human and yeast, and G-patch domain cofactors which regulate DEAH/RHA helicases in human and yeast (Donsbach & Klostermeier, 2021; Sloan & Bohnsack, 2018). In addition, several helicases have specialised cofactors that are not part of bigger families, including DHX37, which is regulated by UTP14A and DHX9, which is regulated by Nup98 and EWS-FLI1 (Capitanio et al., 2017; Choudhury et al., 2019; Donsbach & Klostermeier, 2021; Erkizan et al., 2015; Sloan & Bohnsack, 2018).

The G-patch protein family is a large and diverse class of DEAH/RHA helicase cofactors regulating specific RNA helicases. This family of protein cofactors is characterised by a glycine-rich motif of approximately 50 amino acids that directly binds and activates the helicases. G-patch proteins vary significantly in size (21–264 kDa), domain composition, and cellular localisation (K. E. Bohnsack et al., 2021). However, all have a single G-patch motif embedded in an intrinsically disordered region, and the G-patch proteins show a high prevalence of RNA-binding motifs and domains (Figure 5)(K. E. Bohnsack et al., 2021). Furthermore, almost half of the known G-patch proteins have additional motifs that mediate protein-protein interactions. The G-patch motif is widespread in eukaryotic proteins but is absent from bacteria and archaea. Human cells express more than 20 G-patch proteins, whereas *Saccharomyces cerevisiae* only expresses five (Aravind & Koonin, 1999). In *S. cerevisiae*, four of the five G-patch proteins, i.e. Cmg1, Pxr1, Sqr1, and Spp382, interacts with Prp43 (DHX15 in humans), while only one, Spp2, interacts with Prp2 (DHX16 in humans)(K. E. Bohnsack et al., 2021). In humans, a similar distribution is observed with most of the characterised G-patch proteins, i.e. NKRF, PINX1, RBM5, RBM17, TFIP11, ZGPAT, CMTR1, and CHERP, interacting with DHX15, while only one, GPKOW, interacts with DHX16 (K. E. Bohnsack et al., 2021). The larger total number of human G-patch proteins likely reflects the increased need for regulating RNA metabolism in higher eukaryotes (K. E. Bohnsack et al., 2021; Robert-Paganin et al., 2015), while the similar distribution likely reflects a conserved mode of regulation of Prp43/DHX15 (K. E. Bohnsack et al., 2021, 2022).

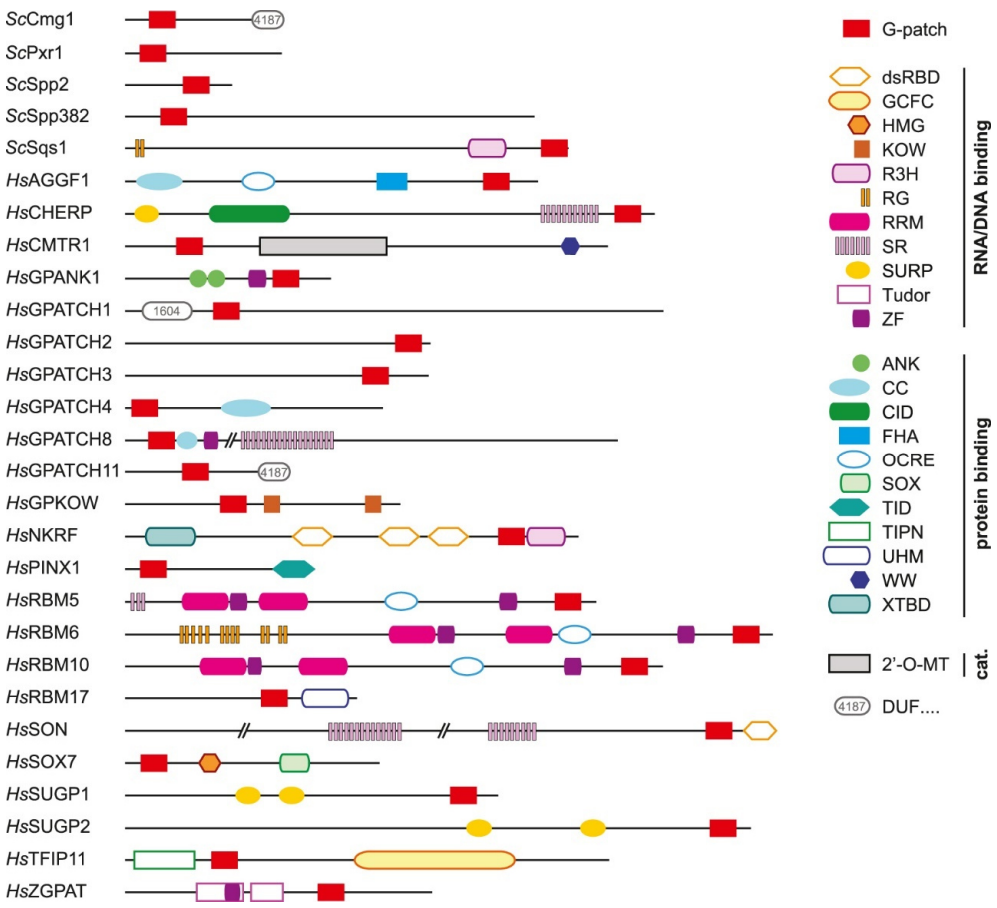


Figure 5: Schematic representation of domain organisation of yeast and human G-patch proteins. *Proteins and their components are drawn to scale except where indicated (// = 650 amino acids). The prefixes Sc and Hs mark proteins from S. cerevisiae and Homo sapiens, respectively. Domains in the figure are annotated as: double-stranded RNA binding domain (dsRBD), GC-rich sequence DNA-binding factor-like protein (GCFC), high-mobility group box (HMG), „Kyprides, Ouzounis, Woese" motif (KOW), R3H motif (R3H), arginine-glycine repeats (RG), RNA recognition motif (RRM), serine-arginine repeats (SR), suppressor-of-white-apricot and PRP21/SPP91 (SURP), Tudor domain (Tudor), Zinc finger (ZF), ankyrin repeat (ANK), coiled-coil (CC), RNA polymerase 2 C-terminal domain interacting domain (CID), forkhead-associated domain (FHA), octamer repeat domain (OCRE), SOX17/18 central domain (SOX), telomerase inhibitory domain (TID), Tuftelin interacting protein N-terminal domain (TIPN), U2AF homology motif (UHM), two tryptophan-containing domain (WW), XRN2 binding domain (XTBD), ribose 2'-O-methyltransferase (2'-O-MT), and Domain of unknown function (DUF). The figure is from Bohnsack et al., 2021.*

Mechanistic and structural information about the interaction between DEAH/RHA helicases and their cognate G-patch cofactors was recently elucidated by cryo-EM structures of DHX15 in complex with the isolated NKRF G-patch motif (NKRF$^{G\text{-patch}}$) and Prp2 (DHX16 homologue) in complex with Spp2$^{G\text{-patch}}$ (Hamann et al., 2020; Studer et al., 2020). When bound to their cognate DEAH/RHA helicase, G-patch motifs increase RNA-binding affinity, as well as ATPase and RNA translocation rates (Hamann et al., 2020; Studer et al., 2020). The cryo-EM structures revealed that the N-terminal brace helix of the G-patch motif packs perpendicular to the long helix of the WH domain, while the C-terminal brace-loop stacks into a hydrophobic pocket on

top of the RecA2 domain (Hamann et al., 2020; Studer et al., 2020). The two points of interaction position the G-patch motif across the back of the RNA-binding channel and tether the WH-domain and RecA2 domain together (Hamann et al., 2020; Studer et al., 2020). The G-patch motif thus functions as a flexible brace that restricts the conformational freedom of its cognate helicase, inducing a configuration compatible with strong RNA binding while allowing sufficient flexibility for the motions required for ATP hydrolysis and RNA translocation (Hamann et al., 2020; Studer et al., 2020).

In human cells, eight different G-patch proteins are known so far to interact with DHX15, of which a subset is functionally characterised, including TFIP11, which likely coordinates DHX15 in spliceosomal disassembly; RMB17 and CHERP, which coordinate DHX15 in splice site regulation; and NKRF, which associates with DHX15 in the context of ribosome assembly (K. E. Bohnsack et al., 2022; de Maio et al., 2018; Han et al., 2017; Hegele et al., 2012; Memet et al., 2017; Yoshimoto et al., 2009). The relationship between DXH15 and G-patch proteins potentially forms the molecular basis for the multifunctionality of DHX15, as the mutually exclusive interactions can effectively partition the helicase between its different target pathways (K. E. Bohnsack et al., 2022; Heininger et al., 2016). For example, in *S. cerevisiae*, perturbations in G-patch cofactors of Prp43, the DXH15 homologue, alter the distribution of the helicase in cellular pathways (K. E. Bohnsack et al., 2022; Heininger et al., 2016).

In contrast to DHX15 and DHX16, most human DEAH/RHA helicases have substitutions at the two G-patch contact sites and are incompatible with conserved G-patch binding (Studer et al., 2020). Therefore, to enable the targeted recruitment and activation of these otherwise non-specific helicases to various pathways, several DEAH/RHA helicases have evolved divergent mechanisms of activation and recruitment by gaining various surface patches or additional N-terminal domains, exemplified by DHX37 and DHX9 (Studer et al., 2020). DHX37 interacts directly with the SSU biogenesis factor UTP14A, which stimulates the helicase's ATPase rate (Boneberg et al., 2019; Choudhury et al., 2019). In contrast, DHX9 has two N-terminal double-stranded RNA-binding domains (dsRBDs) that bridge the WH and RecA2 domains and could functionally replace activation by G-patches by taking over the same domain-tethering role (Prabu et al., 2015).

1.3.2 Regulation of SF2 helicases by E3 ligases

As with RNA helicases, E3 ubiquitin ligases mediate many important cellular functions, but the actions of helicases and E3 ligases have mainly been studied in distinct cellular contexts. Most E3 ligases add a layer of regulation by mediating ubiquitin transfer to target proteins, allowing rapid modulation of the stability and function (Kato et al., 2021; Williams et al., 2019). A three-enzyme ubiquitin transfer cascade, consisting of an E1 activation enzyme, an E2 conjugating enzyme, and an E3 ligase, facilitates the ubiquitination of target proteins. In the cascade, the

E3 ligase performs the final step in the cascade, thus dictating substrate specificity (H. T. Wang & Hur, 2021; Williams et al., 2019)

Several SF2 RNA helicases associate with tripartite motif (TRIM) or TRIM-like E3 ligases, including the three RLRs, RIG-I, MDA5, and DHX58 (Kato et al., 2021), the DEAD-box helicase DDX41 (Kato et al., 2021), and the DEAH/RHA helicases DHX33 and DHX16 (Hage et al., 2022; Weng et al., 2014). This family of E3 ligases is emerging as a key player in innate immunity, and many TRIMs enhance downstream signal transduction through both proteasome-dependent and proteasome-independent mechanisms (Kato et al., 2021; H. T. Wang & Hur, 2021). TRIMs are characterised by three distinct N-terminal domains; the RING domain that binds the E2 conjugating enzyme, two B-box domains and a coiled-coil (CC) domain that facilitates homodimerisation (H. T. Wang & Hur, 2021). The order of domains is the same in all TRIM proteins, but some proteins have linkers that replace one of the three domains (H. T. Wang & Hur, 2021). Although not formally recognised as TRIM E3 ligases, TRIM-like proteins have similar functions and harbour two of the three characteristic domains in the same order. Most TRIM and TRIM-like proteins have C-terminal domains often used to facilitate substrate recognition and binding; the C-terminal domains are less conserved and vary within the family, albeit with some commonalities (H. T. Wang & Hur, 2021). The most common C-terminal domain is the SPRY domain, which is present in approximately half of all known TRIMs (H. T. Wang & Hur, 2021; Williams et al., 2019). In invertebrates, the SPRY domain is often extended N-terminally with a PRY domain to form a PRY/SPRY fusion domain (Williams et al., 2019). The PRY/SPRY domain displays a twisted beta-sandwich fold with a cluster of variable loops (VLs) on one edge; these VLs establish contact with target proteins and thus confer specificity (Kato et al., 2021).

The TRIMs, RIPLET and TRIM65, exemplify the importance of TRIM-meditated regulation of RNA helicases, as they are paramount for activating the RLRs, RIG-I and MDA5. Homodimeric RIPLET and TRIM65 selectively recognise and bind to filamentous oligomeric forms of RIG-I and MDA5, respectively (Kato et al., 2021). Upon interaction between the homodimeric TRIMs and their cognate RLR, the TRIMs facilitate the ubiquitination of the CARDs, thus promoting homo-tetramerisation of the domains. As a result, tetrameric CARDs can stably interact with MAVS, which initiates an antiviral immune response (Kato et al., 2021; Wang & Hur, 2021). This filament-specific recognition requires TRIM bivalency and multiple interactions between the VLs of the PRY/SPRY and the RecA2 domains (Figure 6). The low-affinity but high specificity of these interactions ensures an avidity-dependent recognition where homodimeric TRIMs only recognise and modify their cognate helicase in its filamentous state, guaranteeing high fidelity of the antiviral immune response (Kato et al., 2021).

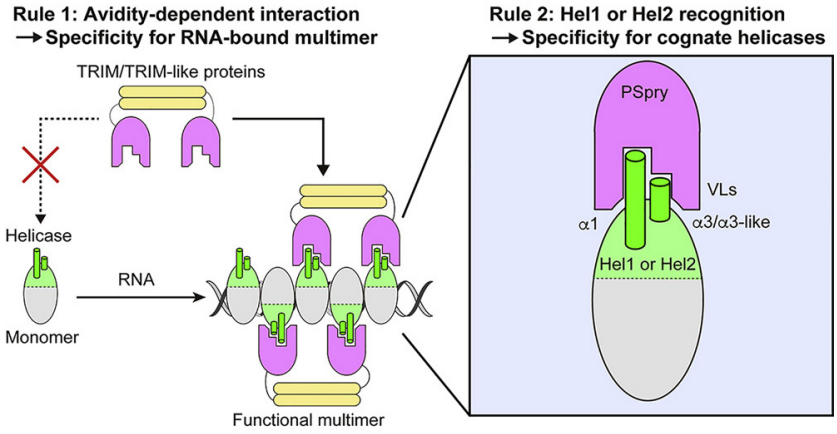


Figure 6: Schematic overview of the two rules of engagement between helicases and TRIM proteins. *Schematic representation of the two evolutionary rules of engagement between helicases and TRIM /TRIM-like proteins. Rule 1: Dimerised TRIM/TRIM-like proteins specifically recognise the RNA-supported filament of their cognate helicase (functional multimer). Rule 2: The PRY/SPRY (PSPRY) domains of TRIM/TRIM-like protein bind to a common epitope, consisting of two alpha-helices and VLs, in the core of their cognate helicase. The figure is from Kato et al., 2021.*

The RLR, DHX58, interacts with TRIM14 in a similar filament-dependent manner; however, as TRIM14 does not have a RING domain and DHX58 does not have CARDs, ubiquitination does not occur (Kato et al., 2021). The DEAD-box helicase, DDX41, similarly associates with TRIM21, TRIM26, and TRIM41 (Kato et al., 2021; Shinriki & Matsui, 2022; Z. Zhang et al., 2013). TRIM21 has been shown to polyubiquitinate DDX41 on lysine 9 (K^{9}) and K^{115} with K^{48}-linked ubiquitin chains (Z. Zhang et al., 2013), but the roles of the TRIM26 and TRIM41 interactions remain unclear. The DEAH/RNA helicases, DHX16 and DHX33, interact with TRIM6 and TRIM33, respectively (Hage et al., 2022; Weng et al., 2014). Overexpression of DHX16 and TRIM6 causes relocation of DHX16 to the cytoplasm, where it enhances RIG-I-dependent antiviral immune response (Hage et al., 2022). Whether, the mode of interaction, is similar to that between RLRs and their cognate TRIMs remains unknown. TRIM33 binds DHX33 and induces K^{63}-linked polyubiquitination of K^{218}; TRIM33 does, however, not have a PRY/SPRY domain and therefore does not interact according to the avidity-dependent model described above (Weng et al., 2014). K63-linked polyubiquitination of DHX33 by TRIM33 is important for RNA-induced activation of the inflammasome and results in proteasomal degradation of DHX33 unless rescued by the deubiquitinase USP36 (Fraile et al., 2018; Weng et al., 2014). Interestingly, DHX40 and DDX24 are substrates of the deubiquitinase USP7 and the knockdown of USP7 causes a decrease in DHX40 and DDX24 protein levels, which is rescued by proteasomal inhibitors, thus indicating that the protein levels of DHX40 and DDX24 are regulated proteasomal degradation (Georges et al., 2018). However, the E3 ligases responsible for the ubiquitination of the helicases remain unidentified.

1.4 Retrotransposition

1.4.1 Transposable elements in humans

Transposable elements (TE) are mobile genetic elements capable of replication. While a typical gene resides at a discrete chromosomal locus, TEs are present in multiple copies at numerous genomic locations. At least 46% of human DNA derives from TEs, and computational algorithms capable of detecting ancient, highly mutated TEs, suggest they account for as much as 70% of the human genomic DNA (gDNA) (Lander et al., 2001; Richardson et al., 2015). However, most TE-derived sequences in the mammalian genome cannot mobilise as they have accumulated mutations and deletions throughout evolution (Richardson et al., 2015). TEs divide into two major classes based on their mobilisation intermediates. Class I elements, also known as retrotransposons, mobilise via RNA intermediates, while class II elements, also known as DNA transposons, mobilise through DNA intermediates (Platt et al., 2018; Richardson et al., 2015). Furthermore, TEs can either be autonomous and encode the proteins necessary for their mobilisation or non-autonomous and thus require proteins encoded by other TEs.

Retrotransposons mobilise by creating new copies of themselves through a 'copy-and-paste' mechanism whereby an RNA intermediate is reverse-transcribed into a complementary DNA (cDNA) and integrated into the genome (Bourque et al., 2018; Platt et al., 2018). Retrotransposons fall into two major groups, the long terminal repeat (LTR) elements, which are structurally similar to retroviruses, and the non-LTR elements (Platt et al., 2018). LTR retrotransposons, including endogenous retroviruses (ERVs), and their non-autonomous derivatives, are present at >450,000 copies in the human genome and comprise approximately 8% of the gDNA (Lander et al., 2001; Richardson et al., 2015). Autonomous LTR elements encode at least a *gag* and *pol* protein, flanked by the 100-300 base pair (bp) direct terminal repeats that give the elements their name (Bourque et al., 2018; Platt et al., 2018). Much like retroviruses, the integration of LTR elements occurs through cleavage and strand transfer catalysed by an integrase (Bourque et al., 2018).

The non-LTR elements include long interspersed elements (LINEs) and short interspersed elements (SINEs), which account for approximately 21% and 13% of the human genome, respectively (Lander et al., 2001; Richardson et al., 2015). Three distinct LINE families exist in the human genome, LINE-1 (L1), LINE-2 (L2), and LINE-3, but only L1 is active (Lander et al., 2001). Autonomous L1s encode two proteins necessary for retrotransposition; ORF1p, a nucleic acid chaperone, and ORF2p, an endonuclease and reverse transcriptase. L1 retrotransposition occurs through an mRNA intermediate that acts as a template for synthesis of a cDNA, which inserts into the genome at a target sequence. The L1-encoded proteins can either act in *cis* and mobilise their genomic element or in *trans* to mobilise non-autonomous

SINEs or cellular mRNAs, with the latter giving rise to processed pseudogenes (Richardson et al., 2015). SINEs, consisting of Alu and SINE-R/VNTR/Alu-like retrotransposons (SVA elements), are a family of exclusively non-autonomous retrotransposable elements freeloading on the proteins expressed by active L1s (Richardson et al., 2015)

Class II elements comprise only 3% of the human genome, and virtually all human DNA transposons are mutated to the point of transposition incompetence (Richardson et al., 2015). DNA transposons are subdivided into two major groups; cut-and-paste transposons and rolling-circle (RC) transposons, also known as Helitron transposons (Kapitonov & Jurka, 2007; Platt et al., 2018). Autonomous 'cut-and-paste' transposons contain terminal inverted repeats (TIRs) surrounding an open reading frame encoding a transposase (Richardson et al., 2015). The transposase facilitates the transposition of a double-stranded DNA intermediate originating from the autonomous transposon or non-autonomous derivatives with similar TIRs (Platt et al., 2018). Autonomous RC transposons encode a RepHel protein, which facilitates the transposition of a single-stranded DNA intermediate through rolling-circle-replication; the details of eukaryotic RC transposition are, however, sparse (Platt et al., 2018).

1.4.2 L1 retrotransposons

The only transposable elements still active in humans are L1, *Alu*, and SVA elements, of which only L1s are autonomous and are therefore also responsible for mobilising the non-autonomous *Alu* and Sva elements (Richardson et al., 2015). L1 retrotransposons account for 17% of the human genome, which includes more than 500,000 copies; however, only approximately 150 copies are full-length and retrotransposition-competent (Lander et al., 2001; Penzkofer et al., 2017). Phylogenetic analysis of human L1 ORFs divides the L1 elements into three distinct lineages: primate-specific $L1PA_{17-1}$ and $L1PB_{3-1}$, and mammalian-specific $L1MA_{4-1}$ (Figure 1) (Khan et al., 2006). The three lineages evolved in parallel in ancestral primate genomes, but ultimately only the $L1PA_{17-1}$ persisted; thus, all retrotransposition-competent L1s belong to the youngest family of the L1PA lineage, i.e. L1PA1 (Khan et al., 2006).

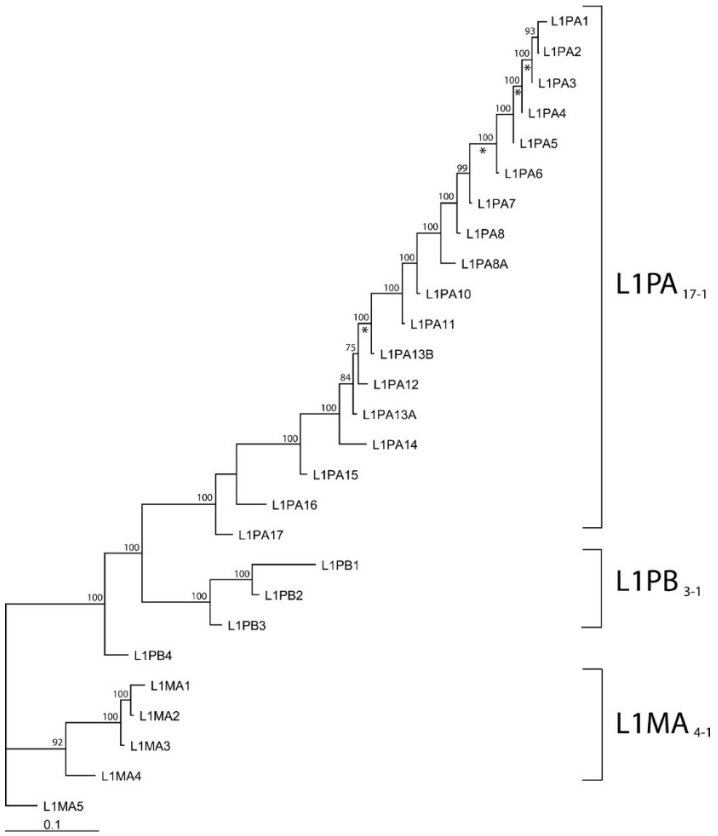


Figure 7: Phylogeny of L1 consensus sequences *A maximum likelihood tree is based on the consensus sequences of the ORF1 and ORF2 of 27 L1 families. The tree shows three separate lineages of the L1, the mammalian-specific L1MA$_{4–1}$ lineage and the two primate-specific lineages, L1PB$_{3–1}$ and L1PA$_{17–1}$. The tree's topology indicates that the three lineages were simultaneously active in ancestral primates. Within the lineages, generally, only one dominant family evolved at the time. The numbers above the nodes indicate the percentages of time the labelled node was present in 1000 bootstrap replicates of the data. Asterisks mark branches on which the free-ratio model assigned estimates of ω >1; a free ratio of significantly >1 indicates that non-synonymous substitutions have reached fixation faster than synonymous substitutions, indicative of positive selection. The figure is from Khan et al., 2006.*

A retrotransposition-competent human L1 is approximately 6 kbp long. It contains a ~900 bp GC-rich 5' UTR with an internal bidirectional RNA polymerase II promoter, two open reading frames (ORF), ORF1 and ORF2, and a 3'UTR with a polyadenylation signal (polyA). In addition, in the antisense direction, opposite the 5'UTR, resides ORF0 and an additional promoter (Figure 2) (Protasova et al., 2021). ORF1 encodes ORF1p, a 40 kDa RBP with nucleic acid chaperone activity. The ORF1p consists of an N-terminal domain (NTD), a CC domain, an RNA recognition motif (RRM), and a CTD (Protasova et al., 2021). The NTD is intrinsically disordered and has a putative RNA binding function. The CC domain contains 14 heptad repeats that facilitate ORF1p homotrimerisation. The RRM and the CTD bind directly to RNA and facilitate nuclear localisation (B. T. Freeman et al., 2019). ORF2 encodes ORF2p, a 150 kDa protein with endonuclease (EN) and reverse transcriptase (RT) activities

(Richardson et al., 2015). The ORF2p consists of an endonuclease domain, a cryptic domain (Cry), a Z-domain, a reverse transcriptase domain, and a CTD with a cysteine-rich region (Christian et al., 2016; Protasova et al., 2021). The ORF0 encodes a 70 amino acid primate-specific L1 peptide (ORF0p) suggested to increase L1 mobility and promote the formation of fusion proteins (Denli et al., 2015).

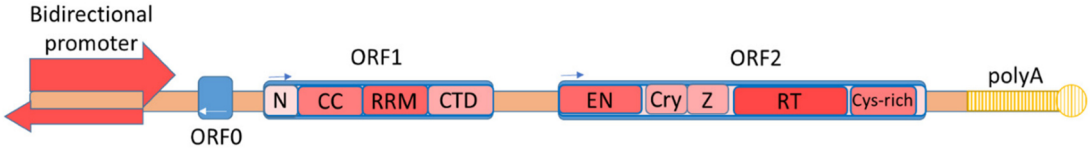


Figure 8: Schematic representation of the canonical L1 retrotransposon. *The full-length L1 retrotransposon is 6 kbp and contains a bi-directional promoter in the 5'UTR, two open reading frames, ORF1 and ORF2, and a 3'UTR with a polyadenylation signal (polyA). The ORF1 encodes a ~40 kDa protein with an N terminal domain (N), a coiled-coiled domain (CC), an RNA recognition motif (RRM), and a C-terminal domain (CTD). ORF2 encodes a ~150 kDa protein with an endonuclease (EN) and reverse transcriptase domain (RT) as well as a cryptic domain (Cry), a Z-domain (Z), and a C-terminal domain with a cysteine-rich sequence (Cys-rich). The antisense ORF0 remains poorly understood but is speculated to increase L1 mobility and promote the formation of fusion proteins. The figure is from Protasova et al., 2021.*

1.4.3 L1 retrotransposition mechanism

The initial step in L1 retrotransposition requires transcription of the retrotransposition-competent genetic element to make an L1 RNA copy; this requires the TATA-less RNA polymerase II promoter located within the ~900 bp 5'UTR (Athanikar et al., 2004). Most of the promoter activity resides within the first 600 bp of the 5'UTR, where the 100 bp 5' terminal region and 390-526 bp regions are important for directing transcription initiation to position +1 of the 5'UTR (Alexandrova et al., 2012; Athanikar et al., 2004). The first ~100 bp contains a conserved 5' terminal Yin Yang 1 (YY1) transcription factor (TF) binding site required for directing transcription initiation to position +1 in the 5'UTR and is, therefore, necessary for the generation of full-length L1 progeny (Athanikar et al., 2004). The 390-526 bp region contains several putative TF binding sites, suggesting the requirement for *cis*-acting TFs in the transcription of L1s (Alexandrova et al., 2012). The full-length L1 RNA copy is subsequently 5' capped and 3' polyadenylated and leaves the nucleus (Athanikar et al., 2004; Belancio et al., 2007)

Translation of the bicistronic mRNA occurs in the cytoplasm leading to the expression of ORF1p and ORF2p. ORF1p proteins form homotrimers through the coiled-coil domain and coat the L1 mRNA at a ratio of one homotrimer per 50 nucleotides of (nt) of RNA (Basame et al., 2006). ORF2p expression is far lower than ORF1p expression, with estimates as low as one ORF2p per 240 ORF1p (Dai et al., 2014; Protasova et al., 2021). In addition to ORF1p

and ORF2p, the formation of the L1 RNP requires several hosts-encoded proteins, including polyadenylate-binding protein 1 (PABPC1) and PABPC4 (Dai et al., 2012; Taylor et al., 2018). The L1 RNP forms discrete foci within the cytoplasm but whether this is important to L1 retrotransposition or an artificial consequence of ORF1p or ORF2p overexpression from the studies is unclear (Doucet et al., 2010).

L1 retrotransposition necessitates L1 RNP transport from the cytosol to the nucleus; however, the mechanism of nuclear import of the L1 RNP remains unclear. ORF1p trimers have been shown to interact with karyopherin Subunit Alpha 2 (KPNA2) and other KPNA family nuclear pore proteins (B. T. Freeman et al., 2019). In addition, the L1 RNP Interact with ESCRT (endosomal sorting complexes required for transport), which, when disrupted, reduces L1 endonuclease activity in the nucleus (Horn et al., 2017). Once in the nucleus, ORF2p makes a single-stranded endonucleolytic cleavage in the gDNA at a degenerate consensus sequence (5′-TTTTT/AA-3′), thus exposing a 3′-hydroxyl (OH) group (Feng et al., 1996; Richardson et al., 2015) (Figure 9). The L1 RNA attaches to the endonuclease recognition site via its polyA tail, and reverse transcription of the L1 RNA by ORF2p occurs by target-primed reverse transcription (TPRT), where the exposed 3′-OH serves as a primer (Cordaux & Batzer, 2009; Cost et al., 2002; Richardson et al., 2015).

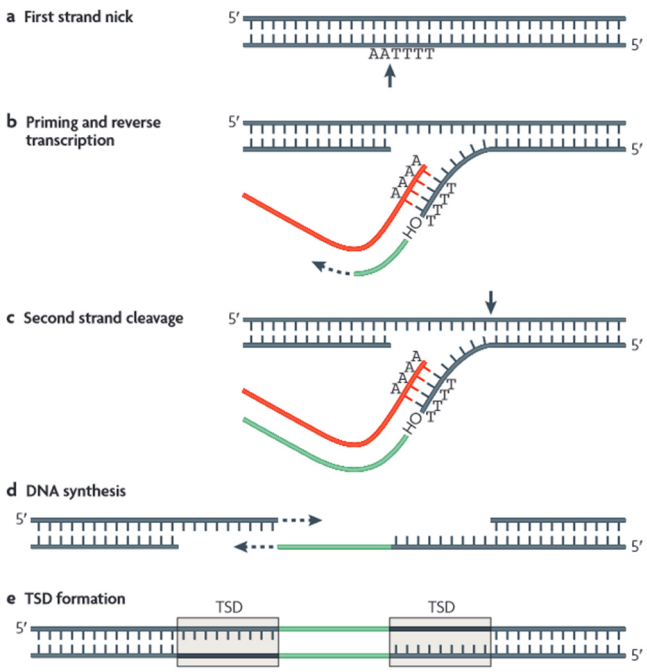


Figure 9: Mechanism of L1 genomic integration by target-primed reverse transcription. *a)L1 endonuclease cleaves the first strand of target DNA at the 5'-TTTTT/AA-3' consensus site. b) The poly-A tail of the retrotransposon anneals to the nicked DNA strand, and the free 3'-OH is used to prime reverse transcription of the L1 RNA. c) The second strand of the target DNA is nicked through a poorly understood mechanism. d) The DNA is repaired by the*

DNA repair machinery. e) Short target site duplications are generated at the L1 insertion site. The figure is from Cordaux et al., 2009.

Several host proteins associated with chromatin regulation, RNA processing, non-homologous end repair (NHEJ), and DNA replication are necessary for L1 retrotransposition (N. Liu et al., 2017). The roles of a few of these have been examined, including poly(ADP-ribose) polymerase 1 (PARP1) and PARP2, which are recruited independently to the ssDNA breaks, after which PARP2 is activated by poly-ADP ribosylation (Miyoshi et al., 2019). Activated PARP2 recruits the replication protein A (RPA) complex, stabilising and protecting the ssDNA (Miyoshi et al., 2019). This RPA interaction with ORF2p is required to facilitate retro-transposition at the TPRT site but, paradoxically, also restricts L1 retrotransposition by recruiting the L1 suppressor, APOBEC3A (Miyoshi et al., 2019).

The reverse transcription and integration process often abort prematurely, resulting in 5' truncated L1 progeny incapable of subsequent retrotransposition. Truncation of the nascent copy is promoted by dsDNA repair factors, including XRCC6, XRCC5, Artemis, and LigIV; however, their exact mechanism of action is unclear (Protasova et al., 2021). Complete integration of the transposable element requires strand ligation, endonucleolytic cleavage of the opposite strand and synthesis of the complementary sequence. This integration process results in a 2-20 bp long target site duplication (TSD) on either side of the retrotransposon and is performed by host proteins involved in DNA replication and repair (Protasova et al., 2021).

1.4.4 L1 propagation under host surveillance

TEs can disrupt coding genes, alter the expression of genes near the integration site, transduce exons, or stimulate recombination, causing a dramatic and rapid rearrangement of the genome that may be greater than the changes caused by point mutations (Platt et al., 2018; X. Zhang et al., 2020). TE insertions are associated with many diseases, including haemophilia, cystic fibrosis, type 2 diabetes mellitus, and several forms of cancer (Hancks & Kazazian, 2016; Platt et al., 2018; X. Zhang et al., 2020). The host cells, therefore, suppress retrotransposition at every stage of the L1 retrotransposition process through various methods. These mechanisms include reducing the availability of L1 DNA; post-transcriptional inhibition by degradation of nascent L1 mRNA; repressing ORF1 and ORF2 translation; sequestering L1 RNPs in the cytosol, and preventing or disrupting integration of the transposon (Protasova et al., 2021). Human cells, therefore, engage L1 retrotransposons in an evolutionary arms race, as L1 retrotransposons hijack cellular processes to promote their spread, and the cell, in turn, attempts to limit retrotransposition. L1 retrotransposons exploit the host cells in numerous ways, and the host cell retaliates in as many (Richardson et al., 2015). However, the mechanisms for many of these processes still lack precise elucidation. Below are a few

examples of how L1 retrotransposition is regulated in human cells where the mode of action is better understood.

1.4.4.1 Transcriptional regulation of L1s

L1 retrotransposons hijack the transcription machinery by binding to several cellular transcription factors. Cells can reduce L1 transcription by limiting L1 DNA accessibility in various ways, including DNA methylations and histone modification (Protasova et al., 2021; Richardson et al., 2015). Several factors regulate chromatin availability in L1 regions, including MEPC2, KRAB-associated protein 1 (KAP1), SIRT7, and the nucleosomal and remodelling deacetylase (NuRD) multiprotein complex. MEPC2 methylates CpG sequences within the L1 5'UTR, thus suppressing L1 transcription (Protasova et al., 2021; Richardson et al., 2015). KAP1 activates KRAB-ZFP leading to H3K9 trimethylation (H3K9me3) by SETDB1, followed by heterochromatin formation. SIRT7 promotes heterochromatin formation through the deacetylation of H3K18 histones, resulting in the association of L1 elements with nuclear lamin A/C and their repression. Finally, the NuRD complex binds to the L1 promoter regions and stimulates heterochromatin formation (Protasova et al., 2021).

1.4.4.2 Translational regulation of L1 protein expression

L1 retrotransposons are regulated by the RNA *N*^{6}-methyladenosine (m^{6}A) modifications (Hwang et al., 2021). Active L1 retrotransposons have a conserved consensus motif recognised by the RNA methyltransferase resulting in m^{6}A modification of A332. m^{6}A332 recruits eIF3, which increases the translation efficiency of ORF1 and thus promotes the formation of L1 RNP. ALKBH5 is a demethylase that can remove m^{6}A modifications and thus antagonises the activity of METTL3 (Hwang et al., 2021). Translation of L1 ORFs can also be suppressed through the actions of the γ-interferon-activated inhibitor of the translation (GAIT) complex. The GAIT complex inhibits L1 mRNA translation through binding, which requires the Condensin II complex, thus repressing the translation initiation factor eIF4G (Protasova et al., 2021; Ward et al., 2017).

1.4.4.3 Regulation of L1 genomic integration

During transposon integration, several cellular processes attempt to truncate and abort retrotransposition in various ways. For example, during transposon integration in the gDNA, transiently exposed cDNA is deaminated by APOBEC3A, resulting in premature termination of retrotransposition (Modenini et al., 2022). APOBEC3A is one of seven enzymes from the APOBEC3 gene family that function as intrinsic host defences that prevent the replication of L1 retroelements and control the spread of retroviruses (Renner et al., 2018; Richardson et al., 2014). The enzyme APOBEC3A is the only member of the family that catalyses the deamination of cytidine nucleotides to uridine nucleotides (C-to-U) in single-strand DNA substrates and is the most effective inhibitor of L1 retrotransposition (Renner et al., 2018;

Richardson et al., 2014). During TPRT, the L1 RNA intermediate protects the cDNA copy. However, during L1 insertion into a new genomic location, the transiently exposed single strand of L1 cDNA becomes susceptible to editing by APOBEC3A. When the cells detect uracil within the DNA, the cellular pathways responsible for DNA degradation and repair are triggered, leading to the termination of retrotransposition (Richardson et al., 2014).

1.5 Aims of this work

RNA helicases are ubiquitous proteins expressed in all three domains of life and some viruses. The term RNA helicases originally referred to proteins that separate RNA duplexes; however, it has come to encompass a group of RNA-binding proteins with shared structural features that display a wide array of RNA remodelling activities. These capabilities make RNA helicases a key driving force in almost all aspects of RNA metabolism. Strict spatial, temporal, and catalytic regulation of RNA helicases is essential in cells, and dysregulation of helicase activity is often associated with tumorigenesis and disease. However, the helicase core of SF2 helicases generally lacks intrinsic specificity, necessitating dedicated strategies to limit the promiscuity of SF2 helicases and to ensure their target specificity.

DHX proteins are a diverse group of SF2 RNA helicases participating in a multitude of cellular processes as RNA/RNP remodellers. Additionally, it has emerged that some DHX proteins participate in multiple cellular processes and therefore require additional layers of regulation. However, the cellular function remains elusive for several DHX proteins, and the mode of regulation remains elusive for most.

The objective of this study was, therefore, to broaden the understanding of the functional and regulatory repertoires of DHX proteins. In this context, the study aimed to:

- Establish a toolbox of resources for exploring DHX protein functions in the cellular context.
- Characterise specific DHX proteins by:
    - Identifying protein interaction partners.
    - Identifying and analysing interactions with substrate RNAs.
    - Demonstrating catalytic activity *in vitro* and exploring regulatory factors.
    - Functionally characterising the proteins in a cellular context, for example, by analysing the effects of depletion on target processes.

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals

General chemicals and laboratory consumables used in this study were purchased from Sigma-Aldrich, Carl Roth, Invitrogen™, Cytiva, or ITW reagents. Specific reagents, kits and lab consumables are specified in Table 2.

Table 2 Reagents and kits

Reagents and kits	Supplier
3xFLAG® peptide	Sigma-Aldrich
4-12% gradient Bis-Tris NuPAGE™ gel	Invitrogen™
Acrylamide 4K solution (30%)	AppliChem
Amersham™ Hybond® P polyvinylidene difluoride (PVDF) 0.45 µm membrane	Cytiva
Amicon® Ultra-4 30K centrifugal filter	Merck Millipore
Anti-FLAG M2 Magnetic Beads	Sigma-Aldrich
ATP-[γ-32P]	PerkinElmer
ATP-lithium salt	Roche
Blasticidin S Hydrochloride	AppliChem
cOmplete His-Tag Purification resin	Roche
cOmplete Mini Protease Inhibitor Cocktail	Roche
cOmplete™ His-tag purification resin	Roche
cOmplete™ Mini Protease Inhibitor Cocktail	Roche
Coomassie Plus (Bradford) assay kit	Thermo Fisher
Dulbecco's Modified Eagle Medium	Gibco™
Fetal Bovine Serum Superior	Sigma-Aldrich
GelPilot® DNA Loading Dye	QIAGEN
GeneRuler 1 kb DNA ladder	Thermo Scientific™
GlycoBlue™	Invitrogen™
Hybond-C™ nitrocellulose membrane	Cytiva
Hygromycin B	AppliChem
Isopropyl-b-D-thiogalactoside (IPTG)	Sigma-Aldrich
LightCycler® 480 SYBR Green I master kit	Roche
Lipofectamine™ 2000 transfection reagent	Invitrogen™
Lipofectamine™ RNAiMAX transfection reagent	Invitrogen™

Luria Broth, according to Miller	Invitrogen™
MetaPhor™ agarose	Lonza
MinElute Gel Extraction kit	Qiagen
Ni-NTA Agarose Resin	Qiagen
NucleoBond® Xtra Midi kit	Macherey-Nagel
NucleoSpin® Gel and PCR Clean-up mini kit	Macherey-Nagel
NucleoSpin® mini kit for plasmid DNA	Macherey-Nagel
NuPAGE™ LDS Sample Buffer	Invitrogen™
Opti-MEM™	Gibco™
PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa	Thermo Scientific™
PageRuler™ Prestained Protein Ladder, 10 to 180 kDa	Thermo Scientific™
PageRuler™ unstained protein ladder	Thermo Scientific™
Penicillin-Streptomycin	Gibco™
Phosphate buffered saline	Gibco™
Phosphoenolpyruvate	Sigma-Aldrich
Pierce™ BCA protein assay kit	Invitrogen™
Pierce™ Coomassie Plus (Bradford) Assay Kit	Thermo Scientific™
Protein G sepharose® 5 Fast Flow	Cytiva
PureLink™ Genomic DNA mini kit	Invitrogen™
Qubit™ dsDNA BR Assay Kit	Invitrogen™
RIPA Buffer (10X)	CST
RNA Clean and Concentrator -5 kit	Zymo Research
RNasin Ribonuclease Inhibitor	Promega
ROX Reference dye	Invitrogen™
SafeView™ Classic	Applied Biological Materials Inc.
Spectra/Por® 4 dialysis tubing, 12-14 kDa MWCO	Spectrum™
TRI Reagent®	Sigma-Aldrich
TrueSeq RA3 3’ end adapter	illumina
TrueSeq RA5 5’ end adapter	illumina
Trypsin-EDTA (0.25%)	Gibco™
X-tremeGene™ 9 DNA transfection reagent	Roche
β-Nicotinamide adenine dinucleotide reduced (NADH)	Sigma-Aldrich

2.1.2 Enzymes

All restriction enzymes and their buffers used for molecular cloning were manufactured by Thermo Scientific™. Specific enzymes for other applications are listed in Table 3.

Table 3 Enzymes

Enzyme	Manufacturer
FastAP Thermosensitive Alkaline Phosphatase	Thermo Scientific™
Phusion® High-fidelity DNA polymerase	Thermo Scientific™
Proteinase K	Roche
Pyruvate kinase/lactic dehydrogenase	Sigma-Aldrich
RNA Ligase 1	NEB
RNace-It ribonuclease cocktail	Agilent Technologies
RNase H	NEB
SuperScript™ III Reverse transcriptase	Invitrogen™
T4 DNA ligase	Thermo Scientific™
T4 Polynucleotide Kinase	Thermo Scientific™
T4 RNA Ligase 1	NEB
T4 RNA Ligase 2 Deletion Mutant	Epicentre
TaKaRa LA Taq DNA polymerase	TaKaRa
TSAP Thermosensitive Alkaline Phosphatase	Promega
TURBO™ DNase	Invitrogen™

2.1.3 Media, buffers and solutions compositions

All media, buffers and solutions used in this study are listed in Table 4 with name, abbreviation (when applicable), and composition.

Table 4 Media, buffers, and solutions

Name	Composition
Coomassie destaining solution	30% (v/v) methanol, 10% (v/v) acetic acid
Coomassie stain	0.1% (w/v) Coomassie R250, 10% (v/v) acetic acid, 40% (v/v) methanol
CRAC Blotting buffer, pH 7.2	25 mM bicine, 25 mM bis-tris, 1.025 mM EDTA, 20% (v/v) methanol
CRAC Wash buffer 1 (CRAC-WB1)	50 mM Tris-HCl pH 7.8, 300 mM NaCl, 10 mM imidazole, 6 M guanidine-HCl, 0.1% (v/v) NP-40, 5 mM 2-mercaptoethanol

CRAC Wash buffer 2 (CRAC-WB2)	50 mM Tris-HCl pH 7.8, 300 mM NaCl, 10 mM imidazole, 0.1% (v/v) NP-40, 5 mM 2-mercaptoethanol
DHX40 elution buffer, pH 8	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM MgCl2, 2 mM 2-mercaptoethanol, 300 mM imidazole, 20% (v/v) glycerol
DHX40 Lysis buffer, pH 8 (D-lysis)	50 mM Tris-HCl pH 7.5, 1000 mM NaCl, 1 mM MgCl2, 2 mM 2-mercaptoethanol, 5 mM imidazole, 0.5% (v/v) Triton X-100, 10% (v/v) glycerol
DHX40 Storage buffer, pH 8	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 2 mM MgCl2, 30% (v/v) glycerol
DHX40 wash buffer 1, pH 8 (D-WB1)	50 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM MgCl2, 2 mM 2-mercaptoethanol, 40 mM imidazole, 10% (v/v) glycerol
DHX40 wash buffer 2, pH 8	50 mM Tris-HCl pH 7.5, 1000 mM NaCl, 1 mM MgCl2, 2 mM 2-mercaptoethanol, 40 mM imidazole, 10% (v/v) glycerol
DNA loading dye (6x)	0.2% (w/v) bromophenol blue, 0.2% (w/v) xylene cyanole, 60% (v/v) glycerol, 60 mM EDTA
Elution buffer	50 mM Tris-HCl pH 7.8, 50 mM NaCl, 150 mM imidazole, 0.1% (v/v) NP-40, 5 mM 2-mercaptoethanol
Immunoprecipitation buffer (IP)	20 mM HEPES-NaOH (pH 7,5), 150 mM NaCl, 0.5 mM EDTA, 10% (v/v) glycerol, 0.5% (v/v) Triton X-100
Laemmli buffer (10x)	250 mM tris, 1.92 M glycine, 0.5% (v/v) SDS
LB7	25 g/L LB powder, 4 g/L D-glucose, 10 mM MgCl2, 200 mM potassium phosphate buffer (pH 7)
MES SDS running buffer, pH 7.3 (20x)	1 M MES, 1M tris , 69.3 mM SDS, 20.5 mM EDTA
Phosphate buffer saline (PBS), pH 7,3	1.37 M NaCl, 27 mM KCl, 80 mM Na2HPO4, 15 mM KH2PO4
PNK buffer	50 mM tris pH 7.8, 10 mM MgCl2, 0.5% (v/v) NP-40, 5 mM 2-merceptoethanol
PPIL4 elution buffer, pH 7.5	50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM MgCl2, 2 mM 2-mercaptoethanol, 250 mM imidazole, 10% (v/v) glycerol
PPIL4 lysis buffer, pH 7.5	50 mM Tris-HCl pH 7.5, 600 mM NaCl, 1 mM MgCl2, 2 mM 2-mercaptoethanol, 5 mM imidazole, 0.5% (v/v) Triton X-100, 10% (v/v) glycerol
PPIL4 storage buffer, pH 7.5	50 mM Tris-HCl pH 7.5, 120 mM NaCl, 2 mM MgCl2, 20% (v/v) glycerol

PPIL4 wash buffer 1, pH 7.5 (P-WB1)	50 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 mM MgCl2, 2 mM 2-mercaptoethanol, 10 mM imidazole, 10% (v/v) glycerol
PPIL4 wash buffer 2, pH 7.5	50 mM Tris-HCl pH 7.5, 1 M NaCl, 1 mM MgCl2, 2 mM 2-mercaptoethanol, 10 mM imidazole, 10% (v/v) glycerol
Resolving gel (10%)	10% acrylamide 4K, 375 mM Tris-HCl pH 8.8, 1% SDS, 1% APS, 0.1% TEMED
Resolving gel (12%)	12% acrylamide 4K, 375 mM Tris-HCl pH 8.8, 1% SDS, 1% APS, 0.1% TEMED
Resolving gel (8%)	8% acrylamide 4K, 375 mM Tris-HCl pH 8.8, 1% SDS, 1% APS, 0.1% TEMED
SDS loading dye (4x)	240 mM Tris-HCl pH 6.8, 40% glycerol, 8% SDS, 0.04% (w/v) bromophenol blue, 5% (v/v) 2-mercaptoethanol
Stacking gel (5%)	5% acrylamide 4K, 125 mM Tris-HCl pH 6.8, 1% SDS, 1% APS, 0.1 % TEMED
TMN150	50 mM Tris-HCl pH 7.8, 150 mM NaCl, 1.5 mM MgCl2, 0.1% (v/v) NP-40, 5 mM 2-merceptoethanol
TMN1000	50 mM Tris-HCl pH 7.8, 1 M NaCl, 1.5 mM MgCl2, 0.1% (v/v) NP-40, 5 mM 2-merceptoethanol
Tris-acetate-EDTA buffer (TAE), pH 8	2 M tris, 0.05 M EDTA, 5.71% (v/v) acetic acid
Tris-buffered Saline (TBS), pH 7,4	50 mM tris pH 7.5, 150 mM NaCl
Western blot transfer buffer (10x)	250 mM tris base, 1.93 M glycine

2.1.4 Oligonucleotides

All oligos used in this study are listed in Table 5 and Table 6 with name, sequence, and in-house identifiers when applicable. All DNA oligonucleotides in Table 5 were synthesised by Sigma Aldrich, and IDT synthesised all oligos in Table 6

Table 5 DNA oligonucleotides

Name	Sequence (5'-3')	ID
DHX16_fwd_AflII_kozak	ATTATTCTTAAGCCACCATGGCGACGCCGGCGGGTCTG	7620
DHX16_qPCR_1_fwd	AGCGCCTACGAGACACTGAT	8147
DHX16_qPCR_1_rev	CGTGGTACCTTGTTCCAGAGT	8148
DHX16_rev_NheI	ATTATTGCTAGCCCCTAGCTCTTCTCGTGTTTTGC	7621

DHX29_Fw_BamHI	ATATATGGATCCATGGGCGGCAAGAACAAGAAACACAAGGC	8381
DHX29_qPCR-2_Fw	ATCTAGGATGGAAGAAAGCACGA	8460
DHX29_qPCR-2_Rv	TCCTGCTGCCTTAACCAACT	8461
DHX29_Rv_NheI	ATATATGCTAGCGTTATTCTCTGTTTTTATCAATTCCGTAATGATCTGCAGAATCTTGTC	8382
DHX29_seq1	GAGGACCCTAAGAGTAAGCC	8291
DHX29_seq2	GCAGTCAGTTCATCAGTTAC	8292
DHX29_seq3	GAGTGCCACTGTGGACAGCGA	8293
DHX29_seq4	GCAGCGCCAGGGAAGAGCTG	8294
DHX29_seq5	GCTGTACTGGTGGCTGGACT	8295
Dhx30_2_Fw	GAACCAGGTGGGATCCTGTG	3962
Dhx30_2_Rv	TCATGGGGATGTTGGAGTGC	3963
DHX30_fw_HindIII	ATATATAAGCTTCAGCCATGTTCAGCCTGGACTCATTCAG	8554
DHX30_rv_Acc65I	ATATATGGGTACCTGAGTCGTCAGCTGTCTTGCGCACATCAAAGCT	8555
DHX30_seq1	CAGATGACGACAGTGCCATT	8296
DHX30_seq2	GCTGGAGCGCTATGTGACCG	8297
DHX30_seq3	GGATCAGAAGGCCATATTCC	8298
DHX30_seq4	GCGTTGGCAGGACCGCAGCT	8299
DHX32_fwd_HindIII_kozak	ATTATTAAGCTTGCCACCATGGAAGAAGAAGGGCTGGAGTG	7622
DHX32_qPCR-1_Fw	GCTGCCTTGACTTGTTGGAAG	8462
DHX32_qPCR-1_Rv	TTTCCACACAGTACTCACTGCTA	8463
DHX32_rev_NheI	ATTATTGCTAGCCTGGAGAGTGCATCTCTGTTCAG	7623
DHX32_seq	TCGGGTGAGAAAGGTGACAT	7643
DHX33_Fw_BamHI	ATATATGGATCCGACATGCCGGAGGAGGCGGGCTTC	8400
DHX33_qPCR-1_Fw	GTGCCAGAGATCCAGAGGTG	8466
DHX33_qPCR-1_Rv	CCGCCTGAATGTGATCTGGA	8467
DHX33_Rv_NheI	ATATATGCTAGCGTTTCTGGCGGTTCTCAGCTTCCTC	8401
DHX33_seq1	GCTACGATGGATGTGGACCT	8300
DHX33_seq2	CCAACTGGACCTGTTAGGTG	8301
DHX36_fw	CACAAAAACCGATGGCCTGG	3918
DHX36_fwd_HindIII_kozak	ATTATTAAGCTTGCCACCATGAGTTATGACTACCATCAGAACTGG	7626

DHX36_rev_NheI	ATTATTGCTAGCGCTGTAATATCCATCCTGGAATCG	7627
DHX36_rv	GGGGAAACCTCTGTGCAGTC	3919
DHX36_seq1	AGTAGCTGCAGAAAGGGCAG	7645
DHX36_seq2	CTCTTCCCAGCCCTCAAAC	7646
DHX38_fwd_HindIII_kozak	ATTATTAAGCTTGCCACCATGGGGGACACCAGTGAGGATG	7628
DHX38_qPCR-1_Fw	GGCAGTGTCAGGTCTACGAA	8474
DHX38_qPCR-1_Rv	GGCTGAGGACCCAATACCTG	8475
DHX38_rev_NheI	ATTATTGCTAGCCAGACCAAAGCGGGCTGG	7629
DHX38_seq1	AGTACTCGGATGACACGCCT	7650
DHX38_seq2	GAAGAGAAAGAGCGAAGCCA	7651
DHX38_seq3	GCTCTCAGGAACAGCGAACT	7652
DHX38_seq4	GTAGGCACAGCAGCAGATCA	7705
DHX40_-699_SalI_rv	ATATATGTGCATTAAAGTCGACCAACTTGGGTAACAAGTC	9616
DHX40_26-_SmaI_fw	ATATATCCCCGGGAAGAGCGGCTCTCGGC	9614
DHX40_E174Q_SDM_fw	CATTATTTTGGATCAAGCCCATGAAAGAACTCTAACTACAGATATC	9427
DHX40_E174Q_SDM_rv	GTTCTTTCATGGGCTTGATCCAAAATAATGACACTGAATTTGGTAAG	9428
DHX40_fwd_Acc65I_kozak	ATTATTGGTACCGCCACCATGTCCCGGTTTCCCGCAG	7630
DHX40_qPCR-2_Fw	GCTCGAAGATCTGTTGGGAGA	8480
DHX40_qPCR-2_Rv	TCTTGCGTAGACTTTGGTGGT	8481
DHX40_rev_NheI	ATTATTGCTAGCGCCTGTTTCCTTTCGTGTGTCAC	7631
DHX40_seq1	AATGGCTGGAGACATCTTGG	7644
DHX40_seq2	TGCTTCCTTCTGGTACTCTGG	7704
DHX40_sgRNA_a_fw	CACCGCCGAGAGCCGCTCTTCCTGG	9003
DHX40_sgRNA_a_rv	AAACCCAGGAAGAGCGGCTCTCGGC	9004
DHX40.sgRNA-a_seq_fw	ATGCAGGGTCCTATGAGGGA	9034
DHX40.sgRNA-a_seq_rv	AAGGCCAAGTGGTGGGTAAC	9035
DHX57_fw	GTGTACCCGCTGGTCTTGTT	3922
DHX57_rv	AGTTCAGCCACCTGATGGGA	3923
DHX58_fw_HindIII	ATATATAAGCTTAGGATGGAGCTTCGGTCCTACCAATGGGAGGTG	8552

DHX58_rv_NheI/Acc 65I	ATATATAGGTACCGCTAGCGTCCAGGGAGAGGTCCGA CAAGTTCTC	8553
DHX58_seq1	CTTGCTGAAGAAGCTCATGG	8307
DHX58_seq2	CGATCAGAGTGTATACGCGTTTG	8308
DHX8_Fw_Acc65I	ATATATGGTACCATGGCTGTGGCTGTAGCCATGGC	8379
DHX8_qPCR-1_Fw	CCAGAATGCCCAAAGCACTTCC	8454
DHX8_qPCR-1_Rv	AAGGTCTCCACAAGCTGAGATG	8455
DHX8_Rv_BamHI	ATATATGGATCCGCGCCGTCGGAAAGCTCGAG	8380
DHX8_seq1	CCAGGAGCAGGAGTCAGAGTC	8282
DHX8_seq2	TTGTCAAGAACCCAGACGG	8283
DHX8_seq3	CTGATGTGCTCTTTGGATTG	8284
DHX8_seq4	CACTCAAGGCCATGGGTATC	8285
DHX9_FW_Acc65I	ATATATGGTACCGAAGAAGACACTTGAATCATGGGTG ACG	8398
DHX9_qPCR1_fw	CTGGCTAAACTCCCCATTGA	3392
DHX9_qPCR1_rv	ATGAAAGGCTCTGGAAAGCA	3393
DHX9_Rv_NheII	ATATATGCTAGCATAGCCGCCACCTCCTCTTCCCTG	8399
DHX9_seq1	CACGAGAACATGGATCAAATAAG	8286
DHX9_seq2	AAGCTGTGGCTACAGCGTTCG	8287
DHX9_seq3	GTCCACAAATATTGCTGAAAC	8288
DHX9_seq4	GACCTGGGAAGCCAAAGTTC	8289
DHX9_seq5	TTGTGCTTGTAGATGACTGG	8290
hsPPIL4_fw_Acc65I	ATATATATGGTACCATGGCGGTTCTACTGGAGACCAC TTTAGGCGAC	9760
hsPPIL4_rv_BamHI	ATATATATTTGGATCCCCCTCTATACTTAGATTTTTCTT TATCTTTGGACTTC	9761
Oligo dT	TTTTTTTTTTTTTTTTTTTTTTTTVN	1624
U6 primer	GAGGGCCTATTTCCCATGATT	8261

Table 6 CRAC oligos *Table of oligos used for the CRAC experiment. All oligos were ordered from IDT. Modification codes: 5' Inverted Dideoxy-T (5InvddT), 5' Adenylation (5rApp), 3' Dideoxy-C (3ddC).*

Name	**Sequence (5'-3')**
RTP_CRAC (DNA)	CCCTGGCACCCGAGAATTCAA
RA3_CRAC (DNA)	/5rApp/TGGAATTCTCGGGCTCCAAGG/3ddC/

RA5_CRAC_N5 (RNA:DNA)	/5InvddT/dAdCdAdGUUCAGAGUUCUACAGUCCGACGAUCNNNNNAGC

2.1.5 Plasmids

Plasmids used in this study are listed in Table 7 with name, in-house ID, consensus coding sequence (CDS), vector backbone, and source.

Table 7 Plasmids

Name	ID	CDS	Vector	Source
PX459-CRISPR-DHX40-sgRNA-a	pMB1794	N/A	PX459	This study
EF06R	Addgene #42940	M80343.1	pCEP4	Addgene
pcDNA5-FRT-TO-DHX16-His_6Prc2xFLAG	pMB1599	CCDS4685.1	pcDNA5-FRT-TO	Bohnsack lab
pcDNA5-FRT-TO-DHX29-His_6Prc2xFLAG	pMB1688	CCDS34158.1	pcDNA5-FRT-TO	Bohnsack lab
pcDNA5-FRT-TO-DHX30-His_6Prc2xFLAG	pMB1691	CCDS2759.1	pcDNA5-FRT-TO	Bohnsack lab
pcDNA5-FRT-TO-DHX32-His_6Prc2xFLAG	pMB1602	CCDS7652.1	pcDNA5-FRT-TO	Bohnsack lab
pcDNA5-FRT-TO-DHX33-His_6Prc2xFLAG	pMB1689	CCDS11072.1	pcDNA5-FRT-TO	Bohnsack lab
pcDNA5-FRT-TO-DHX36-His_6Prc2xFLAG	pMB1601	CCDS3171.1	pcDNA5-FRT-TO	Bohnsack lab
pcDNA5-FRT-TO-DHX38-His_6Prc2xFLAG	pMB1600	CCDS10907.1	pcDNA5-FRT-TO	Bohnsack lab
pcDNA5-FRT-TO-DHX40-E174Q- His_6Prc2xFLAG	pMB1889	N/A	pcDNA5-FRT-TO	Bohnsack lab
pcDNA5-FRT-TO-DHX40-His_6Prc2xFLAG	pMB1598	CCDS11617.1	pcDNA5-FRT-TO	Bohnsack lab
pcDNA5-FRT-TO-DHX58-His_6Prc2xFLAG	pMB1692	CCDS11416.1	pcDNA5-FRT-TO	Bohnsack lab
pcDNA5-FRT-TO-DHX8-His_6Prc2xFLAG	pMB1687	CCDS77040.1	pcDNA5-FRT-TO	Bohnsack lab
pcDNA5-FRT-TO-DHX9-His_6Prc2xFLAG	pMB1690	CCDS41444.1	pcDNA5-FRT-TO	Bohnsack lab
pAcGFP1-N1	PT3716-5	N/A	pAcGFP1-N1	Clontech
A15-ZZ-TEV-PPIL4-His7	pMB1958	CCDS34550.1	A15	Bohnsack lab

A102-MBP-TEV-DHX40_26-699_E174Q-His_{10}	pMB1952	N/A	A102	Bohnsack lab
A102-MBP-TEV-DHX40_26-699-His_{10}	pMB1904	N/A	A102	Bohnsack lab

2.1.6 Cell lines and bacterial strains

Human cell lines, stably transfected HEK293 cell lines and bacterial strains used in this study are listed in Table 8, Table 9, and Table 10, respectively.

Table 8 Human cell lines

Cell line	Specification	Source
HEK293 Flp-In™ T-Rex™	Immortalised human embryonic kidney cells. Contains a single stable integrated FRT site. Used for generation of stably transfected cell lines.	Invitrogen™
HEK293 Flp-In™ T-Rex™ ΔDHX40-1	HEK293 Flp-In™ T-Rex™ with a DHX40 p.Q25Rfs	This study
HEK293 Flp-In™ T-Rex™ ΔDHX40-2	HEK293 Flp-In™ T-Rex™ with a DHX40 p.Q45Pfs	This study
MCF7	Breast cancer cell line. Epithelial cell line from breast tissue. Used for its high expression of L1 retrotransposons.	ECACC

Table 9 Stable expression HEK293 Flp-In™ T-Rex™ cell lines

Background	Transgene	Plasmid	Source
HEK293 Flp-In™ T-Rex™	DHX8-His_6Prc2xFLAG	pMB1687	This study
HEK293 Flp-In™ T-Rex™	DHX9-His_6Prc2xFLAG	pMB1690	This study
HEK293 Flp-In™ T-Rex™	DHX16-His_6Prc2xFLAG	pMB1599	This study
HEK293 Flp-In™ T-Rex™	DHX29-His_6Prc2xFLAG	pMB1688	This study
HEK293 Flp-In™ T-Rex™	DHX30-His_6Prc2xFLAG	pMB1691	This study
HEK293 Flp-In™ T-Rex™	DHX32-His_6Prc2xFLAG	pMB1602	This study
HEK293 Flp-In™ T-Rex™	DHX33-His_6Prc2xFLAG	pMB1689	This study

HEK293 Flp-In™ T-Rex™	DHX36-His_6Prc2xFLAG	pMB1601	This study
HEK293 Flp-In™ T-Rex™	DHX38-His_6Prc2xFLAG	pMB1600	This study
HEK293 Flp-In™ T-Rex™	DHX58-His_6Prc2xFLAG	pMB1692	This study
HEK293 Flp-In™ T-Rex™	DHX40-His_6Prc2xFLAG	pMB1598	This study
HEK293 Flp-In™ T-Rex™	DHX40_E174Q-HisPrc2Flag	pMB1889	This study
HEK293 Flp-In™ T-Rex™ ΔDHX40-1	DHX40-HisPrc2Flag	pMB1598	This study
HEK293 Flp-In™ T-Rex™ ΔDHX40-1	DHX40_E174Q-HisPrc2Flag	pMB1889	This study

Table 10 Bacterial strains

Name	specifications	application	manufacturer
Rosetta™ 2(DE3)	BL21-derivative designed to enhance the expression of eukaryotic proteins. Expresses tRNAs for seven rare codons from a chloramphenicol-resistant plasmid.	PPIL4 recombinant protein expression	Novagen®
SoluBL21™(DE3) RIL	BL21-derivative designed to enhance the expression of insoluble eukaryotic proteins. Expresses tRNAs for seven rare codons on a chloramphenicol-resistant plasmid.	DHX40 recombinant protein expression	AMSBIO
DH5α	Competent cells for general cloning and subcloning.	Propagation of plasmids	Invitrogen™

2.1.7 siRNAs

All siRNAs used in this study are listed in Table 11 with the target gene, sense and anti-sense sequence, the concentration used, and the source of the sequence. siRNAs are ordered from Dharmacon, Qiagen, MWG and Sigma-Aldrich.

Table 11 siRNAs

Name of siRNA	sequence sense /antisense (5'-)	Final conc.	Source of sequence
siDHX8_1	CCCUAAGGUGGAUGAUGAA[dT][dT] / UUCAUCAUCCACCUUAGGG[dA][dG]	50 nM	Qiagen

siDHX8_2	CGAUCAUCAUGUUGGACGA[dT][dT] / UCGUCCAACAUGAUGAUCG[dC][dG]	50 nM	Qiagen
siDHX9_1	GGCUAUAUCCAUCGAAAUU[dT][dT] /	50 nM	(Manojlovic & Stefanovic, 2012)
siDHX9_2	CCAAAGUUCAGCUCAAAGA[dT][dT] /	50 nM	(Manojlovic & Stefanovic, 2012)
siDHX16_1	CUUGGAAGCUUGUGUAGUA[dT][dT] / UACUACACAAGCUUCCAAG[dT][dT]	50 nM	Dharmacon
siDHX16_2	CAUCACUGCUGGUUACUUU[dT][dT] / AAAGUAACCAGCAGUGAUG[dT][dT]	50 nM	Dharmacon
siDHX29_1	UGACCAAUACACUCUUAUA[dT][dT] / UAUAAGAGUGUAUUGGUCA[dT][dG]	50 nM	Qiagen
siDHX29_2	CGUUCUGAUCUACACUUGA[dT][dT] / UCAAGUGUAGAUCAGAACG[dT][dT]	50 nM	Qiagen
siDHX30_1	CGUCAUCCUCCACAGCUAA[dT][dT] / UUAGCUGUGGAGGAUGACG[dT][dT]	50 nM	Qiagen
siDHX30_2	GCGUCACAUAUAGGACCAA[dT][dT] / UUGGUCCUAUAUGUGACGC[dT][dG]	50 nM	Qiagen
siDHX32_1	GGAUCAGGUAACUACUUAA[dT][dT] / UUAAGUAGUUACCUGAUCC[dA][dT]	50 nM	Qiagen
siDHX32_2	GCAAGUAGUGGAUCACCUA[dT][dT] / UAGGUGAUCCACUACUUGC[dT][dG]	50 nM	Qiagen
siDHX33_1	CCUUUGACUUCAUGUCGAA[dT][dT] / UUCGACAUGAAGUCAAAGG[dT][dG]	50 nM	Qiagen
siDHX33_2	GGUUCCUCUUAGCUAUUUA[dT][dT] / UAAAUAGCUAAGAGGAACC[dG][dT]	50 nM	Qiagen
siDHX34_1	GAGCAUCGACUGUACGAAA[dT][dT] / UUUCGUACAGUCGAUGCUC[dC][dT]	50 nM	Qiagen
siDHX34_2	GGAUCCGCUUCGUAGUAGA[dT][dT] / UCUACUACGAAGCGGAUCC[dC][dG]	50 nM	Qiagen
siDHX36_1	GGCUUAUCUAUCACCUAAA[dT][dT] / UUUAGGUGAUAGAUAAGCC[dA][dG]	50 nM	Qiagen
siDHX36_2	CAGUGUUAGUCAUAUCGUA[dT][dT] / UACGAUAUGACUAACACUG[dG][dA]	50 nM	Qiagen
siDHX38_1	GAUCGGGAUUGGUACAUGA[dT][dT] / UCAUGUACCAAUCCCGAUC[dG][dG]	50 nM	Qiagen
siDHX38_2	CGAUUCUGGUUAUUGCAAA[dT][dT] / UUUGCAAUAACCAGAAUCG[dA][dT]	50 nM	Qiagen
siDHX40_1	GGUGUAACUCAACCACGAA[dT][dT] / UUCGUGGUUGAGUUACACC[dA][dA]	50 nM	Qiagen
siDHX40_2	GGCCUUAACUGGUAUCAAA[dT][dT] / UUUGAUACCAGUUAAGGCC[dA][dG]	50 nM	Qiagen

siDHX57_5	AGUUCACCUGAUAACUCAA[dT][dT] / UUGACUUAUCAGGUGAACU[dT][dA]	50 nM	Qiagen
siDHX57_1	GGCGUGCGUGCAAGUUAUA[dT][dT] / UAUAACUUGCACGCACGCC[dT][dT]	50 nM	Qiagen
siDHX58_1	GGGCCGAUCAGAGUGUAUA[dT][dT] / UAUACACUCUGAUCGGCCC[dG][dG]	50 nM	Qiagen
siDHX58_2	CGCACAAGGACACCGUCUA[dT][dT] / UAGACGGUGUCCUUGUGCG[dT][dG]	50 nM	Qiagen
siPPIL4_2	AUACAGAUGUUGUCGACAUUU[dT][dT] / AAAUGUCGACAACAUCUGUAU [dT][dT]	20 nM	Sigma Aldrich
siMETTL3_1	CUGCAAGUAUGUUCACUAUGA[dT][dT] / UCAUAGUGAACAUACUUGCAG[dT][dT]	30 nM	(J. Liu et al., 2014)
siALKBH5_2	CAGUGUAAAUCUUCGCAGU[dT][dT] / ACUGCGAAGAUUUACACUG[dT][dT]	50 nM	Dharmacon
siUSP7_1	CAAAUUAUUCCGCGGCAAA[dT][dT] / UUUGCCGCGGAAUAAUUUG[dG][dG]	3x 10 nM	(Georges et al., 2018)
siNT	CGUACGCGGAAUACUUCGA[dT][dT] / UCGAAGUAUUCCGCGUACG[dT][dT]	10 - 50 nM	(Elbashir et al., 2001)

2.1.8 Antibodies

The primary and secondary antibodies used in this study are listed in Table 11 and respectively. Name, host, manufacturer, manufacturer's ID, and working dilutions for western blotting are provided for all primary antibodies. All secondary antibodies were manufactured by LI-COR.

Table 12 Primary antibodies

Name	Host	manufacturer	ID	Working dilution
ALKBH5 Polyclonal antibody	rabbit	Proteintech®	16837-1-AP	1:2000
DHX16 Polyclonal Antibody	rabbit	Bethyl	A301-537A	1:1000
DHX29 Polyclonal Antibody	rabbit	Bethyl	A300-751A	1:1000
DHX30 Polyclonal Antibody	rabbit	Bethyl	A302-218A	1:2000
DHX32 Polyclonal antibody	rabbit	Abcam	ab235100	1:500
DHX33 Polyclonal antibody	rabbit	Abcam	ab72451	1:1000
DHX34 Polyclonal antibody	rabbit	Abcam	ab94989	1:500
DHX36 Polyclonal antibody	rabbit	Proteintech®	13159-1-AP	1:500
DHX38 Polyclonal antibody	rabbit	Proteintech®	10098-2-AP	1:1000
DHX40 Polyclonal antibody	rabbit	Sigma/Atlas	HPA044350	1:2000
DHX57 Polyclonal antibody	rabbit	Bethyl	A302-300A	1:2000

DHX58/LGP2 Polyclonal antibody	rabbit	Proteintech®	11355-1-AP	1:200
DHX8 Polyclonal Antibody	rabbit	Bethyl	A300-625A	1:2000
DHX9 Polyclonal Antibody	rabbit	Bethyl	A300-854A	1:2000
FLAG® M2 Monoclonal antibody	mouse	Sigma-Aldrich	F3165	1:10,000
HNRNPM Polyclonal antibody	rabbit	Proteintech®	26897-1-AP	1:1000
Line-1-ORF1p monoclonal antibody	mouse	Sigma-Aldrich	MABC1152	1:1000
METTL3 Monoclonal antibody	mouse	Proteintech®	67733-1-Ig	1:5000
PPIL4 Polyclonal antibody	rabbit	Proteintech®	12538-1-AP	1:2000
TRIM27 Polyclonal antibody	rabbit	Proteintech®	12205-1-AP	1:1000
α-Tubulin monoclonal antibody	mouse	Sigma-Aldrich	T6199	1:30,000
USP7 Polyclonal antibody	rabbit	Proteintech®	26948-1-AP	1:1000
XRCC5/Ku80 Polyclonal antibody	rabbit	Proteintech®	16389-1-AP	1:500
XRCC6/KU70 Polyclonal antibody	rabbit	Proteintech®	10723-1-AP	1:2000

Table 13 Secondary antibodies

Name	manufacturer	Working dilution
IRDye® 680RD Donkey anti-Mouse IgG (H + L)	LI-COR	1:10,000 - 1:20,000
IRDye® 680RD Donkey anti-Rabbit IgG (H + L)	LI-COR	1:10,000 - 1:20,000
IRDye® 800CW Donkey anti-Mouse IgG (H + L)	LI-COR	1:10,000 - 1:20,000
IRDye® 800CW Donkey anti-Rabbit IgG (H + L)	LI-COR	1:10,000 - 1:20,000

2.2 Methods

2.2.1 Molecular cloning

Standard molecular cloning procedures were performed according to (Sambrook & Russell, 2001) unless otherwise stated. The frequently used techniques are explained in more detail below. In preparation for each molecular cloning experiment, plasmid maps and DNA primers were designed using DNASTAR Lasergene 10 SeqBuilder.

2.2.1.1 PCR

Molecular cloning was performed using template cDNA from HEK293 (2.2.4.2) or existing plasmids when available (Table 7). The coding sequence of the protein of interest was amplified by polymerase chain reaction (PCR) using Phusion® High-fidelity DNA polymerase with the supplied HF buffer as per the manufacturer's instructions. The standard PCR mix and cycling conditions are given in Table 14 and Table 15, respectively. Three-step PCR was used for primers with an annealing temperature between 50-70°C, and two-step PCR was used for

primers with an annealing temperature >70°C. Annealing temperatures were calculated by NEB Tm Calculator v1.15.0.

Table 14 PCR reaction composition

Component	Conc.
5x Phusion® HF buffer	1X
dNTPs	200 µM
Forward Primer	200 nM
Reverse Primer	200 nM
Template DNA	variable
Phusion® DNA polymerase	0.04 U/µL

Table 15 PCR cycling conditions

PCR cycle conditions (three-step)				**PCR cycle conditions (two-step)**			
Step	Temp.	Time	#Cycles	**Step**	Temp.	Time	#Cycles
Denaturation	98°C	2 min	1x	**Denaturation**	98°C	2 min	1x
Denaturation	98°C	30 s	35x	**Denaturation**	98°C	30 s	35x
Annealing	50-72°C	40 s		**Annealing and elongation**	72°C	20 s/kb	
Elongation	72°C	20 s/kb		**Final elongation**	72°C	10 min	1x
Final elongation	72°C	10 min	1x	**hold**	4°C	-	1x
hold	4°C	-	1x				

The PCR products were checked by agarose gel electrophoresis. 10% of the PCR was supplemented with 6x DNA loading dye (Table 4). Samples were then loaded onto a 1% (w/v) agarose gel in 1x tris-acetate-EDTA buffer (TAE buffer, Table 4) supplemented with SafeView™ Classic and separated at 100 V for 30-60 min in 1x TAE buffer. PCR products were visualised by UV light, and the sizes were estimated by comparison to a GeneRuler 1 kb DNA ladder. PCRs containing the amplicon of the expected size were purified by NucleoSpin® Gel and PCR Clean-up mini kit, according to the manufacturer's recommendations, and eluted in 20 µL water.

2.2.1.2 Restriction enzyme digestion

Plasmids and PCR amplicon were digested with two restriction enzymes to ensure that the insert ligates in the proper orientation (2.2.1.3). The restriction enzyme ratio and recommended

buffer for each digest were determined using the DoubleDigest Calculator (Thermo Scientific™). The DNA was digested for 1-2 hours at 37°C, after which the digested vector backbone was dephosphorylated for 20 min at 37°C using 0.05 U/µL FastAP Thermosensitive Alkaline Phosphatase. Digests were separated by agarose gel electrophoresis, as described above, followed by gel extraction and purification using the NucleoSpin® Gel and PCR Clean-up mini kit according to the manufacturer's recommendations. The quantity and quality of the obtained DNA were estimated using a NanoDrop™ One^{c} microvolume UV-Vis Spectrophotometer (Thermo Scientific™).

2.2.1.3 Ligation

The ligation of DNA fragments and linearised plasmids with compatible sticky ends was performed using T4 DNA ligase with the manufacturer's buffer. The ligation reaction, containing 50 ng linearised plasmid and three to five times molar excess of the digested amplicon, was incubated overnight at 4°C. The reaction containing the ligated plasmid was directly used to transform competent *Escherichia coli* (*E. coli*).

2.2.1.4 Transformation of E. coli

Chemically competent *E. coli* were transformed by heat shock. Chemically competent *E. coli* were incubated with either ligation reaction or plasmid for 20 min on ice, followed by 45 s heat shock at 42°C and 2 min on ice. Transformed *E. coli* were grown in Luria Broth according to Miller (LB) for 1 hour at 37°C at 600 rpm. Bacteria containing the plasmid were selected by overnight growth at 37°C on LB-agar plates containing appropriate antibiotics considering the bacterial strain (Table 10) and the plasmid (Table 7). Depending on the application, different E. coli strains were utilised. For plasmid propagation and cloning, DH5α was used, SoluBL21™(DE3) RIL was used for the purification of recombinant DHX40 proteins (2.2.6.1), and Rosetta™ 2(DE3) was used for the purification of recombinant PPIL4 (2.2.6.1).

2.2.1.5 Extraction of plasmid DNA and sequencing

For plasmid propagation, single colonies grown on LB-agar, according to Miller, were used to inoculate LB supplemented with appropriate selection antibiotics. Inoculated cultures were grown overnight at 37°C at 150 rpm. Plasmids were extracted using NucleoSpin® mini kit for plasmid DNA or NucleoBond® Xtra Midi kit according to the manufacturer's protocol. The plasmid DNA was eluted in water, and concentration and quality were determined using a NanoDrop™ One^{c} microvolume UV-Vis Spectrophotometer. Purified plasmids were sanger sequenced by Eurofins Genomics using primers annealing to the vector backbone (Table 5). For sequencing of plasmid with longer inserts, additional insert-specific sequencing primers were designed (Table 5). Sanger sequencing results were analysed using DNASTAR Lasergene 10 SeqMan.

2.2.1.6 Site-directed mutagenesis

Constructs with mutations encoding a single amino acid substitution were obtained by site-directed mutagenesis using two partially overlapping primers (Zheng et al., 2004). Each primer was designed to contain the desired mutation and flank the mutated region by at least 10 canonical base pairs. PCR was carried out using the components and conditions specified in Table 16 and Table 17. After amplifying the mutated DNA, the parental plasmid was digested using 0.2 U/µL DpnI for 3 hours at 37°C. The reaction mix was used directly to transform competent *E. coli,* as described in 2.2.1.4. For multiple substitutions or longer indels, site-directed mutagenesis was performed, according to (Edelheit et al., 2009). Primers were designed similarly as above but were used individually in two single-primer reactions (Table 18 and Table 19). After the PCR, equal amounts of each reaction were combined and denatured at 95°C, followed by cooling to randomly anneal PCR products and parental plasmid stands (Table 20). Plasmids containing methylated template strands were digested using DpnI for 3 hours at 37°C. DpnI-digested reactions were used to transform *E. coli,* as described in section 2.2.1.4. Plasmids were extracted, and Sanger sequenced as described in 2.2.1.5.

Table 16 Site-directed mutagenesis PCR composition

Component	Conc.
5x Phusion HF buffer	1X
dNTPs	200 µM
Forward Primer	200 nM
Reverse Primer	200 nM
Template DNA	10-50 ng
DMSO	3%
Phusion DNA polymerase	0.04 U/µL

Table 17 SDM PCR conditions

Step	Temperature	Time	#Cycles
Denaturation	98°C	2 min	1x
Denaturation	98°C	30 s	12-16x
Annealing	50-55°C	60 s	
Elongation	68°C	60 s/kb	
Elongation	68°C	5 min	1x
hold	37°C	-	1x

Table 18 SDM single primer PCR composition

Component	Reaction 1	Reaction 2
5x Phusion HF buffer	1x	1x
dNTPs	200 µM	200 µM
Forward Primer	400 nM	-
Reverse Primer	-	400 nM
Template DNA	10 ng/µL	100 ng/µL
DMSO	3%	3%
Phusion DNA polymerase	0.04 U/µL	0.04 U/µL

Table 19 SDM single primer PCR conditions

Step	Temperature	Time	#Cycles
Denaturation	98°C	2 min	1x
Denaturation	98°C	30 s	30x
Annealing	50-55°C	60 s	
Elongation	72°C	30-40 s/kb	
Elongation	72°C	5 min	1x

Table 20 Plasmid annealing

Step	Temperature	Time
1	95°C	5 min
2	90°C	1 min
3	80°C	1 min
4	70°C	30 s
5	60°C	30 s
6	50°C	30 s
7	40°C	30 s
8	37°C	hold

2.2.2 Cell culture

2.2.2.1 Culture conditions and passaging, preparing of stocks

HEK293 Flp-In™ T-Rex™ and MCF7 cells were cultured at 37°C, and 5% CO_2 in a humified atmosphere in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% (v/v) foetal bovine serum (FBS) superior and 100 µg/mL penicillin-streptomycin. Cells were grown

until 80-90% confluency, at which point they were passaged at a ratio of 1:10 for HEK293 and 1:3 for MCF7 cells. Cells were passaged by washing with 1x phosphate buffered saline (PBS) and dissociated from their plate by treatment with 0.25% Trypsin-EDTA. Detached cells were diluted in complete DMEM and seeded onto a fresh plate. For long-term storage in liquid nitrogen, human cell stocks were prepared by pelleting ~300,000 cells at 1000 g for 5 min and resuspending in 1 mL complete DMEM supplement with 10% (v/v) dimethyl sulfoxide (DMSO). Culture stocks were prepared in cryotubes and cooled at -1°C/min in a Mr. Frosty™ freezing container at -80°C. Stocks were moved to liquid nitrogen storage after at least 24 hours at -80°C.

2.2.2.2 Generation of stably transfected HEK293 Flp-In cell lines

Stably transfected cell lines for the inducible expression of proteins with a C-terminal His_6Prc2FLAG were generated by co-transfecting HEK293 Flp-In™ T-Rex™ cells with a pcDNA5/FRT/TO-derived construct containing the gene of interest (pcDNA5:GOI, Table 7) and a vector for expression of the Flp recombinase, pOG44. For each construct, two wells of a 6-well plate were seeded with approximately 100,000 cells and transfected after 24 h. Each well was treated with 100 µL transfection complex containing 91 µL Opti-MEM™, 9 µL X-tremeGene™ 9 DNA transfection reagent, 0.6 µg pcDNA5:GOI, 1.8 µg pOG44. Transfected cells were selected 48 hours after transfection by adding 100 µg/mL Hygromycin B and 10 µg/mL Blasticidin S Hydrochloride to the medium. Selection pressure was maintained for 2-3 weeks before the expression of the proteins of interest was confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and western blotting as described in section 2.2.3.1. To induce the expression of the tagged proteins, the cells were incubated for 24 hours in media containing 1 µg/mL tetracycline unless otherwise stated.

2.2.2.3 Transient transfection of plasmids

For transient expression of proteins from plasmids, HEK293 cells were transfected with plasmid DNA using Lipofectamine™ 2000 transfection reagent. 200,000 HEK293 cells were seeded in a well in a 6-well plate and transfected after 24 h. Each well was transfected with 300 µL transfection complex consisting of 1.5 µg plasmid DNA and 1.5 µL Lipofectamine™ 2000 diluted in Opti-MEM™. The medium was replaced with complete DMDM 24 hours after transfection, and cells were grown for at least 48 hours thereafter.

2.2.2.4 RNA interference-mediated protein depletion

For RNA interference (RNAi)-mediated depletion of proteins, HEK293 or MCF7 cells were transfected with siRNAs (Table 11). Transfection of human cells with siRNAs was performed using Lipofectamine™ RNAiMAX transfection reagent according to the manufacturer's protocol for reverse transfection in a 6-well plate. In general, 10-50 nM siRNA and 5 µL transfection reagent were diluted in 500 µL Opti-MEM and incubated for 15-20 min before

addition to the cells. The medium was replaced after 24 hours to remove excess transfection reagent, and the cells were harvested 48-96 hours after transfection, depending on the siRNA. RNAi-mediated depletion of USP7 was carried out as described previously with a few modifications (Georges et al., 2018). HEK293 cells were reverse transfected with 10 nM siUSP7-1 as described above. 24 hours and 48 hours after the initial transfection, the culture medium was replaced, and the cells were forward transfected with 10 nM siUSP7-1 using 5 µL Lipofectamine™ RNAiMAX according to the manufacturer's protocol for forward transfection. The culture media was replaced 24 hours after the last transfection, and cells were harvested 24 hours thereafter. Knockdown efficiency was determined by qPCR (2.2.4.2) or SDS-PAGE and western blotting (2.2.3.1.2)

2.2.2.5 CRISPR-Cas9-mediated genome editing

The clustered regularly interspaced short palindromic repeats (CRISPR)/cas9 genome editing system was used to generate DHX40 knockout cell lines. The CRISPR/cas9 system utilizes single guide RNAs (sgRNAs) to guide endonucleolytic cleavage within the coding sequence of the gene of interest, followed by imperfect DNA repair by host proteins (Ran et al., 2013). CRISPR/cas9 knockout of DHX40 followed the Zhang Lab protocol with modifications (Cong et al., 2013; Ran et al., 2013).

2.2.2.5.1 Cloning of constructs

Target-specific sgRNAs were designed based on the human GRCh38/hg38 genome using the CRISPR track within the UCSC genome browser (Kent et al., 2002). The CRISPR track shows *Staphylococcus pyogenes* Cas9 (SpCas9) target sites and annotates them with specificity and efficiency predicted by CRISPOR (Concordet & Haeussler, 2018). Only guide sequences with targets within the first two exons of DHX40 with a >70% Cutting frequency determination (CFD) specificity score were considered (Doench et al., 2016). Guides with high efficiency scores and frameshift probabilities were prioritised when guides with similar specificity were compared. Guide sequences were inserted into the sgRNA scaffold described by Zhang lab to add the DNA overhangs necessary for ligation into the PX459 vector (Table 7). PX459 was linearised by BpiI according to the manufacturer's protocol, and constructs for transfection were generated by inserting the sgRNA sequence into the linearised PX459 as described previously (Ran et al., 2013). The plasmid was propagated and extracted as described in sections 2.2.1.4 and 2.2.1.5, respectively. The construct sequence was confirmed by sanger sequencing using a U6 targeting primer (Table 5).

2.2.2.5.2 Transfection of cells

Transfection of HEK293 Flp-In™ T-Rex™ cells with CRISPR constructs was performed as described previously (see section 2.2.2.2) except for excluding pOG44. The culture medium was changed 24 hours after transfection, at which point selection using 1 µg/mL puromycin

was initiated. At 90% confluency, the wells were expanded to a 10 cm tissue culture dish. Clonal isolation was achieved by single-cell dilution. Selected cells were counted using a Neubauer-improved counting chamber and diluted to a final concentration of 5 cells/mL. 200 µL of this cell suspension was added to each well in a 96-well plate to a final average seeding density of 1 cell/well. Cell growth was monitored until the formation of monoclonal colonies within the well. Wells containing more than one colony were discarded. Monoclonal colonies were expanded to 12-well plates at confluency. Knockout of DHX40 was confirmed by SDS-PAGE and western blotting (2.2.3.1). The genomic mutation achieved by gene editing was determined by extraction and analysis of gDNA (2.2.2.5.4).

2.2.2.5.3 Extraction of genomic DNA and analysis

Knockout cell lines generated using the CRISPR/cas9 system were sequenced at the genomic level by gDNA extraction and sequencing of the loci targeted by the sgRNAs. gDNA from monoclonal cell lines was extracted using the PureLink™ Genomic DNA mini kit according to the manufacturer's instructions. A 500-800 bp region containing the predicted mutation was amplified by PCR as described previously (2.2.1.1) using primers specific to the region (Table 5). The amplicon was Sanger sequenced using the same primers as for the PCR. The chromatogram from the Sanger sequencing was visualised with 4Peaks, and the sequence was analysed and aligned with the annotated mRNA sequence by Serial Cloner 2-6-1.

2.2.3 Protein-based methods

2.2.3.1 SDS-PAGE and western blotting

2.2.3.1.1 Total protein sample preparation

Ice-cold radioimmunoprecipitation assay (RIPA) lysis buffer supplemented with cOmplete™ mini protease inhibitor cocktail was added to human cell pellets to a final volume of 100 µL/1x10^6 cells and incubated for 30 min at 4°C with agitation. The RIPA lysate was cleared by centrifugation at 16,000 g for 20 min at 4°C and transferred to a new tube. The protein concentration in the cleared lysate was determined by Pierce™ BCA protein assay kit, and Proteins were precipitated from by addition of 12.5 % trichloroacetic acid (TCA) followed by thorough vortexing, 20 min incubation on ice, and centrifugation for 20 min at 16,000 g at 4°C. Precipitated proteins were washed with 5x volumes of ice-cold acetone to remove residual TCA and dried before adding 2x SDS loading dye (Table 4) buffered with 200 mM Tris-HCl (pH 8.8) to a final concentration of 5 µg/µL total protein. The protein sample was resuspended and denatured by 20 min incubation at 37°C while shaking vigorously, followed by boiling at 95°C for 10 min.

2.2.3.1.2 SDS- polyacrylamide electrophoresis

Protein samples were separated according to their size and charge by SDS-PAGE according to Laemmli (Laemmli, 1970). Proteins were separated in a discontinuous system consisting of a 5% polyacrylamide (PAA) stacking gel and an 8-12 % PAA resolving gel (Table 4). Electrophoresis was performed in 1x Laemmli buffer (Table 4) at 15 mA/gel in the stacking gel and 30 mA/gel in the resolving gel until the desired separation was achieved. To estimate protein size, the samples were separated alongside a PageRuler™ protein ladder (Table 2). The proteins were visualised either by Coomassie staining (2.2.3.1.2), silver staining (2.2.3.1.3), or western blotting (2.2.3.1.4)

2.2.3.1.3 Coomassie staining

Coomassie staining of proteins after SDS-PAGE was done by incubating the gel in Coomassie stain (Table 4) until thoroughly saturated. The stained gel was destained by several washes in a Destaining solution (Table 4) until the proteins could be visualised. Gels were scanned using an Odyssey® CLx imaging system (LI-COR) or a flatbed scanner.

2.2.3.1.4 Silver staining

Silver staining of gels after SDS-PAGE was performed with the SilverQuest™ silver staining kit according to the manufacturer's protocol using the basic staining protocol. Gels were scanned using a flatbed scanner.

2.2.3.1.5 Western blot

Samples separated by SDS-PAGE were transferred onto a 0.45 µm Hybond P® polyvinylidene difluoride (PVDF) membrane using a wet-transfer system. The transfer was carried out for 75 min at 100 V in cold western blot transfer buffer (Table 4). The elevated temperature during transfer was prevented by adding a cold pack to the buffer reservoir or running the transfer at 4°C. After transfer, the membranes were blocked for 1 hour at room temperature in 10% milk in tris-buffered Saline (TBS, Table 4). Blocked membranes were incubated with primary antibodies (Table 12) diluted in 5% milk in TBS supplemented with 0.1% Tween-20 (TBS-T) either overnight at 4°C or for 1-3 hours at room temperature. Next, the membranes were washed three times for 10 min in TBS-T before incubation with IRDye-conjugated secondary antibodies (Table 13) diluted in 5% milk in TBS-T supplemented with 0.01% SDS for 1 hour at room temperature while protected from light. Membranes were washed as above before signal detection in an Odyssey® CLx imaging system (LI-COR). Images were analysed using Image Studio Lite (version 5.2.5, LI-COR)

2.2.3.2 Immunoprecipitation (IP) of protein complexes

2.2.3.2.1 Anti-FLAG IP

Two 15 cm culture dishes of HEK293 Flp-In™ T-Rex™ cells expressing either a His_6Prc2xFLAG-tagged protein or only the His_6Prc2xFLAG peptide were grown to 80-90% confluency at the time of harvest. Expression of His_6Prc2xFLAG-tagged proteins or His_6Prc2xFLAG control was induced by adding 1 µg/mL tetracycline to the growth medium 24 hours before harvest unless otherwise stated. Cells were harvested by dissociation from the plate (2.2.2.1), followed by centrifugation at 1000 g for 5 min. Cell pellets were resuspended in 2 mL cold IP buffer (Table 4) supplemented with cOmplete™ Mini Protease Inhibitor Cocktail and kept on ice. Cells were lysed by sonication (3 cycles of 15 sec with 0.3 sec on/0.7 off, 17% amplitude) using a Branson Digital Sonifier® 450 (Branson Ultrasonics Corporation) with a microtip probe, followed by centrifugation at 20,000 g for 10 min at 4°C to clear the lysate. The cleared lysate was transferred to a new tube, and 0.4-0.8% of the volume was saved as input. The remaining lysate was added to 20 µL of pre-equilibrated anti-FLAG® M2 Magnetic beads and incubated for 2 hours at 4°C while rotating head-over-tail. Where indicated, the lysate was treated with 100 µg/mL RNase A, 10 U/mL RNase T1, or 10 U/mL RNase H during incubation with beads. After incubation, the beads were washed five times with IP buffer to remove lysate and unbound complexes. His_6Prc2xFLAG-tagged proteins were eluted from the beads through competitive elution by adding 200 µg/mL 3xFLAG® peptide in IP buffer. TCA was added to eluate and input samples to a final concentration of 20%, followed by thorough vortexing and 20 min incubation on ice. Precipitated proteins were pelleted by centrifugation at 20,000 g for 20 min at 4°C, after which the protein pellet was washed in cold acetone and air-dried briefly. For samples intended for mass spectrometry (2.2.3.3), the pellets were resuspended in 35 µL 1x NuPAGE™ LDS Sample Buffer supplemented with 50 mM DTT and denatured at 70°C for 20 min. For samples intended for SDS-PAGE and western blotting (2.2.3.1), the pellets were resuspended in 35 µL 2x SDS loading dye (Table 4) buffered with 200 mM Tris-HCl (pH 8.8) and boiled at 95°C for 10 min.

2.2.3.2.2 Anti-PPIL4 IP

50 µL Protein G sepharose® 5 Fast Flow slurry was washed and equilibrated in PBS supplemented with 0.1% Triton™ X-100 (PBS-T). The sepharose was resuspended in 1 mL PBS-T and incubated with either 2 µg anti-PPIL4 or 2 µg Immunoglobulin G (IgG) isotype (Table 12) for three hours at 4°C while rotating head-over-tail. After incubation, the IgG-coupled sepharose was washed and equilibrated in IP buffer (Table 4). Four 15 cm culture dishes of HEK293 or HEK293 ΔDHX40-1 cells were grown to 80-90% confluency at the time of harvest. Cells were harvested as described in 2.2.3.2.1, resuspended in 3 mL IP buffer supplemented with 3x cOmplete™ Mini Protease Inhibitor Cocktail and sonicated as described

in 2.2.3.2.1. After sonication, the lysate was supplemented with 1.5 mM $MgCl_2$ and pre-cleared by adding 50 µL pre-equilibrated Protein G sepharose® 5 Fast Flow beads. Beads were removed by centrifugation at 20,000 g for 10 min at 4°C, and 5% of the cleared lysate was saved as input. The remaining lysate was divided equally between the antibody- and IgG-coupled sepharose and incubated for 2 hours at 4°C while rotating head-over-tail. After binding, the sepharose was washed three times with IP buffer supplement with 1x cOmplete™ Mini Protease Inhibitor Cocktail to remove unbound complexes. The bound complexes were eluted by adding 50 µL 4x SDS-loading dye (Table 4) lacking 2-mercaptoethanol to the sepharose and heating to 55°C for 10 min. The SDS-loading dye was removed from the sepharose, supplemented with 5% (v/v) 2-mercaptoethanol, and the samples were heated to 95°C for 10 min. Input samples were supplemented with 4x SDS-loading dye and treated as eluate samples. Samples were analysed by SDS-PAGE and western blotting (2.2.3.1)

2.2.3.3 Affinity purifications mass spectrometry and SAINT data analysis

Proteins samples obtained as described in 2.2.3.2.1 were sent for affinity purification analysis (AP-MS) at the Proteomics Service Facility (University Medical Center Göttingen). SDS-PAGE separations of proteins, fractionations into 11 equidistant gel slices, and in-gel Trypsin digestion, followed by liquid chromatography with tandem mass spectrometry (LC-MS/MS), protein identification, and quantification, performed at the aforementioned facility. Significance analysis of interactome (SAINT) express was used as a statistical tool to assign confidence scores to protein-protein interactors identified within the proteomics spectral counts dataset (Teo et al., 2014). Input files were prepared using the R package artMS (Jimenez-Morales et al., 2022), and the SAINTexpress analysis was performed using recommended settings.

2.2.4 RNA-based methods

2.2.4.1 Total RNA extraction

Total RNA was extracted from HEK293 cells using TRI Reagent® following the manufacturer's protocol for monolayer cells with a few modifications. HEK293 cells, grown to 80-90% confluency in a 6-well plate, were washed briefly with 1x PBS before adding 500 µL TRI reagent® directly to each well. The lysate was homogenised by several passes through a pipette before it was transferred to a tube and allowed to incubate for 5 min at room temperature. Phase-separation was achieved by adding 400 µL chloroform to the lysate, followed by thorough vortexing, 5 min incubation at room temperature, and finally centrifugation at 12,000 g for 15 min at 4°C. Next, the upper aqueous phase containing RNA was transferred to a new tube, and RNAs were precipitated by adding 500 µL isopropanol, vortexing, 10 min incubation at room temperature, and centrifugation at 20,000 g for 20 min at 4°C. Finally, the extracted RNA pellet was washed with 1 mL cold 75% ethanol, briefly air-

dried and resuspended in 20-30 µL nuclease-free water. RNA concentration and quality were determined using a NanoDrop™ One^{c} microvolume UV-Vis Spectrophotometer.

2.2.4.2. cDNA synthesis and quantitative PCR

cDNA synthesis from total RNA was performed by SuperScript™ III Reverse transcriptase primed by oligo(dT) (Table 5) according to the manufacturer's protocol. Quantitative PCR (qPCR) primers (see Table 5) were designed using Primer-BLAST (Ye et al., 2012). Default Primer-BLAST settings were used, except for PCR product size, which was set to 70-200 bp, and exon junction span selection which was set to "Primer must span an exon-exon junction". Before using primers for quantification, primer efficiency was tested by performing a qPCR standard curve on a dilution series of cDNA. Only primers with an efficiency between 80-110% were deemed acceptable. Melting curve analysis was performed to determine the specificity of the used primers, and only primers with a single product were used. qPCR was performed using the LightCycler® 480 SYBR Green I master kit according to the manufacturer's recommendations. The final reaction consisted of 1x SYBR Green mix, 300 µM of each forward and reverse primer and 3 µL diluted cDNA (variable). Each sample was pipetted in technical triplicates within the same 96-well qPCR plate. The qPCRs were amplified using either a LightCycler® 480 system (Roche) or an MX3000p® system (Agilent Stratagene) with the cycling conditions listed in Table 21. When using the MX3000p® system, 0.1x ROX Reference dye was added to each well.

Table 21 qPCR program

Step	Temp.	Time.	Signal detection	#Cycles
Denaturation	95°C	5 min	-	1x
Denaturation	95°C	10 sec	-	50-70x
Annealing	58°C	20 sec	End cycle	
Elongation	72°C	15 sec	-	
Melt curve analysis	95°C	10 sec	-	1x
	55°C	1 min	-	1x
	55°C-97°X	-	throughout	1x

The quantification cycle value (Ct), i.e. the cycle at which the amplification was detectable above the background, was determined by the system-related software using the second derivative maximum method. For knockdown experiments, the relative mRNA level for the protein of interest was calculated using the ΔΔCt method (Equation 1). The ΔΔCt method calculates the expression level of an mRNA of interest in a treated sample relative to an untreated sample after normalization to a reference gene.

Equation 1

$$2^{-\Delta\Delta ct} = 2^{-(\Delta Ct_{treated} - \Delta Ct_{untreated})}$$

$$\Delta Ct = Ct_{target} - Ct_{reference}$$

Ct_{gene}: Quantification cycle value for which the amplification was detectable above the background. Either for the target gene or reference housekeeping gene.

ΔCt_{sample}: Difference between target and reference gene expression within the same sample. Used to normalize samples

$2^{-\Delta\Delta ct}$: Relative expression of the gene of interest in the treated sample. Relative to untreated sample.

2.2.5 Next-generation sequencing-based methods and data analysis

2.2.5.1 Crosslinking and analysis of cDNA (CRAC)

Specify protein-RNA interactions were analysed by UV-crosslinking and analysis of cDNA as described previously with some modifications (Haag et al., 2017). Three 15 cm culture dishes of stably transcfected HEK293 cells with inducible expression of either DHX40-His_6Prc2xFLAG or the His_6Prc2xFLAG peptide were grown to 80-90% confluency at time crosslinking. 24 hours before crosslinking, expression of the His_6Prc2xFLAG-tagged protein was induced by adding 1 µg/mL tetracycline to the growth medium. The medium was removed, and cells were washed with cold PBS before crosslinking three times at 254 nm with 800 mJ/cm^2 in a Stratalinker 2400 UV crosslinker (Stratagene). Crosslinked cells were collected and resuspended in 1 mL TMN150 buffer (Table 4) supplemented with 3x cOmplete® Mini Protease Inhibitor Cocktail. Cells were lysed by sonication (4 cycles of 15 sec with 0.5 sec on/0.5 off, 40% amplitude) using a Sonopuls™ HD 2070 Homogenizer (Bandelin Electronic™) with a microtip probe, followed by centrifugation at 20,000 g for 15 min at 4°C to clear the lysate. His_6Prc2xFLAG tagged complexes were enriched by anti-FLAG affinity purification by incubation with 50 µL pre-equilibrated anti-FLAG® M2 Magnetic Beads for 2-4 hours at 4°C rotating head-over-tail. Unbound lysate was removed by washes with TMN1000 (Table 4), followed by equilibration of the magnetic bead complexes with TMN150. FLAG-tagged proteins were eluted from the beads through competitive elution by the addition of 150 µg/mL 3xFLAG® peptide in TMN150 and incubation overnight at 4°C. The RNAs in the eluate were partially digested by treatment with ~0.16 U/mL RNace-It ribonuclease cocktail for 30 sec at 37°C after which the reaction was stopped by supplementing with 6 M guanidine hydrochloride, 300 mM NaCl and 10 mM imidazole. Protein RNA complexes were immobilized by incubation with 50 µL Ni-NTA agarose pre-equilibrated with CRAC wash buffer 1 (CRAC-WB1, Table 4) for 2 hours at 4°C, after which

unbound complexes were removed by several washing steps with CRAC-WB1 and PNK buffer (Table 4). Enriched RNAs were dephosphorylated, ligated to a 3' end adapter, 5' phosphorylated with ^{32}P, and finally ligated to a 5' end adapter as follows. Dephosphorylation was performed using 10 U/µL TSAP Thermosensitive Alkaline Phosphatase in PNK buffer for 30 min at 37°C, after which the resin was washed with CRAC-WB1 and PNK buffer. Ligation of TruSeq RA3 3' adapter (Table 6) was performed in PNK buffer containing 1 µM RA3 3' adapter, 10 U/µL T4 RNA ligase 2 deletion mutant, 1 U/µL RNasin and 10% PEG8000 overnight at 16°C. The resin was washed as described above, and phosphorylation with 1 U/µL T4 Polynucleotide Kinase and 0.5 µCi/µL ATP-[γ-^{32}P] was performed at 37°C for 40 min, followed by addition of 1.25 mM ATP-lithium salt and an additional 20 min incubation at 37°C. The resin was washed as described above. Ligation of the TrueSeq RA5 (N5) 5′ adapter (Table 6) was performed overnight at 16°C in 1x PNK buffer supplemented with 1.25 µM RA5 5' adapter, 0.5 U/µL T4 RNA Ligase 1 and 1 mM ATP. Finally, the resin was thoroughly washed in CRAC wash buffer 2 (CRAC-WB2, Table 4), and the crosslinked protein-RNA complexes were eluted in elution buffer (Table 4). The eluate was supplemented with 5 µg/mL BSA and 75 µg/mL GlycoBlue™, and protein-RNA complexes were precipitated by adding TCA to a final concentration of 20%, vortexing thoroughly and incubating on ice for 30 min. Precipitated protein-RNA complexes were centrifuged at 20,000 g for 30 min at 4°C. The pellets were washed in cold acetone, air-dried briefly and resuspended in 30 µL 1x NuPAGE™ LDS Sample Buffer supplemented with 50 mM DTT and denatured at 70°C for 10 min. Samples were separated on a 4-12% gradient Bis-tris NuPAGE™ gel in MES SDS running buffer (Table 4) at 100 V followed by transfer onto a Hybond-C™ membrane using a wet transfer system at 80 V for 2 hours at 4°C in Blotting buffer (Table 4). The membrane was exposed to an X-ray film to detect radioactive signals by autoradiography. Regions of membranes containing crosslinked RNA-protein complexes, as well as the corresponding regions in the control lane, were identified and excised. Excised pieces of the membrane were incubated overnight at 55°C with CRAC-WB1 supplemented with 1% SDS, 5 mM EDTA and 250 µg/ml Proteinase K to elute RNA. The RNA was recovered by adding an equal volume of phenol:chloroform:isoamyl alcohol (PCI, 25:24:1) in the presence of 0.3 M sodium acetate pH 5.2. Phase separation was achieved by centrifugation for 20 min at 20,000 g and 4°C, after which the upper aqueous phase was transferred to a new tube. RNA was precipitated by adding two volumes of 100% ethanol and 15 µg/mL GlycoBlue™ and overnight incubation at -20°C. The RNA was pelleted by centrifugation at 20,000 g for 20 min at 4°C, after which the RNA pellet was washed with 70% ethanol and air-dried. The pellet was resuspended in Superscript® III reverse transcription reaction mixture (Table 22), and reverse transcription was performed according to the manufacturer's protocol using a primer specific for the 3' end adapter. The resulting cDNA was amplified by PCR using the TaKaRa LA Taq DNA

polymerase using the conditions shown in Table 23 and Table 24. A sufficient amount of amplicon was achieved by performing the PCR in triplicates which were then pooled. The cDNA library was precipitated using PCI as described above, after which the pellet was resuspended in 1x gel loading dye (Qiagen), and the PCR products were separated on a 3% MetaPhor™ agarose gel in 1x tris-borate-EDTA buffer (TBE, Table 4) at 100 mA/gel in 1x TBE. Gel segments containing appropriate PCR amplification products were excised, and the amplicons were purified using the MinElute Gel Extraction kit according to the manufacturer's protocol. The concentration of the final cDNA library was determined using the Qubit dsDNA BR Assay Kit and a Qubit Fluorometer (Thermo Fischer Scientific) according to the manufacturer's recommendations. The cDNA library was sent for next-generation-sequencing (NGS) at NGS Integrative Genomics Core Unit (University Medical Center Göttingen). Single read (50 bp) sequencing was conducted using a HiSeq 2500 system (illumina). Post-processing and bioinformatics analysis was performed as follows. Barcodes were removed using pyCRAC/PyBarcodeFilter.py, after which the adaptors were removed using Flexbar 3.5.0, and PCR duplicates were collapsed by pyCRAC/pyFactQDuplicateRemover.py (Dodt et al., 2012; Webb et al., 2014). The reads were aligned to the GRCh38 (release 28) genome using STAR 2.7.0d setting multimapper = 100 (Dobin et al., 2013). Peak calling and selection of peaks with a fold change ≥ 2 were performed using Peakachu and R/DESeq2, respectively (Holmqvist et al., 2018; Love et al., 2014). The original BAM was subsetted by fold change ≥ 2 using SAMtools (Danecek et al., 2021). Feature counts was performed using FeatureCounts v2.0.0 to annotate biotype (GENCODE GTF), intron/exon (in-house GFT), Metagene (GENCODE GFT), and transposable elements (TE GFT) (Frankish et al., 2021; Jin et al., 2015; Liao et al., 2014). CRAC experiments and bioinformatics analyses were performed in collaboration with Philipp Hackert and Nicolás Lemus, MD PhD, respectively.

Table 22 Reverse transcription reaction mixture

component	concentration
First-strand buffer	1x
Reverse transcription primer (RTP_CRAC)	0.5 µM
dNTP mix (Roche)	0.5 mM
DTT	5 mM
Rnasin	2 U/µL
Superscript™ II reverse transcriptase	10 U/µL

Table 23 PCR composition (CRAC)

PCR components	
TaKaRa LA Taq (5 units/µl)	0.5 µL
10x LA PCR Buffer ll (Mg2+ plus)	5 µL
dNTP Mixture (2.5 mM each)	2.5 µL
Forward primer (10 µM)	1 µL
Reverse primer (10 µM)	1 µL
cDNA template	2 µL
H2O	up to 50 µL

Table 24 PCR conditions (CRAC)

PCR conditions			
Step	Temp.	Time	#Cycles
initial denaturation	98°C	2 min	x1
denaturation	98°C	30 s	x30
annealing	60°C	40 s	
elongation	68°C	40 s	
final elongation	72°C	5 min	x1

2.2.5.2 RNA-seq

Total RNA, extracted from human cells as described in 2.2.4.1, was further cleaned and concentrated to archive a quality suitable for NGS. RNA was cleaned using the RNA Clean and Concentrator -5 kit following the manufacturer's instructions for DNase digestion and total RNA clean-up. The RNA was eluted in 12 µL of nuclease-free water, and the concentration and quality were determined using a NanoDrop™ one^{c} UV-Vis Spectrophotometer. The RNA samples were then sent to the NGS Integrative Genomics Core Unit (University Medical Center Göttingen) for library preparation and NGS. Library preparation and rRNA depletion of 2 µg of the RNA sample was done according to the TrueSeq Stranded Total RNA kit. Single read (50 bp) sequencing was conducted using a HiSeq 4000 (illumina). Quality control of the raw sequencing data was done using FastQC v0.11.9 (Andrews, 2010). Read alignment was performed STAR 2.7.0d (Dobin et al., 2013). First, the reads were aligned to rRNA (U13369.1), after which non-aligned reads were aligned to the GRCh38 (release 28) genome. Annotation of transposable elements and differential expression (DE) analysis was done by TEtranscripts 2.2.1, and analysis after DE analysis was done in R (Jin et al., 2015; Team, 2021). Bioinformatics analysis was performed in collaboration with Nicolás Lemus, MD PhD.

2.2.6 *In vitro* methods

2.2.6.1 Recombinant protein expression and purification

Plasmids for recombinant expression of DHX40 wild type and mutants (Table 7) were generated using the protein-coding sequence from HEK293 cDNA and in-house available protein expression vectors. Disordered regions, as predicted by MetaDisorder and Swiss-Model, were omitted (Kozlowski & Bujnicki, 2012; Waterhouse et al., 2018). An N-terminal maltose binding protein (MBP) tag was introduced to increase solubility, and a C-terminal tag consisting of ten histidines (His10) was introduced to enable tag-affinity purification. For expression and purification of recombinant DHX40 protein from *E. coli*, SoluBL21™(DE3) RIL *E. coli* were transformed with the recombinant protein expression plasmid as described in 2.2.4.1. For selection, the transformed E. coli were plated on LB agar plates with 34 µg/mL chloramphenicol and 100 µg/mL ampicillin. The growth medium for recombinant protein expression (LB7) consisted of LB buffered with 200 mM potassium phosphate buffer (pH 7) and supplemented with 10 mM $MgCl_2$ and 4 g/L D-glucose (Azatian et al., 2019; Christensen et al., 2017). Starter cultures, inoculated with *E. coli* colonies, were grown overnight at 37°C at 150 rpm in LB7 supplemented with 100 µg/mL ampicillin and 34 µg/mL chloramphenicol. Expression cultures, consisting of LB7 supplemented with 100 µg/mL ampicillin and 17 µg/mL chloramphenicol, were inoculated 1:50 with a starter culture. Expression cultures were grown in baffled 5 L flasks at 37°C and 80 rpm until the early stationary phase (Ou et al., 2004). At the stationary phase, the growth temperature was reduced to 16°C; the cultures were allowed to adapt to the lower temperature for 4-5 h, and recombinant protein expressions were induced by the addition of 200 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) to the expression medium (Kandror et al., 2002; Prasad et al., 2011). The cultures were harvested after 20-24 hours by centrifugation at 4200 g for 20 min, after which the pellets were washed in cold PBS supplemented with 0.1 mM phenylmethylsulfonyl fluoride (PMSF) and snap-frozen in liquid nitrogen. All protein purification steps were done at 4°C with buffers cooled to ≤ 4°C. Cells were suspended in 10 mL DHX40 lysis buffer (D-lysis, Table 4) per gram of cell pellet. The cells were then lysed in an EmulsiFlex-C3 (Avestin) by four passes at 15,000 psi. The lysate was cleared by centrifugation at 30,000 g for 20 min at 4°C. Thereafter, the cleared lysate was incubated with 20 µL cOmplete® His-tag purification resin per gram of pellet and incubated overnight while rotating head-over-tail. After binding, the resin was loaded onto a gravity flow column and washed with a series of buffers to remove contaminants. At each washing step, the column bed was disturbed, and the wash was allowed to sit for 5 min before removal by gravity flow. The series of washes were as follows, 100 column volumes (CV) D-lysis, 100 CV DHX40 wash buffer 1 (D-WB1), 100 CV DHX40 wash buffer 2, 100 CV D-WB1, and finally, 100 CV D-WB1. The recombinant protein was eluted with DHX40 elution buffer (Table 4) in 4 CV fractions until no more protein was eluted. The protein content of each fraction was

estimated by dot blot, and fractions containing the highest protein concentration were pooled together. The pooled protein was dialysed overnight in Spectra/Por® 4 dialysis tubing with 12-14 kD molecular weight cut-off (MWCO) against storage buffer (Table 4) and concentrated in an Amicon® Ultra-4 30K centrifugal filter. The protein concentration was estimated using the Pierce™ Coomassie Plus (Bradford) Assay Kit, and the protein purity was visualised by SDS-PAGE followed by Coomassie staining as described in 2.2.3.1.2. and 2.2.3.1.2.

Plasmid for recombinant expression of PPIL4 wild (Table 7) was generated using the protein-coding sequence from HEK293 cDNA and an in-house available protein expression vector. The expression plasmid included an N-terminal ZZ-tag and a C-terminal tag consisting of seven histidines (His7). For expression and purification of recombinant PPIL4 protein, Rosetta™ 2 (DE3) *E. coli* was transformed with recombinant protein expression plasmid as described in 2.2.4.1. For selection, the transformed E. coli were plated on LB agar plates with 34 µg/mL chloramphenicol and 100 µg/mL ampicillin. Starter cultures, inoculated with *E. coli* colonies, were grown overnight at 37°C at 150 rpm in LB supplemented with 100 µg/mL ampicillin. The starter culture was used for 1:100 inoculation of expression cultures consisting of LB supplemented with 100 µg/mL ampicillin. Expression cultures were grown in baffled 5 L flasks at 37°C and 80 rpm until the mid-logarithmic phase when the cultures were cooled down to 18°C. The cooled cultures were supplemented with 250 mM IPTG to induce recombinant protein and then grown overnight at 18°C. Cells were harvested by centrifugation at 4200 g for 20 min, after which the pellets were washed in cold PBS. All protein purification steps were done at 4°C with buffers cooled to ≤ 4°C. Cells were suspended in 20 mL PPIL4 lysis buffer (Table 4) per litre of expression culture and lysed in an EmulsiFlex-C3 by four passes at 15,000 psi. The lysate was cleared by centrifugation at 30,000 g for 20 min at 4°C. The cleared lysate was incubated with 500 µL cOmplete® His-tag purification resin per litre of expression culture and incubated for 4 hours while rotating head-over-tail. After binding, the resin was loaded onto a gravity flow column and washed with a series of buffers to remove contaminants. The series of washes were as follows: 20 mL PPIL4 wash buffer 1 (P-WB1), 20 mL PPIL4 wash buffer 2, and 20 mL P-WB1. The recombinant protein was eluted with PPIL4 elution buffer (Table 4) in 500 µL fractions until no more protein was eluted. The protein content of each fraction was estimated by dot blot, and fractions containing the highest protein concentration were pooled together. The pooled protein was dialysed overnight in Spectra/Por® 4 dialysis tubing with 12-14 kD MWCO against storage buffer (Table 4). The protein concentration was estimated using Pierce™ Coomassie Plus (Bradford) Assay Kit, and the protein purity was visualised by SDS-PAGE followed by Coomassie staining as described in 2.2.3.1.2. and 2.2.3.1.2. Purification of PPIL4 recombinant protein was performed with help from Philipp Hackert.

2.2.6.2 NADH-coupled ATPase assay

The ATPase activity of recombinant proteins was determined by an nicotinamide adenine dinucleotide (NADH)-coupled ATPase assay performed as described previously with modifications (Scharschmidt et al., 1979; L. Wang et al., 2020). In short, reactions containing 50 mM Tris-HCl pH 7.4, 25 mM NaCl, 2 mM MgCl2, 360 µM NADH, 4 mM ATP, 1.5 mM phosphoenolpyruvate (PEP), 20 U/mL pyruvate kinase/ lactic dehydrogenase (PK/LDH) was set up and supplemented with 1.5 µM DHX40, 1.5 µM PPIL4, or 1.5 µM 25S H23/24 RNA (5'-GAAAGAUGAAAAGAACUUAAACAAAACAAAAC-3') as indicated. The assay was carried out at 30°C and NADH oxidation to NAD+, a measure of ATPase activity, was measured by a decrease in fluorescence at $\lambda_{ex} = 355$ nm and $\lambda_{em} = 480$ nm. Fluorescence was measured at 50-sec intervals for 90-120 min using an Appliscan Multimode Plate Reader (Thermo Scientific™) and the compatible software SkanIt® (version 2.3, Thermo Scientific™). An NADH standard curve ranging from 300 µM to 18.75 µM NADH was prepared in the same reaction buffer and used to convert fluorescence to NADH concentration. The slope in fluorescence decay was corrected for background NADH decay by subtracting the rate of decay in a protein-free reaction.

2.2.7 Retrotransposition assay

HEK293 cells were transfected either with an L1 plasmid or a plasmid encoding green fluorescent protein (GFP) (Table 7). The L1 plasmid, EF06R, functioned as a reporter for L1 retrotransposition, as insertion into the host genome would induce expression of GFP (Farkash et al., 2006). The GFP plasmid, pAcGFP1-N1, was used to measure transfection efficiency. HEK293 cells were transfected with the EF06R or pAcGFP1-N1 plasmid, as described in 2.2.2.3. After 24 h, the cells were trypsinised, seeded in 6-well plates at 400,000 cells per well and transfected with siRNAs as described in 2.2.2.4. After 48 h, cells were trypsinised, split equally into two wells and grown for an additional 24 hours, at which point cells transfected with pAcGFP1-N1 were harvested and measured by Fluorescence-activated Cell Sorting (FACS) as described in 2.2.7.1.2 and cells transfected with EF06R were treated with 1 µg/mL puromycin to start selection. EF06R-transfected cells were under selection for 72 hours before they were harvested and measured by FACS as described in 2.2.7.1.2.

2.2.7.1 FACS sample preparation and analysis

Cells were trypsinised as described in 2.2.2.1 and centrifuged at 500 g for 3 min. Cells were washed three times in cold PBS, resuspended in 300 µL cold PBS, and kept on ice. Samples were measured immediately after preparation. Flow cytometry was performed using a BD FACSCanto II system (BD Biosciences). The GFP was excited by a 488 nm laser and detected using the 502LP-530/30 filter. Cell counting was done using the BD FACSdiva software (v6.1.1, BD biosciences). Within the BD FACSdiva software, the main population and GFP

signal threshold were selected using non-treated HEK293 cells as a reference. The number of GFP-positive cells within a sample only transfected with EF06R was used for scaling between different experiments to adjust for differences in transfection efficiency. Scaling and statistical analysis were performed in excel and PRISM 8, respectively.

3. Results

3.1 Establishment of tools for analysing DHX proteins in cells

DHX proteins are a diverse group of RNA helicases participating in a multitude of cellular processes through their functions in RNA/RNP remodelling and mediating protein-protein interactions. However, the cellular function remains elusive for a few, and the mode of regulation remains elusive for most. Additionally, it is evident that some DHX proteins participate in multiple cellular processes, and it is important to recognise this propensity for multifunctionality when studying RNA helicases. Therefore, to obtain a better understanding of DHX proteins and their function in vivo, a toolbox including stably transfected cell lines, verified DHX protein knockdowns and tested antibodies was established.

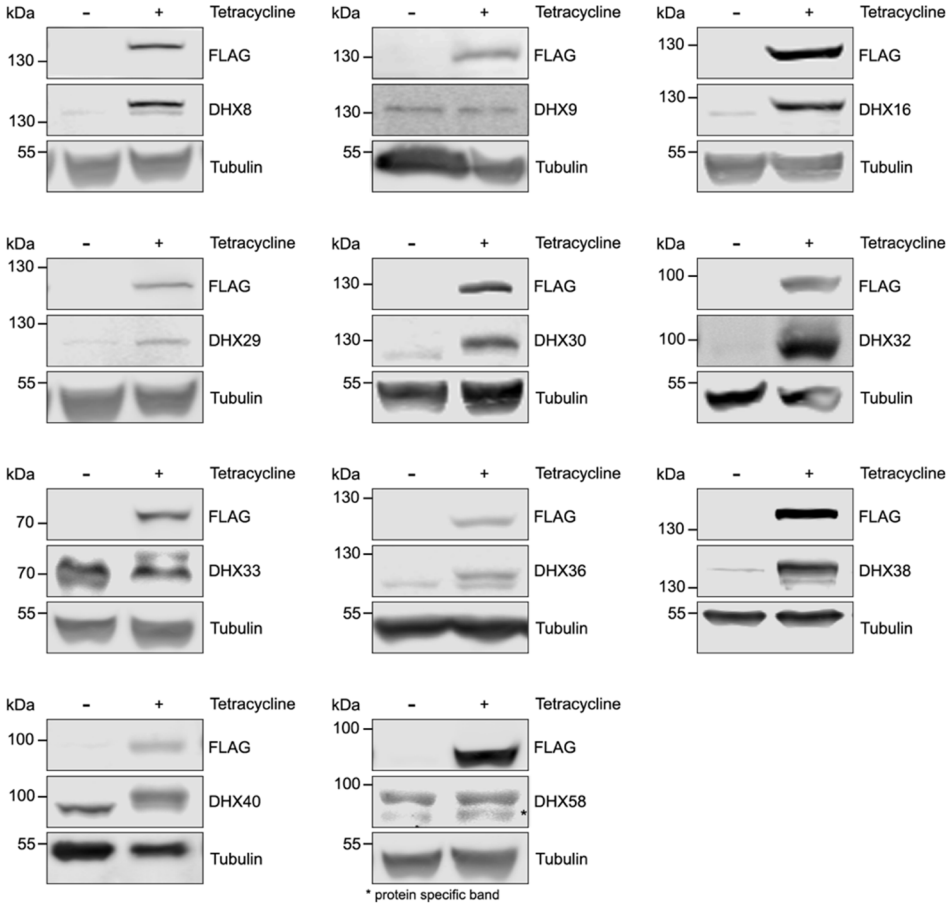


Figure 10 Expression of C-terminally His_6Prc2xFLAG-tagged DHX proteins from HEK293 Flp-In™ T-Rex™ stably transfected cell lines. *HEK293 Flp-In™ T-Rex™ were transfected with plasmids encoding C-terminally His_6Prc2xFLAG-tagged DHX proteins to generate inducible stably transfected cell lines. Expression of the transgene was induced with 1 µg/mL tetracycline for 24 hours, after which cell extracts were analysed by western blotting. Expressed proteins were detected with an anti-FLAG antibody and antibodies against the endogenous proteins. Anti-tubulin was used as a loading control.*

First, the coding sequences of 11 DHX proteins were cloned into a pcDNA5-derived mammalian expression vector (Table 7), allowing inducible expression of a DHX protein with a C-terminal His_6Prc2xFLAG tag. The expression vectors were used to generate HEK293 Flp-In™ T-Rex™ (referred to as HEK293) stably transfected cell lines (Table 8 and Table 9). Expression of C-terminally tagged DHX proteins was induced with tetracycline, and expression was confirmed by western blotting with an anti-FLAG antibody and antibodies against the endogenous proteins (Table 12; Figure 10)

Next, the RNAi-mediated knockdown of 10 DHX proteins was established in HEK293 cells using two siRNAs for each DHX protein. DHX8 was excluded as knockdowns could not be achieved with available siRNAs. Knockdowns were performed by reverse transfecting HEK293 cells for 72 hours with a non-target siRNA (siNT) or siRNAs targeting the indicated DHX proteins. The knockdown efficiency of each siRNA was estimated at mRNA and protein levels (Figure 11 and Figure 12). The level of mRNA expression in knockdown samples relative to non-target control samples was determined by reverse transcription and qPCR (qPCR) and normalised to the mRNA levels of the housekeeping genes GAPDH and EMC7 (Figure 11). The RNAi-mediated knockdowns reduced the expression of the target mRNA by >80% of all DHX proteins except DHX33 and DHX38.

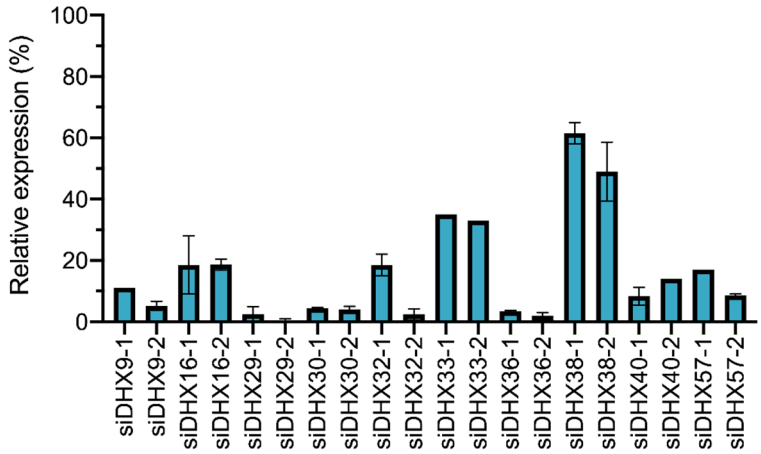


Figure 11 Relative expression of DHX mRNAs after RNAi-mediated knockdown. *HEK293 cells were reverse-transfected for 72 hours with siNT or siRNAs targeting indicated DHX proteins. Levels of mRNA expression in knockdown samples relative to expression in siNT-treated cells were estimated by reverse transcription and qPCR. In addition, expression levels were normalised to the expression of the housekeeping genes, GAPDH and EMC7. Results are the mean of biological duplicates or triplicates; error bars show the standard error of the mean (SEM). Data is available in Supplementary table 1*

To estimate DHX protein levels after RNAi-mediated knockdowns, cell extracts were analysed by western blotting using antibodies targeting the endogenous protein (Figure 12). Western blots confirmed that the DHX proteins tested were depleted following transfection with

indicated siRNAs. However, for the DHX38 knockdowns, the western blot analysis indicated an almost complete DHX38 knockdown, whereas the qPCR-measured a knockdown efficiency of ~40%. Several factors could cause this discrepancy between the results, including the siRNAs affecting the internal normalisation. Western blotting against endogenous DHX protein after generation of stably transfected cell lines (Figure 10) and RNAi-mediated knockdowns (Figure 12) furthermore served as quality controls for the available antibodies. During the assembly of the toolbox, eight antibodies were found to work well, i.e. the antibodies against DHX8, DHX16, DHX29, DHX30, DHX36, DHX38, DHX40, and DHX57.

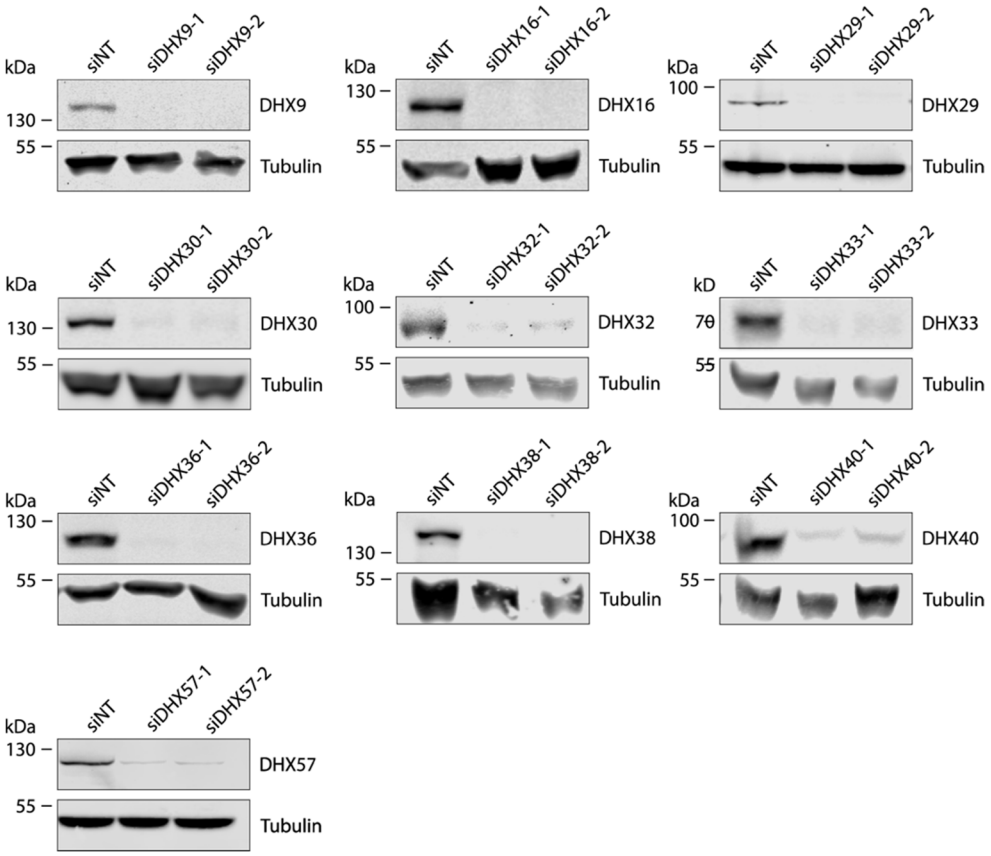


Figure 12 Expression of DHX proteins after RNAi-mediated knockdown. *HEK293 cells were reverse-transfected for 72 hours with siNT or siRNAs targeting the indicated DHX proteins. Cell extracts were analysed by western blotting using antibodies against the endogenous DHX proteins. Anti-tubulin was used as a loading control.*

After evaluating the available tools and consulting the current literature, DHX40 was chosen as a promising protein for further study. DHX40-His_6Prc2xFLAG could be readily detected with the anti-DHX40 and the anti-FLAG antibody, although the protein could not be overexpressed more than ~2-3 fold even at high concentrations of tetracycline (Figure 10). The greatest knockdown of DHX40 was achieved with siDHX40-1, which reduced DHX40 expression by ~90%. DHX40 was of specific interest as there was little to no available information about the function and mode of regulation of this putative RNA helicase.

3.2 Protein interaction partners of DHX40

3.2.1 Establishing knockout-rescue cell lines

Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 knockout and knockout-rescue systems are valuable tools for elucidating protein functions in a cellular context (Ran et al., 2013). The CRISPR/cas9 system utilises single guide RNA (sgRNA)-guided Cas9 endonucleolytic cleavage of target DNA followed by imperfect repair by host proteins to generate premature stop codons through frameshift mutations. Two sgRNA sequences targeting DHX40 were designed using the UCSC genome viewer, and the efficiency of the sgRNAs was estimated using CRISPOR (Concordet & Haeussler, 2018; Kent et al., 2002). The two sgRNAs were designed to target exon 1 (sgRNA-1) and exon 2 (sgRNA-2) of DHX40, respectively. The guide sequences were inserted into a CRISPR/Cas9 system plasmid (PX459), which was then transiently transfected into HEK293 cells (Table 8 and Table 9)(Ran et al., 2013). Monoclonal knockout cell lines were obtained by isolating cells through single-cell dilution. The knockout was then detected by western blotting (Figure 13A), and the achieved genomic mutation was identified by Sanger sequencing of the gDNA locus (Figure 13B).

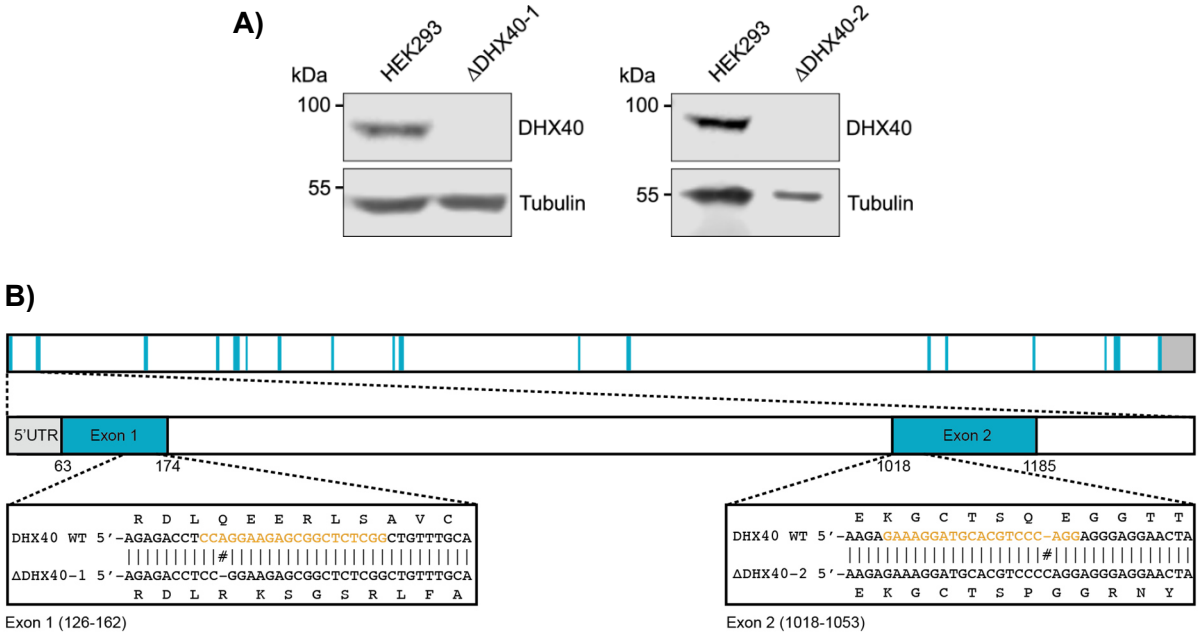
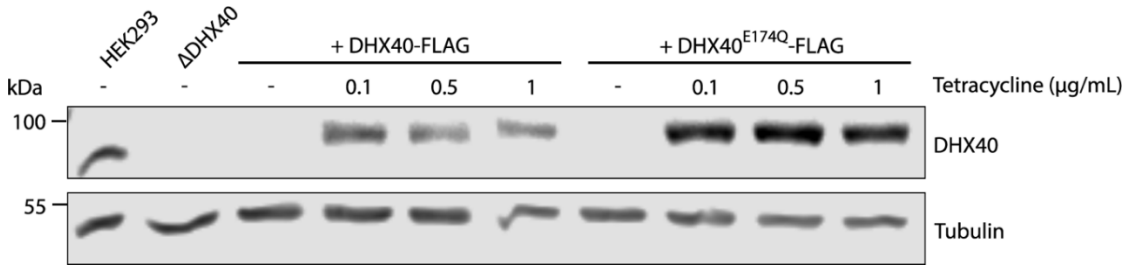


Figure 13 DHX40 knockout cell lines generated using the CRISPR/Cas9 genome editing system. A) *The CRISPR/Cas9 knockout cell lines, ΔDHX40-1 (frameshift in exon 1) and ΔDHX40-2 (frameshift in exon 2), were confirmed by western blotting using anti-DHX40 and anti-tubulin as a loading control.* ***B)*** *DHX40 gene architecture, exons are shown in blue, introns in white, and UTRs in grey. The first magnified view shows the genomic region around exon 1 and exon 2. The second magnified view shows the genomic sequence around the sgRNA-targeted sequences (orange) and the alignment of each mutated sequence to WT gDNA.*

Transfection with sgRNA-1 and sgRNA-2 generated the monoclonal knockdown cell lines ΔDHX40-1 and ΔDHX40-2, respectively. Western blotting confirmed reduced or disrupted expression of DHX40 in both cell lines (Figure 13). In addition, Sanger sequencing of genomic DNA extracted from the cell lines revealed that the ΔDHX40-1 cell line is homozygous for a single nucleic acid deletion in exon 1 of DHX40, while the ΔDHX40-2 cell line is homozygous for a single nucleic acid insertion in exon 2. The frameshift deletion in exon one changed the amino acid sequence from position 24 onward, whereas the frameshift mutation changed the amino acids sequence from position 43 onward, producing a premature stop codon at position 91. Both frameshifts mutations where predicted by SIFT Indel to cause NMD of the DHX40 mRNA (Hu & Ng, 2012).

Next, the ΔDHX40-1 was used to generate stably transfected cell lines as described above (Section 3.1). The generation of stably transfected cell lines from the HEK293-derived ΔDHX40-1 cell line established DHX40 knockout-rescue systems, allowing for the inducible expression of tagged DHX40 while limiting artificial consequences of overexpression. Knockout-rescue cell lines for the tetracycline-inducible expression of DHX40-His_6Prc2xFLAG or the $DHX40^{E174Q}$-His_6Prc2xFLAG (Figure 14) in the ΔDHX40-1 background were generated. The E174Q is a single amino acid substitution in Motif II, which is involved in NTP-binding and hydrolysis. Similar amino acid substitutions have been shown to impair NTP hydrolysis in SF2 helicases, a pre-requisite for DEAH/RHA helicase activity (Hamann et al., 2019; Y. He et al., 2017; Studer et al., 2020).

A)



B)

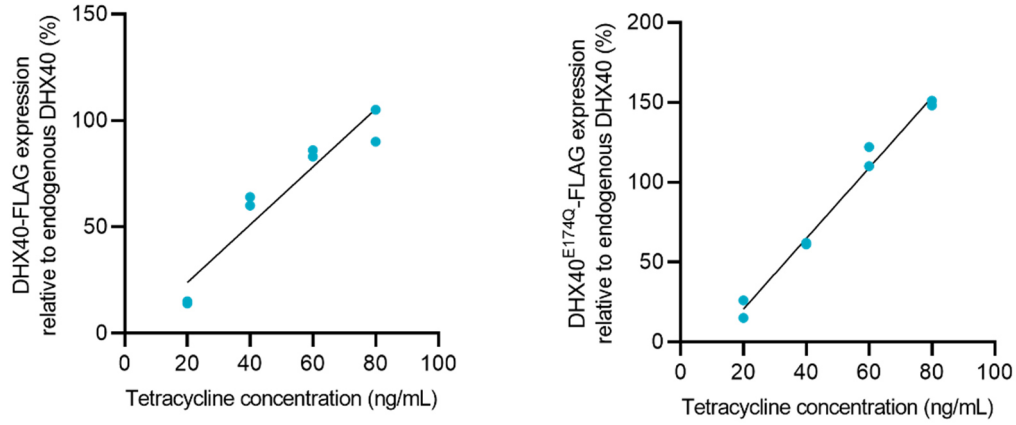


Figure 14 DHX40 Knockout-rescue systems derived from HEK293. A) *Stably transfected cell lines with inducible expression of DHX40-His$_6$Prc2xFLAG (DHX40-FLAG) or DHX40^{E174Q}-His$_6$Prc2xFLAG (DHX40^{E174Q}-FLAG) were generated from ΔDHX40-1. Expression of tagged proteins was induced for 24 hours with the indicated amounts of tetracycline. Cell extracts were analysed by western blotting using anti-DHX40. Anti-tubulin was used as a loading control.* ***B)*** *DHX40-His$_6$PrcFLAG2 and DHX40^{E174Q}-His$_6$Prc2xFLAG expression were induced for 24 hours with indicated tetracycline concentrations. Cell extracts from induced samples and a HEK293 control were analysed as technical duplicates by western blotting using anti-DHX40 and anti-Tubulin. The DHX40 signal was quantified by LI-COR image studio, normalised to the Tubulin, and the expression relative to HEK293 control was calculated. Linear regression was used to calculate the required concentration for the expression of DHX40-His$_6$PrcFLAG2 (~75 ng/mL) and DHX40^{E174Q}-His$_6$Prc2xFLAG (~40 ng/mL) at close to the endogenous level of DHX40. Quantified signals are available in Supplementary table 2-3.*

DHX40-His$_6$Prc2xFLAG and DHX40^{E174Q}-His$_6$Prc2xFLAG expression were induced for 24 hours with different tetracycline concentrations to determine the appropriate concentration of tetracycline to induce protein expression to close to the endogenous level of DHX40. As induction with tetracycline concentrations above 100 ng/mL caused overexpression (Figure 14A), the tetracycline titration was repeated with concentrations <100 ng/mL. Cell extracts were analysed by western blotting, and the quantified signal of DHX40 was used to determine the tetracycline concentration required to induce the transgene expression to close to endogenous levels for DHX40 (Figure 14B). Endogenous expression levels of DHX40-

His_6Prc2xFLAG and DHX40^{E174Q}-His_6Prc2xFLAG were achieved with ~75 ng/mL and ~40 ng/mL tetracycline, respectively. There was no detectable expression of tagged proteins in the absence of tetracycline (Figure 14A).

3.2.2 Affinity purification and mass spectrometry

To better understand DHX40 interactions in the cells, C-terminally-tagged DHX40 proteins were expressed in ΔDHX40-1 cell lines, followed by the enrichment of DHX40 and associated proteins by anti-FLAG immunoprecipitation. Expression of DHX40-His_6Prc2xFLAG and DHX40^{E174Q}-His_6Prc2xFLAG was induced for 24 hours using 75 ng/mL and 40 ng/mL, respectively, to achieve expression close to endogenous levels for DHX40. In addition, HEK293 cells expressing the His_6Prc2xFLAG peptide were used as a control. Proteins recovered by anti-FLAG immunoprecipitation were separated by SDS-PAGE and visualised by silver staining (Figure 15). The silver stain confirmed the enrichment of tagged DHX40 and additional co-precipitated proteins. The immunoprecipitation was repeated in biological duplicates, and the recovered proteins were identified and quantified by liquid chromatography with tandem mass spectrometry (LC-MS/MS). To identify non-specific interactions, the total spectral counts of each protein recovered with DHX40 were compared to those recovered with the FLAG peptide control.

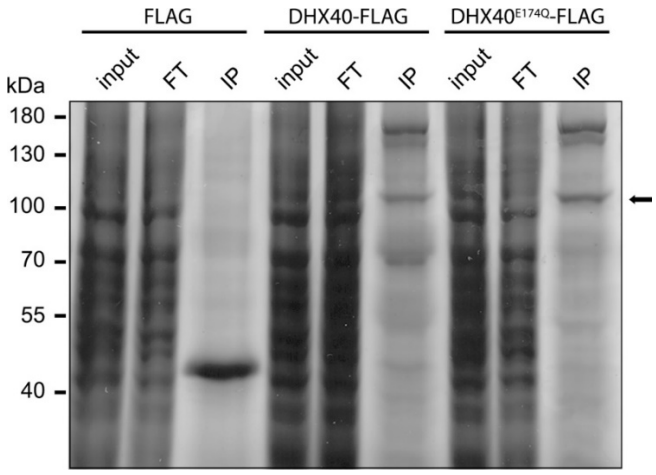


Figure 15 Anti-FLAG immunoprecipitation of C-terminally tagged DHX40 and DHX40^{E174Q} with associated proteins. *DHX40-His_6Prc2xFLAG (DHX40-FLAG) and DHX40^{E174Q}-His_6Prc2xFLAG (DHX40^{E174Q}-FLAG) were expressed from ΔDHX40-1 cells for 24 hours with 75 ng/mL and 40 ng/mL tetracycline, respectively. In addition, His6Prc2xFLAG peptide expression was induced in HEK293 cells for 24 hours with 75 ng/mL tetracycline and used as a control. FLAG-tagged proteins and associated interactors were recovered with anti-FLAG magnetic beads. Bait (DHX40) protein, prey proteins, and inputs (0.8%) were visualised by SDS-PAGE and silver staining. Input, flowthrough (FT), and immunoprecipitation (IP) samples are shown for all three cell lines. The band corresponding to DHX40-FLAG and DHX40E174Q-FLAG is marked with an arrow.*

The spectral count data were analysed using SAINTexpress to normalise the datasets and to separate the true interaction partners from the background (Choi et al., 2012; Teo et al., 2014). The statistical model in SAINT postulates that prey proteins are either “true interactors” of the bait or non-specific binders or contaminants, i.e. “false interactors”. In the case of true interactors, the protein abundance must be significantly higher in the sample versus negative controls. The probability score, SAINT score, represents the confidence level of putative interactions (Teo et al., 2014). Proteins with a SAINT score >0.8 and more than 20 spectral counts averaged over duplicates were considered true interactors. A total of 318 proteins fulfilling these criteria were identified (Figure 16); 233 were found in both samples, and 54 and 31 proteins were unique to DHX40^{E174Q} and DHX40, respectively.

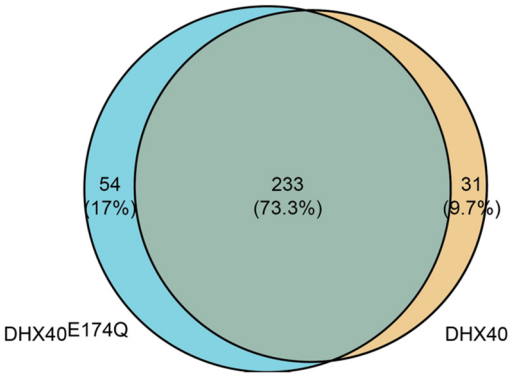


Figure 16 Overlap between DHX40 and DHX40^{E17Q4} associated proteins. *Proteins co-enriched by immuno-precipitation of tagged proteins and identified by LC-MS/MS were analysed by SAINTexpress. Proteins with a SAINT score > 0.8 and Average spectral counts (AvgSpc) > 20 were considered true interactors. 73.3% of true interactors were present in both IPs, while 17% and 9.7% were unique to DHX40^{E174Q} and DHX40, respectively.*

Most unique protein interactors had low abundance and were below the spectral count threshold in the other sample. Therefore, it appears that the protein interactomes of wildtype DHX40 and DHX40 carry the E174Q amino acid substitution are very similar. The Average spectral counts (AvgSpc) for true protein interactors were compared for DHX40 and DHX40^{E174Q} (Figure 17A-B). Several interaction partners were of particular interest either due to their abundance in the immunoprecipitations or due to existing literature (Table 25).

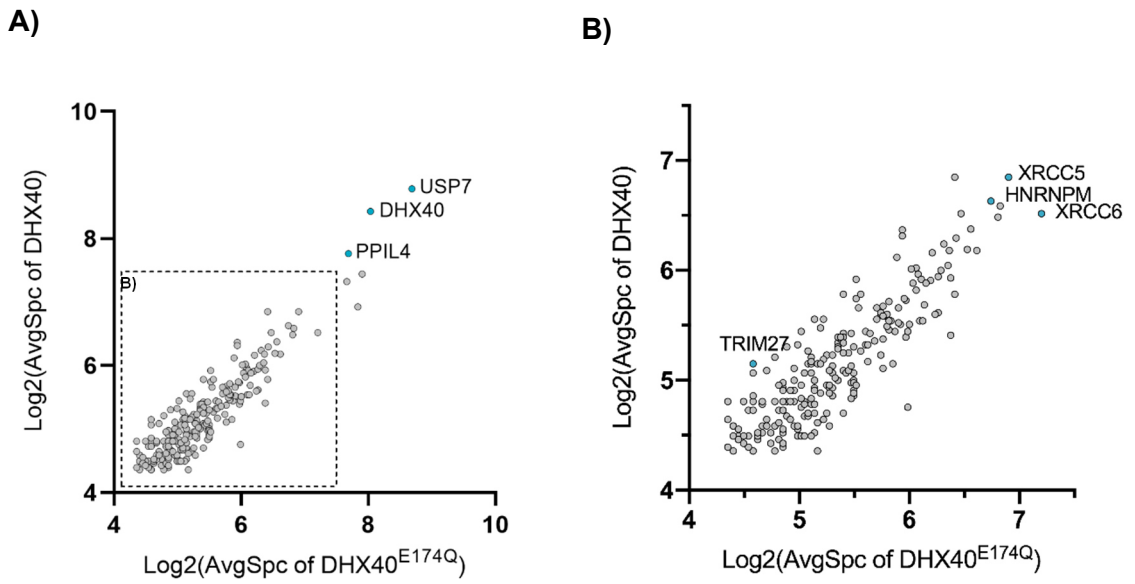


Figure 17 Comparison of Log2 average spectral counts between DHX40 and DHX40^{E174Q} samples. A) *Comparison of Log2(AvgSpc) for all true interactors of DHX40 and DHX40^{E174Q}. The ratio between total spectral counts of true interactors was used for scaling the two datasets. Selected proteins of interest are labelled.* ***B)*** *Magnified view of a region of A). Selected proteins are labelled*

Consistent with previous studies, USP7 was highly abundant in DHX40 and DHX40^{E174Q} immunoprecipitations, indicating a stable interaction with DHX40, independent of its ATPase activity. USP7 is a deubiquitinase previously reported to interact with and rescue DHX40 from proteasomal degradation (Georges et al., 2018). In addition, the mass spectrometry uncovered several previously unreported interaction partners, including PPIL4, XRCC5, XRCC6, HNRNPM and TRIM27 (Table 25).

Table 25 SAINT analysis results for selected proteins associated with DHX40 and DHX40^{E174Q}. *SAINT score and average spectral counts for selected proteins identified by LC-MS/MS as prey proteins of DHX40. An extended list of identified proteins is available in Supplementary table 4-5.*

Identified proteins	Average Spectral Counts			SAINT score	
	DHX40	DHX40^{E174Q}	FLAG	DHX40	DHX40^{E174Q}
USP7	441	412.19	19.5	1	1
DHX40	344.5	262.11	0	1	1
PPIL4	217	205.74	2	1	1
XRCC5	115	119.78	21	1	1
HNRNPM	99	107.1	20.5	1	1
XRCC6	91.5	147.26	15.5	1	1
TRIM27	35.5	23.96	1	1	1

TRIM27 is an E3 ligase associated with cell proliferation, apoptosis, innate immune response, and transcriptional regulation (Yu et al., 2022). This protein was particularly interesting because previous studies identified conversed interactions between SF2 helicases and TRIM proteins (Kato et al., 2021). XRCC5 and XRCC6 are key players in DNA NHEJ, where they contribute by forming a heterodimer that binds DNA double-stranded breaks (Kragelund et al., 2016). HNRNPM is an RBP that binds pre-mRNAs and is associated with pre-mRNA splicing (Ho et al., 2021). Finally, PPIL4 is a putative propyl isomerase (PPIase) that is largely uncharacterised. However, PPIL4 was recently shown to be an RBP that binds to L1 elements (van Nostrand et al., 2020). Interestingly, several other highly enriched proteins are implicated in regulating L1 retrotransposition, including XRCC5, XRCC6, and HNRNPM (N. Liu et al., 2017), potentially suggesting a role for DHX40 in L1 regulation.

3.3 Regulation of DHX40 expression level

TRIM E3 ligases have been reported to bind SF2 helicases through a conserved interaction between the helicase core and the Pry/Spry domain of TRIM E3 ligases (Kato et al., 2021). Additionally, DHX40 has been confirmed as a substrate of USP7, a deubiquitinase that rescues DHX40 from proteasomal degradation (Georges et al., 2018; Kato et al., 2021). The interactions of DHX40 proteins with USP7 and TRIM27, identified by immunoprecipitation and LC-MS/MS, were validated by repeating the above anti-FLAG immunoprecipitations (Section 3.2.2) and analysing the recovered proteins by western blotting using antibodies against the endogenous proteins (Figure 18).

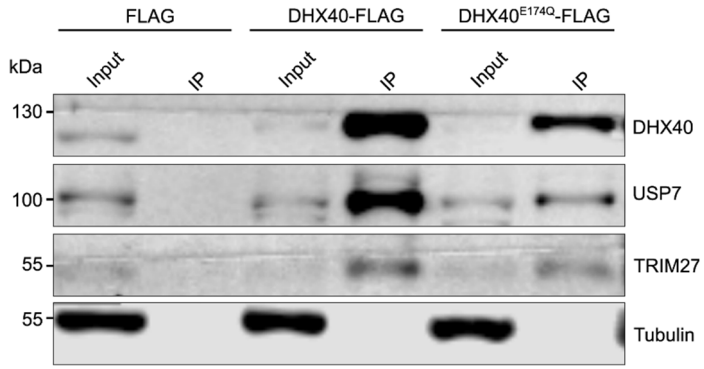


Figure 18. Association of USP7 and TRIM27 with DHX40 detected by anti-FLAG immunoprecipitation and western blotting. *Anti-FLAG IPs were performed with ΔDHX40-1 rescue systems expressing either DHX40-His$_6$Prc2xFLAG (DHX40-FLAG) or DHX40_E174Q-His$_6$Prc2xFLAG (DHX40^{E174Q}-FLAG) at close to endogenous levels of DHX40. FLAG-tagged proteins were immunoprecipitated with anti-FLAG magnetic beads, and recovered proteins and inputs (0.4%) were analysed by western blotting using antibodies against the indicated proteins. Tubulin was used as a control.*

Western blotting confirmed that USP7 and TRIM27 are specifically recovered in anti-FLAG immunoprecipitation of C-terminally-tagged DHX40 proteins but not with the His$_6$Prc2xFLAG

peptide alone. Furthermore, neither the TRIM27 interaction nor the USP7 interactions were affected by the E174Q amino acid substitution. Next, to eliminate the possibility that the interactions observed were bridged by RNA, immunoprecipitation experiments in the presence of RNases were performed (Figure 19). DHX40-His_6Prc2xFLAG expression was induced in HEK293 cells by 1 µg/mL tetracycline for 24 hours, and the tagged DHX40 and associated proteins were recovered on anti-FLAG beads in the presence of RNase A, RNase T1, and RNase H (Figure 19). RNase A and RNase T1 hybridise the phosphodiester bond 3' to a pyrimidine nucleotide and guanine nucleotide, respectively, whereas RNAse H hydrolyses the phosphodiester bonds of RNA hybridised to DNA.

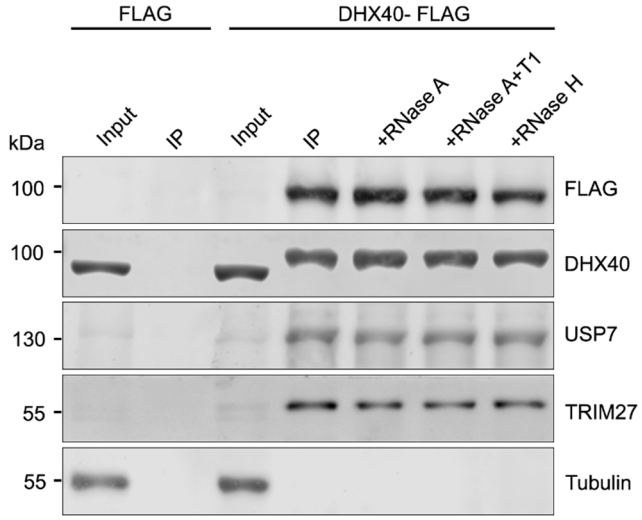


Figure 19 Recovery of TRIM27 and USP7 by immunoprecipitation of DHX40 in the presence of RNases. *DHX40-His_6Prc2xFLAG expression was induced with 1 µg/mL tetracycline for 24 hours in HEK293 cells. Immunoprecipitations were performed on the DHX40-FLAG-expressing cells in the presence of RNase A, RNases A+T1 and RNase H, as indicated. Recovered proteins and inputs (0.4%) were analysed by western blotting using the indicated antibodies. Tubulin was used as a control.*

Western blot analysis of the recovered proteins revealed that DHX40 interactions with USP7 and TRIM27 were not affected by the nuclease treatments. These results indicate that the interactions are not bridged by RNA and are, thus, protein-protein interactions. USP7 has been reported to deubiquitinate and stabilise many interaction partners, including DHX40 (Kato et al., 2021). To confirm that USP7 stabilise DHX40 in HEK293 cells, USP7 was depleted by RNAi-mediated knockdown, and cell extracts were analysed by western blotting using antibodies against DHX40 and USP7 (Figure 20). As expected, the depletion of USP7 reduced DHX40 protein levels, thus confirming that DHX40 is a substrate for deubiquitinylation by USP7. This result suggests that DHX40 is marked for proteasomal degradation by ubiquitination; however, the ubiquitinated lysine has not been reported.

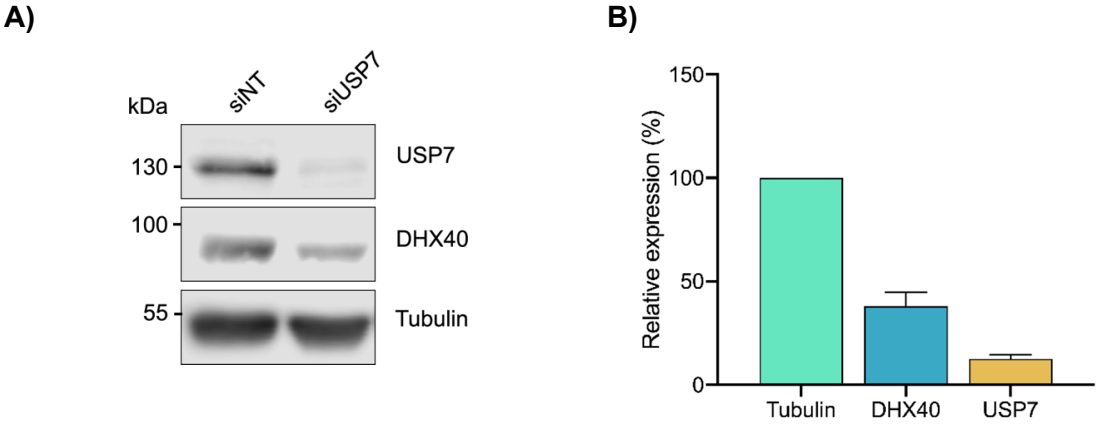


Figure 20 Effect of RNAi-mediated depletion of USP7 on DHX40 protein levels. A) *HEK293 cells were transfected with siNT or siUSP7. Cell extracts were analysed by western blotting using the indicated antibodies. Tubulin was used as a loading control.* ***B)*** *Knockdown of USP7 was performed as in A and analysed by western blotting. The signal was quantified by LI-COR Image studio and normalised to the tubulin signal. The expression of indicated proteins after siUSP7 treatment is relative to the signal in siNT-treated samples. The results are the mean of three biological duplicates, and the error bars show SEM. Data is available in Supplementary table 6-7*

A recent study demonstrated a conserved mode of interaction between the Pry/Spry domain of TRIM E3 ligases and the helicase core of cognate SF2 helicases. Furthermore, the study identified several amino acid residues important for DDX41:TRIM26 and DDX41:TRIM41 interaction through mutagenesis analysis (Kato et al., 2021). To determine if this mode of interaction is conserved in the interaction between DHX40 and TRIM27, a model of the DHX40 RecA1 domain, generated using the SWISS modelling tool, was structurally aligned with the crystal structure of the DDX41 DEAD domain (Figure 21). Based on this structural alignment, two regions were chosen for mutagenesis analysis of DHX40. The alpha mutation (αMut) disrupts a small alpha helix by substituting the ^{95}FSQH^{98} amino acid sequence with GGGG (Figure 21A). In the N-terminal deletion mutant (ΔNterm), the first 61 amino acids in DHX40 were truncated; this region includes a long disordered loop and a short N-terminal helix (Figure 21B). The two variants were furthermore combined in a double mutant.

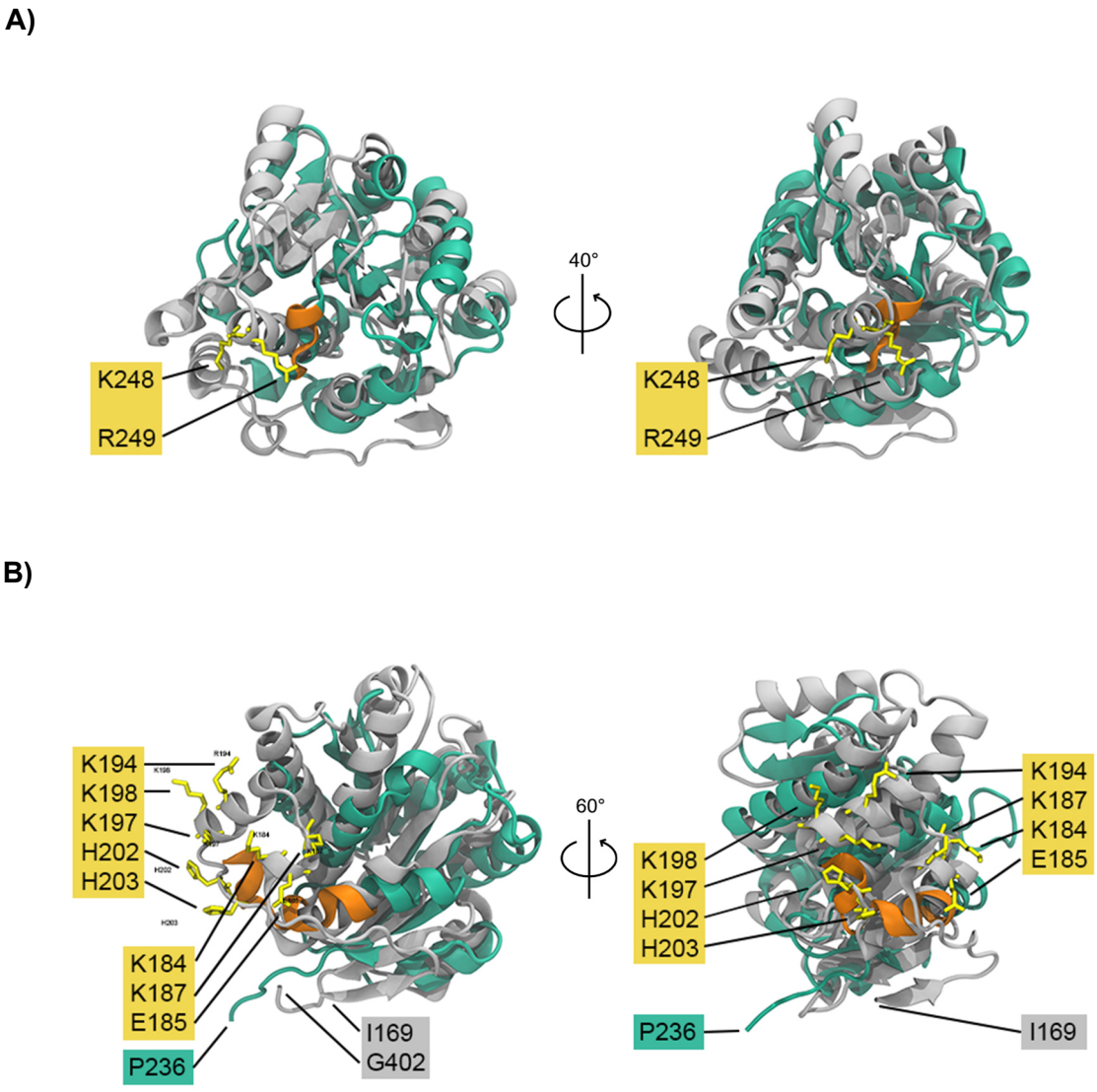


Figure 21 Structural alignment of DDX41^{DEAD} (PDB: 5GVR) and DHX40^{RecA1}(6HYS, theoretical SWISS-Prot model, 01/10/2021). *The RecA1 domain of DHX40 (cyan; amino acids 49-236) was structurally aligned with the DEAD domain of DDX41 (grey; amino acids 169-402). Amino acids implicated in the interaction between DDX41 and its cognate TRIM proteins are marked in yellow (Kato et al., 2021).* ***A)*** *DHX40 αMut: A single alpha helix (orange) was disrupted in DHX40 by replacing ^{95}FSQH^{98} with ^{95}GGGG^{98}.* ***B)*** *DHX40 ΔNterm: amino acid residues 1-61, which include a disordered loop and a small alpha helix, were removed (residues 49-61 are marked in orange, residues 1-48 were not included in the structure). The visible termini of DHX40 and DDX41 are marked in cyan and grey boxes, respectively. Structural alignment and visualisation were in VMD (Eargle et al., 2006; Humphrey et al., 1996; Stone, 1995).*

Plasmids encoding the C-terminally His_6Prc2xFLAG-tagged DHX40 mutants were constructed by site-directed mutagenesis of the plasmid encoding DHX40-His_6Prc2xFLAG. Stably transfected HEK293 cell lines for the inducible expression of the DHX40 variants were generated as described previously (Section 3.1). Expression of C-terminally-tagged DHX40 and DHX40 variants was induced in HEK293 cells for 24 hours using 1 µg/mL tetracycline.

Tagged DHX40 and associated proteins were recovered by anti-FLAG immunoprecipitation, as described previously.

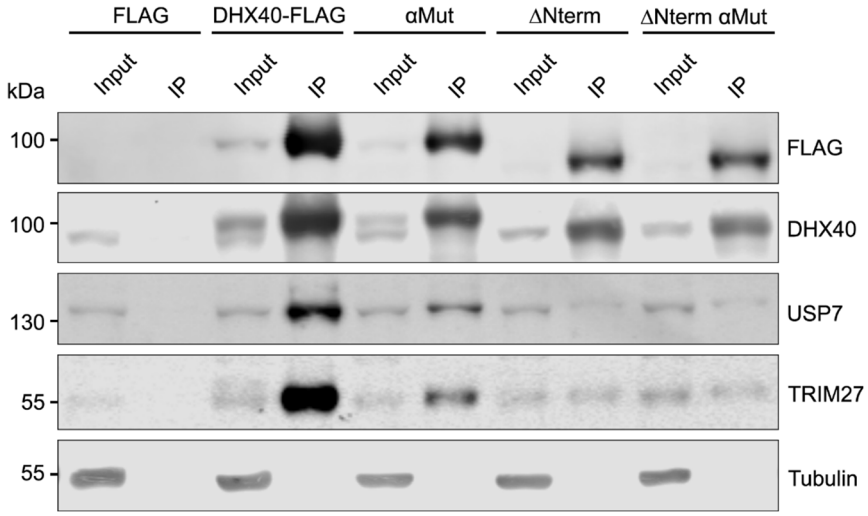


Figure 22 Anti-FLAG immunoprecipitation of tagged DHX40 and DHX40 mutants to determine requirements of TRIM27 interaction. *Tagged DHX40 and mutants (αMut, ΔNterm, and ΔNterm-αMut) were expressed from HEK293 cells and recovered by anti-FLAG immunoprecipitation. The recovered proteins and inputs (0.4%) were subsequently analysed by the indicated antibodies. Tubulin was used as a control.*

The western blot analysis of proteins recovered by immunoprecipitation of tagged DHX40 wildtype and mutants revealed that USP7 and TRIM27 were recovered at detectable levels regardless of mutation. However, the αMut and ΔNterm mutations separately greatly reduced the interaction between DHX40 and TRIM27, with the ΔNterm mutation causing the greatest reduction of the enrichment of TRIM27. The double mutant showed a similar pattern to the ΔNterm mutation, suggesting that the decrease in binding between DHX40 and TRIM27 was mainly caused by ΔNterm mutation. The mutations furthermore caused a slight reduction in DHX40 interaction with USP7; Whether this reduction is due to disruption of the interaction site or due to reduced ubiquitination requires further investigation.

3.4 PPIL4 as a DHX40 interactor

3.4.1 DHX40 interaction with PPIL4

The interaction between DHX40 and PPIL4, identified by immunoprecipitations and LC-MS/MS, was validated by repeating the above anti-FLAG immunoprecipitations (Section 3.2.2) and analysing the recovered proteins by western blotting using antibodies against the PPIL4 and DHX40 (Figure 23).

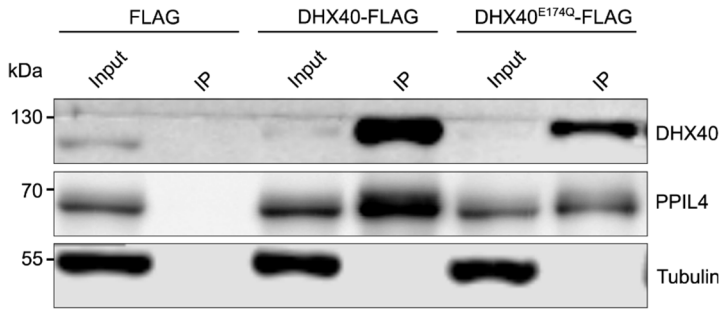


Figure 23 Association of PPIL4 with DHX40 detected by anti-FLAG immunoprecipitation and western blotting. *Anti-FLAG IPs were performed with ΔDHX40-1 rescue systems expressing either DHX40-His$_6$Prc2xFLAG (DHX40-FLAG) or DHX40_E174Q-His$_6$Prc2xFLAG (DHX40^{E174Q}-FLAG) at close to endogenous levels of DHX40. FLAG-tagged proteins were immunoprecipitated with anti-FLAG magnetic beads, and recovered proteins and inputs (0.4%) were analysed by western blotting using antibodies against the indicated proteins. Tubulin was used as a control.*

Western blotting confirmed that PPIL4 is specifically recovered in anti-FLAG immuno-precipitation of C-terminally-tagged DHX40 proteins but not in immunoprecipitations of the His$_6$Prc2xFLAG peptide alone. Furthermore, the interaction between DHX40 and PPIL4 was not affected by the E174Q amino acid substitution.

PPIL4 is a putative cyclophilin PPIas consisting of a PPIase domain, an RRM, and several disordered regions (Figure 24). Although still functionally uncharacterised, PPIL4 has been confirmed to bind L1 RNAs and it is therefore, possible that the interaction between DHX40 and PPIL4 is bridged by RNA (van Nostrand et al., 2020).

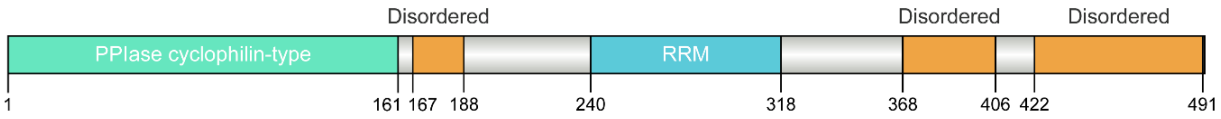


Figure 24 Schematic representation of the PPIL4 protein drawn to scale. *PPIL4 consist of a PPIase cyclophilin-type domain (green), which is conserved among PPIase, three disordered regions (orange) and an RRM (Blue). The representation was created using IBS 2.0 (Xie et al., 2022)*

Therefore, to investigate the possibility that the interaction between DHX40 and PPIL4 is bridged by RNA, immunoprecipitation experiments in the presence of RNases were performed as described in Section 3.3 (Figure 25). Western blot analysis confirmed that none of the nuclease treatments affected PPIL4 recovery in anti-FLAG immunoprecipitation of His_6Prc2xFLAG-tagged DHX40, suggesting a direct protein-protein interaction between DHX40 and PPIL4.

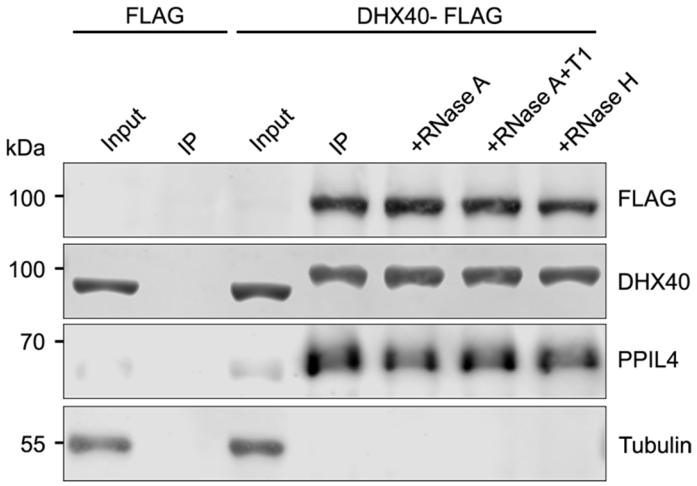


Figure 25 Recovery of PPIL4 by immunoprecipitation of DHX40 in the presence of RNases. *DHX40-His_6Prc2xFLAG expression was induced with 1 µg/mL tetracycline for 24 hours in HEK293 cells. IPs were performed on the DHX40-His_6Prc2xFLAG (DHX40-FLAG) expressing cells in the presence of RNase A, RNases A+T1 and RNase H, as indicated. Recovered proteins and inputs (0.4%) were analysed by western blotting using the indicated antibodies. Tubulin was used as a control.*

Interestingly, PPIL4 was reported as an interaction partner of USP7. The study reported that PPIL4 was recovered by affinity purification of USP7 but that PPIL4 was not a substrate for deubiquitination (Georges et al., 2018). To assess if the interaction between USP7 and PPIL4 is bridged by DHX40, anti-PPIL4 IPs were performed in ΔDHX40-1 and HEK293 cells (Figure 26). Endogenous PPIL4 and associated proteins were precipitated from the cells by anti-PPIL4 coupled to protein G resin. Immunoprecipitated proteins were then analysed by western blotting (Figure 26). Western blotting of immunoprecipitated proteins revealed that USP7 coprecipitates with PPIL4 in HEK293 cells but not in ΔDHX40-1. These results indicate that DHX40 bridges the interaction between USP7 and PPIL4 and that PPIL4 is a direct interaction partner of DHX40 rather than USP7.

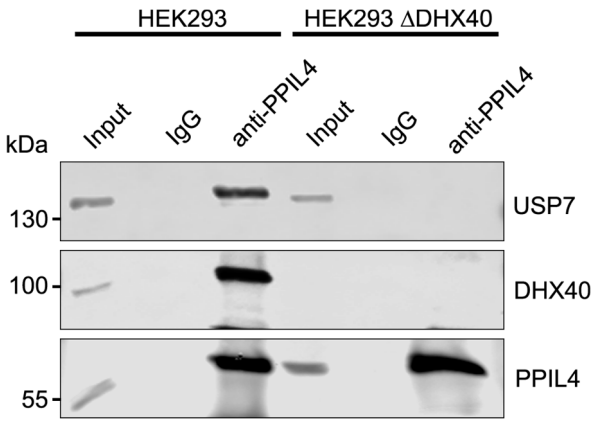


Figure 26 Recovery of USP7 by anti-PPIL4 immunoprecipitation from ΔDHX40-1 and HEK293 cells. *Protein G resin was coupled to anti-PPIL4 or IgG isotype control. The coupled resin immunoprecipitated PPIL4 and associated proteins from ΔDHX40-1 and HEK293 cells. Recovered proteins, inputs (2.5%), and IgG controls were analysed by western blotting using the indicated antibodies.*

3.4.2 Regulation of DHX40 ATPase activity

DEAH/RHA helicases are processive helicases in which NTP hydrolysis is coupled to translocation (Hamann et al., 2019; Sloan & Bohnsack, 2018; Studer et al., 2020). In addition, several helicases from this subfamily have been reported to be regulated by protein cofactors (K. E. Bohnsack et al., 2021; Boneberg et al., 2019; Choudhury et al., 2019).To determine if DHX40 is an active ATPase, a prerequisite for DEAH/RHA helicase translocation, recombinantly expressed proteins were purified and used to determine the steady-state ATPase activity of DHX40 *in vitro*. The steady-state ATPase assay investigated the effect of PPIL4 on DHX40 ATPase activity. First, a construct for expressing an N- and C-terminally truncated DHX40 variant was generated to remove disordered regions, likely improving protein solubility and folding. In addition, an N-terminal maltose binding protein (MBP) tag was introduced to improve solubility, and a C-terminal polyhistidine tag (His_{10}) was added to allow affinity purification by immobilised metal affinity chromatography (IMAC). Next, a similar DHX40 construct with an E174Q amino acid substitution was generated to asses if the mutation impaired ATPase activity. Additionally, a construct for the expression of full-length PPIL4 with an N-terminal ZZ-tag and a C-terminal His_7-tag was generated.

For recombinant expression and purification of DHX40 variants, competent E. coli cells were transformed with DHX40 expression plasmid and grown to stationary phase in buffered LB media with added magnesium and glucose (Azatian et al., 2019). At stationary phase, the cells were cooled, and the expression of DHX40 variants was induced with IPTG. DHX40 recombinant expression was verified by SDS-PAGE and Coomassie staining (Figure 27A, “non-induced” lane and “induced” lane). Next, cells expressing the protein were harvested and lysed, after which insoluble proteins were removed by centrifugation; a substantial amount of

recombinantly expressed protein precipitated, indicating low solubility (Figure 27A “insoluble”). Next, the cleared lysate was incubated with Ni-NTA resin to recover His-tagged proteins specifically. The Ni-NTA resin was a limiting reagent, i.e. the amount used was expected to be insufficient to bind all expressed DHX40. The deficiency was by design to reduce the binding of non-specific contaminants. Unbound proteins, including *E. coli* proteins and recombinantly expressed DHX40, remained in the flow-through (Figure 27A “Flow-through”). While immobilised on the resin, the protein was washed by a series of washing steps to remove contaminants (not shown). Finally, His-tagged proteins were eluted with imidazole (Figure 27A “Elution 1-4”). Both DHX40 constructs were recombinantly expressed and purified using the same protocol. PPIL4 was grown in log-phase rather than stationary phase but was similarly purified. Purified proteins were dialysed to remove imidazole, and the quality of the final purified proteins was visualised by SDS-PAGE and Coomassie staining (Figure 27B).

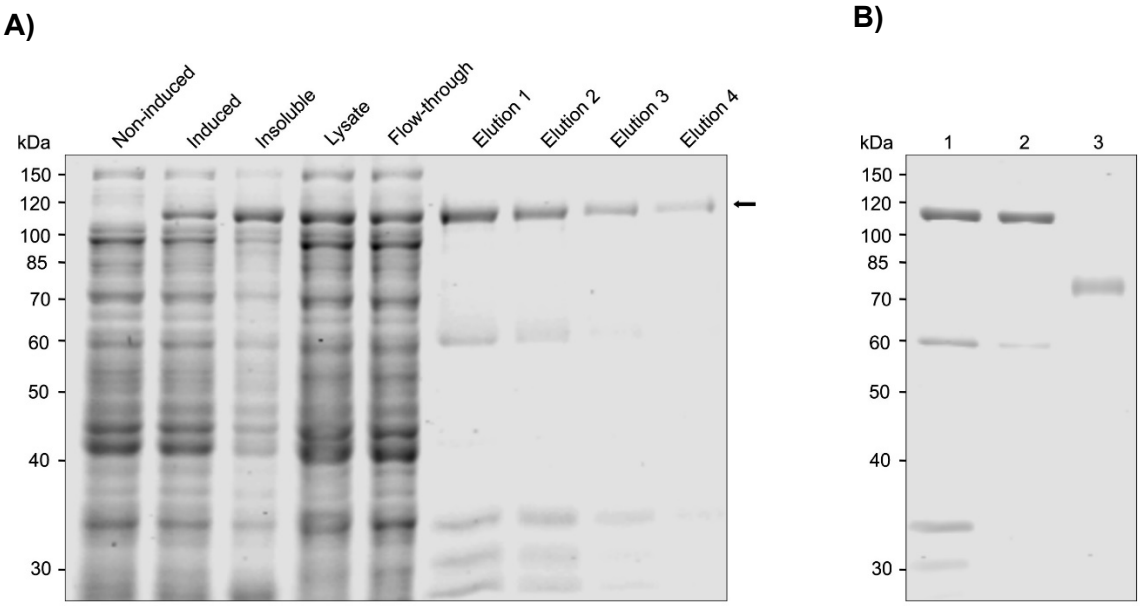


Figure 27 Recombinant expression of DHX and affinity purification of DHX40 and PPIL4 proteins. A) *Overview of purification steps for MBP-DHX40_26-699-His$_{10}$ by coomassie-stained SDS-PAGE. MBP-DHX40_26-699-His$_{10}$ was recombinantly expressed in E. coli, purified by batch purification on ni-NTA resin and eluted with imidazole.* ***B)*** *Recombinantly expressed proteins purified by IMAC followed by dialysis. Lane 1 is MBP-DHX40_26-699-His$_{10}$, (~110 kDa) lane 2 is MBP-DHX40_26-699_E174Q-His$_{10}$, (~110 kDa) and lane 3 is ZZ-PPIL4-His$_7$ (~70 kDa). Purification of the PPIL4 variant was performed by Philipp Hackert.*

The additional proteins co-purified with DHX40 variants were, based on size, speculated to be common IMAC contaminants, e.g. Hsp60 (57.0 kDa), YadF (25.0 kDa), and CAT (25.5 kDa) (Bolanos-Garcia & Davies, 2006). These contaminants are prevalent in IMAC protein purifications of lowly expressed His-tagged proteins as they compete for binding to the resin (Bolanos-Garcia & Davies, 2006). Although recombinantly expressed and purified by the same protocol, the final protein preparation of the DHX40^{E174Q} variant was slightly purer than the

DHX40 variant (Figure 27B lane 1-2). The PPIL4 protein variant was highly expressed and could be purified with few to no contaminants (Figure 27B lane 3).

The steady-state ATP hydrolysis activity of DHX40 was analysed by *in vitro* NADH-coupled ATPase assay in the presence or absence of a model RNA (Figure 28A-B). In this assay, the ATP hydrolysis is coupled to NADH oxidation by the intermediate proteins, Pyruvate kinase (PK) and Lactate dehydrogenase (LDH), and the intermediate substrate phosphoenolpyruvate (PEP), which is present in excess (Figure 28A) (Scharschmidt et al., 1979). The consumption rate of NADH can be measured by monitoring the fluorescence at λ_{ex} = 355 nm and λ_{em} = 480 nm and transforming the fluorescence to NADH concentration using a standard curve (L. Wang et al., 2020). This method, therefore, provides a convenient way of measuring ATPase hydrolysis and limits product inhibition by ADP.

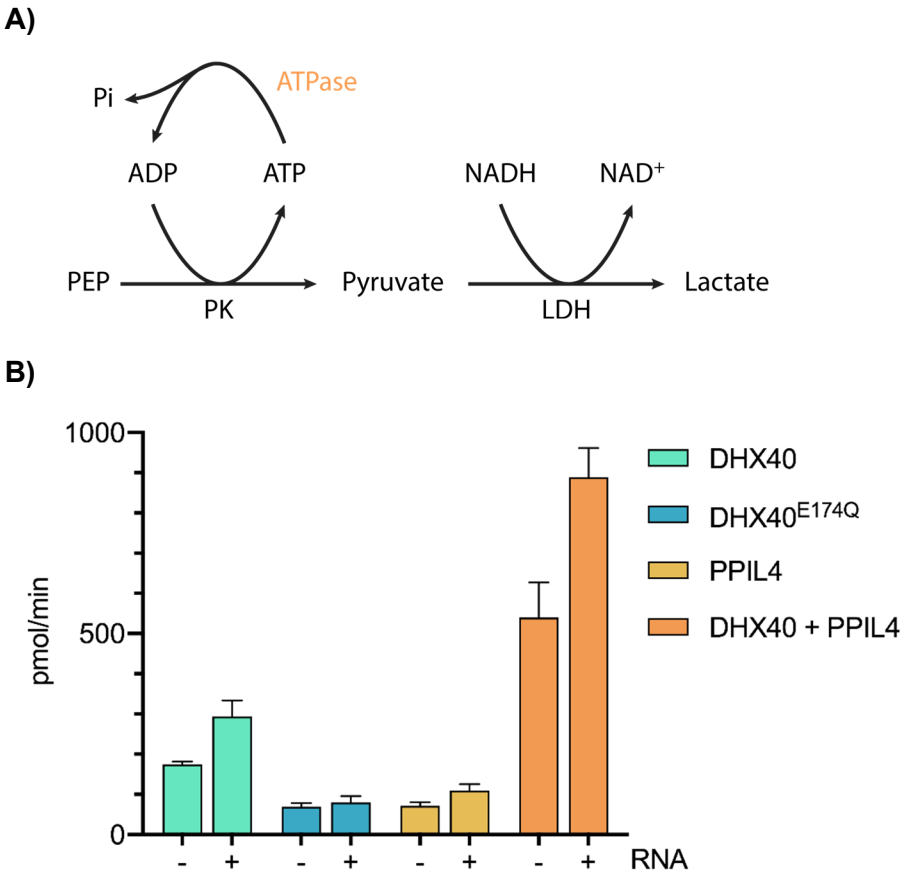


Figure 28 NADH-coupled ATPase assay using purified recombinant DHX40 and PPIL4 proteins. A) *Schematic overview of NADH-coupled ATPase assay. ATP hydrolysis is coupled to the conversion of PEP to pyruvate by PK; pyruvate is subsequently converted to lactate by LDH, which simultaneously oxidises NADH to NAD⁺. The NADH consumption is monitored by fluorescence at λ_{ex} = 355 nm and λ_{em} = 480 nm.* ***B)*** *The in vitro NADH-coupled ATPase assay was used to measure the ATP hydrolysis activity of purified recombinant DHX40_26-699-His10, MBP-DHX40_26-699_E174Q-His10, and ZZ-PPIL4-His7 in the presence or absence of a model RNA. The ATPase rate is given as pmol hydrolysed ATP per minute. NADH decay is accounted for by substrating the*

NADH hydrolysis rate in a negative control without protein from every other reaction. The results are the mean of three independent experiments, and error bars show ± standard error of the mean (s.e.m). Plotted numbers and statistical analysis is available in Supplementary table 8-11

In the *in vitro* setup, recombinant DHX40 displayed ATPase activity stimulated by RNA. The E174Q amino acid substitution significantly impaired DHX40 ATPase. As expected, PPIL4 did not display ATPase activity in the presence or absence of RNA. Compared to the activity of helicase alone, the addition of PPIL4 significantly increased DHX40 ATPase activity and elevated the stimulatory effect of RNA. Together, these results demonstrate that DHX40 is an RNA-stimulated ATPase and that PPIL4 is a stimulatory cofactor of DHX40.

3.5 RNA substrates of DHX40

Identifying direct RNA targets is a critical step in uncovering the function of RNA-binding proteins, such as RNA helicases. Several methods are frequently used for recovering RNAs associated with a target protein, including RNA immunoprecipitation (RIP), crosslinking and immunoprecipitation (CLIP), and crosslinking and analysis of cDNA (CRAC) (Hafner et al., 2021). For DHX40, the CRAC method was used, as this method has allowed the identification of RNA substrates and binding sites for several RNA helicases (M. T. Bohnsack et al., 2012; Choudhury et al., 2020).

DHX40-His_6Prc2xFLAG or His_6Prc2xFLAG peptide expression was induced with 1 µg/mL in HEK293 stably transfected cells for 24 hours. Then, the cells were irradiated with ultraviolet light (UV) at 254 nm to crosslink RNA-binding proteins to their RNA substrates (Figure 29A). After crosslinking, C-terminally tagged DHX40 was recovered under native conditions on anti-FLAG beads, and crosslinked RNA was partially digested. Next, a second purification step was performed by IMAC under denaturing conditions, followed by on-bead adaptor ligation and radioactive labelling of RNA fragments. Subsequently, the eluted protein-RNA complexes were separated by NuPAGE, and radioactive RNAs were detected by autoradiography (Figure 29B).

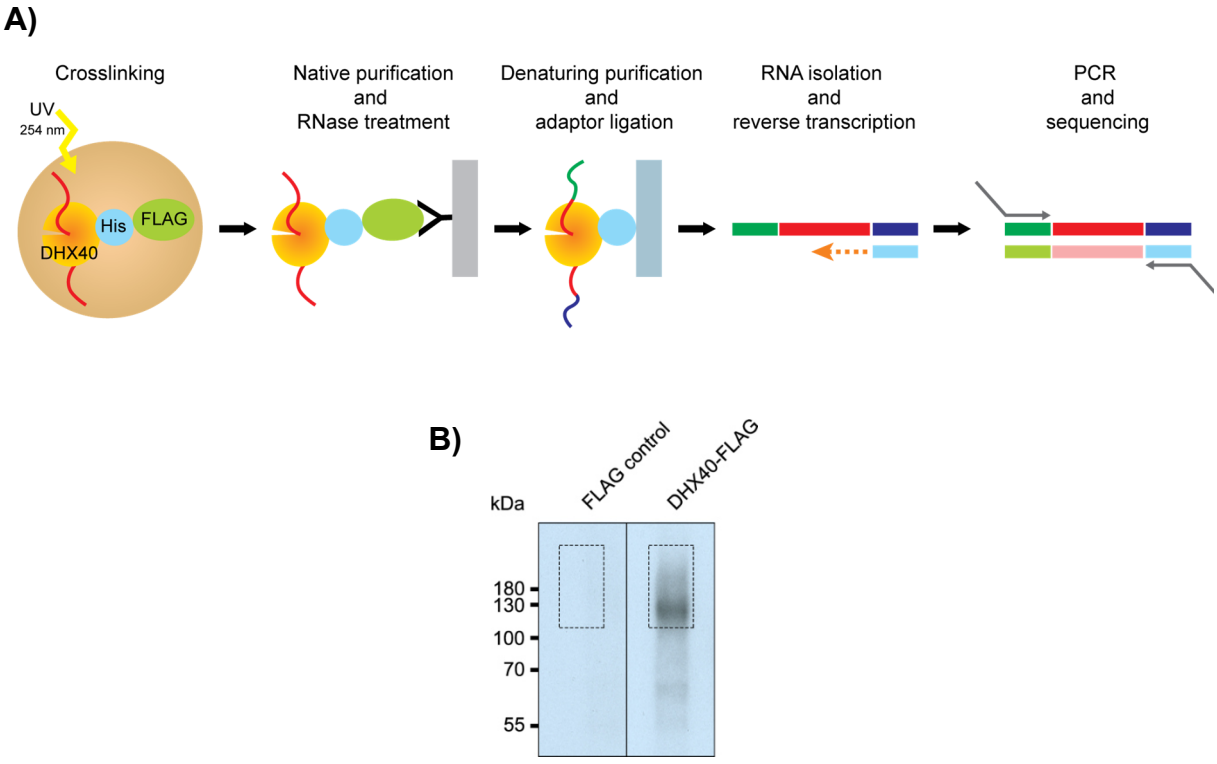


Figure 29 Overview of UV crosslinking and analysis of cDNA (CRAC) for identification of DHX40 RNA substrates. *A) Schematic outline of the CRAC method. DHX40-His$_6$Prc2xFLAG was crosslinked to RNA substrates by irradiation with UV light at 254 nm. Protein-RNA complexes were purified by tandem affinity purification. First, the proteins-RNA complexes were immobilised on anti-FLAG beads under native conditions, after which RNAs co-eluted with the FLAG-tagged protein were partially digested. Second, the proteins-RNA complexes were purified by IMAC under denaturing conditions, followed by on-bead adaptor ligation and radioactive labelling of recovered RNAs. Next, protein-RNA complexes were separated by NuPAGE, and RNAs were extracted from the membrane region corresponding to the size of DHX40 RNA complexes (see B). Finally, isolated RNAs were reverse transcribed, and the cDNA library was amplified by PCR and subjected to next-generation sequencing (NGS). The schematic outline is based on Haag et al., Meth. Mol. Biol. 2017. B) Autoradiography of 32P-labeled protein-RNA complexes after separation by NuPAGE and transfer to nitrocellulose membrane. Cells expressing DHX40-His$_6$Prc2xFLAG (DHX40-FLAG) or His$_6$Prc2xFLAG peptide (FLAG control) were treated as in A. The autoradiography scan of DHX40-FLAG was compared to the FLAG control to confirm the specificity of the signal. RNAs were isolated from the indicated regions, encompassing DHX40 in complexes with RNAs and the equivalent region in the control lane. The CRAC experiment was performed by Philipp Hackert.*

Autoradiography detected a radioactive signal in the DHX40-His$_6$Prc2xFLAG sample but not in the control sample, indicating that DHX40 interacts directly and specifically with cellular RNAs. The membrane region containing the DHX40-RNA complexes and the equivalent region in the control lane was excised and used isolation of RNAs (Figure 29B). First, the RNAs were released from the protein complexes by proteinase K digestion, after which RNAs were extracted. Finally, the extracted RNAs were reverse transcribed to generate a cDNA library, which was subsequently amplified by PCR and subjected to NGS.

Obtained sequencing reads were mapped to the GRCh38 human genome, and peak calling was performed using the adaptive approach implemented in the tool PEAKachu (Holmqvist et al., 2018). In this approach, peaks, i.e. genomic regions with clustered reads, are defined using the blockbuster algorithm (Langenberger et al., 2009). Peaks were defined for the DHX40 library and compared to the peptide control to generate a count matrix for each genomic region. Normalisation and fold change calculations for each genomic region were performed in the DESeq2 package in R (Love et al., 2014). Genomic regions with > 50 read counts and fold change (FC) > 5 in the DHX40 sample relative to peptide control were defined as significant peaks. Only reads assigned to significant peaks were considered in the downstream bioinformatic analysis.

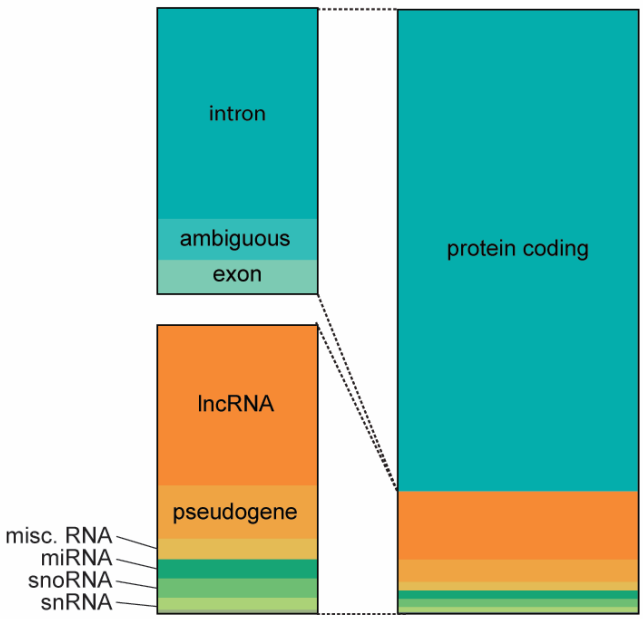


Figure 30 Relative biotype distribution of significant peaks in the DHX40 CRAC dataset. *Sequencing reads from the DHX40 CRAC were mapped to the GRCh38 human genome GRCh38 human genome, and peak calling was performed using the adaptive approach implemented in the tool PEAKachu. The DHX40 dataset was subset by significant peaks (FC>5 and read counts >50) to eliminate nonspecific RNAs and contaminants. 493171 normalised reads from significant peaks mapped to annotated genomic features. Most reads were mapped to "protein coding" (79.78%), followed by "lncRNA" (11.24%) and "pseudogene" (3.72%). The protein-coding reads could additionally be mapped to either introns (73.69%) or exons (11.85%) or were ambiguous (14.46%). Reads were furthermore mapped to miscellaneous RNAs (misc. RNA), i.e. unclassifiable non-coding RNAs, miRNAs, snoRNAs, and snRNAs. The number of normalised reads mapped to each biotype is available in the Supplementary table 12-14. The CRAC bioinformatic analysis was performed in collaboration with Nicolás, Lemus MD PhD*

Meta-gene analysis of significant peaks revealed enrichment of peaks mapping to genomic regions annotated as protein-coding RNAs, lncRNAs and pseudogenes (Figure 30). Interestingly, the majority of reads mapping to protein-coding RNAs map to introns. This

distribution may reflect a role in regulating pre-mRNA splicing but could also indicate DHX40 binding pre-mRNAs near TEs, as 60% of TEs are located in intronic sequences (Sela et al., 2007). Also, lncRNAs and pseudogenes are associated with transposable elements as 83% of lncRNAs contain at least one TE and retrotransposition of mRNAs can generate processed pseudogenes (Fort et al., 2021; Troskie et al., 2021). Furthermore, many top protein interactors of DHX40 have been shown to regulate L1 retrotransposition, and two, PPIL4 and HNRNPM, bind L1 RNAs directly (N. Liu et al., 2017; van Nostrand et al., 2020).

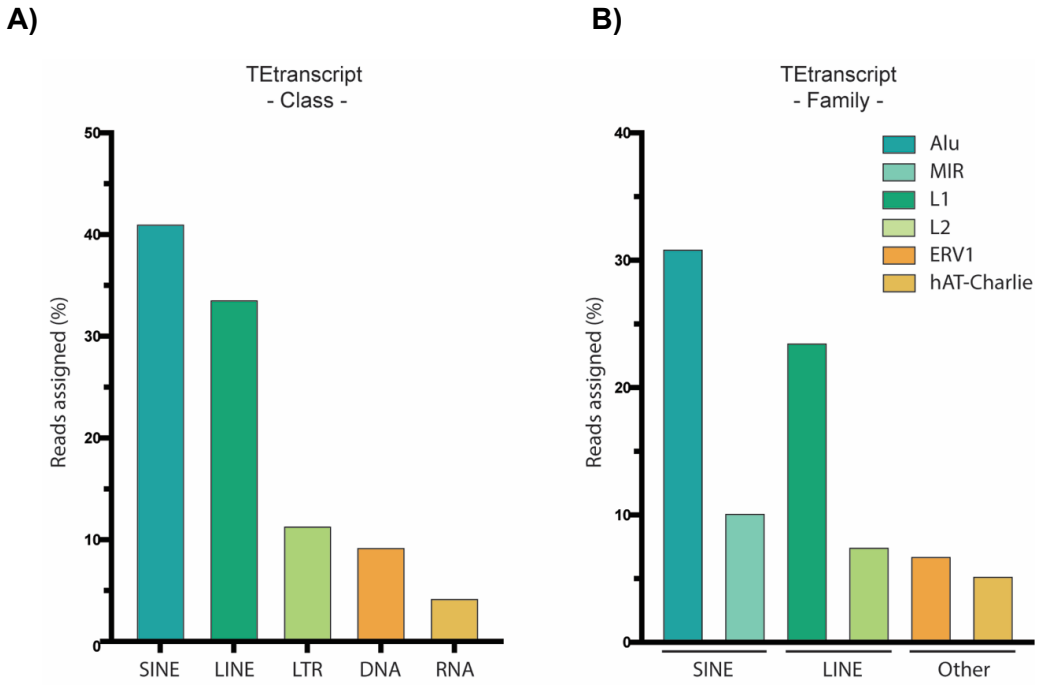


Figure 31 Distribution of significant DHX40 CRAC reads mapping to TE classes and families. *DHX40 CRAC sequencing reads were assigned to TE classes (A) or families (B) based on the GTF file from TEtranscript.* ***A)*** *104066 normalised reads mapped to different TE classes, including SINEs (41.07%), LINEs (33.53%), LTRs (11.39%), DNA repeat elements (9.27%), RNA repeat elements (4.27%), and satellite repeats (0.36%, not shown).* ***B)*** *104066 normalised reads mapped to different TE families. TE families with >5% assigned reads are shown and grouped by TE classes. Most reads mapped to Alu elements (30.91%), L1s (23.54%), mammalian-wide interspersed repeats (MIRs, 10.16%), or L2s (7.5%). The number of normalised reads mapped to each TE is available in Supplementary table 15-16. The CRAC bioinformatic analysis was performed in collaboration with Nicolás Lemus, MD. PhD*

To gain insight into DHX40 putative association with RNAs derived from TEs, reads were assigned to different TE classes and families based on the annotation by TEtranscript (Jin et al., 2015) (Figure 31). After normalisation, 10368 reads were assigned to TE classes (Figure 31A), and 104066 were assigned to TE families (Figure 31B).

Most reads mapping to TE classes mapped either to SINE (~41%) or LINE (~34%) derived sequences, with a few mapping to either LTRs or other RNA or DNA TE-derived sequences. SINEs and LINEs belong to non-LTR TEs, and are, unlike LTR elements, still active in humans. Although defective, TE-derived sequences are often retained in the genome and can be actively transcribed, thus accounting for DHX40 binding to inactive TEs. Most TE family reads mapped to Alu sequences (~31%), belonging to SINEs, or L1 sequences (~24%), belonging to LINEs. Although reads mapping Alu element-derived sequences were more abundant in the DHX40 CRAC dataset, priority in the subsequent analysis was given to L1-derived reads as Alu TE elements are non-autonomous and require L1-encoded proteins.

To gain insight into the DHX40 association with L1 TEs, the distribution of reads from the significant and specific peaks in the DHX40 CRAC dataset along the canonical L1 transposable element was visualised (Figure 32). As illustrated by the heatmap and histogram, reads cluster at three locations within ORF2, indicating that DHX40 is more likely to crosslink to these sites within the L1-derived RNAs. Interestingly, it has been observed that many L1 suppressors bind within ORF1 and ORF2 (Attig et al., 2018). Moreover, as illustrated by the heatmap (Figure 32), DHX40 crosslink more to L1 RNAs derived from younger subfamilies, e.g. L1PA1 and L1PA2, than to older subfamilies. Additionally, DHX40 bind to the 5'UTR of a few L1 RNAs, including those of the active subfamily L1PA1, indicating that these are potentially full-length elements and are possibly retrotransposition competent.

Next, to explore a potential function of DHX40 in regulating L1 retrotransposition, total RNAs from biological duplicates of ΔDHX40-1 cells, DHX40 knockdown cells, and corresponding controls were subjected to RNA-seq. Acquired reads were mapped to the GRCh38 human genome allowing multi-mapping, and differential expression analysis was performed using TEtranscript (Jin et al., 2015).

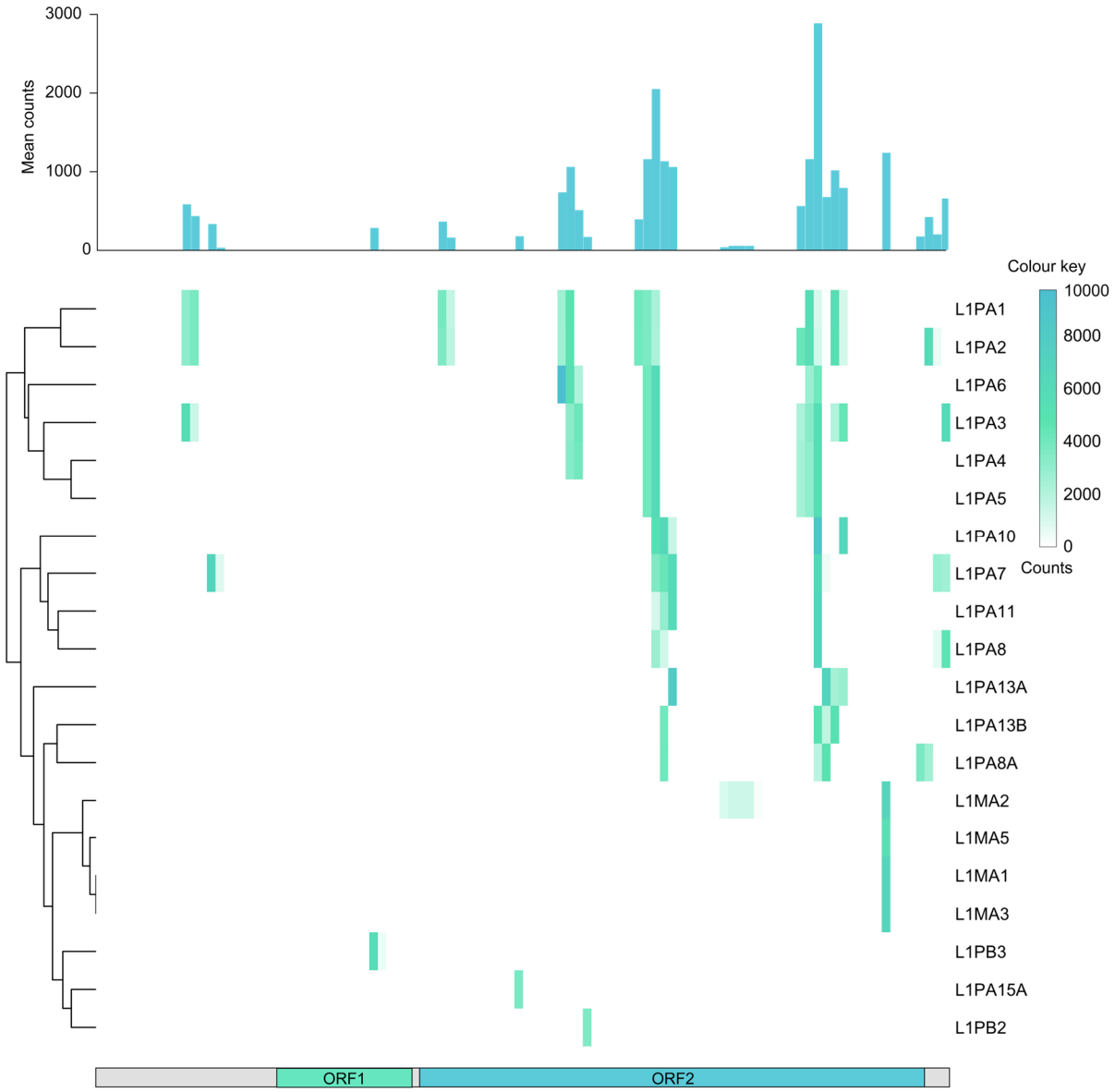


Figure 32 Distribution of reads from the DHX40 CRAC dataset along the canonical L1 RNA intermediate. *Reads from the DHX40 CRAC dataset, mapping to L1-derived sequences, were mapped to the canonical L1 RNA. The heatmap displays the distribution of reads within L1s of different subfamilies. The histogram displays the condensed data of the heatmap. The CRAC bioinformatic analysis was performed in collaboration with Nicolás Lemus, MD PhD*

The differential expression analysis confirmed that DHX40 mRNA levels were reduced in both knockdown and knockout samples (Figure 33A). Furthermore, reads mapping to DHX40 mRNA in knockout samples mapped to the 5'UTR, upstream of the frameshift mutation; the number of DHX40-mapped reads in the knockout sample was, therefore, not indicative of the expression of full-length DHX40. In addition, several protein-coding RNAs were significantly up- or down-regulated following DHX40 deletion/depletion (Figure 33A); However, gene ontology (GO) enrichment analysis for biological processes, involving the proteins encoded by

the differentially expressed transcripts, did not reveal a consistent pattern between chronic and acute depletion of DHX40 (Supplementary figure 1-4). In addition, differential expression was observed for several other RNAs mapped to different biotypes, including L1s (Figure 33B), lncRNAs, and LTRs (Supplementary figure 5).

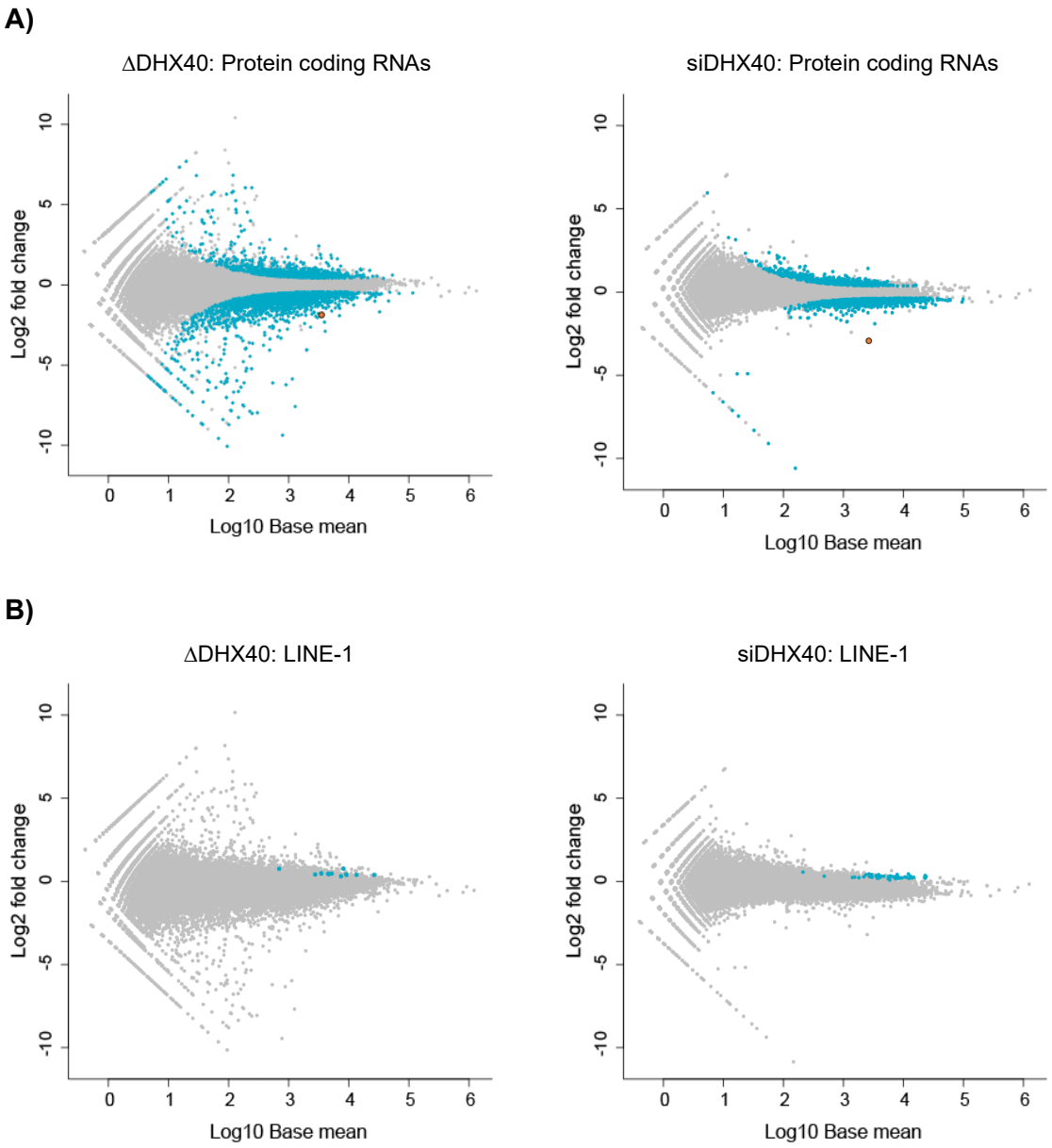


Figure 33 MA plots for differential expression analysis of RNA-seq of DHX40 knockout and knockdown samples. *The plots represent the log2 fold change (FC) versus normalised log10 base mean between DHX40 depleted samples and control. A) MA plots for differentially expressed protein-coding RNAs. Significantly (adjusted p-value (padj) < 0.05) upregulated (FC>2) or downregulated (FC<-2) protein-coding RNAs are marked in blue. DHX40 mRNA is marked in orange.* ***B)*** *MA plots for differentially expressed L1 RNAs. Significantly (padj < 0.05) upregulated (FC>2) or downregulated (FC<-2) L1 RNAs are marked in blue. The complete dataset is available in Supplementary figure 5. The bioinformatic analysis was performed in collaboration with Nicolás Lemus, MD PhD*

L1 RNAs were significantly upregulated in DHX40 knockdown and knockout cells (Figure 33B). Despite the higher number of total reads in the knockout sample, more L1 RNAs were significantly upregulated in the knockdown samples than in the knockout sample. The difference in relative distribution between L1 and total reads in the knockout versus knockdown samples suggests that the knockdown cells have adapted to the DHX40 deletion and limit L1 retrotransposition by alternative means. The fold change of L1 RNAs in the depletion/deletion sample versus control was largely unaffected by the means or duration of DHX40 depletion, i.e. chronic and acute. This result is within expectation for HEK293 cells as L1 mRNA expression levels depend primarily on cell type (B. Freeman et al., 2022).

3.6 Functional analysis of DHX40 in regulating L1 retrotransposition

The regulation of L1 activity is complex and not completely understood, although several proteins have been observed to affect L1 retrotransposition (Protasova et al., 2021). Regulators of L1 retrotransposition are generally divided into activators and suppressors (N. Liu et al., 2017). Activators are proteins hijacked by the L1 retrotransposon to facilitate or promote retrotransposition, whereas suppressors limit retrotransposition, thus reducing the threat to genome stability (N. Liu et al., 2017). Several L1 regulators were identified as DHX40 interactors by LC-MS/MS, including PPIL4, HNRNPM, XRCC5, and XRCC6. The interactions of DHX40 with the beforementioned proteins were validated by repeating the anti-FLAG immunoprecipitations and analysing the recovered proteins by western blotting (Figure 34). As previously demonstrated (Figure 23), PPIL4 is highly enriched in the immunoprecipitation of DHX40 compared to controls.

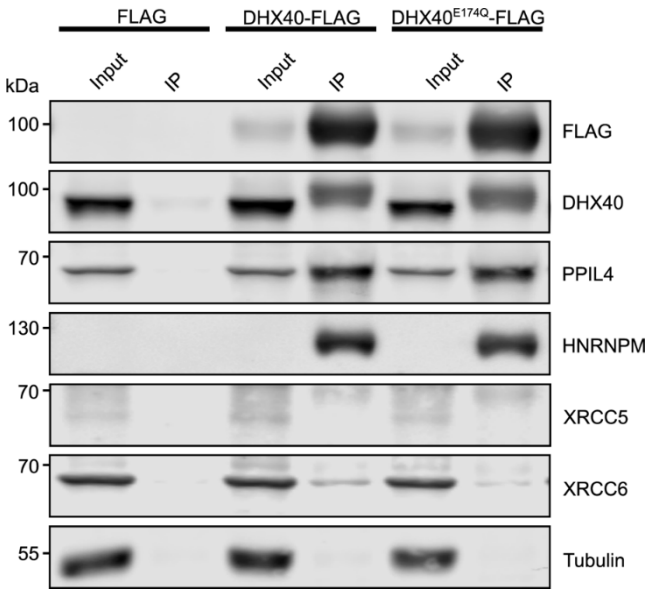


Figure 34 Association of L1 regulators with DHX40 detected by immunoprecipitation and western blotting. *Anti-FLAG immunoprecipitations (IPs) were performed with ΔDHX40-1 rescue systems expressing either DHX40-His_6Prc2xFLAG (DHX40-FLAG) or DHX40_E174Q-His_6Prc2xFLAG ($DHX40^{E174Q}$-FLAG) at close to endogenous levels of DHX40. FLAG-tagged proteins were immunoprecipitated with anti-FLAG magnetic beads, and recovered proteins and inputs (0.4%) were analysed by western blotting using antibodies against the indicated proteins. Tubulin was used as a control.*

The heterogeneous nuclear ribonucleoprotein (hnRNP), HNRNPM, is similar to PPIL4, specifically enriched in immunoprecipitations versus control. HNRNPM has been shown to bind strongly to antisense L1s but has not been confirmed to repress L1 retrotransposition (van Nostrand et al., 2020). However, several hnRNPs have been reported to participate in L1 inhibition (Protasova et al., 2021). XRCC6 and XRCC5 are activators of L1 retrotransposition, although they have also been suggested to be partially responsible for abortive L1 retrotransposition leading to 5' truncated progeny (N. Liu et al., 2017; Protasova et al., 2021). XRCC6 could readily be confirmed as an interactor with DHX40; however, the western blotting results for XRCC5 were ambiguous (Figure 34).

Next, to determine whether DHX40 is a regulator of L1 retrotransposition, the effects of DHX40 knockdown on L1 retrotransposition were evaluated using a cell-based L1-reporter assay (Figure 35). Knockdown of DHX40 was prioritised over knockout, as according to the RNA-seq, the knockdown caused upregulation of a greater number of L1-derived RNAs (relative to total assigned reads) than the knockout. Furthermore, several proteins have been identified as L1 regulators by measuring the effect of their knockdown in a similar retrotransposition assay, thus providing a good selection of appropriate controls for the assay.

In the assay, HEK293 cells were transiently transfected with a plasmid encoding a retrotransposition-competent L1 with a reporter gene and a puromycin resistance cassette.

The reporter gene consisted of an antisense enhanced green fluorescent protein (EGFP) gene interrupted by a γ-globin intron in the opposite orientation to the EGFP gene (Farkash et al., 2006; Ostertag et al., 2000). Expression of the EGFP gene only occurs after the removal of the intron during successful retrotransposition.

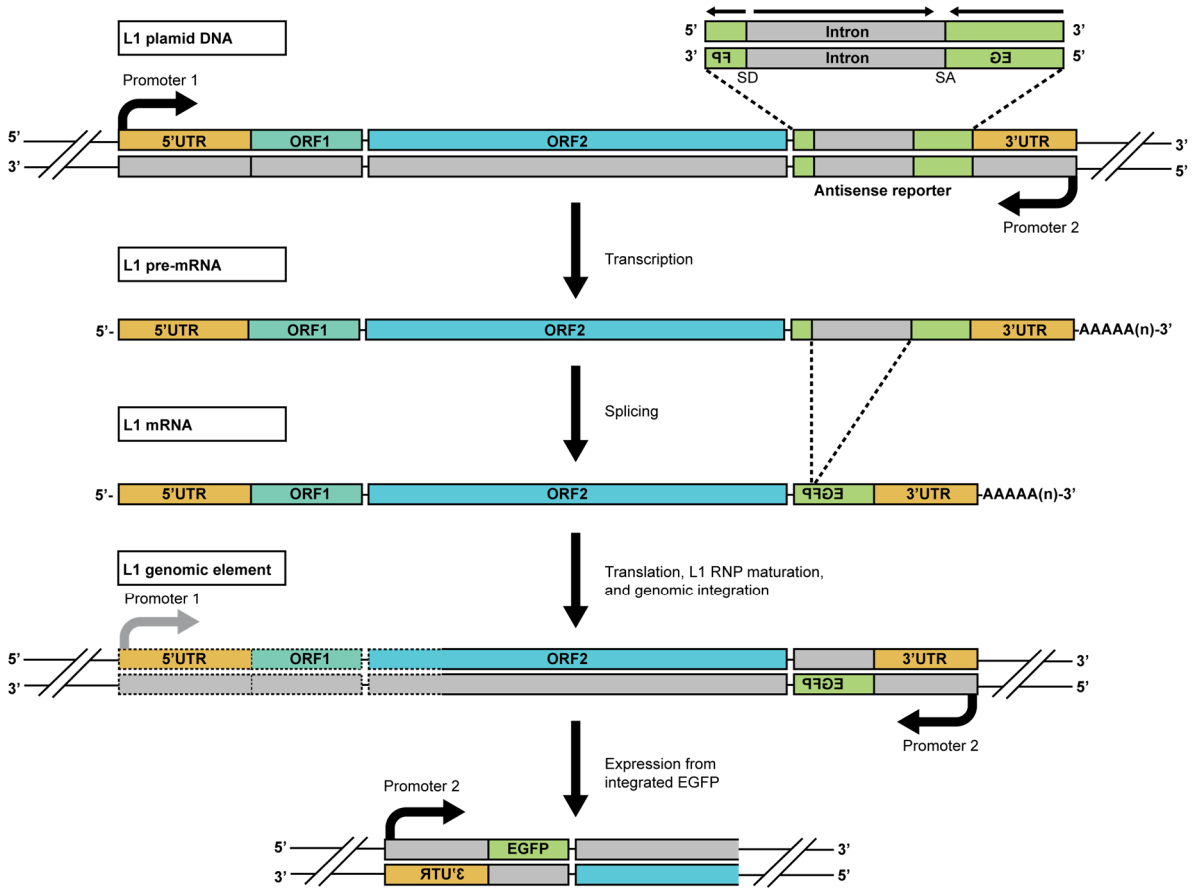


Figure 35 Schematic overview of the cell-based L1 retrotransposition assay. *The L1 reporter plasmid encodes a retrotransposition-competent L1 with an antisense EGFP reporter gene in the 3' UTR. The EGFP reporter gene is interrupted by a γ-globin intron in the opposite orientation to the EGFP gene. Transcription of the L1 is initiated by the endogenous promoter (promoter 1) in the 5'UTR, and the intron is excised by pre-mRNA splicing. ORF1 and ORF2 are translated from the L1 mRNA, and the L1 RNP is assembled from L1-encoded proteins and host proteins. The L1 sequence is integrated into the host genome through L1 retrotransposition. This process is often abortive, leading to 5' truncation (indicated by dashed outline) and retrotransposition of incompetent progeny. Successful retrotransposition leads to the integration of spliced EGFP gene into the host gene and subsequent expression from the reporter's CMV promoter (promoter 2).*

METTL3 and ALKBH5 have recently been reported to regulate L1 retrotransposition through regulating m^{6}A modification of the L1 5'UTR (Hwang et al., 2021). In this context, METTL3 methylates the 5'UTR leading to increased translation efficiency of the L1 mRNA; ALKBH5 is an antagonist to METTL3 function and removes the m^{6}A modification, thus decreasing L1 translation. METTL3 and ALKBH5 are, therefore, good controls for the cell-based L1

retrotransposition assay as the knockdown of these proteins decrease and increase retrotransposition, respectively (Hwang et al., 2021).

In preparation for the assay, RNAi-mediated depletion of DHX40, PPIL4, METTL3, and ALKBH5 was established in similar culture conditions as the assay. The knockdowns were confirmed by analysing cell extracts by SDS-PAGE and western blotting (Figure 36A). In addition, the detection of ORF1p by immunoblotting was attempted to determine if depletion of any of these factors affected L1-encoded protein levels. ORF1p was, however, not expressed at detectable levels in HEK293 cells. ORF1p expression varies between cell lines, and MCF7 cells display higher ORF1p expression than HEK293 (Philippe et al., 2016). Therefore, the effect of DHX40 and PPIL4 depletion on ORF1p expression was determined in MCF7 cells (Figure 36B). Differences in ORF1p expression after DHX40 and PPIL4 depletion were not detectable by western blotting.

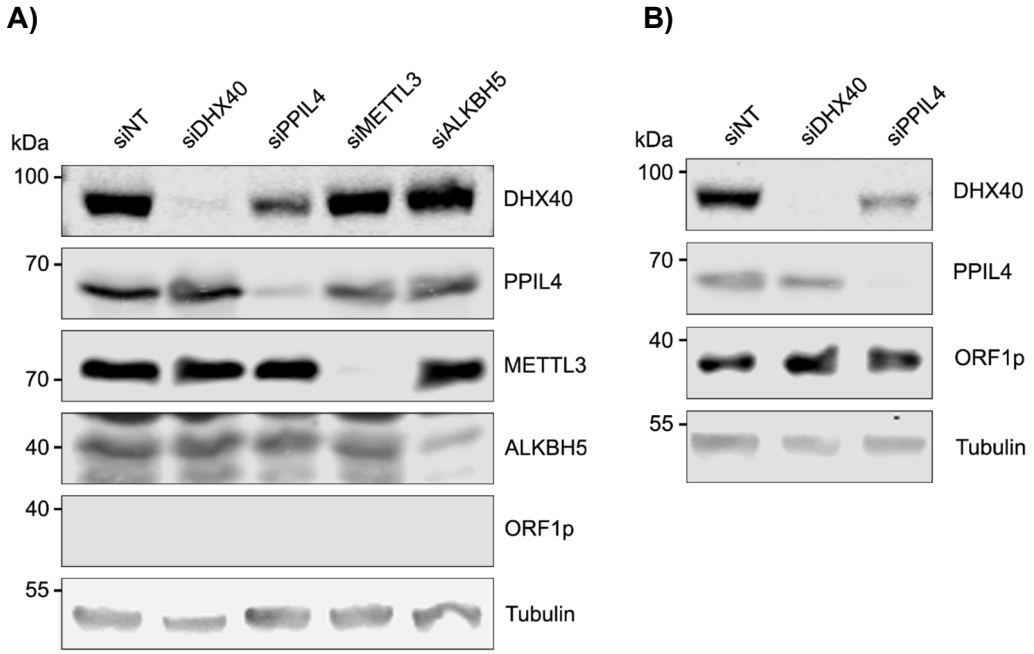


Figure 36 Establishing knockdown conditions in HEK293 and MCF7 cells. *A) HEK293 cells were reverse transfected with siNT, siDHX40-1, siPPIL4, siMETTL3, or siALKBH5 for 72 hours. Knockdowns were confirmed by western blotting using the indicated antibodies. In addition, an antibody against ORF1p was used to detect the expression of L1-encoded proteins in the host cell line. B) MCF7 cells were reverse transfected with siNT, siDHX40-1, or siPPIL4 for 72 hours. The knockdowns and their effect on ORF1p expression in MCF7 cells were detected by western blotting using antibodies against indicated proteins.*

In the cell-based retrotransposition assay, HEK293 cells were transiently transfected with the L1 reporter plasmid for 24 hours, after which the cells were left untreated or transfected with siNT or siRNAs targeting mRNAs of the described proteins. Then, 72 hours after transfection, cells were selected with puromycin for 72 hours to eliminate untransfected cells. Finally,

relative retrotransposition frequency was measured by Fluorescence-activated Cell Sorting (FACS) (Figure 37).

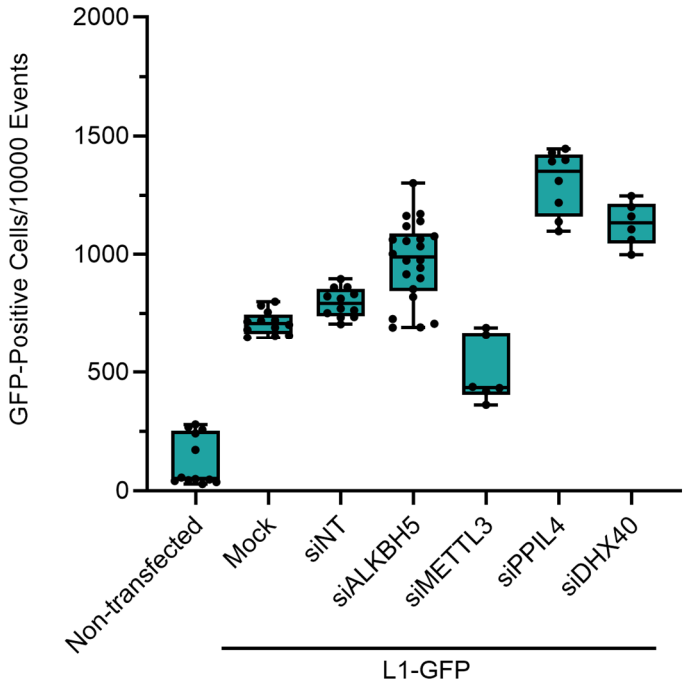


Figure 37 Effect of knockdowns on retrotransposition frequency estimated by L1 retrotransposition assay. *HEK293 cells were transfected with an L1 reporter plasmid (L1-GFP) and siRNAs, as indicated. After the elimination of untransfected cells by puromycin selection, the expression of GFP was measured by FACS. The number of GFP-expressing cells per 10,000 cells was used to measure retrotransposition frequency. The plot represents the results from three or more biological replicates per sample and technical duplicates per biological replicate. Numbers and statistical analysis are available in Supplementary table 17-19.*

For cells not transfected with the reporter plasmid, <250 / 10,000 cells were positive, i.e. they had a green fluorescent signal above the threshold value due to autofluorescence. Cells transfected with the L1 reporter but not siRNAs (Mock) had a relative frequency of ~700 positive cells per 10,000 cells, whereas cells transfected with siRNA had a relative frequency of ~800 /10,000 positive cells, indicating that the siRNA transfection procedure slightly, but not significantly, increased the retrotransposition frequency. As expected, the depletion of ALKBH5 and METTL3 increased (~1,000/10,000) and decreased (~500/10,000) the relative retrotransposition, respectively (Hwang et al., 2021) (Figure 37). Depletion of PPIL4 and DHX40 significantly increased relative L1 retrotransposition frequency with ~1100 positive cells and ~1300 positive cells per 10,000 recorded cells, respectively. The retrotransposition assay, therefore, confirmed previous observations that PPIL4 plays a role in regulating retrotransposition (van Nostrand et al., 2020) and established DHX40 as an L1 suppressor

4. Discussion

RNA helicases comprise a group of RNA-binding proteins with shared structural features and a wide array of functions in RNA remodelling, making them a key driving force in almost all aspects of RNA metabolism. Therefore, strict spatial, temporal, and catalytic regulation of RNA helicases is essential. However, the helicase core of SF2 helicases generally lacks intrinsic specificity, necessitating dedicated strategies to limit their promiscuity and ensure their target specificity (Sloan & Bohnsack, 2018). DHX proteins are a diverse group of SF2 RNA helicases participating in a multitude of cellular processes as RNA/RNP remodellers. Notably, some DHX proteins demonstrate a propensity for multifunctionality and thus require additional layers of regulation to ensure their appropriate distribution between different cellular processes (K. E. Bohnsack et al., 2021; Sloan & Bohnsack, 2018). However, the functions and modes of regulation of most DHX proteins require further investigation. Therefore, the objective of this study was to broaden our understanding of the functional and regulatory repertoires of DHX proteins.

In this context, I first established a toolbox for exploring DHX protein functions in a cellular context, including verified siRNAs, stably transfected cell lines for inducible expression of tagged DHX proteins, and tested antibodies. The toolbox for DHX40, a putative helicase of unknown function, was expanded with CRISPR/Cas9 knockout cell lines, CRISPR/Cas9-based rescue systems, constructs for recombinant protein expression, and stably transfected cell lines for inducible expression of DHX40 variants. The expanded toolbox allowed new insights into the function of DHX40. The generated data confirmed that DHX40 is a substrate of the deubiquitinase USP7 and revealed that it binds to the PPIL4, a putative prolyl isomerase, TRIM27, an E3 ligase, and several L1 regulators. *In vitro* experiments showed that DHX40 is an RNA-dependent ATPase stimulated by PPIL4 and that its ATPase activity is inhibited by an amino acid substitution in motif II in the helicase core. The role of DHX40 in RNA metabolism was investigated by CRAC and RNA-seq, which revealed that DHX40 binds to L1 RNAs and regulates their abundance in HEK293 cells, respectively. Finally, a cell-based L1 retrotransposition assay showed that PPIL4 and DHX40 are suppressors of L1 retrotransposition.

4.1 Toolbox for functional characterisation of DHX protein

Deciphering the functions of a protein requires various techniques and tools to investigate subcellular localisation, interactions (protein, RNA, DNA, ligands), and potential substrates. In this study, I first sought to establish a toolbox for the functional characterisation of DHX proteins. This toolbox served as an initial screen of available tools and established a good

foundation for further investigating DHX proteins. In the screen, I reviewed commercially available resources, i.e. siRNAs and antibodies, and developed protein-specific tools for characterisation. A promising target for further characterisation (DHX40) was then selected based on the quality of the available tools and gaps in the current literature.

4.1.1 Ectopic expression of DHX proteins

Ectopic expression is widely used to investigate protein functions in cells, and a regulatable system allowing the expression of a transgene sequence was an essential component of our toolbox. The ectopic expression of proteins can be achieved by transfection, i.e. the introduction of foreign nucleic acids into cells, a powerful cell culture technique that enables the study of gene products and their functions in eukaryotic cells. Generally, transfections can be classified into two types, namely stable and transient transfections (Chong et al., 2021). Stable transfection refers to the integration of foreign DNA into the host genome, thus allowing the host system to sustain the long-term expression of a transgene. In contrast, transiently transfected nucleic acids are only present in the cell for a limited time and will eventually be lost, as they are not integrated into the genome (Chong et al., 2021; Fus-Kujawa et al., 2021).

Transient transfection is usually applied in short-term studies to investigate the effect of the knock-in/knockdown of a particular gene (Chong et al., 2021). Transient transfection can be very efficient, making it an excellent technique for small-scale experiments that do not require high biomass (Fus-Kujawa et al., 2021). However, transfection can induce cellular stress, and transfected cells may not fully recover within the time window of transgene expression. In contrast, stable transfection is more labour-intensive and time-consuming but does not require repeated transfection once a cell line has been generated. Stable transfection furthermore results in a more reproducible expression of the transgene than transient transfection; stably transfected cells can be selected, e.g. by antibiotics, whereas the transiently transfected cell population is heterogeneous as a result of incomplete transfection efficiency. Finally, stably transfected cell lines are better suited for large-scale experiments, as up-scaling does not require optimisation of the transfection procedure, and large-scale cultures do not require costly transfection agents (Fus-Kujawa et al., 2021). For this study, the ideal approach for ectopic expression of proteins was, therefore, stable transfection.

Multiple methods for generating stably transfected cells exist, with the most widely used being integration by lentiviruses, referred to as viral transduction, and chemical transfection of cells engineered to include a recombinase-dependent gene integration system with a compatible vector (Szczesny et al., 2018). Viral transduction is widely recognised as a highly effective method for transfecting difficult-to-transfect cells; however, it is associated with higher cytotoxicity than non-viral transfection and poses a risk of viral infection (Chong et al., 2021). Furthermore, viral transduction using lentiviruses integrates the transgene randomly into the

host cell genome, often at multiple locations, and is therefore associated with a high risk of insertional mutagenesis and gene disruption (Tong & Yin, 2021). Furthermore, because viral transduction generates a polyclonal population with different phenotypes, the generation of stably transfected cell lines requires clonal selection, monitoring of cell lines, and cell line phenotyping. Chemical transfection does not require specialised labs and equipment. However, this approach has a low transfection efficiency compared to viral transduction and requires parental cell lines engineered to improve transgene integration (Szczesny et al., 2018). The ideal transfection approach should be selected depending on the cell type and experimental needs. Applied methods should aim to have high transfection efficiency, low cytotoxicity, minimal side effects on cell morphology, ease of use, and, most importantly, high reproducibility.

In this study, I generated eleven stably transfected cell lines for the expression of C-terminally His_6Prc2xFLAG-tagged variants of DHX8, DHX9, DHX16, DHX29, DHX30, DHX32, DHX33, DHX36, DHX38, DHX40, and DHX58 using FLP recombinase-based methodology. The genome of the pre-engineered cell line contains an FLP recombinase target (FRT) sequence recognised by yeast FLP recombinase (Jensen et al., 2020; Karimi et al., 2007; Szczesny et al., 2018). Co-transfection with a compatible vector containing the transgene and complementary FRT sites and a construct encoding the recombinase ensures that the transgene is only integrated at a specific locus as the parental cell line is isogenic regarding the FTR sites; this methodology thus prevents gene disruption and makes clonal selection unnecessary (Jensen et al., 2020; Karimi et al., 2007; Szczesny et al., 2018). Stably transfected cell lines generally display constant expression of the protein of interest at levels depending on the number of integration sites, integration locus, or promoter. This, in turn, puts selection pressure on the transfected cells to adapt to the expression of the transgene, thereby affecting the analyses of the cell line. Inducible transgene expression prevents this problem by allowing control over the expression level, time, and onset. The inducible promoter used in this study was the T-REx promoter, in which constitutive repression of transgene transcription is alleviated by tetracycline (Szczesny et al., 2018). Induction allows for the determination of the effects of different transgene expression levels in a cellular context and can reduce unintentional effects of overexpression, which can otherwise affect protein-protein interactions, substrate interactions, subcellular localisation, and the cell's metabolic state.

Another factor to consider is the induction time, as depending on the translation efficiency, PTMs, localisation, and integration in macromolecular complexes and cellular pathways, this might differ between not only transgene-encoded proteins but also applications of the same cell line. Longer induction times might be required if the proteins function in the context of a macromolecular complex. For example, DHX15 and DHX37 are known to function in ribosome

biogenesis; it would therefore be important to induce transgene expression of the tagged proteins for long enough so that they are not only expressed but also in the functional context required for the investigation (K. E. Bohnsack et al., 2022; Boneberg et al., 2019; Choudhury et al., 2019). In contrast, the DHX40 protein level appears to be strongly controlled; it might therefore be appropriate to induce expression for only a short time to avoid triggering proteasomal degradation or mechanisms that limit expression.

Ectopic transgene expression also gives control over the expressed sequence, allowing for functional analyses of different variants of the same gene, including different isoforms, diseased variants, catalytically inactive enzymes, or proteins lacking one or more domains. All stably transfected cells in the toolbox – excluding the expanded elements of the DHX40 toolbox – used the longest expressed isoform without any changes to the protein sequence other than the addition of a fusion epitope tag. The expression of a fusion tag can aid subsequent analyses and should be selected depending on the planned applications. Epitope-tagging with short peptides allows for the detection and identification of transgene-encoded proteins and are, owing to their small size, less likely to disrupt the function and localisation of the protein (Palmer & Freeman, 2004). As good-quality antibodies targeting the protein of interest are sometimes unavailable, epitope tagging may be necessary if recognition by antibodies is required. Antibodies against common epitope tags are commercially available and are often reliable, enabling immunoblotting, immunofluorescence, and affinity purification. Additional options include fluorescent or enzymatic tags, e.g. fluorescent proteins and biotin ligases, which can aid in determining localisation, cellular pathways, proximal proteins, and expression levels (Palmer & Freeman, 2004; Szczesny et al., 2018).

Another consideration is fusion tag placement, as the tag can interfere with protein localisation, lead to misfolding of terminal protein domains, or interfere with transmembrane regions. Especially larger N-terminal tags, such as fluorescent proteins, can disrupt protein localisation, exemplified by the mislocalisation of ~50% of N-terminally GFP-tagged proteins in HEK293T (Palmer & Freeman, 2004). Therefore, despite the relatively small size of the His_6Prc2xFLAG tag, all DHX proteins were C-terminally tagged to avoid mislocalisation. Moreover, as there are several examples of C-terminally tagged DEAH/RHA helicases that display catalytic activity in cells, it was unlikely that the C-terminal tag would affect the structure or function of the conserved C-terminal domains of the DEAH/RHA helicases in this study (Choudhury et al., 2019; Lessel et al., 2017; Sweeney et al., 2021). Additionally, the N-termini of some DEAH/RHA helicases contains sequences or structures necessary for their function and localisation (Tauchert et al., 2017). N-terminal tagging of these proteins would therefore risk disrupting their structure or function, as seen for DHX30, where N-terminal tagging with GFP prevents import into the mitochondrial and sequesters the protein in the cytosol (Lessel et al.,

2017). A C-terminal tag furthermore has the advantage of only tagging fully translated transgenes products.

His_6Prc2xFLAG epitope tag used in this study was designed specifically for the CRAC protocol, as this double epitope, in addition to benefits conferred by epitope tags in general, allowed tandem affinity purification of protein-RNA complexes. The 2xFLAG moiety enables anti-FLAG immunoprecipitation followed by either competitive elution with 3xFLAG peptides or proteolytic cleavage of the Prc site, whereas the His_6 moiety enables native and denaturing affinity purification by IMAC with either pH-dependent elution or elution by titration with imidazole (Block et al., 2009).

The stably transfected cell lines used in this study were derived from the HEK293 parental cell line; this well-established human embryonic kidney cell line is easy to maintain in culture, has a high growth rate, and is widely used because it is easy to transfect (Thomas & Smart, 2005). It was, therefore, a good initial choice for the generation of toolbox cell lines. However, working in other cell lines might have advantages depending on the application or the investigated gene. For example, for L1 retrotransposition assays that became relevant when characterising DHX40, it could be preferable to use the MCF7 cell line, which has been shown to have the highest expression of L1 mRNA amongst commonly used human cell lines (B. Freeman et al., 2022; Philippe et al., 2016). This cell line is, therefore, widely used in studies on L1 retrotransposition. Another important consideration when choosing the model cell line for the characterisation of proteins is the expression of the studied and auxiliary proteins. For example, DHX34 and DHX57 cell lines were not represented in this study as I could not amplify the genes from HEK293 cDNA. DHX34 and DHX57 have low expression in kidneys but are enriched in testis; it might be better to study these proteins in the Tera-1 cell line as they might have tissue-specific functions and interaction partners (Uhlén et al., 2015).

4.1.1 Depletion of DHX protein in HEK293 cells

One of the most important approaches in functional studies is the downregulation or deactivation of genes so that their respective products are depleted. This can be achieved by several experimental strategies that differ in downregulation efficiency and longevity. In this study, both CRISPR/Cas9-mediated knockout and RNAi-mediated knockdowns were utilised. The initial toolbox includes two verified siRNAs for each of the ten DHX proteins. All siRNAs for RNAi-mediated knockdowns, except siRNAs targeting DHX38 and DHX33, had an excellent knockdown efficiency. However, there was a discrepancy in the knockdown efficiency when estimated for DHX38 and DHX33 with qPCR versus western blotting. This discrepancy could be due to the siRNAs affecting the internal normalisation controls or may reflect differences in mRNA and protein turnover rates.

The development of CRISPR/Cas9 methodology has opened many possibilities, as complete inactivation of a gene by deletion or insertional inactivation has a strong advantage in many cases compared to downregulation; this method is, however, restricted to non-essential genes (Ran et al., 2013; Szczesny et al., 2018). Additionally, CRISPR/Cas9 knockout cell lines may adapt to the lack of the protein of interest, reducing or modifying the phenotype otherwise observed upon acute downregulation. Finally, some gene disruptions may be hazardous or oncogenic, resulting in accumulated genomic changes to the cell lines. Alternatively, a variation of the genome editing system, CRISP/Cas9 mediated homology-directed repair (CRISPR/Cas9-HDR), can be used to fuse the endogenous protein of interest with auxin-inducible destabilising domain, also called a degron (S. Li et al., 2019). The system requires co-transfection of a CRISP/Cas9 plasmid with appropriate sgRNAs and a donor plasmid with the degron coding sequence flanked by homology arms. CRISPR/Cas9-HDR uses a sgRNA specific to a protospacer adjacent motif (PAM) at the 3' or 5' of the coding sequence to induce a double-stranded break (Ran et al., 2013; J. P. Zhang et al., 2017). The double-stranded break is then repaired by homology-directed repair (HDR) using the sequence from the donor plasmid, thus adding the degron sequence to the protein (J. P. Zhang et al., 2017). Exogenously expressed F-box protein TIR1, in a chimeric SKP1-CUL1-F-Box ubiquitin E3 ligase, then targets the degron-tagged protein for proteasomal degradation upon the addition of the plant hormone auxin (X. Li et al., 2022). This method has the advantage of rapidly targeting protein for degradation; this method is, therefore, suitable for proteins with otherwise long half-lives.

Another widely used technique for protein downregulated is RNA interference (RNAi), which involves transfection with short RNAs complementary to the target mRNA, thus repressing its translation or targeting it for degradation (Chong et al., 2021). This method is advantageous if targeting specific isoforms is required, but it can have off-target effects. The most straightforward and widely used strategy is to transiently transfect cells with target-specific siRNA, but similar to transfection for transgene-encoded protein expression, this can also be performed by stable transfection. For inducible downregulation of target genes, the Flp-In T-Rex system can, similarly to transgene-encoded protein expression, be used to integrate a gene encoding a short hairpin RNA (shRNA) under an inducible promotor into the genome (Chong et al., 2021; Fus-Kujawa et al., 2021). This methodology has the same advantages as the stable expression discussed above, i.e. reproducibility and easy up-scaling at the cost of time-consuming generation of cell lines.

Induction of DHX40-His6Prc2xFLAG in the HEK293 cell line suggested that DHX40 is regulated at the protein level since the transgene proved difficult to overexpress more than 2-3 fold and the endogenous protein and the transgene-encoded protein were present at close

to equal amounts. Protein level regulation by degradation removes proteins from their functional interactions and increases their interaction with proteins responsible for their degradation. A widely used approach to prevent the undesired effects of overexpression is using a rescue system. Rescue systems combine the downregulation/inactivation of endogenous genes and ectopic expression of a replacement transgene encoding a tagged or variant version of the protein of interest. Rescue systems can furthermore aid in deciphering the requirements for a gene function in a cellular context by testing the ability of different transgene variants to rescue a downregulation phenotype. Endogenous DHX40 inactivated by CRISPR/Cas9 was therefore replaced by DHX40-His_6Prc2xFLAG and $DHX40^{E174Q}$-His_6Prc2xFLAG to prevent the undesired effects of overexpression from influencing the immunoprecipitation results, while simultaneously investigating the requirement of ATPase activity.

In the course of this study, it was revealed that DHX40 is involved in cellular processes related to genome stability, i.e. L1 retrotransposition. Therefore, it is possible that cell lines with CRISPR/Cas9 mediated inactivation of DHX40 may either have adapted to the absence of DHX40 and accumulated genomic mutations due to upregulated L1 retrotransposition. In this case, relying on RNAi-mediated knockdowns in combination with the expression of DHX40 variants encoded by RNAi-resistant transgenes might be preferable. Several different approaches can realise such knockdown rescue systems. One approach would be to generate stably transfected cell lines with inducible expression of DHX40 RNAi-resistant transgenes and transiently transfect these with siRNAs targeting the endogenous DHX40 mRNA. Another approach would be to generate cell lines by stable transfection using a plasmid with an inducible bi-directional promoter that allows the simultaneous expression of two genes encoding an shRNA and the complementing RNAi-insensitive transgene, respectively. This approach would enable inducible downregulation of the endogenous DHX40 and concomitant expression of the transgene-encoded DHX40 variants (Szczesny et al., 2018).

In addition to the DHX40 toolbox, which has been extensively used and expanded, the DHX30 toolbox was used in a study presenting a method to silence mitochondrial gene expression (Cruz-Zaragoza et al., 2021); here, it was confirmed that C-terminally tagged DHX30 is imported into the mitochondria and is responsible for priming mt-mRNAs for translation (Cruz-Zaragoza et al., 2021). However, the function and regulation of DHX32 remain elusive, and the role of DHX33 in translation is still poorly understood. Future investigation into the role and regulation of these proteins could be aided by the toolbox assembled in this study.

4.2 Helicase regulation by E3 ligase

The covalent attachment of ubiquitin is a prominent post-translational modification regulating protein activity, stability, interactions, and localisation and controlling intercellular signalling (Komander, 2009). The attachment of the C-terminal glycine of ubiquitin to the target lysines of substrate proteins is mediated by a three-enzyme ubiquitin transfer cascade consisting of an E1 activation enzyme, an E2 conjugating enzyme, and an E3 ligase, which facilitates the ubiquitin transfer to target lysines, thus dictating substrate specificity (Sun & Zhang, 2022; Williams et al., 2019). Ubiquitination is reversible, and ubiquitin molecules are removed from substrates by a family of proteins named deubiquitinases (Sun & Zhang, 2022). Cellular (de-)ubiquitination events are orchestrated by two E1 enzymes, approximately 40 E2 enzymes, more than 1000 E3 ligases, and approximately 100 deubiquitinases encoded by the human genome (Sun & Zhang, 2022). Additionally, ubiquitinated proteins are recognised by at least 29 highly specialised ubiquitin-binding domains (Komander, 2009; Radley et al., 2019). The ubiquitin molecule contains seven lysines and one N-terminal methionine that provide free NH_2 (amine)-groups that can be ubiquitinated, thus facilitating the multi-step formation of ubiquitin chains of different lengths and linkages. The fate of ubiquitinated proteins depends on the number and location of ubiquitinated lysines, the length of chains, linkage types, e.g. K^{11}, K^{29}, and K^{48}, and the composition of linkage types, i.e. homotypic, mixed, and branched (Komander, 2009; Sun & Zhang, 2022). Ubiquitination is, therefore, incredibly versatile due to its reversibility and the diverse repertoire of ubiquitin chains that can be assembled on the target proteins and influence their fate.

The deubiquitinase USP7 is primarily located in the nucleus and has been shown to regulate the stability of multiple proteins involved in the DNA damage response, transcription, epigenetic control, immune response, and viral transfection (Pozhidaeva & Bezsonova, 2019). USP7 consists of an N-terminal TRAF-like domain, a catalytic domain, and five C-terminal ubiquitin-like domains (ubl1-5), of which the TRAF-like domain and the ubl2 have been shown to facilitate interactions with protein targets (Pfoh et al., 2015). A previous study into targets of USP7 in gastric carcinoma cells revealed that DHX40 is a target for deubiquitination by USP7 in AGS, CNE2Z, and HCT116 cells (Georges et al., 2018). Furthermore, they showed that the effect of siUSP7 knockdown or USP7 inhibition on DHX40 protein levels differed between the investigated cell lines and that some targets for USP7 deubiquitination, e.g. PPM1G and USP11, are cell line-specific. They furthermore showed that the effect of USP7 knockdown on DHX40 protein level could be rescued by the proteasome inhibitor MG132, confirming that DHX40 is targeted for proteasomal degradation and rescued by USP7 (Georges et al., 2018).

Through immunoprecipitation of tagged DHX40 variants, I confirmed that USP7 is co-precipitated with DHX40 from HEK293 cells. I furthermore confirmed that RNAi-mediated

knockdown of USP7 decreases the protein level of DHX40, indicating that DHX40 is a substrate of USP7 deubiquitination in HEK293 cells. These results suggest that DHX40 is ubiquitinated, although the ubiquitinated sites, chain lengths and types, and responsible E3 ligase are unidentified.

According to PhosphoSitePlus® (v6.7.0.1), 18 ubiquitinated lysines have been identified in DHX40 by high-throughput studies, i.e. the sites have not yet been detected with methods other than mass spectrometry (Hornbeck et al., 2015). Ten of the identified ubiquitination sites are within RecA1, three are within RecA2, four are in the C-terminal domains, and one is assigned N-terminal to the core domains. Most of the identified ubiquitination sites have only been identified in one paper; however, two sites in RecA1 and two sites in RecA2 have been identified in three to seven high-throughput studies making these most convincing. However, further investigation is required to confirm the ubiquitination status of the lysines, for example, by substituting the individual lysines for arginine and detecting the difference in ubiquitination by western blotting using ubiquitin-specific antibodies (Emmerich & Cohen, 2015). However, this would only provide information about ubiquitination sites, not chain length or type. Further insight into the chain types present on DHX40 is not provided by the link to USP7, as USP7 deubiquitinates both mono- and poly-ubiquitinated substrates and can cleave all ubiquitin linkages except M^1-linkages (Pozhidaeva & Bezsonova, 2019). Furthermore, proteasomal degradation can be signalled by various ubiquitin patterns, with the most efficient and best studied being multiple short K^{48}-linked chains or branched structures with K^{11}- and K^{48}-linkages (Swatek & Komander, 2016). Together, this suggests that DHX40 is ubiquitinated with chains containing K^{48}-linkages but that USP7 deubiquitination is not necessarily limited to chains that target DHX40 for degradation.

Characterising protein ubiquitination is challenging, partially due to the simultaneous modification of one or several lysine residues in the target protein, which significantly increases the difficulty of localising the ubiquitination sites using traditional methods. Additionally, the dynamic changes and complexity of ubiquitin chains, which vary in length, linkage, and overall architecture, complicate the characterisation of identified ubiquitination sites (Sun & Zhang, 2022). A widely used semi-quantitative analysis of ubiquitination events is western blotting, using antibodies that target specific ubiquitin linkages. However, this method is low throughput and should optimally be performed with mutagenesis analysis of different lysines to determine the type of ubiquitin chains associated with the target and to which lysine they are covalently attached (Emmerich & Cohen, 2015; Sun & Zhang, 2022). Furthermore, detecting ubiquitin chains targeting proteins for degradations requires treatment with MG132 to prevent proteasomal degradation and, in the case of DHX40, additional treatment with the USP7-specific inhibitor AD04, could be performed to determine the ubiquitins removed by this

deubiquitinase (Emmerich & Cohen, 2015; Georges et al., 2018). Mass spectrometry-based proteomics for ubiquitination profiling plays an important role in identifying ubiquitin events and is a valuable tool in combination with conventional western blotting. Ubiquitination profiling is enabled by the enrichment of ubiquitinated substrates from the whole cell to increase the identification sensitivity of protein ubiquitination (Emmerich & Cohen, 2015; Sun & Zhang, 2022). In addition, middle-down proteomics, i.e. with restricted digestion resulting in longer peptides, can be used with linkage-specific enrichment using aptamers or antibodies to gain insight into the architecture of ubiquitin polymers attached to a specific protein. However, considering the high activity of trypsin, the digestion process in middle-down proteomics can be hard to control. Alternatively, middle-down proteomics can be performed using ubiquitin-clipping, a strategy using an engineered viral protease, Lb^{pro*}, with preferable cleavage towards all types of di-ubiquitin to quantify branch-point ubiquitin (Sun & Zhang, 2022).

Further insight into ubiquitination events on DHX40 would require the identification of E3 ligases facilitating ubiquitin transfer to DHX40. Recently a study revealed a conserved mode of interaction between the helicases core of SF2 helicases and the Pry/Spry domain of TRIM E3 ligases (Kato et al., 2021). Furthermore, the study showed that the dimerised TRIM E3 proteins bind specifically to homo-oligomers of their cognate helicase. Such homo-oligomers are arranged as an RNA-supported helicase filament in which multiple copies of the helicase are bound to a single RNA molecule forming a tightly packed RNA-helicase structure (Kato et al., 2021). The MDA5 interaction with TRIM65 and the RIG-I interaction with RIPLET resulted in avidity-dependent ubiquitination of their CARD domains, while the interaction between DHX58 and TRIM14 was avidity-dependent but ubiquitin-independent (Kato et al., 2021). Using cryo-EM and mutagenesis analysis, they identified regions of the helicase core that are required for or strengthen the interaction with the cognate TRIM proteins. To show that the mode of interaction was conserved among SF2 helicases and not specific to RLRs, they investigated the DEAD-box helicase DDX41 and showed that it interacted with TRIM41 and TRIM26 in an avidity-dependent manner using a similar epitope (Kato et al., 2021).

Through immunoprecipitation, followed by LC-MS/MS and SAINTexpress analysis of obtained spectral counts, I showed that DHX40 interacts with TRIM27, an E3 ubiquitin ligase with a Pry/Spry domain. Additionally, a previous study has shown that immunoprecipitation of TRIM27 pulls down USP7 (Cai et al., 2018). Therefore, further investigation into the configuration of this TRIM27-USP7-DHX40 complex is required, e.g. by immunoprecipitation of USP7 in HEK293 cells and DHX40 knockout cells, which would determine if DHX40 bridges the TRIM27-USP7 interaction. The intriguing finding that TRIM27 might be a DHX40 interaction partner raised the possibility that DHX40 might be regulated through a similar interaction as observed previously for other TRIM proteins. To test this possibility, the predicted structure of

the DHX40 RecA1 (DHX40^{RecA1}) and DDX41 DEAD domain (DDX41^{DEAD}, PDB:5GVR) was structurally aligned to identify regions required for TRIM27 interaction (Kato et al., 2021; Omura et al., 2016). Subsequent mutagenesis analysis showed that truncation of DHX40 N-terminal regions and disruption of a small alpha-helix with RecA1 decrease TRIM27 interaction.

According to the mutagenesis analysis, the N-terminal truncation had the greatest effect on TRIM27 binding, consistent with the structural alignment, which situated the DHX40 N-terminal region close to three separate regions in the DDX41^{DEAD} involved in binding to its cognate TRIMs. However, mutation of these three regions in DDX41^{DEAD} did not equally affect the TRIM26 $^{Pry/SPry}$ and TRIM41 $^{Pry/SPry}$ binding, evident by persistent although severely weakened TRIM26$^{Pry/Spry}$ binding to all mutants, and unaffected TRIM41 $^{Pry/SPry}$ to one of the mutants (Kato et al., 2021). These results indicate that although the avidity requirements and general binding regions are conserved, the involved structures within the binding region differ not only between different helicases but also between helicases and their different TRIMs. The truncated N-terminal region in DHX40 encompassed a small alpha-helix and three lysines listed on PhosphoSitePlus® as ubiquitination sites (K^{40}, K^{57}, and K^{61}), one of which had three high-throughput papers associated with it (K^{57}). It is currently unclear if TRIM27 ubiquitinates these lysines and if the removal or mutation of a TRIM protein's substrate lysine would affect their binding to their target protein. Investigation of the effect of amino acid substitution of these N-terminal lysines would be required to determine their role in TRIM27 binding. The alpha helix-disrupting mutation affected the TRIM27 binding less than the N-terminal truncation; however, considering the relative effect on the DHX40 sequence, this mutation greatly affected TRIM27 binding. Mutations in the equivalent region in DDX41^{DEAD} completely abolished the interaction between DDX41^{DEAD} and dimerised TRIM41$^{Pry/Spry}$ and TRIM26$^{Pry/Spry}$ in pull-down experiments using pre-purified proteins (Kato et al., 2021). DHX40 interacts with USP7 through a binding pocket in the latter's ubl2 domain, which has been shown to require a KxxxK motif for binding (Georges et al., 2018; Pfoh et al., 2015). DHX40 contain four KxxxK motifs, two in the C-terminal domains, one in RecA2, and one of which is removed with the N-terminal truncation (^{57}KQRKK^{61}). The removal of this site could potentially account for the loss of interaction between DHX40 and USP7 in the N-terminal truncated variant. However, a reciprocal experiment performed in another study showed that mutation of the ubl2 of USP7 almost completely abolished DHX40 binding (Georges et al., 2018). Furthermore, K^{57} and K^{61} are both listed as ubiquitination sites, and these PTMs would impede the binding of the ubl2 domain of USP7. The lysines in the three remaining KxxxK motifs in DHX40 are not listed as ubiquitination sites. Finally, the reduction in USP7 binding to DHX40 could potentially be caused by reduced ubiquitination of DHX40 as an effect of reduced TRIM27 interaction.

As the interactions with TRIM proteins can be ubiquitin-dependent and independent, as exemplified by DHX58 interaction with TRIM14, further investigation would be required to ascertain the nature of the interaction between DHX40 and TRIM27. A ubiquitination assay using a reconstituted ubiquitin transfer cascade can determine if a specific E3 ligase can facilitate ubiquitin transfer to a protein (Cadena et al., 2019). However, special consideration will have to be taken to account for the proposed bivalency-dependent interaction. Therefore, the ubiquitin assay should be performed in the presence of RNAs of different lengths. Together, this ubiquitin assay and the techniques for characterising ubiquitination events in cells discussed above would provide valuable insight into DHX40 regulation by ubiquitination and ubiquitin modifiers.

The proposed conserved mode of interaction between TRIM27 and DHX40 raises the question about the requirement for RNA filament formation by DHX40. The RNA-binding pattern of DHX40 on L1 RNA, revealed by CRAC, and the parallels between viral infection and retrotransposition suggest that RNA filament formation could potentially be required for signal transduction analogous to signal transduction by RIG-I and MDA5 (Cadena et al., 2019; Kato et al., 2021). In this model, the requirement for RNA-filament formation would serve as a safety mechanism and prevent unchecked signal transduction that would otherwise lead to unspecific RNA degradation. Given that at least 46% of human DNA derives from TEs, stringent regulation and multiple checkpoints would likely be required to avoid uncontrolled degradation of RNAs containing TE-derived sequences (Lander et al., 2001).

RNA-supported filament formation by DHX40 could be tested by native gel mobility shift assay of DHX40 proteins incubated with labelled RNAs of different lengths (Kato et al., 2021). This experimental approach would reveal if DHX40 forms filaments on RNAs and could provide insight into RNA substrate preferences, e.g. length and GC content. Furthermore, assuming DHX40 RNA filament formation, the bivalency requirement for TRIM interaction could be investigated by incubating the DHX40-RNA filament with dimerised TRIM27$^{Pry/Spry}$.

4.3 DEAH/RHA regulation by cofactors

4.3.1 DHX40 stimulation by PPIL4

As the helicase core of SF2 helicases interacts primarily with the sugar-phosphate backbone of substrate RNAs, it generally lacks intrinsic specificity. This non-specificity necessitates dedicated strategies to limit the promiscuity of SF2 helicases and to ensure timely activation (K. E. Bohnsack et al., 2021; Donsbach & Klostermeier, 2021). The predominant mechanism of RNA helicase regulation is interactions with cofactor proteins, which generally recruit their cognate RNA helicases to specific RNA substrates and often double as enhancers of otherwise poor helicase activity (K. E. Bohnsack et al., 2021; Hamann et al., 2020; Studer et al., 2020).

The largest class of DEAH/RHA helicase cofactors is the G-patch protein family (K. E. Bohnsack et al., 2021). Members of this diverse family of protein cofactors have a single G-patch motif embedded in an intrinsically disordered region and show a high prevalence of RNA-binding motifs and domains. Furthermore, almost half of the known G-patch proteins have additional motifs that mediate protein-protein interactions. In human cells, eight G-patch proteins are known to interact with DHX15, while DHX16 and DHX35 are regulated by one G-patch protein each (K. E. Bohnsack et al., 2021, 2022).

Structural characterisation of human DHX15 in complex with NKRF G-patch motif ($NKRF^{G\text{-}patch}$) and Prp2 from C. thermophilum (homologue of DHX16) in complex with $Spp2^{G\text{-}patch}$ identified two sites required for G-patch contact, which in most human DEAH/RHA helicases, including DHX40, have accumulated substitutions that are incompatible with conserved G-patch binding (Hamann et al., 2020; Studer et al., 2020). Therefore, to enable the targeted recruitment and activation of these otherwise non-specific helicases to various pathways, non-G-patch protein-interacting DEAH/RHA helicases have likely evolved divergent mechanisms of activation and recruitment. As a result, these proteins could either have specialised cofactors or be regulated by a yet unidentified family of protein cofactors (Studer et al., 2020). Examples of helicases regulated by specialised cofactors not known to be part of bigger families include DHX37, which is regulated by UTP14A and DHX9, which is regulated by NUP98 and EWS-FLI1 (Capitanio et al., 2017; Choudhury et al., 2019; Donsbach & Klostermeier, 2021; Erkizan et al., 2015; Sloan & Bohnsack, 2018).

In this study, I identified PPIL4 as a stimulatory cofactor of DHX40, thus expanding our knowledge about DEAH/RHA regulation. Immunoprecipitation followed by western blots or LC-MS/MS showed that PPIL4 interacts with DHX40 in HEK293 cells, and an *in vitro* assay showed that PPIL4 stimulates DHX40 ATPase activity.

PPIL4 is an uncharacterised member of the cyclophilin PPIases, for which a PPIase domain is a defining feature. The PPIase domain can catalyse the cis-trans isomerisation of proline imidic peptide bonds in oligopeptides, a process involved in protein folding, chromatin modification, and protein regulation (Davis et al., 2010). PPIL4 furthermore has an RRM, an RNA-binding domain involved in the binding of ssRNA, and three disordered regions (Krepl et al., 2022). Strikingly, the composition of domains is similar to G-patch proteins, which usually have both RNA and protein binding domains. This raises the question if the disordered regions of PPIL4 are required for the stimulation of DHX40. The disordered regions of G-patch proteins have been proposed to "reel in" interaction partners by a "fly-casting" mechanism in which binding between the disordered G-patch and the cognate helicase binding partner induces a helical conformation of the G-patch, leading to spatial compression (K. E. Bohnsack et al., 2021; Hamann et al., 2020; Shoemaker et al., 2000)

Structural studies and interaction mapping approaches would be required to ascertain how PPIL4 influences DHX40 activity, as similar interactions have not been reported in the literature. X-ray crystallography of DHX40 in complex with PPIL4, and potentially RNA, would structurally characterise the interaction between this helicase-cofactor pair and potentially ascertain how PPIL4 stimulates DHX40 ATPase activity (Hamann et al., 2019, 2020). However, the disordered and highly flexible region of the proteins would be inaccessible to crystallography. Crosslinking mass spectrometry could reveal the identity of linked residues and, thus, the interacting regions between DHX40 and PPIL4, including contact sites in flexible regions (Leitner et al., 2020). Crosslinking data would therefore be able to provide information on conformational states that are too dynamic or unstable to be revealed by X-ray crystallography (Leitner et al., 2020). Considering the long, disordered C-terminal region and the extended flexible region between the PPIase and RRM motif in PPIL4, Crosslinking mass spectrometry might provide valuable insight into the PPIL4-DHX40 interaction.

The requirement for the PPIL4 disordered regions could also be determined by recombinant expression and purification of either a PPIL4 PPIase deletion mutant (PPIL4$^{\Delta PPIase}$) or any of the isolated disordered regions and performing ATPase assay with DHX40. Similar ATPase assays have previously been performed for several G-patch regulated DEAH/RHA helicases which showed that the isolated G-patch domain could specifically stimulate its cognate helicase (Studer et al., 2020; Tauchert et al., 2017).

The presence of the PPIase domain also raises the possibility that DHX40 is a substrate of PPIL4, and the structural characterisation and interaction mapping approaches discussed above could shed light on this possibility. In this case, isomerisation of targeted prolines could stabilise DHX40 in a conformation favourable for ATPase activity. However, PPIL4 is largely uncharacterised and has no identified substrates, so PPIase activity has not been measured. Therefore, the requirement for PPIase activity for DHX40 stimulation could potentially be explored in different ways. An important first step would be to determine PPIL4 PPIase activity, followed by mutagenesis analysis to design a catalytically inactive mutant (PPIL4^{inactive}). The PPIase activity against chromogenic- or fluorogenic-labelled model tetrapeptide substrates could be determined by a protease-coupled PPIase assay using prolyl-isomer-specific proteases (Fischer & Aumüller, 2003; Mori et al., 2009). If PPIase activity is observed, an option would be to perform the DHX40 ATPase assay with increasing concentrations of the general cyclophilin inhibitor cyclosporin (Mori et al., 2009); this experimental approach would determine if DHX40 and cyclosporin compete for the PPIase active site. Additionally, DHX40 ATPase assays with PPIL4 or PPIL4^{inactive} would establish if binding alone is enough to stimulate ATPase activity or if prolyl isomerisation is required. A further option would be to

incubate DHX40 with immobilised PPIL4 and remove PPIL4 prior to subsequent ATPase assays. As prolyl isomerisation is a slow process without a catalyst, the prolines in DHX40 affected by PPIL4 would likely start in an "active" conformation and, during the assay, slowly return to an energetically favourable state, causing DHX40 ATPase activity to decrease over the course of the experiment (Schmidpeter & Schmid, 2015). In this experimental setup, the binding of PPIL4 variants to DHX40 and the potential consequence of the interaction (i.e. proline isomerisation) would be uncoupled so their influences on DHX40 ATPase activity could be independently assessed.

PPIL4 has been confirmed to be an RBP in a large-scale screen that determined the RNA interactomes of 150 RBPs in K562 and HepG2 cell lines. This screen showed that PPIL4 interacts with L1 sense sequences (van Nostrand et al., 2020). Considering the stimulatory effect of PPIL4 on DHX40 ATPase activity and the association of both these proteins with L1 elements, PPIL4 may be responsible for or contribute to the recruitment of DHX40 to its RNA substrates. If this is the case, PPIL4 would act similarly to other helicase cofactors by coupling substrate targeting with activation. The functional implications of DHX40 and PPIL4 in L1 retrotransposition will be discussed in Section 4.4.

4.3.2 DEAH/RHA regulation by cyclyophilin PPIases

The discovery that PPIL4 regulates DHX40 ATPase activity brings forth the question if this is a specialised interaction or if other DEAH/RHA helicases may also be regulated by cognate PPIases. Strikingly, eight PPIase, all belonging to the cyclophilin family, are known to be components of the human spliceosome (Zhan et al., 2018). These isomerases are recruited in the B, B^{act}, and C/C* complexes and the U4/U6.U5 tri-snRNP and are thought to be responsible for RNP conformational changes in the later stages of the splicing process (Wahl et al., 2009). The precise functions of these PPIases have not yet been established; however, as these proteins are not present in yeast, it is speculated that they might function specifically in higher eukaryotes to enhance the flexibility of the spliceosome (Wahl et al., 2009). Furthermore, given that pre-mRNA splicing requires four DEAH/RHA helicases that, similar to the PPIases, function in the late stages of splicing, these helicases could potentially be PPIase targets (Jarmoskaite & Russell, 2014).

The DEAH/RHA proteins DHX38 and DHX8 are present in the C and P complexes of the pre-mRNA spliceosome (Wilkinson et al., 2020). However, these helicases have not been associated with any protein cofactors, and their potential modes of regulation are, therefore, unexplored. Notably, DHX38 is positioned near the PPIases PPWD1 and PPIG in the C complex (Rajiv & Davis, 2018; Zhan et al., 2018). The position of DHX8 in relation to spliceosomal PPIases is unclear, but it is present in the same complexes as PPIE, PPIG,

PPIL1, PPIL3, and PPWD1 (Rajiv & Davis, 2018). The two remaining characterised spliceosomal DEAH/RHA helicases, DHX16 and DHX15, are regulated by G-patch proteins. However, DHX16 is present in the B^{act} complex together with three PPIase, namely PPIE, PPIL1, and PPIL2, and only DHX15 does not seem to be employed in the same complexes as any of the spliceosomal PPIase (Rajiv & Davis, 2018). Of the spliceosomal PPIases, only PPIE have an RNA-binding domain, namely an RRM, and three of the spliceosomal PPIase (PPIH, PPIL1, and PPIL3) are minimal PPIase only consisting of the titular domain (Rajiv & Davis, 2018).

The interaction between DHX40 and PPIL4 might represent an example of a broader principle of DEAH/RHA helicases regulation by cyclophilin PPIases domains. Further investigation into this intriguing possibility might shed light on the functions of PPIases in pre-mRNA splicing and the regulation of DEAH/RHA proteins.

4.4 DHX40 role in the regulation of transposable elements

At least 46% of human DNA is derived from TEs, i.e. transposon and retrotransposons, and computational algorithms capable of detecting ancient, highly mutated TEs, suggest they account for as much as 70% of the human gDNA (Lander et al., 2001; Richardson et al., 2015). L1 retrotransposons account for 17% of the human genome, which includes more than 500,000 copies; however, only approximately 150 copies are full-length and retrotransposition-competent (Lander et al., 2001; Penzkofer et al., 2017). Phylogenetic analysis of human L1 RNAs divides the L1 elements into three distinct lineages: primate-specific $L1PA_{17-1}$ and $L1PB_{3-1}$ and mammalian-specific $L1MA_{4-1}$ (Khan et al., 2006). All retrotransposition-competent L1s belong to the youngest family of the L1PA lineage, i.e. L1PA1 (Khan et al., 2006). L1 retrotransposition occurs through an mRNA intermediate that acts as a template for synthesis of cDNA, which inserts into the genome at a target sequence. The L1-encoded proteins can either act in *cis* and mobilise their genomic element or in *trans* to mobilise non-autonomous SINEs, consisting of Alu and SVA elements, or cellular mRNAs, with the latter giving rise to processed pseudogenes (Richardson et al., 2015).

In this study, I discovered that DHX40 interacts with a cohort of L1 interacting proteins in HEK293 cells, including HNRNPM, XRCC5, XRCC6, and PPIL4. Previous studies investigating the consequences of XRCC5 (also known as Ku80) and XRCC6 (also known as Ku70) silencing on L1 retrotransposition in K562 and HeLa cells showed that the proteins are L1 retrotransposition activators (N. Liu et al., 2017). Interestingly, XRCC5 and XRCC6 interact with human immunodeficiency virus 1 (HIV-1) 5'UTR and can bind to and regulate transcription from HIV-1 LTR. Moreover, XRCC5 downregulation impairs HIV-1 integration and transcription (Stake et al., 2015). Additionally, a recent study showed that XRCC5 interacts with L1 ORF2p

in a DNA-dependent but RNA-independent interaction, suggesting that XRCC5 participates in L1 integration into the genome (Miyoshi et al., 2019). HNRNPM was recently shown to bind selectively to antisense L1 sequences, but its role in L1 regulation has not been confirmed (van Nostrand et al., 2020). Interestingly, HNRNPM has been suggested to monitor RLR activation and prevent an excessive innate antiviral immune response by antagonising RLR RNA-binding (Cao et al., 2019). PPIL4 has previously been shown to bind specifically to sense L1 sequences, and silencing of PPIL4 increases L1 retrotransposition frequency (N. Liu et al., 2017; van Nostrand et al., 2020).

After discovering that DHX40 interacts with L1 regulators and interactors, the DHX40 CRAC dataset was analysed with a focus on repetitive elements. This approach revealed that DHX40 primarily crosslinks to introns of protein-coding transcripts, lncRNAs, and pseudogenes and was additionally found to crosslink to transcript sequences derived from Alu elements and L1s. The enrichment of intronic reads in our CRAC dataset for DHX40 is in line with a role in regulating transposable elements, as 60% of TEs are located with intronic sequences (Sela et al., 2007). Additionally, Alu elements are highly abundant within coding transcripts due to an integration preference into gene-rich regions. The prevalence of Alu elements means that most primary nuclear transcripts will have Alu sequences in one or both orientations within introns (Deininger, 2011). The enrichment of Alu-derived sequences in the DHX40 CRAC dataset, therefore, potentially accounts for a substantial fraction of the reads assigned to introns. Moreover, retrotransposition-competent Alu RNAs are synthesised by RNA polymerase III and are generally expressed at low levels (Deininger, 2011). Most of the Alu RNA-derived reads in the dataset are, therefore, most likely not from retrotransposons but rather Alu-derived sequences embedded in RNA polymerase II transcripts. Similarly, sequences derived from defective TEs are retained in the genome if they are not disruptive and can be embedded in other transcripts (Lander et al., 2001; Sultana et al., 2017, 2019). This could account for the crosslinking of DHX40 to transcripts with sequences derived from inactive TE families. The enrichment of lncRNAs could likewise be due to interaction with TEs as 75-83% of lncRNAs contain at least one TE, and on average, 26-33% of a lncRNA sequence originates from TEs. The TE-derived sequences in lncRNAs can affect their synthesis, localisation and function (Fort et al., 2021). Furthermore, TE-derived sequences embedded into lncRNAs show higher conservation than the remaining sequence; in contrast, TE-derived sequences outside gene bodies generally lack sequence conservation (Fort et al., 2021).

Owing to the binding of DHX40 to L1 regulators and interactors, the crosslinking between DHX40 and L1-derived transcripts was analysed in more depth. Due to the library preparation strategy used in the CRAC approach, it was impossible to determine if DHX40 crosslinked to the sense or antisense sequence of L1s. However, DHX40 crosslinking positions were plotted

along the canonical model of L1 RNA to identify possible binding sites and divided into different L1 families. The coverage of crosslinking sites shows that DHX40 mainly crosslinks within three clusters in ORF2 of younger L1 subfamilies, i.e. L1PA1-3. Additionally, DHX40 crosslinking sites cluster within a single region of the 5'UTR of L1PA1-3. The crosslinking pattern of DHX40 to L1s indicates specific RNA-binding sites within the L1 sequence could be recognised by DHX40. The reduced binding of DHX40 to specific L1 regions in older L1 families likely reflects the evolution of L1 elements. Phylogenetic analysis of L1 families shows that ORF2, and to a lesser degree ORF1, have remained relatively conserved; meanwhile, the L1 5'UTRs have accumulated non-synonymous mutations through evolution, and frequently, the L1 lineages have acquired radically different 5'UTRs (Khan et al., 2006). The coiled-coiled domain of ORF1 underwent an episode of adaptive evolution during the evolution of L1PA5-3, evident by a large number of amino acid substitutions and three in-frame indels in this region (Khan et al., 2006). The apparent crosslinking of DHX40 to L1 elements from older L1 subfamilies within the ORF2 is likely due to this region's higher sequence conservation.

Additionally, the discovery of crosslinking sites in the 5'UTR of L1PA1 could suggest that DHX40 associates with full-length and potentially retrotransposition-competent L1 RNAs. The youngest L1 family, L1PA1, is divided into subfamilies based on their shared sequence variants (SSVs). Only members of the two youngest subfamilies of L1PA1 defined by the ACA/G and ACG/G SSVs in their 3'UTRs display retrotransposition activity in vivo (Penzkofer et al., 2017). There are ~150 potentially active L1s in the human reference genome, with a few "hot" L1s accounting for the bulk of retrotransposition events (Beck et al., 2010; Penzkofer et al., 2017). As the CRAC experimental approach prioritises binding sites, and DHX40 does not bind L1 SSVs, it was not possible to determine the subfamilies and, thereby, the retrotransposition potential of the crosslinked L1s. The CRAC approach could be complemented by RNA immunoprecipitation followed by RNA-seq (RIP-Seq) using longer paired-end reads. The RIP-Seq approach would allow the identification of the subfamilies of interaction L1 but would not provide accurate information on binding sites.

The discovery that DHX40 crosslinks L1-derived sequences supported the notion that DHX40 may function as a regulator of L1 retrotransposition. This discovery was especially intriguing considering the recently described binding of PPIL4 to L1 RNAs and the stimulatory effect of PPIL4 on DHX40 ATPase activity (see Section 4.3). As various DEAH/RHA helicases are recruited to appropriate RNAs by their cognate protein cofactors, PPIL4 could potentially contribute to DHX40 binding to its substrates, thereby endowing target specificity to the likely promiscuous DXH40 helicase core (K. E. Bohnsack et al., 2021, 2022; Choudhury et al., 2019). Similar to G-patch proteins in the regulation of their cognate DEAH/RHA protein, PPIL4 might

not only contribute to substrate recognition but also stimulate DHX40 catalytic activity in a context-appropriate manner.

As a complement to defining the RNA interactome of DHX40, the transcriptome of DHX40 knockdown and knockout cells was investigated to uncover differential expression of transcripts related to retrotransposition. Unfortunately, there was generally little consensus between our RNA-seq data for the knockout and knockdown samples regarding protein-coding transcripts. As DHX40 was later confirmed to be an L1 retrotransposition suppressor, the discrepancy between samples could be due to genomic instability in the knockout cell line arising from a chronic lack of DHX40. It is possible that the cell lines adapted to the loss of DHX40, thereby reducing the effects that might be observed upon acute depletion.

Several different types of pseudogenes were significantly differentially expressed in the DHX40 knockout cells relative to the control. Notably, differential expression of pseudogenes can suggest both genomic instability and regulation at post-transcriptional levels as they are potential sources of endogenous siRNAs, can act as competitive inhibitors of translation, and sequester miRNAs (Kalyana-Sundaram et al., 2012). Moreover, previously published RNA-seq data from 293 samples encompassing 13 tissue types and 248 cancer samples have provided evidence of lineage and cancer-specific pseudogene expression patterns (Kalyana-Sundaram et al., 2012). The differential expression analysis of cells lacking DHX40 revealed that snRNAs and snoRNAs were significantly upregulated in the DHX40 knockout samples. Together these data suggest that the CRISP/Cas9-mediated inactivation of DHX40 changed the transcriptome and metabolic state of the knockout cell line compared to the HEK293 control.

In agreement with the mass spectrometry and CRAC data, the differential expression analysis revealed a general increase in the levels of TE-derived sequences arising from DNA transposable elements, LTRs and LINEs regardless of the means and duration of the DHX40 depletion. However, significantly upregulated LTR-derived sequences were more abundant in the knockout sample, while significantly upregulated sequences derived from LINEs and DNA transposable elements were more abundant in the knockdown. Alu-derived sequences could not be detected due to the experimental setup. It was, therefore, impossible to determine if the Alu elements observed in the CRAC experiments indicated a role for DHX40 in regulating Alu elements. The discovery that DHX40 depletion, regardless of means and duration, increased the level of L1-derived sequences supported the notion that DHX40 may function as a suppressor of L1 retrotransposition. Additionally, due to the unpredictable changes to transcriptome and the proposed genomic instability of the knockout cell lines, the knockdown system was regarded as a more stable model for our L1 retrotransposition assay.

4.4.1 DHX40 as a supressor of L1 retrotransposition

Using a cell-based retrotransposition assay, I showed that the knockdown of DHX40 and PPIL4 significantly increased the retrotransposition frequency of an L1 reporter in HEK293 cells, thereby demonstrating that PPIL4 and DHX40 are suppressors of L1 retrotransposition. Intriguingly, the increase in retrotransposition frequency was similar for DXH40 and PPIL4 knockdown, and PPIL4 knockdown slightly decreased DHX40 protein levels in both HEK293 and MCF7 cells. Together with the robust interaction between PPIL4 and DHX40, these results could suggest that DHX40 and PPIL4 work in concert to repress L1 retrotransposition.

Considering that PPIL4 specifically binds sense L1 elements and the proposed promiscuity of the DHX40 helicase core, investigating the effect of PPIL4 on DHX40 L1 RNA binding and catalytic activity could provide valuable insight into how DHX40 is specifically recruited to L1 RNAs by PPIL4 and the mechanism of L1 suppression (van Nostrand et al., 2020). An important first step would be to extend the biochemical characterisation of DHX40 by performing helicase activity assays; this would determine if the already established ATPase activity, a "hallmark" for helicase, is coupled to the unwinding of nucleic acid substrates (Hamann et al., 2019). The helicase activity of DHX40 can readily be measured using a fluorescence-based unwinding assay (Hamann et al., 2021). The assay can additionally be performed using different substrates, i.e. dsRNAs, DNAs, and RNA:DNA hybrids with overhangs of different lengths, and in the presence of PPIL4 to determine the preferred substrate and the stimulatory effect of PPIL4 on DHX40 helicase activity. Assuming helicase activity, the next step would be to perform mutagenesis analyses to determine the requirements for catalytic activity and to generate different catalytically inactive or non-PPIL4 binding variants for investigation in cellular contexts.

The consequence of disrupted DHX40-PPIL4 activity or impaired biochemical function of DHX40 or PPILs on L1 suppressions could additionally be investigated in cell-based assays. A limiting factor is, however, the low copy number of L1 transcripts in HEK293 cells, as only highly sensitive experimental approaches can detect the minute amounts of crosslinked L1 RNAs. The amount of crosslinked L1 RNAs could be increased by using the MCF7 cell line, which is widely used in retrotransposition studies due to its high production of L1 RNAs (B. Freeman et al., 2022; Philippe et al., 2016). The concentration of L1 RNAs in the cells could also be increased by transfection with the L1-reporter plasmid used in the retrotransposition assay performed here or a similar plasmid with a more potent promoter, e.g. a CAG promoter (Xie et al., 2011). Engineering a system with transient higher or inducible L1 expression could make it feasible to detect DHX40 binding to L1 RNAs in several different conditions by crosslinking followed by Northern blotting or qPCR.

A system capable of greater L1 RNA expression would furthermore be advantageous for future L1 retrotransposition assay as minor effects on retrotransposition frequency would be more readily detected. In the interest of broadening our understanding of the role of DHX40 in regulating retrotransposition, several variations of the assay could be performed with different conditions, including overexpression of DHX40 or DHX40 variants, DHX40 knockdown with simultaneous expression of DHX40 variants at endogenous level, and expression of L1 RNAs with silent mutations in the ORF2. DHX40 variants to investigate would include DHX40^{E174Q}, additional variants with mutations within the core motifs, and non-PPIL4 interacting as variants, as these would help determine the requirements for DHX40 catalytic activity and PPIL4-binding for L1 retrotransposition suppression. The UPF1 family helicase MOV10 is a potent retrotransposition suppressor that has been shown to colocalise with the ORF1p in cytoplasmic granules. Previous studies on the requirements of MOV10 in the suppression of retrotransposition have shown that mutations in motifs I, II, III, IV, and VI in the helicase core drastically reduced the ability of MOV10 to reduce retrotransposition (Goodier et al., 2012). MOV10, therefore, provides an example of helicase-dependent suppression of L1 retrotransposition. To provide insight into sequence requirements for L1 retrotransposition regulation by DHX40, a retrotransposition assay using plasmids encoding L1 RNAs with silent mutations in the ORF2 could be performed. Unlike an *in vitro* binding assay using synthesised RNAs and purified proteins, this setup would determine the binding requirement in a functional context.

Regulation of retrotransposition can happen at several stages, including L1 transcription, mRNA maturation, translation, cytoplasmic L1 RNP assembly, nuclear import, and genomic integration (Protasova et al., 2021). However, considering the interaction partners of DHX40 and the localisation reported previously (Thul et al., 2017), it most likely that DHX40 regulate retrotransposition in the nucleus, i.e. during transcription RNA maturation or integration. Furthermore, our crosslinking data shows that DHX40 binds intronic regions, which could suggest a regulation at transcription or RNA maturation levels. However, multi-mapping in the CRAC analysis allows the sequencing reads to be mapped to the distinct TE families and the intronic regions in which TE-derived sequences are embedded. The enrichments of intronic reads could, therefore, be a consequence of TE binding. It would additionally be possible that DHX40 binds to L1-derived sequences regardless of the transcript. i.e. bind to both retrotransposable L1 RNAs and intronic regions containing the ORF2. Therefore, our CRAC dataset alone does not provide sufficient information to determine the timing of L1 regulation by DHX40.

Interestingly, DHX40 has previously been reported to interact with the histone post-translation modifications: H3K27ac and H3K4me3 (Ji et al., 2015). This interaction supports the model

that DHX40 regulates retrotransposition during transcription or integration. H3K27ac and H3K4me3 are associated with enhancers and promoters and are both enriched at full-length L1s in mouse embryonic stem cells (J. He et al., 2019). Notably, de novo insertions of L1 in humans have a modest but significant association with several chromatin marks, including H3K27ac and H3K4me3, and ENCODE-defined chromatin segmentation states, including "weak transcription" and "weak enhancer" (Sultana et al., 2019). Moreover, the H3K27ac and H3K4me3 histone marks are significantly enriched near active L1 promotors in MCF7, HEK293, and HeLa cells (B. Freeman et al., 2022). Therefore, the epigenetic markers that DHX40 is associated with appear to affect both L1 transcription and integration. Assuming a role in suppressing L1s at transcription, DHX40 could potentially contribute to premature abortion of transcription upon recognition of the ORF2 or target L1 RNAs immediately after transcription, thus preventing maturation and export to the nucleus. Regulation at a transcriptional level would likely be reflected in the protein expression level of the L1-encoded protein: ORF1p. No differences in ORF1p levels following DHX40 depletion in MCF7 cells have been observed so far, but optimisation and quantification of the western blotting would be required to confirm that ORF1p levels are not increased following the knockdown of DHX40.

Non-histone chromatin immunoprecipitation, using DHX40 as bait, followed by sequencing of associated sequences (ChIP-seq), and ChIP followed by mass spectrometry (ChIP-MS) could provide valuable insight into the association between DHX40 and epigenetic markers and sequences and how this association might affect L1 regulation (Park, 2009). ChIP-seq data could be complemented by RNA-seq data (Paired-end, 100 bp reads) for DHX40 knockdowns and control, followed by differential expression analysis. In contrast to the RNA-seq performed in this study, RNA-seq with longer paired-end reads would allow for unique alignment of L1s to different genomic loci. With complimentary ChIP-seq and RNA-seq data, it would be possible to determine the effect of DHX40 chromatin association on L1 transcription. This data could additionally be complemented by *in vitro* binding assays to determine the binding affinity of DHX40 towards RNA, DNA, and RNA:DNA hybrids. The identification of the preferred and possible nucleic acid substrates would provide valuable mechanistic insight into the cellular function of DHX40.

XRCC5 and XRCC6 are both modest activators of L1 retrotransposition (N. Liu et al., 2017). Their role in retrotransposition is unclear; however, XRCC5 have been shown to interact with ORF2p, and the protein has been associated with the retroviral integration of HIV-I (Miyoshi et al., 2019; Stake et al., 2015). XRCC5 and XRCC6 have also been shown to affect HIV-I transcription. Therefore, the interaction of DHX40 with these proteins could support a role in either L1 transcription or integration. DHX40 could potentially recognise RNA:DNA hybrids and antagonise L1 integration by displacing L1-encoded proteins or host-proteins required for DNA

repair in a helicase activity-dependent manner. Assuming recognition of RNA:DNA hybrids DHX40 could potentially also prevent retrotransposition of Alu elements and mRNAs. The nuclear protein HNRNPM was recently shown to specifically bind to antisense L1 RNA in a screen investigating the RNA interactomes of RBPs but has not been characterised as a regulator of L1 retrotransposition (van Nostrand et al., 2020). However, it has been suggested that HNRNPM binding to the antisense L1 prevents the use of cryptic splice sites introduced by the integration of the retrotransposon (van Nostrand et al., 2020). Moreover, HNRNPM has been shown to antagonise antiviral response mediated by MDA5 and RIG-I. Upon viral infection, HNRNPM is exported to the cytoplasm, where it binds to the viral RNA and impairs the RNA-binding of the MDA5 and RIG-I (Cao et al., 2019). This function of HNRNPM has been proposed to monitor and prevent uncontrolled antiviral signalling.

Intriguingly, the strong interaction between HNRNPM and DHX40 could potentially suggest that DHX40 is a retrotransposon-specific RNA sensor within the nucleus, akin to viral RNA sensing by RLRs in the cytoplasm. Assuming a similar role in retrotransposition, HNRNPM might displace antisense L1s from DHX40 to prevent DHX40 from unspecifically targeting these RNAs. By this potential mechanism as an L1 RNA sensor, DHX40 would target L1 RNAs for degradation. Given the prevalence of L1-derived sequences in coding transcripts, stringent control mechanisms would be required to prevent uncontrolled degradation of non-L1 RNAs. A potential control mechanism could be the requirement for avidity-dependent signalling facilitated by RNA-supported DHX40 filament formation. A role as an RNA sensor in the nucleus could explain the hinted association of DHX40 with both L1 transcription and integration. Insight into the effect of HNRNPM on DHX40 RNA binding could be provided by the knockdown of HNRNPM followed by the analysis of RNAs crosslinked to DHX40.

4.5 Conclusions

This study provided a valuable toolbox for the future functional characterisation of DHX proteins in a cellular context. The usefulness of this tool was exemplified by the functional characterisation of DHX30 performed by Cruz-Zaragoza et al. 2021, and the functional characterisation of DHX40. In this study, I showed that DHX40 is an RNA-dependent ATPase stimulated by PPIL4 and confirmed that USP7 regulates DHX40 protein levels. Furthermore, with a comprehensive analysis of the RNA interactome and a cell-based L1 retrotransposition assay, we showed that PPIL4 and DHX40 are suppressors of L1 retrotransposition in HEK293.

References

Alexandrova, E. A., Olovnikov, I. A., Malakhova, G. v., Zabolotneva, A. A., Suntsova, M. v., Dmitriev, S. E., & Buzdin, A. A. (2012). Sense transcripts originated from an internal part of the human retrotransposon LINE-1 5' UTR. *Gene*, *511*(1), 46–53. https://doi.org/10.1016/J.GENE.2012.09.026

Andrews, S. (2010). FastQC - A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. *Babraham Bioinformatics*.

Antonicka, H., & Shoubridge, E. A. (2015). Mitochondrial RNA Granules Are Centers for Posttranscriptional RNA Processing and Ribosome Biogenesis. *Cell Reports*, *10*(6), 920–932. https://doi.org/10.1016/J.CELREP.2015.01.030

Aratani, S., Fujii, R., Oishi, T., Fujita, H., Amano, T., Ohshima, T., Hagiwara, M., Fukamizu, A., & Nakajima, T. (2001). Dual roles of RNA helicase A in CREB-dependent transcription. *Molecular and Cellular Biology*, *21*(14), 4460–4469. https://doi.org/10.1128/MCB.21.14.4460-4469.2001

Aravind, L., & Koonin, E. v. (1999). G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends in Biochemical Sciences*, *24*(9), 342–344. https://doi.org/10.1016/S0968-0004(99)01437-1

Athanikar, J. N., Badge, R. M., & Moran, J. v. (2004). A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Research*, *32*(13), 3846. https://doi.org/10.1093/NAR/GKH698

Attig, J., Agostini, F., Gooding, C., Chakrabarti, A. M., Singh, A., Haberman, N., Zagalak, J. A., Emmett, W., Smith, C. W. J., Luscombe, N. M., & Ule, J. (2018). Heteromeric RNP Assembly at LINEs Controls Lineage-Specific RNA Processing. *Cell*, *174*(5), 1067. https://doi.org/10.1016/J.CELL.2018.07.001

Azatian, S. B., Kaur, N., & Latham, M. P. (2019). Increasing the buffering capacity of minimal media leads to higher protein yield. *Journal of Biomolecular NMR*, *73*(1–2), 11. https://doi.org/10.1007/S10858-018-00222-4

Bao, P., Höbartner, C., Hartmuth, K., & Lührmann, R. (2017). Yeast Prp2 liberates the 5′ splice site and the branch site adenosine for catalysis of pre-mRNA splicing. *RNA*, *23*(12), 1770–1779. https://doi.org/10.1261/RNA.063115.117/-/DC1

Barber, K. W., & Rinehart, J. (2018). The ABCs of PTMs. *Nature Chemical Biology*, *14*(3), 188. https://doi.org/10.1038/NCHEMBIO.2572

Basame, S., Wai-Lun Li, P., Howard, G., Branciforte, D., Keller, D., & Martin, S. L. (2006). Spatial assembly and RNA binding stoichiometry of a LINE-1 protein essential for

retrotransposition. *Journal of Molecular Biology*, *357*(2), 351–357. https://doi.org/10.1016/J.JMB.2005.12.063

Beck, C. R., Collier, P., Macfarlane, C., Malig, M., Kidd, J. M., Eichler, E. E., Badge, R. M., & Moran, J. v. (2010). LINE-1 Retrotransposition Activity in Human Genomes. *Cell*, *141*(7), 1159. https://doi.org/10.1016/J.CELL.2010.05.021

Belancio, V. P., Whelton, M., & Deininger, P. (2007). Requirements for polyadenylation at the 3′ end of LINE-1 elements. *Gene*, *390*(1–2), 98–107. https://doi.org/10.1016/J.GENE.2006.07.029

Bertram, K., Agafonov, D. E., Liu, W. T., Dybkov, O., Will, C. L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., & Lührmann, R. (2017). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature*, *542*(7641), 318–323. https://doi.org/10.1038/NATURE21079

Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J., & Schäfer, F. (2009). Chapter 27 Immobilized-Metal Affinity Chromatography (IMAC): A Review. *Methods in Enzymology*, *463*(C), 439–473. https://doi.org/10.1016/S0076-6879(09)63027-5

Bohnsack, K. E., & Bohnsack, M. T. (2019). Uncovering the assembly pathway of human ribosomes and its emerging links to disease. *The EMBO Journal*, *38*(13). https://doi.org/10.15252/EMBJ.2018100278

Bohnsack, K. E., Ficner, R., Bohnsack, M. T., & Jonas, S. (2021). Regulation of DEAH-box RNA helicases by G-patch proteins. *Biological Chemistry*, *402*(5), 561–579. https://doi.org/10.1515/HSZ-2020-0338/ASSET/GRAPHIC/J_HSZ-2020-0338_FIG_002.JPG

Bohnsack, K. E., Kanwal, N., & Bohnsack, M. T. (2022). Prp43/DHX15 exemplify RNA helicase multifunctionality in the gene expression network. *Nucleic Acids Research*, *50*(16), 9012–9022. https://doi.org/10.1093/NAR/GKAC687

Bohnsack, M. T., Tollervey, D., & Granneman, S. (2012). Identification of RNA helicase target sites by UV cross-linking and analysis of cDNA. *Methods in Enzymology*, *511*, 275–288. https://doi.org/10.1016/B978-0-12-396546-2.00013-9

Bolanos-Garcia, V. M., & Davies, O. R. (2006). Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochimica et Biophysica Acta*, *1760*(9), 1304–1313. https://doi.org/10.1016/J.BBAGEN.2006.03.027

Boneberg, F. M., Brandmann, T., Kobel, L., den Heuvel, J. van, Bargsten, K., Bammert, L., Kutay, U., & Jinek, M. (2019). Molecular mechanism of the RNA helicase DHX37 and its

activation by UTP14A in ribosome biogenesis. *RNA (New York, N.Y.)*, *25*(6), 685–701. https://doi.org/10.1261/RNA.069609.118

Bosco, B., Rossi, A., Rizzotto, D., Hamadou, M. H., Bisio, A., Giorgetta, S., Perzolli, A., Bonollo, F., Gaucherot, A., Catez, F., Diaz, J. J., Dassi, E., & Inga, A. (2021). DHX30 Coordinates Cytoplasmic Translation and Mitochondrial Function Contributing to Cancer Cell Survival. *Cancers*, *13*(17). https://doi.org/10.3390/CANCERS13174412

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology 2018 19:1*, *19*(1), 1–12. https://doi.org/10.1186/S13059-018-1577-Z

Brosius, J., & Raabe, C. A. (2016). What is an RNA? A top layer for RNA classification. *RNA Biology*, *13*(2), 140. https://doi.org/10.1080/15476286.2015.1128064

Cadena, C., Ahmad, S., Xavier, A., Willemsen, J., Park, S., Park, J. W., Oh, S. W., Fujita, T., Hou, F., Binder, M., & Hur, S. (2019). Ubiquitin-dependent and -independent roles of E3 ligase RIPLET in innate immunity. *Cell*, *177*(5), 1187. https://doi.org/10.1016/J.CELL.2019.03.017

Cai, J., Chen, H. Y., Peng, S. J., Meng, J. L., Wang, Y., Zhou, Y., Qian, X. P., Sun, X. Y., Pang, X. W., Zhang, Y., & Zhang, J. (2018). USP7-TRIM27 axis negatively modulates antiviral type I IFN signaling. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *32*(10), 5238–5249. https://doi.org/10.1096/FJ.201700473RR

Cao, P., Luo, W. W., Li, C., Tong, Z., Zheng, Z. Q., Zhou, L., Xiong, Y., & Li, S. (2019). The heterogeneous nuclear ribonucleoprotein hnRNPM inhibits RNA virus-triggered innate immunity by antagonizing RNA sensing of RIG-I-like receptors. *PLoS Pathogens*, *15*(8). https://doi.org/10.1371/JOURNAL.PPAT.1007983

Capitanio, J. S., Montpetit, B., & Wozniak, R. W. (2017). Human Nup98 regulates the localization and activity of DExH/D-box helicase DHX9. *ELife*, *6*. https://doi.org/10.7554/ELIFE.18825

Caterino, M., & Paeschke, K. (2022). Action and function of helicases on RNA G-quadruplexes. *Methods*, *204*, 110–125. https://doi.org/10.1016/J.YMETH.2021.09.003

Chen, M. C., & Ferré-D'Amaré, A. R. (2017). Structural Basis of DEAH/RHA Helicase Activity. *Crystals 2017, Vol. 7, Page 253*, *7*(8), 253. https://doi.org/10.3390/CRYST7080253

Choi, H., Glatter, T., Gstaiger, M., & Nesvizhskii, A. I. (2012). SAINT-MS1: Protein-protein interaction scoring using label-free intensity data in affinity purification-mass spectrometry

experiments. *Journal of Proteome Research*, *11*(4), 2619–2624. https://doi.org/10.1021/PR201185R

Chong, Z. X., Yeap, S. K., & Ho, W. Y. (2021). Transfection types, methods and strategies: A technical review. *PeerJ*, *9*. https://doi.org/10.7717/PEERJ.11165/SUPP-1

Choudhury, P., Hackert, P., Memet, I., Sloan, K. E., & Bohnsack, M. T. (2019). The human RNA helicase DHX37 is required for release of the U3 snoRNP from pre-ribosomal particles. *RNA Biology*, *16*(1), 54–68. https://doi.org/10.1080/15476286.2018.1556149

Choudhury, P., Kretschmer, J., Hackert, P., Bohnsack, K. E., & Bohnsack, M. T. (2020). The DExD box ATPase DDX55 is recruited to domain IV of the 28S ribosomal RNA by its C-terminal region. *Https://Doi.Org/10.1080/15476286.2020.1829366*, *18*(8), 1124–1135. https://doi.org/10.1080/15476286.2020.1829366

Christensen, D. G., Orr, J. S., Rao, C. v., & Wolfe, A. J. (2017). Increasing Growth Yield and Decreasing Acetylation in Escherichia coli by Optimizing the Carbon-to-Magnesium Ratio in Peptide-Based Media. *Applied and Environmental Microbiology*, *83*(6). https://doi.org/10.1128/AEM.03034-16

Christian, C. M., Deharo, D., Kines, K. J., Sokolowski, M., & Belancio, V. P. (2016). Identification of L1 ORF2p sequence important to retrotransposition using Bipartile Alu retrotransposition (BAR). *Nucleic Acids Research*, *44*(10), 4818. https://doi.org/10.1093/NAR/GKW277

Concordet, J. P., & Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research*, *46*(W1), W242–W245. https://doi.org/10.1093/NAR/GKY354

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (New York, N.Y.)*, *339*(6121), 819. https://doi.org/10.1126/SCIENCE.1231143

Cooper, G. M. (2019). The Cell: A Molecular Approach 8th Ed. In *The Cell A molecular Approach*.

Cordaux, R., & Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nature Reviews. Genetics*, *10*(10), 691. https://doi.org/10.1038/NRG2640

Cost, G. J., Feng, Q., Jacquier, A., & Boeke, J. D. (2002). Human L1 element target-primed reverse transcription in vitro. *The EMBO Journal*, *21*(21), 5899–5910. https://doi.org/10.1093/EMBOJ/CDF592

Cruz-Zaragoza, L. D., Dennerlein, S., Linden, A., Yousefi, R., Lavdovskaia, E., Aich, A., Falk, R. R., Gomkale, R., Schöndorf, T., Bohnsack, M. T., Richter-Dennerlein, R., Urlaub, H.,

& Rehling, P. (2021). An in vitro system to silence mitochondrial gene expression. *Cell*, *184*(23), 5824-5837.e15. https://doi.org/10.1016/J.CELL.2021.09.033

Dai, L., LaCava, J., Taylor, M. S., & Boeke, J. D. (2014). Expression and detection of LINE-1 ORF-encoded proteins. *Mobile Genetic Elements*, *4*(3), e29319. https://doi.org/10.4161/MGE.29319

Dai, L., Taylor, M. S., O'Donnell, K. A., & Boeke, J. D. (2012). Poly(A) binding protein C1 is essential for efficient L1 retrotransposition and affects L1 RNP formation. *Molecular and Cellular Biology*, *32*(21), 4323–4336. https://doi.org/10.1128/MCB.06785-11

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), 1–4. https://doi.org/10.1093/GIGASCIENCE/GIAB008

Davis, T. L., Walker, J. R., Campagna-Slater, V., Finerty, P. J., Finerty, P. J., Paramanathan, R., Bernstein, G., Mackenzie, F., Tempel, W., Ouyang, H., Lee, W. H., Eisenmesser, E. Z., & Dhe-Paganon, S. (2010). Structural and biochemical characterization of the human cyclophilin family of peptidyl-prolyl isomerases. *PLoS Biology*, *8*(7). https://doi.org/10.1371/journal.pbio.1000439

de Bortoli, F., Espinosa, S., & Zhao, R. (2021). DEAH-Box RNA Helicases in Pre-mRNA Splicing. *Trends in Biochemical Sciences*, *46*(3), 225–238. https://doi.org/10.1016/J.TIBS.2020.10.006

de Maio, A., Yalamanchili, H. K., Adamski, C. J., Gennarino, V. A., Liu, Z., Qin, J., Jung, S. Y., Richman, R., Orr, H., & Zoghbi, H. Y. (2018). RBM17 Interacts with U2SURP and CHERP to Regulate Expression and Splicing of RNA-Processing Proteins. *Cell Reports*, *25*(3), 726-736.e7. https://doi.org/10.1016/J.CELREP.2018.09.041

Deininger, P. (2011). Alu elements: Know the SINEs. *Genome Biology*, *12*(12), 1–12. https://doi.org/10.1186/GB-2011-12-12-236/FIGURES/5

Denli, A. M., Narvaiza, I., Kerman, B. E., Pena, M., Benner, C., Marchetto, M. C. N., Diedrich, J. K., Aslanian, A., Ma, J., Moresco, J. J., Moore, L., Hunter, T., Saghatelian, A., & Gage, F. H. (2015). Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell*, *163*(3), 583–593. https://doi.org/10.1016/J.CELL.2015.09.025

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/BIOINFORMATICS/BTS635

Dodt, M., Roehr, J. T., Ahmed, R., & Dieterich, C. (2012). FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology*, *1*(3), 895–905. https://doi.org/10.3390/BIOLOGY1030895

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology 2015 34:2*, *34*(2), 184–191. https://doi.org/10.1038/nbt.3437

Donsbach, P., & Klostermeier, D. (2021). Regulation of RNA helicase activity: Principles and examples. *Biological Chemistry*, *402*(5), 529–559. https://doi.org/10.1515/HSZ-2020-0362/XML

Doucet, A. J., Hulme, A. E., Sahinovic, E., Kulpa, D. A., Moldovan, J. B., Kopera, H. C., Athanikar, J. N., Hasnaoui, M., Bucheton, A., Moran, J. v., & Gilbert, N. (2010). Characterization of LINE-1 Ribonucleoprotein Particles. *PLoS Genetics*, *6*(10), 1–19. https://doi.org/10.1371/JOURNAL.PGEN.1001150

Duic, I., Tadakuma, H., Harada, Y., Yamaue, R., Deguchi, K., Suzuki, Y., Yoshimura, S. H., Kato, H., Takeyasu, K., & Fujita, T. (2020). Viral RNA recognition by LGP2 and MDA5, and activation of signaling through step-by-step conformational changes. *Nucleic Acids Research*, *48*(20), 11664–11674. https://doi.org/10.1093/NAR/GKAA935

Duss, O., Stepanyuk, G. A., Puglisi, J. D., & Williamson, J. R. (2019). Transient Protein-RNA Interactions Guide Nascent Ribosomal RNA Folding. *Cell*, *179*(6), 1357-1369.e16. https://doi.org/10.1016/J.CELL.2019.10.035

Eargle, J., Wright, D., & Luthey-Schulten, Z. (2006). Multiple Alignment of protein structures and sequences for VMD. *Bioinformatics (Oxford, England)*, *22*(4), 504–506. https://doi.org/10.1093/BIOINFORMATICS/BTI825

Edelheit, O., Hanukoglu, A., & Hanukoglu, I. (2009). Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnology*, *9*. https://doi.org/10.1186/1472-6750-9-61

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., & Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, *411*(6836), 494–498. https://doi.org/10.1038/35078107

Emmerich, C. H., & Cohen, P. (2015). Optimising methods for the preservation, capture and identification of ubiquitin chains and ubiquitylated proteins by immunoblotting. *Biochemical and Biophysical Research Communications*, *466*(1), 1–14. https://doi.org/10.1016/J.BBRC.2015.08.109

Erkizan, H. V., Schneider, J. A., Sajwan, K., Graham, G. T., Griffin, B., Chasovskikh, S., Youbi, S. E., Kallarakal, A., Chruszcz, M., Padmanabhan, R., Casey, J. L., Uren, A., & Toretsky,

J. A. (2015). RNA helicase A activity is inhibited by oncogenic transcription factor EWS-FLI1. *Nucleic Acids Research*, *43*(2), 1069–1080. https://doi.org/10.1093/NAR/GKU1328

Fairman-Williams, M. E., Guenther, U. P., & Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. *Current Opinion in Structural Biology*, *20*(3), 313–324. https://doi.org/10.1016/J.SBI.2010.03.011

Farkash, E. A., Kao, G. D., Horman, S. R., & Prak, E. T. L. (2006). Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. *Nucleic Acids Research*, *34*(4). https://doi.org/10.1093/nar/gkj522

Feng, Q., Moran, J. v., Kazazian, H. H., & Boeke, J. D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, *87*(5), 905–916. https://doi.org/10.1016/S0092-8674(00)81997-2

Fidaleo, M., de Paola, E., & Paronetto, M. P. (2016). The RNA helicase A in malignant transformation. *Oncotarget*, *7*(19), 28711. https://doi.org/10.18632/ONCOTARGET.7377

Fischer, G., & Aumüller, T. (2003). Regulation of peptide bond cis/trans isomerization by enzyme catalysis and its implication in physiological processes. *Reviews of Physiology, Biochemistry and Pharmacology*, *148*, 105–150. https://doi.org/10.1007/S10254-003-0011-3

Fort, V., Khelifi, G., & Hussein, S. M. I. (2021). Long non-coding RNAs and transposable elements: A functional relationship. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1868*(1), 118837. https://doi.org/10.1016/J.BBAMCR.2020.118837

Fraile, J. M., Campos-Iglesias, D., Rodríguez, F., Astudillo, A., Vilarrasa-Blasi, R., Verdaguer-Dot, N., Prado, M. A., Paulo, J. A., Gygi, S. P., Martín-Subero, J. I., Freije, J. M. P., & López-Otín, C. (2018). Loss of the deubiquitinase USP36 destabilizes the RNA helicase DHX33 and causes preimplantation lethality in mice. *The Journal of Biological Chemistry*, *293*(6), 2183. https://doi.org/10.1074/JBC.M117.788430

Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J. E., Mudge, J. M., Sisu, C., Wright, J. C., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Boix, C., Sala, S. C., Cunningham, F., Domenico, T. di, Donaldson, S., Fiddes, I. T., Girón, C. G., … Flicek, P. (2021). GENCODE 2021. *Nucleic Acids Research*, *49*(D1), D916–D923. https://doi.org/10.1093/NAR/GKAA1087

Freeman, B. T., Sokolowski, M., Roy-Engel, A. M., Smither, M. E., & Belancio, V. P. (2019). Identification of charged amino acids required for nuclear localization of human L1 ORF1 protein. *Mobile DNA*, *10*(1), 1–17. https://doi.org/10.1186/S13100-019-0159-2/FIGURES/7

Freeman, B., White, T., Kaul, T., Stow, E. C., Baddoo, M., Ungerleider, N., Morales, M., Yang, H., Deharo, D., Deininger, P., & Belancio, V. P. (2022). Analysis of epigenetic features characteristic of L1 loci expressed in human cells. *Nucleic Acids Research*, *50*(4), 1888–1907. https://doi.org/10.1093/NAR/GKAC013

Fuller-Pace, F. v. (2006). DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. *Nucleic Acids Research*, *34*(15), 4206–4215. https://doi.org/10.1093/NAR/GKL460

Fus-Kujawa, A., Prus, P., Bajdak-Rusinek, K., Teper, P., Gawron, K., Kowalczuk, A., & Sieron, A. L. (2021). An Overview of Methods and Tools for Transfection of Eukaryotic Cells in vitro. *Frontiers in Bioengineering and Biotechnology*, *9*, 634. https://doi.org/10.3389/FBIOE.2021.701031/BIBTEX

Ganser, L. R., Kelly, M. L., Herschlag, D., & Al-Hashimi, H. M. (2019). The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews. Molecular Cell Biology*, *20*(8), 474–489. https://doi.org/10.1038/S41580-019-0136-0

Georges, A., Marcon, E., Greenblatt, J., & Frappier, L. (2018). Identification and Characterization of USP7 Targets in Cancer Cells. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-34197-x

Goodier, J. L., Cheung, L. E., & Kazazian, H. H. (2012). MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. *PLoS Genetics*, *8*(10). https://doi.org/10.1371/JOURNAL.PGEN.1002941

Gorbalenya, A. E., & Koonin, E. v. (1993). Helicases: amino acid sequence comparisons and structure-function relationships. *Current Opinion in Structural Biology*, *3*(3), 419–429. https://doi.org/10.1016/S0959-440X(05)80116-2

Haag, S., Kretschmer, J., Sloan, K. E., & Bohnsack, M. T. (2017). Crosslinking methods to identify RNA methyltransferase targets in vivo. In *Methods in Molecular Biology* (Vol. 1562). https://doi.org/10.1007/978-1-4939-6807-7_18

Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D., Ule, J., & Zavolan, M. (2021). CLIP and complementary methods. *Nature Reviews Methods Primers 2021 1:1*, *1*(1), 1–23. https://doi.org/10.1038/s43586-021-00018-1

Hage, A., Bharaj, P., van Tol, S., Shi, P.-Y., Freiberg, A. N., & Correspondence, R. R. (2022). The RNA helicase DHX16 recognizes specific viral RNA to trigger RIG-I-dependent innate antiviral immunity. *CellReports*, *38*, 110434. https://doi.org/10.1016/j.celrep.2022.110434

Hamann, F., Enders, M., & Ficner, R. (2019). Structural basis for RNA translocation by DEAH-box ATPases. *Nucleic Acids Research*, *47*(8), 4349–4362. https://doi.org/10.1093/NAR/GKZ150

Hamann, F., Schmitt, A., Favretto, F., Hofele, R., Neumann, P., Xiang, S. Q., Urlaub, H., Zweckstetter, M., & Ficner, R. (2020). Structural analysis of the intrinsically disordered splicing factor Spp2 and its binding to the DEAH-box ATPase Prp2. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(6), 2948–2956. https://doi.org/10.1073/PNAS.1907960117/SUPPL_FILE/PNAS.1907960117.SAPP.PDF

Hamann, F., Zimmerningkat, L. C., Becker, R. A., Garbers, T. B., Neumann, P., Hub, J. S., & Ficner, R. (2021). The structure of Prp2 bound to RNA and ADP-BeF3- reveals structural features important for RNA unwinding by DEAH-box ATPases. *Acta Crystallographica. Section D, Structural Biology*, *77*(Pt 4), 496–509. https://doi.org/10.1107/S2059798321001194

Han, H., Braunschweig, U., Gonatopoulos-Pournatzis, T., Weatheritt, R. J., Hirsch, C. L., Ha, K. C. H., Radovani, E., Nabeel-Shah, S., Sterne-Weiler, T., Wang, J., O'Hanlon, D., Pan, Q., Ray, D., Zheng, H., Vizeacoumar, F., Datti, A., Magomedova, L., Cummins, C. L., Hughes, T. R., … Blencowe, B. J. (2017). Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. *Molecular Cell*, *65*(3), 539-553.e7. https://doi.org/10.1016/J.MOLCEL.2017.01.011

Hancks, D. C., & Kazazian, H. H. (2016). Roles for retrotransposon insertions in human disease. *Mobile DNA 2016 7:1*, *7*(1), 1–28. https://doi.org/10.1186/S13100-016-0065-9

Hashem, Y., des Georges, A., Dhote, V., Langlois, R., Liao, H. Y., Grassucci, R. A., Hellen, C. U. T., Pestova, T. v., & Frank, J. (2013). Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. *Cell*, *153*(5), 1108. https://doi.org/10.1016/J.CELL.2013.04.036

He, J., Fu, X., Zhang, M., He, F., Li, W., Abdul, M. M., Zhou, J., Sun, L., Chang, C., Li, Y., Liu, H., Wu, K., Babarinde, I. A., Zhuang, Q., Loh, Y. H., Chen, J., Esteban, M. A., & Hutchins, A. P. (2019). Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nature Communications 2019 10:1*, *10*(1), 1–13. https://doi.org/10.1038/s41467-018-08006-y

He, Y., Staley, J. P., Andersen, G. R., & Nielsen, K. H. (2017). Structure of the DEAH/RHA ATPase Prp43p bound to RNA implicates a pair of hairpins and motif Va in translocation along RNA. *RNA (New York, N.Y.)*, *23*(7), 1110–1124. https://doi.org/10.1261/RNA.060954.117

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C. L., Pena, V., Lührmann, R., & Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. *Molecular Cell*, *45*(4), 567–580. https://doi.org/10.1016/J.MOLCEL.2011.12.034

Heininger, A. U., Hackert, P., Andreou, A. Z., Boon, K. L., Memet, I., Prior, M., Clancy, A., Schmidt, B., Urlaub, H., Schleiff, E., Sloan, K. E., Deckers, M., Lührmann, R., Enderlein, J., Klostermeier, D., Rehling, P., & Bohnsack, M. T. (2016). Protein cofactor competition regulates the action of a multifunctional RNA helicase in different pathways. *RNA Biology*, *13*(3), 320–330. https://doi.org/10.1080/15476286.2016.1142038

Herschlag, D. (1995). RNA chaperones and the RNA folding problem. *Journal of Biological Chemistry*, *270*(36), 20871–20874. https://doi.org/10.1074/jbc.270.36.20871

Ho, J. S. Y., di Tullio, F., Schwarz, M., Low, D., Incarnato, D., Gay, F., Tabaglio, T., Zhang, J., Wollman, H., Chen, L., An, O., Chan, T. H. M., Hickman, A. H., Zheng, S., Roudko, V., Chen, S., Karz, A., Ahmed, M., He, H. H., … Guccione, E. (2021). Hnrnpm controls circrna biogenesis and splicing fidelity to sustain cancer cell fitness. *ELife*, *10*. https://doi.org/10.7554/ELIFE.59654

Holmqvist, E., Li, L., Bischler, T., Barquist, L., & Vogel, J. (2018). Global Maps of ProQ Binding In Vivo Reveal Target Recognition via RNA Structure and Stability Control at mRNA 3′ Ends. *Molecular Cell*, *70*(5), 971-982.e6. https://doi.org/10.1016/J.MOLCEL.2018.04.017

Horn, A. v., Celic, I., Dong, C., Martirosyan, I., & Han, J. S. (2017). A conserved role for the ESCRT membrane budding complex in LINE retrotransposition. *PLOS Genetics*, *13*(6), e1006837. https://doi.org/10.1371/JOURNAL.PGEN.1006837

Hornbeck, P. v., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., & Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Research*, *43*(Database issue), D512–D520. https://doi.org/10.1093/NAR/GKU1267

Hu, J., & Ng, P. C. (2012). Predicting the effects of frameshifting indels. *Genome Biology*, *13*(2), 1–11. https://doi.org/10.1186/GB-2012-13-2-R9/FIGURES/4

Hug, N., Aitken, S., Longman, D., Raab, M., Armes, H., Mann, A. R., Rio-Machin, A., Fitzgibbon, J., Rouault-Pierre, K., & Cáceres, J. F. (2022). A dual role for the RNA helicase DHX34 in NMD and pre-mRNA splicing and its function in hematopoietic differentiation. *RNA*, *28*(9), 1224–1238. https://doi.org/10.1261/RNA.079277.122

Hug, N., & Cáceres, J. F. (2014). The RNA helicase DHX34 activates NMD by promoting a transition from the surveillance to the decay-inducing complex. *Cell Reports*, *8*(6), 1845–1856. https://doi.org/10.1016/J.CELREP.2014.08.020

Hug, N., Longman, D., & Cáceres, J. F. (2016). Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research*, *44*(4), 1483. https://doi.org/10.1093/NAR/GKW010

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5

Hwang, S. Y., Jung, H., Mun, S., Lee, S., Park, K., Baek, S. C., Moon, H. C., Kim, H., Kim, B., Choi, Y., Go, Y. H., Tang, W., Choi, J., Choi, J. K., Cha, H. J., Park, H. Y., Liang, P., Kim, V. N., Han, K., & Ahn, K. (2021). L1 retrotransposons exploit RNA m6A modification as an evolutionary driving force. *Nature Communications 2021 12:1*, *12*(1), 1–14. https://doi.org/10.1038/s41467-021-21197-1

Ilagan, J. O., Chalkley, R. J., Burlingame, A. L., & Jurica, M. S. (2013). Rearrangements within human spliceosomes captured after exon ligation. *RNA (New York, N.Y.)*, *19*(3), 400–412. https://doi.org/10.1261/RNA.034223.112

Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. *Trends in Biochemical Sciences*, *36*(1), 19–29. https://doi.org/10.1016/J.TIBS.2010.07.008

Jankowsky, E., & Fairman, M. E. (2007). RNA helicases - one fold for many functions. *Current Opinion in Structural Biology*, *17*(3), 316–324. https://doi.org/10.1016/J.SBI.2007.05.007

Jankowsky, E., & Fairman, M. E. (2008). Duplex unwinding and RNP remodeling with RNA helicases. *Methods in Molecular Biology*, *488*, 343–355. https://doi.org/10.1007/978-1-60327-475-3_22

Jarmoskaite, I., & Russell, R. (2014). RNA helicase proteins as chaperones and remodelers. *Annual Review of Biochemistry*, *83*, 697. https://doi.org/10.1146/ANNUREV-BIOCHEM-060713-035546

Jensen, O., Ansari, S., Gebauer, L., Müller, S. F., Lowjaga, K. A. A. T., Geyer, J., Tzvetkov, M. v., & Brockmöller, J. (2020). A double-Flp-in method for stable overexpression of two genes. *Scientific Reports 2020 10:1*, *10*(1), 1–14. https://doi.org/10.1038/s41598-020-71051-5

Ji, X., Dadon, D. B., Abraham, B. J., Lee, T. I., Jaenisch, R., Bradner, J. E., & Young, R. A. (2015). Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(12), 3841–3846. https://doi.org/10.1073/PNAS.1502971112/-/DCSUPPLEMENTAL

Jimenez-Morales, D., Rosa Campos, A., von Dollen, J., Krogan, N., & Swaney, D. (2022). *artMS: Analytical R tools for Mass Spectrometry* (R Package version 1.16.0, http://artms.org.).

Jin, Y., Tam, O. H., Paniagua, E., & Hammell, M. (2015). TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics (Oxford, England)*, *31*(22), 3593–3599. https://doi.org/10.1093/BIOINFORMATICS/BTV422

Kalyana-Sundaram, S., Kumar-Sinha, C., Shankar, S., Robinson, D. R., Wu, Y. M., Cao, X., Asangani, I. A., Kothari, V., Prensner, J. R., Lonigro, R. J., Iyer, M. K., Barrette, T., Shanmugam, A., Dhanasekaran, S. M., Palanisamy, N., & Chinnaiyan, A. M. (2012). Expressed Pseudogenes in the Transcriptional Landscape of Human Cancers. *Cell*, *149*(7), 1622. https://doi.org/10.1016/J.CELL.2012.04.041

Kandror, O., DeLeon, A., & Goldberg, A. L. (2002). Trehalose synthesis is induced upon exposure of Escherichia coli to cold and is essential for viability at low temperatures. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(15), 9727. https://doi.org/10.1073/PNAS.142314099

Kapitonov, V. v., & Jurka, J. (2007). Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends in Genetics*, *23*(10), 521–529. https://doi.org/10.1016/J.TIG.2007.08.004

Karimi, M., Goldie, L. C., Ulgiati, D., & Abraham, L. J. (2007). Integration site-specific transcriptional reporter gene analysis using Flp recombinase targeted cell lines. *BioTechniques*, *42*(2), 217–224. https://doi.org/10.2144/000112317/ASSET/IMAGES/LARGE/FIGURE6.JPEG

Kato, K., Ahmad, S., Zhu, Z., Young, J. M., Mu, X., Park, S., Malik, H. S., & Hur, S. (2021). Structural analysis of RIG-I-like receptors reveals ancient rules of engagement between diverse RNA helicases and TRIM ubiquitin ligases. *Molecular Cell*, *81*(3), 599-613.e8. https://doi.org/10.1016/J.MOLCEL.2020.11.047

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, and D. (2002). The human genome browser at UCSC. *Genome Research*, *12*(6), 996–1006. https://doi.org/10.1101/GR.229102

Khan, H., Smit, A., & Boissinot, S. (2006). Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Research*, *16*(1), 78. https://doi.org/10.1101/GR.4001406

Khemici, V., & Linder, P. (2018). RNA helicases in RNA decay. *Biochemical Society Transactions*, *46*(1), 163. https://doi.org/10.1042/BST20170052

Komander, D. (2009). The emerging complexity of protein ubiquitination. *Biochemical Society Transactions*, *37*(Pt 5), 937–953. https://doi.org/10.1042/BST0370937

Kozlowski, L. P., & Bujnicki, J. M. (2012). MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics*, *13*(1), 1–11. https://doi.org/10.1186/1471-2105-13-111/TABLES/7

Kragelund, B. B., Weterings, E., Hartmann-Petersen, R., & Keijzers, G. (2016). The Ku70/80 ring in Non-Homologous End-Joining: easy to slip on, hard to remove. *Frontiers in Bioscience (Landmark Edition)*, *21*(3), 514–527. https://doi.org/10.2741/4406

Krepl, M., Pokorná, P., Pokorná, P., Mí Ynsk´yynsk´ Ynsk´y, V., Stadlbauer, P., & Jiříšponer, J. J. (2022). Spontaneous binding of single-stranded RNAs to RRM proteins visualized by unbiased atomistic simulations with a rescaled RNA force field. *Nucleic Acids Research*, *50*(21), 12480–12496. https://doi.org/10.1093/NAR/GKAC1106

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature 1970 227:5259*, *227*(5259), 680–685. https://doi.org/10.1038/227680a0

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., … Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature 2001 409:6822*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Langenberger, D., Bermudez-Santana, C., Hertel, J., Hoffmann, S., Khaitovich, P., & Stadler, P. F. (2009). Evidence for human microRNA-offset RNAs in small RNA sequencing data. *Bioinformatics*, *25*(18), 2298–2301. https://doi.org/10.1093/BIOINFORMATICS/BTP419

Leitner, A., Bonvin, A. M. J. J., Borchers, C. H., Chalkley, R. J., Chamot-Rooke, J., Combe, C. W., Cox, J., Dong, M. Q., Fischer, L., Götze, M., Gozzo, F. C., Heck, A. J. R., Hoopmann, M. R., Huang, L., Ishihama, Y., Jones, A. R., Kalisman, N., Kohlbacher, O., Mechtler, K., … Rappsilber, J. (2020). Toward Increased Reliability, Transparency, and Accessibility in Cross-linking Mass Spectrometry. *Structure (London, England : 1993)*, *28*(11), 1259–1268. https://doi.org/10.1016/J.STR.2020.09.011

Lessel, D., Schob, C., Küry, S., Reijnders, M. R. F., Harel, T., Eldomery, M. K., Coban-Akdemir, Z., Denecke, J., Edvardson, S., Colin, E., Stegmann, A. P. A., Gerkes, E. H., Tessarech, M., Bonneau, D., Barth, M., Besnard, T., Cogné, B., Revah-Politi, A., Strom, T. M., … Kreienkamp, H. J. (2017). De Novo Missense Mutations in DHX30 Impair Global Translation and Cause a Neurodevelopmental Disorder. *The American Journal of Human Genetics*, *101*(5), 716–724. https://doi.org/10.1016/J.AJHG.2017.09.014

Li, S., Prasanna, X., Salo, V. T., Vattulainen, I., & Ikonen, E. (2019). An efficient auxin-inducible degron system with low basal degradation in human cells. *Nature Methods*, *16*(9), 866–869. https://doi.org/10.1038/s41592-019-0512-x

Li, X., Sun, B., Qian, H., Ma, J., Paolino, M., & Zhang, Z. (2022). A high-efficiency and versatile CRISPR/Cas9-mediated HDR-based biallelic editing system. *Journal of Zhejiang University. Science. B*, *23*(2), 141. https://doi.org/10.1631/JZUS.B2100196

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/BIOINFORMATICS/BTT656

Liu, H.-L., & Cheng, S.-C. (2012). The interaction of Prp2 with a defined region of the intron is required for the first splicing reaction. *Molecular and Cellular Biology*, *32*(24), 5056–5066. https://doi.org/10.1128/MCB.01109-12

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., Dai, Q., Chen, W., & He, C. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nature Chemical Biology*, *10*(2), 93–95. https://doi.org/10.1038/NCHEMBIO.1432

Liu, N., Lee, C. H., Swigut, T., Grow, E., Gu, B., Bassik, M. C., & Wysocka, J. (2017). Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature 2018 553:7687*, *553*(7687), 228–232. https://doi.org/10.1038/nature25179

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9

Lyu, K., Chen, S.-B., Chow, E. Y.-C., Zhao, H., Yuan, J.-H., Cai, M., Shi, J., Chan, T.-F., Tan, J.-H., & Kwok, C. K. (2022). An RNA G-Quadruplex Structure within the ADAR 5'UTR Interacts with DHX36 Helicase to Regulate Translation. *Angewandte Chemie (International Ed. in English)*. https://doi.org/10.1002/ANIE.202203553

Manojlovic, Z., & Stefanovic, B. (2012). A novel role of RNA helicase A in regulation of translation of type I collagen mRNAs. *RNA (New York, N.Y.)*, *18*(2), 321–334. https://doi.org/10.1261/RNA.030288.111

Martin, R., Straub, A. U., Doebele, C., & Bohnsack, M. T. (2013). DExD/H-box RNA helicases in ribosome biogenesis. *RNA Biology*, *10*(1), 4–18. https://doi.org/10.4161/RNA.21879

Memet, I., Doebele, C., Sloan, K. E., & Bohnsack, M. T. (2017). The G-patch protein NF-κB-repressing factor mediates the recruitment of the exonuclease XRN2 and activation of the RNA helicase DHX15 in human ribosome biogenesis. *Nucleic Acids Research*, *45*(9), 5359–5374. https://doi.org/10.1093/NAR/GKX013

Miyoshi, T., Makino, T., & Moran, J. v. (2019). Poly(ADP-Ribose) Polymerase 2 Recruits Replication Protein A to Sites of LINE-1 Integration to Facilitate Retrotransposition. *Molecular Cell*, *75*(6), 1286-1298.e12. https://doi.org/10.1016/J.MOLCEL.2019.07.018

Modenini, G., Abondio, P., & Boattini, A. (2022). The coevolution between APOBEC3 and retrotransposons in primates. *Mobile DNA*, *13*(1). https://doi.org/10.1186/S13100-022-00283-1

Mori, T., Itami, S., Yanagi, T., Tatara, Y., Takamiya, M., & Uchida, T. (2009). Use of a real-time fluorescence monitoring system for high-throughput screening for prolyl isomerase inhibitors. *Journal of Biomolecular Screening*, *14*(4), 419–424. https://doi.org/10.1177/1087057109333979

Murat, P., Marsico, G., Herdy, B., Ghanbarian, A., Portella, G., & Balasubramanian, S. (2018). RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biology*, *19*(1). https://doi.org/10.1186/S13059-018-1602-2

Nakajima, T., Uchida, C., Anderson, S. F., Chee-Gun, L., Hurwitz, J., Parvin, J. D., & Montminy, M. (1997). RNA helicase A mediates association of CBP with RNA polymerase II. *Cell*, *90*(6), 1107–1112. https://doi.org/10.1016/S0092-8674(00)80376-1

Nishimura, K., Cho, Y., Tokunaga, K., Nakao, M., Tani, T., & Ideue, T. (2019). DEAH box RNA helicase DHX38 associates with satellite I noncoding RNA involved in chromosome segregation. *Genes to Cells*, *24*(8), 585–590. https://doi.org/10.1111/GTC.12707

Nogueira, G., Fernandes, R., García-Moreno, J. F., & Romão, L. (2021). Nonsense-mediated RNA decay and its bipolar function in cancer. *Molecular Cancer 2021 20:1*, *20*(1), 1–19. https://doi.org/10.1186/S12943-021-01364-0

Oborská-Oplová, M., Fischer, U., Altvater, M., & Panse, V. G. (2022). Eukaryotic Ribosome assembly and Nucleocytoplasmic Transport. *Methods in Molecular Biology*, *2533*, 99–126. https://doi.org/10.1007/978-1-0716-2501-9_7/TABLES/3

Omura, H., Oikawa, D., Nakane, T., Kato, M., Ishii, R., Ishitani, R., Tokunaga, F., & Nureki, O. (2016). Structural and functional analysis of DDX41: A bispecific immune receptor for DNA and cyclic dinucleotide. *Scientific Reports*, *6*. https://doi.org/10.1038/SREP34756

Ostertag, E. M., Luning Prak, E. T., DeBerardinis, R. J., Moran, J. v., & Kazazian, H. H. (2000). Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Research*, *28*(6), 1418. https://doi.org/10.1093/NAR/28.6.1418

Ou, J., Wang, L., Ding, X., Du, J., Zhang, Y., Chen, H., & Xu, A. (2004). Stationary phase protein overproduction is a fundamental capability of Escherichia coli. *Biochemical and Biophysical Research Communications*, *314*(1), 174–180. https://doi.org/10.1016/j.bbrc.2003.12.077

Palmer, E., & Freeman, T. (2004). Investigation Into the use of C- and N-terminal GFP Fusion Proteins for Subcellular Localization Studies Using Reverse Transfection Microarrays. *Comparative and Functional Genomics*, *5*(4), 342. https://doi.org/10.1002/CFG.405

Park, P. J. (2009). ChIP–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics 2009 10:10*, *10*(10), 669–680. https://doi.org/10.1038/nrg2641

Pearce, S. F., Rebelo-Guiomar, P., D'Souza, A. R., Powell, C. A., van Haute, L., & Minczuk, M. (2017). Regulation of Mammalian Mitochondrial Gene Expression: Recent Advances. *Trends in Biochemical Sciences*, *42*(8), 625–639. https://doi.org/10.1016/J.TIBS.2017.02.003

Peña, C., Hurt, E., & Panse, V. G. (2017). Eukaryotic ribosome assembly, transport and quality control. *Nature Structural & Molecular Biology 2017 24:9*, *24*(9), 689–699. https://doi.org/10.1038/nsmb.3454

Penzkofer, T., Jäger, M., Figlerowicz, M., Badge, R., Mundlos, S., Robinson, P. N., & Zemojtel, T. (2017). L1Base 2: more retrotransposition-active LINE-1s, more mammalian genomes. *Nucleic Acids Research*, *45*(Database issue), D68. https://doi.org/10.1093/NAR/GKW925

Pfoh, R., Lacdao, I. K., Georges, A. A., Capar, A., Zheng, H., Frappier, L., & Saridakis, V. (2015). Crystal Structure of USP7 Ubiquitin-like Domains with an ICP0 Peptide Reveals a Novel Mechanism Used by Viral and Cellular Proteins to Target USP7. *PLOS Pathogens*, *11*(6), e1004950. https://doi.org/10.1371/JOURNAL.PPAT.1004950

Philippe, C., Vargas-Landin, D. B., Doucet, A. J., van Essen, D., Vera-Otarola, J., Kuciak, M., Corbin, A., Nigumann, P., & Cristofari, G. (2016). Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *ELife*, *5*(MARCH2016). https://doi.org/10.7554/ELIFE.13926

Platt, R. N., Vandewege, M. W., & Ray, D. A. (2018). Mammalian transposable elements and their impacts on genome evolution. *Chromosome Research*, *26*(1), 25. https://doi.org/10.1007/S10577-017-9570-Z

Pozhidaeva, A., & Bezsonova, I. (2019). USP7: structure, substrate specificity, and inhibition. *DNA Repair*, *76*, 30. https://doi.org/10.1016/J.DNAREP.2019.02.005

Prabu, J. R., Müller, M., Thomae, A. W., Schüssler, S., Bonneau, F., Becker, P. B., & Conti, E. (2015). Structure of the RNA Helicase MLE Reveals the Molecular Mechanisms for Uridine Specificity and RNA-ATP Coupling. *Molecular Cell*, *60*(3), 487–499. https://doi.org/10.1016/J.MOLCEL.2015.10.011

Prasad, S., Khadatare, P. B., & Roy, I. (2011). Effect of chemical chaperones in improving the solubility of recombinant proteins in Escherichia coli. *Applied and Environmental Microbiology*, *77*(13), 4603–4609. https://doi.org/10.1128/AEM.05259-11

Protasova, M. S., Andreeva, T. V., & Rogaev, E. I. (2021). Factors Regulating the Activity of LINE1 Retrotransposons. *Genes*, *12*(10). https://doi.org/10.3390/GENES12101562

Pyle, A. M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. *Annual Review of Biophysics*, *37*, 317–336. https://doi.org/10.1146/ANNUREV.BIOPHYS.37.032807.125908

Radley, E. H., Long, J., Gough, K. C., & Layfield, R. (2019). The 'dark matter' of ubiquitin-mediated processes: opportunities and challenges in the identification of ubiquitin-binding domains. *Biochemical Society Transactions*, *47*(6), 1949–1962. https://doi.org/10.1042/BST20190869

Rajiv, C., & Davis, T. L. (2018). Structural and Functional Insights into Human Nuclear Cyclophilins. *Biomolecules*, *8*(4). https://doi.org/10.3390/BIOM8040161

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols 2013 8:11*, *8*(11), 2281–2308. https://doi.org/10.1038/nprot.2013.143

Ranji, A., & Boris-Lawrie, K. (2010). RNA helicases: Emerging roles in viral replication and the host innate response. *RNA Biology*, *7*(6), 775. https://doi.org/10.4161/RNA.7.6.14249

Renner, T. M., Bélanger, K., Goodwin, L. R., Campbell, M., & Langlois, M. A. (2018). Characterization of molecular attributes that influence LINE-1 restriction by all seven human APOBEC3 proteins. *Virology*, *520*, 127–136. https://doi.org/10.1016/J.VIROL.2018.05.015

Richardson, S. R., Doucet, A. J., Kopera, H. C., Moldovan, J. B., Garcia-Perez, J. L., & Moran, J. v. (2015). The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. *Microbiology Spectrum*, *3*(2). https://doi.org/10.1128/MICROBIOLSPEC.MDNA3-0061-2014

Richardson, S. R., Narvaiza, I., Planegger, R. A., Weitzman, M. D., & Moran, J. v. (2014). APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *ELife*, *2014*(3). https://doi.org/10.7554/ELIFE.02008

Robert-Paganin, J., Réty, S., & Leulliot, N. (2015). Regulation of DEAH/RHA Helicases by G-Patch Proteins. *BioMed Research International*, *2015*. https://doi.org/10.1155/2015/931857

Sales-Lee, J., Perry, D. S., Bowser, B. A., Diedrich, J. K., Rao, B., Beusch, I., Yates, J. R., Roy, S. W., & Madhani, H. D. (2021). Coupling of spliceosome complexity to intron diversity. *Current Biology : CB*, *31*(22), 4898-4910.e4. https://doi.org/10.1016/J.CUB.2021.09.004

Sambrook, J., & Russell, D. (2001). Molecular Cloning: A Laboratory Manual 3rd edn Plainview. *New York: Cold Spring Harbor Laboratories*.

Sauer, M., Juranek, S. A., Marks, J., de Magis, A., Kazemier, H. G., Hilbig, D., Benhalevy, D., Wang, X., Hafner, M., & Paeschke, K. (2019). DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nature Communications 2019 10:1*, *10*(1), 1–15. https://doi.org/10.1038/s41467-019-10432-5

Scharschmidt, B. F., Keeffe, E. B., Blankenship, N. M., & Ockner, R. K. (1979). Validation of a recording spectrophotometric method for measurement of membrane-associated Mg- and NaK-ATPase activity. *The Journal of Laboratory and Clinical Medicine*, *93*(5).

Schmidpeter, P. A. M., & Schmid, F. X. (2015). Prolyl Isomerization and Its Catalysis in Protein Folding and Protein Function. *Journal of Molecular Biology*, *427*(7), 1609–1631. https://doi.org/10.1016/J.JMB.2015.01.023

Sela, N., Mersch, B., Gal-Mark, N., Lev-Maor, G., Hotz-Wagenblatt, A., & Ast, G. (2007). Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. *Genome Biology*, *8*(6). https://doi.org/10.1186/GB-2007-8-6-R127

Semlow, D. R., Blanco, M. R., Walter, N. G., & Staley, J. P. (2016). Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites. *Cell*, *164*(5), 985–998. https://doi.org/10.1016/J.CELL.2016.01.025

Shinriki, S., & Matsui, H. (2022). Unique role of DDX41, a DEAD-box type RNA helicase, in hematopoiesis and leukemogenesis. *Frontiers in Oncology*, *12*, 4677. https://doi.org/10.3389/FONC.2022.992340/BIBTEX

Shoemaker, B. A., Portman, J. J., & Wolynes, P. G. (2000). Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(16), 8868–8873. https://doi.org/10.1073/PNAS.160259697/ASSET/F4B3AB7E-2FFC-4FC0-B3B8-56ED10BA6B5D/ASSETS/GRAPHIC/PQ1602596006.JPEG

Singleton, M. R., Dillingham, M. S., & Wigley, D. B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annual Review of Biochemistry*, *76*, 23–50. https://doi.org/10.1146/ANNUREV.BIOCHEM.76.052305.115300

Sloan, K. E., & Bohnsack, M. T. (2018). Unravelling the Mechanisms of RNA Helicase Regulation. *Trends in Biochemical Sciences*, *43*(4), 237–250. https://doi.org/10.1016/J.TIBS.2018.02.001

Stake, M., Singh, D., Singh, G., Marcela Hernandez, J., Kaddis Maldonado, R., Parent, L. J., & Boris-Lawrie, K. (2015). HIV-1 and two avian retroviral 5′ untranslated regions bind orthologous human and chicken RNA binding proteins. *Virology*, *486*, 307–320. https://doi.org/10.1016/J.VIROL.2015.06.001

Steimer, L., & Klostermeier, D. (2012). RNA helicases in infection and disease. *Https://Doi.Org/10.4161/Rna.20090*, *9*(6), 751–771. https://doi.org/10.4161/RNA.20090

Stok, J. E., Oosenbrug, T., Haar, L. R., Gravekamp, D., Bromley, C. P., Zelenay, S., Reis e Sousa, C., & Veen, A. G. (2022). RNA sensing via the RIG-I-like receptor LGP2 is

essential for the induction of a type I IFN response in ADAR1 deficiency. *The EMBO Journal*, *41*(6). https://doi.org/10.15252/EMBJ.2021109760

Stone, J. (1995). An Efficient Library for Parallel Ray Tracing and Animation. *In Intel Supercomputer Users Group Proceedings*.

Strittmatter, L. M., Capitanchik, C., Newman, A. J., Hallegger, M., Norman, C. M., Fica, S. M., Oubridge, C., Luscombe, N. M., Ule, J., & Nagai, K. (2021). psiCLIP reveals dynamic RNA binding by DEAH-box helicases before and after exon ligation. *Nature Communications 2021 12:1*, *12*(1), 1–15. https://doi.org/10.1038/s41467-021-21745-9

Studer, M. K., Ivanovic, L., Weber, M. E., Marti, S., & Jonas, S. (2020). Structural basis for DEAH-helicase activation by G-patch proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(13), 7159–7170. https://doi.org/10.1073/PNAS.1913880117/SUPPL_FILE/PNAS.1913880117.SAPP.PDF

Sultana, T., van Essen, D., Siol, O., Bailly-Bechet, M., Philippe, C., Zine El Aabidine, A., Pioger, L., Nigumann, P., Saccani, S., Andrau, J. C., Gilbert, N., & Cristofari, G. (2019). The Landscape of L1 Retrotransposons in the Human Genome Is Shaped by Pre-insertion Sequence Biases and Post-insertion Selection. *Molecular Cell*, *74*(3), 555-570.e7. https://doi.org/10.1016/J.MOLCEL.2019.02.036

Sultana, T., Zamborlini, A., Cristofari, G., & Lesage, P. (2017). Integration site selection by retroviruses and transposable elements in eukaryotes. *Nature Reviews Genetics 2017 18:5*, *18*(5), 292–308. https://doi.org/10.1038/nrg.2017.7

Sun, M., & Zhang, X. (2022). Current methodologies in protein ubiquitination characterization: from ubiquitinated protein to ubiquitin chain architecture. *Cell & Bioscience 2022 12:1*, *12*(1), 1–17. https://doi.org/10.1186/S13578-022-00870-Y

Swatek, K. N., & Komander, D. (2016). Ubiquitin modifications. *Cell Research 2016 26:4*, *26*(4), 399–422. https://doi.org/10.1038/cr.2016.39

Sweeney, T. R., Dhote, V., Guca, E., Hellen, C. U. T., Hashem, Y., & Pestova, T. v. (2021). Functional role and ribosomal position of the unique N-terminal region of DHX29, a factor required for initiation on structured mammalian mRNAs. *Nucleic Acids Research*, *49*(22), 12955–12969. https://doi.org/10.1093/NAR/GKAB1192

Szczesny, R. J., Kowalska, K., Klosowska-Kosicka, K., Chlebowski, A., Owczarek, E. P., Warkocki, Z., Kulinski, T. M., Adamska, D., Affek, K., Jedroszkowiak, A., Kotrys, A. v., Tomecki, R., Krawczyk, P. S., Borowski, L. S., & Dziembowski, A. (2018). Versatile approach for functional analysis of human proteins and efficient stable cell line generation

using flp-mediated recombination system. *PLoS ONE*, *13*(3). https://doi.org/10.1371/journal.pone.0194887

Tanner, N. K., & Linder, P. (2001). DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Molecular Cell*, *8*(2), 251–262. https://doi.org/10.1016/S1097-2765(01)00329-X

Tauchert, M. J., Fourmann, J. B., Lührmann, R., & Ficner, R. (2017). Structural insights into the mechanism of the DEAH-box RNA helicase Prp43. *ELife*, *6*. https://doi.org/10.7554/ELIFE.21510

Taylor, M. S., Altukhov, I., Molloy, K. R., Mita, P., Jiang, H., Adney, E. M., Wudzinska, A., Badri, S., Ischenko, D., Eng, G., Burns, K. H., Fenyö, D., Chait, B. T., Alexeev, D., Rout, M. P., Boeke, J. D., & LaCava, J. (2018). Dissection of affinity captured LINE-1 macromolecular complexes. *ELife*, *7*. https://doi.org/10.7554/ELIFE.30094

Team, R. C. (2021). *R: A language and environment for statistical computing v. 3.6. 1 (R Foundation for Statistical Computing, Vienna, Austria, 2019)*. Scientific Reports.

Teo, G., Liu, G., Zhang, J., Nesvizhskii, A. I., Gingras, A. C., & Choi, H. (2014). SAINTexpress: improvements and additional features in Significance Analysis of Interactome software. *Journal of Proteomics*, *100*, 37. https://doi.org/10.1016/J.JPROT.2013.10.023

Thomas, P., & Smart, T. G. (2005). HEK293 cell line: a vehicle for the expression of recombinant proteins. *Journal of Pharmacological and Toxicological Methods*, *51*(3), 187–200. https://doi.org/10.1016/J.VASCN.2004.08.014

Thul, P. J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L. M., Bäckström, A., Danielsson, F., Fagerberg, L., Fall, J., Gatto, L., Gnann, C., Hober, S., Hjelmare, M., Johansson, F., … Lundberg, E. (2017). A subcellular map of the human proteome. *Science*, *356*(6340). https://doi.org/10.1126/SCIENCE.AAL3321/SUPPL_FILE/AAL3321_THUL_SM_TABLE_S9.XLSX

Tong, C., & Yin, Y. (2021). Localization of RNAs in the nucleus: cis- and trans- regulation. *RNA Biology*, *18*(12), 2073–2086. https://doi.org/10.1080/15476286.2021.1894025

Troskie, R. L., Faulkner, G. J., & Cheetham, S. W. (2021). Processed pseudogenes: A substrate for evolutionary innovation. *BioEssays*, *43*(11), 2100186. https://doi.org/10.1002/BIES.202100186

Tuck, A. C., Rankova, A., Arpat, A. B., Liechti, L. A., Hess, D., Iesmantavicius, V., Castelo-Szekely, V., Gatfield, D., & Bühler, M. (2020). Mammalian RNA Decay Pathways Are Highly Specialized and Widely Linked to Translation. *Molecular Cell*, *77*(6), 1222-1236.e13. https://doi.org/10.1016/J.MOLCEL.2020.01.007

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I. M., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A. K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., … Pontén, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science (New York, N.Y.)*, *347*(6220). https://doi.org/10.1126/SCIENCE.1260419

van Nostrand, E. L., Pratt, G. A., Yee, B. A., Wheeler, E. C., Blue, S. M., Mueller, J., Park, S. S., Garcia, K. E., Gelboin-Burkhart, C., Nguyen, T. B., Rabano, I., Stanton, R., Sundararaman, B., Wang, R., Fu, X. D., Graveley, B. R., & Yeo, G. W. (2020). Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biology*, *21*(1). https://doi.org/10.1186/S13059-020-01982-9

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*, *136*(4), 701–718. https://doi.org/10.1016/J.CELL.2009.02.009

Walbott, H., Mouffok, S., Capeyrou, R., Lebaron, S., Humbert, O., van Tilbeurgh, H., Henry, Y., & Leulliot, N. (2010). Prp43p contains a processive helicase structural architecture with a specific regulatory domain. *The EMBO Journal*, *29*(13), 2194–2204. https://doi.org/10.1038/EMBOJ.2010.102

Wang, H. T., & Hur, S. (2021). Substrate recognition by TRIM and TRIM-like proteins in innate immunity. *Seminars in Cell & Developmental Biology*, *111*, 76. https://doi.org/10.1016/J.SEMCDB.2020.09.013

Wang, L., Johnson, Z. L., Wasserman, M. R., Levring, J., Chen, J., & Liu, S. (2020). Characterization of the kinetic cycle of an ABC transporter by single-molecule and cryo-EM analyses. *ELife*, *9*. https://doi.org/10.7554/eLife.56451

Ward, J. R., Vasu, K., Deutschman, E., Halawani, D., Larson, P. A., Zhang, D., Willard, B., Fox, P. L., Moran, J. v., & Longworth, M. S. (2017). Condensin II and GAIT complexes cooperate to restrict LINE-1 retrotransposition in epithelial cells. *PLoS Genetics*, *13*(10). https://doi.org/10.1371/JOURNAL.PGEN.1007051

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/NAR/GKY427

Webb, S., Hector, R. D., Kudla, G., & Granneman, S. (2014). PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biology*, *15*(1). https://doi.org/10.1186/gb-2014-15-1-r8

Weng, L., Mitoma, H., Tricot, C., Bao, M., Liu, Y., Zhang, Z., & Liu, Y.-J. (2014). The E3 ubiquitin ligase TRIM33 is essential for cytosolic RNA-induced NLRP3 inflammasome activation. *Journal of Immunology (Baltimore, Md.: 1950)*, *193*(7), 3676. https://doi.org/10.4049/JIMMUNOL.1401448

Wilkinson, M. E., Charenton, C., & Nagai, K. (2020). RNA Splicing by the Spliceosome. *Https://Doi.Org/10.1146/Annurev-Biochem-091719-064225*, *89*, 359–388. https://doi.org/10.1146/ANNUREV-BIOCHEM-091719-064225

Williams, F. P., Haubrich, K., Perez-Borrajero, C., & Hennig, J. (2019). Emerging RNA-binding roles in the TRIM family of ubiquitin ligases. *Biological Chemistry*, *400*(11), 1443–1464. https://doi.org/10.1515/HSZ-2019-0158/ASSET/GRAPHIC/J_HSZ-2019-0158_FIG_003.JPG

Xie, Y., Li, H., Luo, X., Li, H., Gao, Q., Zhang, L., Teng, Y., Zhao, Q., Zuo, Z., & Ren, J. (2022). IBS 2.0: an upgraded illustrator for the visualization of biological sequences. *Nucleic Acids Research*, *50*(W1), W420–W426. https://doi.org/10.1093/NAR/GKAC373

Xie, Y., Rosser, J. M., Thompson, T. L., Boeke, J. D., & An, W. (2011). Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. *Nucleic Acids Research*, *39*(3), e16–e16. https://doi.org/10.1093/NAR/GKQ1076

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*. https://doi.org/10.1186/1471-2105-13-134

Yoshimoto, R., Kataoka, N., Okawa, K., & Ohno, M. (2009). Isolation and characterization of post-splicing lariat-intron complexes. *Nucleic Acids Research*, *37*(3), 891–902. https://doi.org/10.1093/NAR/GKN1002

Yu, C., Rao, D., Wang, T., Song, J., Zhang, L., & Huang, W. (2022). Emerging roles of TRIM27 in cancer and other human diseases. *Frontiers in Cell and Developmental Biology*, *10*. https://doi.org/10.3389/FCELL.2022.1004429

Zhan, X., Yan, C., Zhang, X., Lei, J., & Shi, Y. (2018). Structure of a human catalytic step I spliceosome. *Science (New York, N.Y.)*, *359*(6375), 537–545. https://doi.org/10.1126/SCIENCE.AAR6401

Zhang, J. P., Li, X. L., Li, G. H., Chen, W., Arakaki, C., Botimer, G. D., Baylink, D., Zhang, L., Wen, W., Fu, Y. W., Xu, J., Chun, N., Yuan, W., Cheng, T., & Zhang, X. B. (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biology*, *18*(1), 1–18. https://doi.org/10.1186/S13059-017-1164-8/FIGURES/6

Zhang, X., Zhang, R., & Yu, J. (2020). New Understanding of the Relevant Role of LINE-1 Retrotransposition in Human Disease and Immune Modulation. *Frontiers in Cell and Developmental Biology*, *8*, 657. https://doi.org/10.3389/FCELL.2020.00657/BIBTEX

Zhang, Y., You, J., Wang, X., & Weber, J. (2015). The DHX33 RNA Helicase Promotes mRNA Translation Initiation. *Molecular and Cellular Biology*, *35*(17), 2918–2931. https://doi.org/10.1128/MCB.00315-15

Zhang, Z., Bao, M., Lu, N., Weng, L., Yuan, B., & Liu, Y. J. (2013). The E3 ubiquitin ligase TRIM21 negatively regulates the innate immune response to intracellular double-stranded DNA. *Nature Immunology*, *14*(2), 172–178. https://doi.org/10.1038/NI.2492

Zheng, L., Baumann, U., & Reymond, J. L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research*, *32*(14). https://doi.org/10.1093/nar/gnh110

Zhou, Z., & Reed, R. (1998). Human homologs of yeast Prp16 and Prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. *EMBO Journal*, *17*(7), 2095–2106. https://doi.org/10.1093/EMBOJ/17.7.2095

Zhu, Q., Tan, P., Li, Y., Lin, M., Li, C., Mao, J., Cui, J., Zhao, W., Wang, H. Y., & Wang, R. F. (2018). DHX29 functions as an RNA co-sensor for MDA5-mediated EMCV-specific antiviral immunity. *PLoS Pathogens*, *14*(2). https://doi.org/10.1371/JOURNAL.PPAT.1006886

Supplementary information

Supplementary table 1 Relative expressions of DHX mRNAs after RNAi-mediated knockdown.

	Relative expression (%)				
siRNA	Replicate 1	Replicate 2	Replicate 3	Mean	SEM
siDHX9-1	11.03	11.02	11.06	11.0	0.0
siDHX9-2	2.8	7.88	5.03	5.2	1.5
siDHX16-1	9	28		18.5	9.5
siDHX16-2	22	18	16	18.7	1.8
siDHX29-1	0.017	4.93		2.5	2.5
siDHX29-2	0.05	0.98		0.5	0.5
siDHX30-1	4	4	5	4.3	0.3
siDHX30-2	3	3	6	4.0	1.0
siDHX32-1	15	22		18.5	3.5
siDHX32-2	0	1	6	2.3	1.9
siDHX33-1	35	35		35.0	0.0
siDHX33-2	33	33		33.0	0.0
siDHX36-1	3	4	3	3.3	0.3
siDHX36-2	1	4	1	2.0	1.0
siDHX38-1	58	65		61.5	3.5
siDHX38-2	64	52	31	49.0	9.6
siDHX40-1	3	13	9	8.3	2.9
siDHX40-2	14	14		14.0	0.0
siDHX57-1	17	17		17.0	0.0
siDHX57-2	9	8		8.5	0.5

Supplementary table 2 Western blot quantification for tetracycline induction of DHX40-His$_6$Prc2xFLAG in ΔDHX40-1 *The anti-DHX40 signal was quantified by Image Studio and normalised to the anti-tubulin signal.*

Image Name	Name	Tetracycline conc.	Normalized DHX40 signal
0011100_02	HEK293	-	73007.52
0011100_02	HEK293	-	76785.12
0011100_02	ΔDHX40-1_DHX40-FLAG	20	10537.06
0011100_02	ΔDHX40-1_DHX40-FLAG	20	11165.43
0011100_02	ΔDHX40-1_DHX40-FLAG	40	44700.71
0011100_02	ΔDHX40-1_DHX40-FLAG	40	47778.67
0011100_02	ΔDHX40-1_DHX40-FLAG	60	64709.35
0011100_02	ΔDHX40-1_DHX40-FLAG	60	62311.21
0011100_02	ΔDHX40-1_DHX40-FLAG	80	67237.17
0011100_02	ΔDHX40-1_DHX40-FLAG	80	78418.99

Supplementary table 3 Western blot quantification for tetracycline induction of DHX40_E174Q-His$_6$Prc2xFLAG in ΔDHX40-1 *The anti-DHX40 signal was quantified by Image Studio and normalised to the anti-tubulin signal.*

Image Name	Name		Normalized DHX40 signal
0011099_03	HEK293	-	50183.31
0011099_03	HEK293	-	44968.94
0011099_03	ΔDHX40-1_DHX40_E174Q-FLAG	20	12505.05
0011099_03	ΔDHX40-1_DHX40_E174Q-FLAG	20	11346.84
0011099_03	ΔDHX40-1_DHX40_E174Q-FLAG	40	46123.66
0011099_03	ΔDHX40-1_DHX40_E174Q-FLAG	40	45435.30
0011099_03	ΔDHX40-1_DHX40_E174Q-FLAG	60	91492.81
0011099_03	ΔDHX40-1_DHX40_E174Q-FLAG	60	82607.69
0011099_03	ΔDHX40-1_DHX40_E174Q-FLAG	80	112782.98
0011099_03	ΔDHX40-1_DHX40_E174Q-FLAG	80	111007.47

Supplementary table 4: Top 50 prey proteins from DHX40-His6Prc2xFLAG anti-FLAG Immunoprecipitation. *Normalisation and analysis were performed in SAINTexpress. The top 50 prey protein, according to AvgSpec, are shown; the complete table is available in the supplementary zip file.*

Protein	Spec	SpecSum	AvgSpec	ctrlCounts	SaintScore	logOddsScore	FoldChange	BFDR	ENSEMBL
USP7	437|445	882	441	19|20	1	485.75	22.62	0	ENSG00000187555
DHX40	364|325	689	344.5	0|0	1	465.2	3445	0	ENSG00000108406
PPIL4	228|206	434	217	2|2	1	278.31	108.5	0	ENSG00000131013
NCL	176|171	347	173.5	27|42	1	26.6	5.03	0	ENSG00000115053
DDX21	164|156	320	160	0|11	1	14.12	29.09	0	ENSG00000165732
GPATCH1	131|112	243	121.5	0|0	1	164.86	1215	0	ENSG00000076650
XRCC5	123|107	230	115	22|20	1	61.37	5.48	0	ENSG00000079246
RPL4	107|123	230	115	1|2	1	143.05	76.67	0	ENSG00000174444
HNRNPM	102|96	198	99	17|24	1	53.94	4.83	0	ENSG00000099783
RPS3	103|89	192	96	22|24	1	37.14	4.17	0	ENSG00000149273
XRCC6	112|71	183	91.5	14|17	1	37.37	5.9	0	ENSG00000196419
RPS3A	91|92	183	91.5	11|15	1	65.66	7.04	0	ENSG00000145425
MYBBP1A	72|107	179	89.5	0|3	1	24.83	59.67	0	ENSG00000132382
PDCD11	80|86	166	83	2|7	1	25.52	18.44	0	ENSG00000148843
RPL3	73|92	165	82.5	0|1	1	103.86	165	0	ENSG00000100316
RPL7A	78|81	159	79.5	0|1	1	110.96	159	0	ENSG00000148303

SSB	78|79	157	78.5	5|5	1	80.75	15.7	0	ENSG00000138385
GTPBP4	81|70	151	75.5	0|0	1	105.25	755	0	ENSG00000107937
HADHA	68|78	146	73	11|21	1	10.61	4.56	0	ENSG00000084754
RPL5	77|69	146	73	10|19	1	13.71	5.03	0	ENSG00000122406
NOP2	76|69	145	72.5	0|6	1	12.32	24.17	0	ENSG00000111641
ILF3	75|70	145	72.5	18|24	1	21.16	3.45	0	ENSG00000129351
RRP1B	71|72	143	71.5	0|0	1	106.67	715	0	ENSG00000160208
RPL7	76|63	139	69.5	6|10	1	49.21	8.69	0	ENSG00000147604
RSL1D1	70|62	132	66	0|0	1	93.85	660	0	ENSG00000171490
H4C9	58|72	130	65	11|7	1	40.23	7.22	0	ENSG00000276180
RPS8	72|57	129	64.5	3|2	1	66.07	25.8	0	ENSG00000142937
NPM1	59|69	128	64	19|19	1	14.15	3.37	0	ENSG00000181163
SERBP1	59|67	126	63	20|17	1	15.65	3.41	0	ENSG00000142864
DHX30	54|71	125	62.5	8|13	1	33.73	5.95	0	ENSG00000132153
NAT10	61|62	123	61.5	0|0	1	92.42	615	0	ENSG00000135372
DHX35	58|64	122	61	0|0	1	88.14	610	0	ENSG00000101452
RPS9	65|56	121	60.5	1|1	1	75.08	60.5	0	ENSG00000170889
RPS18	58|63	121	60.5	2|1	1	74.08	40.33	0	ENSG00000231500
EXOSC10	61|59	120	60	1|11	1	7.75	10	0	ENSG00000171824
YBX3	59|59	118	59	2|8	1	14.08	11.8	0	ENSG00000060138

YBX1	57|61	118	59	3|8	1	20.98	10.73	0	ENSG00000065978
HNRNPA3	56|59	115	57.5	19|20	1	9.01	2.95	0	ENSG00000170144
YTHDC2	52|61	113	56.5	2|10	1	8.79	9.42	0	ENSG00000047188
RPS2	54|56	110	55	1|4	1	32.29	22	0	ENSG00000140988
RPS7	58|52	110	55	6|8	1	39.57	7.86	0	ENSG00000171863
RBM28	57|53	110	55	1|6	1	13.98	15.71	0	ENSG00000106344
RPS6	52|55	107	53.5	2|2	1	62.26	26.75	0	ENSG00000137154
DDX54	53|54	107	53.5	0|0	1	81	535	0	ENSG00000123064
DDX24	53|53	106	53	0|0	1	81	530	0	ENSG00000089737
PARP1	54|51	105	52.5	10|20	0.99	3.8	3.5	0	ENSG00000143799
DHX36	45|60	105	52.5	0|2	1	26.66	52.5	0	ENSG00000174953
IGF2BP1	60|43	103	51.5	10|13	1	17.48	4.48	0	ENSG00000159217
DDX18	49|53	102	51	0|0	1	75.27	510	0	ENSG00000088205
RPL23A	53|48	101	50.5	1|1	1	63.74	50.5	0	ENSG00000198242

Supplementary table 5: Top 50 prey proteins from DHX40_E174Q-His6Prc2xFLAG anti-FLAG Immunoprecipitation. *Normalisation and analysis were performed in SAINTexpress. The top 50 prey protein, according to AvgSpec, are shown; the complete table is available in the supplementary zip file.*

Prey	Spec	SpecSum	AvgSpec	ctrlCounts	SaintScore	logOddsScore	FoldChange	BFDR	ENSEMBL
USP7	242|343	825	412.7	19|20	1	227.13	21.16	0	ENSG00000187555
DHX40	178|194	524	262.44	0|0	1	258.12	2624.38	0	ENSG00000108406
NCL	161|178	478	239.16	27|42	1	23.95	6.93	0	ENSG00000115053
GPATCH1	143|181	457	228.58	0|0	1	208.7	2285.75	0	ENSG00000076650
PPIL4	107|185	412	206	2|2	1	139.16	103	0	ENSG00000131013
DDX21	137|150	404	202.47	0|11	1	12.85	36.81	0	ENSG00000165732
XRCC6	83|126	294	147.44	14|17	1	49.83	9.51	0	ENSG00000196419
XRCC5	86|84	239	119.93	22|20	1	37.05	5.71	0	ENSG00000079246
RPS3	73|88	227	113.58	22|24	1	18.76	4.94	0	ENSG00000149273
MYBBP1A	70|89	224	112.17	0|3	1	24.22	74.78	0	ENSG00000132382
HNRNPM	66|86	214	107.23	17|24	1	19.03	5.23	0	ENSG00000099783
NOP2	74|65	196	98.06	0|6	1	11.77	32.69	0	ENSG00000111641
PDCD11	58|76	189	94.53	2|7	1	18.21	21.01	0	ENSG00000148843
RPL5	62|69	184	92.42	10|19	1	11.12	6.38	0	ENSG00000122406
RPS3A	63|63	177	88.89	11|15	1	35.14	6.84	0	ENSG00000145425
SSB	51|71	172	86.07	5|5	1	45.69	17.21	0	ENSG00000138385
RPL4	45|76	170	85.36	1|2	1	55.85	56.9	0	ENSG00000174444

RBM28	67|54	170	85.36	1|6	1	14.22	24.4	0	ENSG00000106344
WDR83	55|63	166	83.25	0|0	1	83.86	832.46	0	ENSG00000123154
DHX35	45|73	166	83.25	0|0	1	69.53	832.46	0	ENSG00000101452
ILF3	63|54	165	82.54	18|24	0.93	1.85	3.94	0	ENSG00000129351
RSL1D1	51|65	163	81.84	0|0	1	78.13	818.35	0	ENSG00000171490
GTPBP4	43|70	159	79.72	0|0	1	66.66	797.19	0	ENSG00000107937
NPM1	53|58	156	78.31	19|19	1	6.9	4.12	0	ENSG00000181163
SYNCRIP	37|72	153	76.9	9|10	1	16.3	8.1	0	ENSG00000135316
NAT10	46|63	153	76.9	0|0	1	70.97	768.97	0	ENSG00000135372
NOP56	46|61	150	75.49	0|0	1	70.97	754.86	0	ENSG00000101361
RRP1B	52|53	148	74.08	0|0	1	79.56	740.75	0	ENSG00000160208
EXOSC10	42|62	146	73.37	1|11	1	5.95	12.23	0	ENSG00000171824
YBX3	44|57	142	71.25	2|8	1	10.42	14.25	0	ENSG00000060138
IGF2BP1	37|63	141	70.55	10|13	1	10.54	6.14	0	ENSG00000159217
RPS11	35|64	139	69.84	3|8	1	11.47	12.7	0	ENSG00000142534
RPS18	34|64	138	69.14	2|1	1	40.51	46.1	0	ENSG00000231500
MKI67	46|51	136	68.43	0|0	1	70.97	684.31	0	ENSG00000148773
DHX30	45|51	135	67.73	8|13	1	23.65	6.45	0	ENSG00000132153
H4C9	34|61	134	67.02	11|7	1	14.33	7.45	0	ENSG00000276180
YTHDC2	44|51	134	67.02	2|10	1	7.51	11.17	0	ENSG00000047188

YBX1	47|46	131	65.61	3|8	1	16.21	11.92	0	ENSG00000065978
RPS8	33|59	129	64.9	3|2	1	33.59	25.96	0	ENSG00000142937
RPS16	38|53	128	64.2	5|8	1	25.13	9.88	0	ENSG00000105193
RPL10A	32|58	126	63.49	4|5	1	24.33	14.11	0	ENSG00000198755
CHD4	38|52	126	63.49	4|6	1	29.74	12.7	0	ENSG00000111642
DDX24	37|52	125	62.79	0|0	1	58.03	627.88	0	ENSG00000089737
RBM19	47|42	125	62.79	1|5	1	15.04	20.92	0	ENSG00000122965
DDX54	33|55	124	62.08	0|0	1	52.25	620.82	0	ENSG00000123064
RPL3	38|49	122	61.38	0|1	1	53.96	122.75	0	ENSG00000100316
RPL7A	34|53	122	61.38	0|1	1	48.22	122.75	0	ENSG00000148303
CDC5L	40|46	121	60.67	4|14	1	5.33	6.74	0	ENSG00000096401
GNL3	29|56	119	59.97	0|0	1	46.45	599.66	0	ENSG00000163938
SRRM2	36|49	119	59.97	5|16	0.95	2.15	5.71	0	ENSG00000167978

Supplementary table 6 Western blot quantification of anti-DHX40 and anti-USP7 signal in siUSP7-treated samples and control. *The anti-DHX40 and anti-USP7 signals were quantified by Image Studio and normalised to the anti-tubulin signal. Expression in the treated sample relative to siNT control was calculated for the indicated proteins.*

Image Name	Antibody	Sample (replicate)	Signal	Normalised signal	Relative expression
0015382_01	DHX40	siNT(1)	1210000	1210000.00	100%
0015382_01	DHX40	siUSP7 (1)	301000	598395.21	49%
0015382_01	DHX40	siNT (2)	1440000	2034382.98	100%
0015382_01	DHX40	siUSP7 (2)	389000	535883.82	26%
0015382_01	DHX40	siNT (3)	1490000	1766714.29	100%
0015382_01	DHX40	siUSP7 (3)	643000	675556.96	38%
0015382_01	USP7	siNT (1)	3460000	3460000.00	100%
0015382_01	USP7	siUSP7 (1)	265000	526826.35	15%
0015382_01	USP7	siNT (2)	3790000	5354382.98	100%
0015382_01	USP7	siUSP7 (2)	341000	469759.34	9%
0015382_01	USP7	siNT (3)	3330000	3948428.57	100%
0015382_01	USP7	siUSP7 (3)	515000	541075.95	14%

Supplementary table 7 Summary of western blot quantification of anti-DHX40 and anti-USP7 signal in siUSP7-treated samples and control. *Summary of the data in Supplementary table 6.*

	Relative signal in siUSP7-treated sample		
	Tubulin	**DHX40**	**USP7**
Replicate 1	100.0%	49.5%	15.2%
Replicate 2	100.0%	26.3%	8.8%
Replicate 3	100.0%	38.2%	13.7%
Mean	**100.0%**	**38%**	**13%**
SEM	**0.0%**	**7%**	**2%**

Supplementary table 8 ATPase activity measured for recombinantly expressed and purified proteins. *The table shows the ATPase activity for three independent replicates.*

	ATPase activity (pmol/min)											
RNA	DHX40			DHX40^{E174Q}			PPIL4			DHX40 + PPIL4		
(-)	169.3	167.1	188.0	86.6	63.4	58.3	63.2	89.8	62.0	673.0	376.4	570.5
(+)	373.2	247.6	258.9	105.4	83.5	51.9	128.6	77.2	121.8	1012.4	894.7	760.8

Supplementary table 9 Summary of ATPase activity measured for recombinantly expressed and purified proteins. *Summary of the data in Supplementary table 8. Analysis was perfomed in GraphPad Prism.*

	DHX40			DHX40^{E174Q}			PPIL4			DHX40 + PPIL4		
RNA	Mean	SEM	N	Mean	SEM	N	Mean	SEM	N	Mean	SEM	N
(-)	174.82	6.62	3	69.43	8.70	3	71.63	9.06	3	539.97	86.98	3
(+)	293.23	40.13	3	80.25	15.54	3	109.19	16.10	3	889.30	72.67	3

Supplementary table 10 Two-way ANOVA analysis of ATPase activities measured for recombinantly expressed and purified proteins. *Two-way ANOVA analysis of the data in Supplementary table 8. Analysis was perfomed in GraphPad Prism.*

Two-way ANOVA	Ordinary				
Alpha	0,05				
Source of Variation	% of total variation	P value	P value summary	Significant?	
Interaction	5,57	<0,01	**	Yes	
RNA	5,23	<0,01	***	Yes	
Protein	84,4	<0,01	****	Yes	
ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Interaction	106482	3	35494	F (3, 16) = 6,24	P<0,01
RNA	99898	1	99898	F (1, 16) = 17,6	P<0,01
Protein	1614201	3	538067	F (3, 16) = 94,7	P<0,01
Residual	90954	16	5685		

Difference between row means					
Mean of - RNA	214				
Mean of + RNA	343				
Difference between means	-129				
SE of difference	30,8				
95% CI of difference	-194 to -63,8				
Data summary					
Number of columns (Protein)	4				
Number of rows (RNA)	2				
Number of values	24				

Supplementary table 11 Tukey's multiple comparisons test for ATPase activities measured for recombinantly expressed and purified proteins. *Analysis of the data in Supplementary table 8 and Supplementary table 10. Analysis was perfomed in GraphPad Prism.*

Tukey's multiple comparisons test	Mean Diff,	95,00% CI of diff,	Below threshold?	Summary	Adjusted P Value
- RNA					
DHX40 vs. DHX40^{E174Q}	105	-70,8 to 281	No	ns	0.35
DHX40 vs. PPIL4	103	-73,0 to 279	No	ns	0.37
DHX40 vs. DHX40 + PPIL4	-365	-541 to -189	Yes	***	<0,01
DHX40^{E174Q} vs. PPIL4	-2.23	-178 to 174	No	ns	>0,99
DHX40^{E174Q} vs. DHX40 + PPIL4	-471	-647 to -294	Yes	****	<0,01
PPIL4 vs. DHX40 + PPIL4	-468	-644 to -292	Yes	****	<0,01
+ RNA					
DHX40 vs. DHX40^{E174Q}	213	36,8 to 389	Yes	*	0.02
DHX40 vs. PPIL4	184	7,91 to 360	Yes	*	0.04
DHX40 vs. DHX40 + PPIL4	-596	-772 to -420	Yes	****	<0,01
DHX40^{E174Q} vs. PPIL4	-28.9	-205 to 147	No	ns	0.96
DHX40^{E174Q} vs. DHX40 + PPIL4	-809	-985 to -633	Yes	****	<0,01
PPIL4 vs. DHX40 + PPIL4	-780	-956 to -604	Yes	****	<0,01

Supplementary table 12 Relative biotype distribution of significant peaks in the DHX40 CRAC dataset. *The table shows the number and the relative distribution of reads assigned to signification peaks. A gene ID list is available in the supplementary zip file. The CRAC bioinformatic analysis was performed in collaboration with Nicolás Lemus, MD PhD*

Biotype	Normalised reads	Normalised reads (%)
protein_coding	393439.29	79.78%
lncRNA	55441.58	11.24%
pseudogene	18365.06	3.72%
misc_RNA	7088.57	1.44%
miRNA	6699.74	1.36%
snoRNA	6647.08	1.35%
snRNA	4135.54	0.84%
scRNA	600.7	0.12%
TEC	322.89	0.07%
scaRNA	254.5	0.05%
ribozyme	171.5	0.03%
rRNA	4.36	0.00%
Total	**493170.81**	**100%**

Supplementary table 13 Relative distribution of significant peaks in different pseudogenes in the DHX40 CRAC dataset. *The table shows the number and the relative distribution of reads assigned to pseudogenes. A gene ID list is available in the supplementary zip file. The CRAC bioinformatic analysis was performed in collaboration with Nicolás Lemus, MD PhD*

Biotype: Pseudogenes	Normalised reads	Normalised reads (%)
unprocessed_pseudogene	2163.82	11.78%
transcribed_unprocessed_pseudogene	6968.93	37.95%
transcribed_processed_pseudogene	1491.18	8.12%
transcribed_unitary_pseudogene	375.25	2.04%
rRNA_pseudogene	3172.89	17.28%
processed_pseudogene	4192.99	22.83%
Total	**18365.06**	**100.00%**

Supplementary table 14 Relative distribution of significant peaks within exon and introns in the DHX40 CRAC dataset. *The table shows the number and the relative distribution of reads assigned to exon and intron. A gene ID list is available in the supplementary zip file. The CRAC bioinformatic analysis was performed in collaboration with Nicolás Lemus, MD PhD*

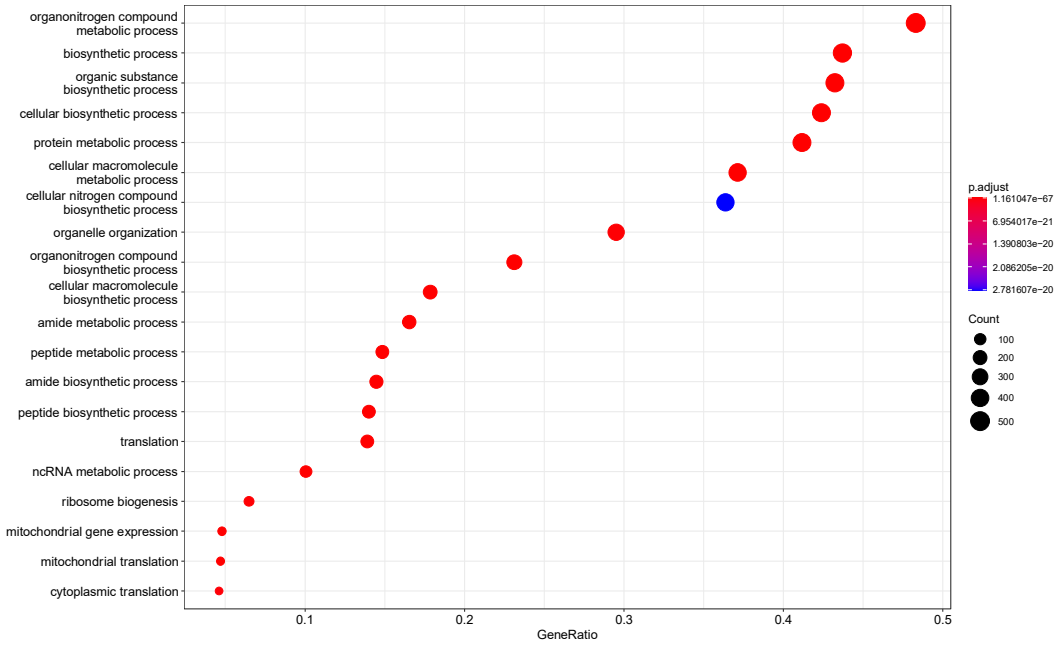
Biotype: protein coding	**Normalised reads**	**Normalised reads (%)**
intron	289933.9	73.69%
ambigous	56873.46	14.46%
exon	46631.93	11.85%
Total	**393439.29**	**100.00%**

Supplementary table 15 Distribution of significant DHX40 CRAC reads mapping to TE classes and families. *The table shows the number and the relative distribution of reads assigned to TE classes. Reads were annotated using hg38_rmsk_TE.gtf. A gene ID list is available in the supplementary zip file. The CRAC bioinformatic analysis was performed in collaboration with Nicolás Lemus, MD PhD*

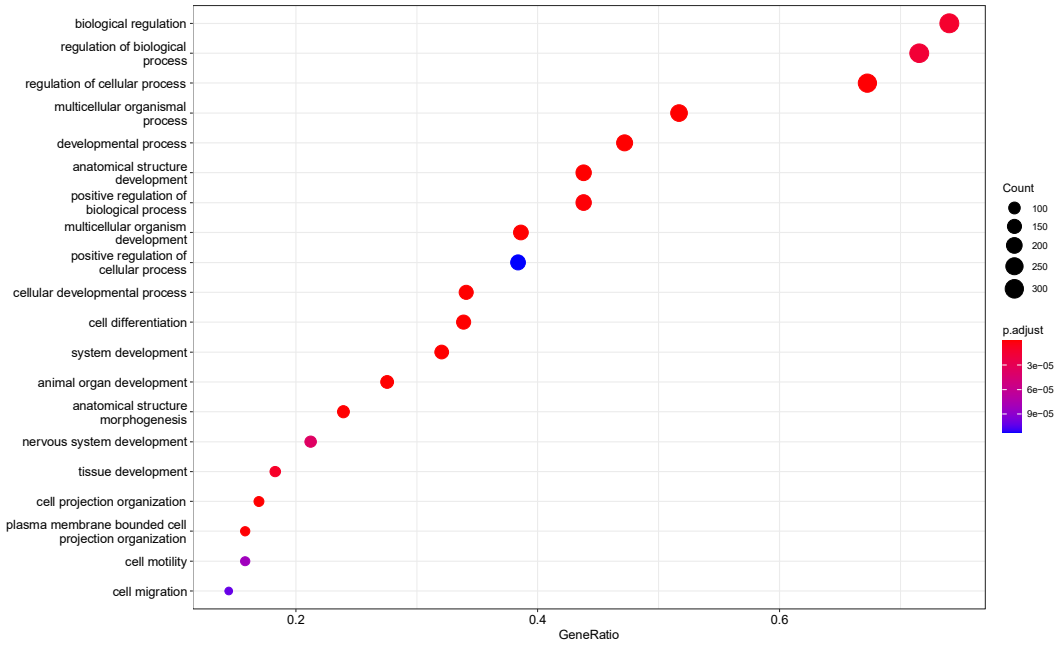
TE - class	**Normalised reads**	**Normalised reads (%)**
SINE	42742.31	41.07%
LINE	34998.62	33.63%
LTR	11857.34	11.39%
DNA	9647.29	9.27%
RNA	4443.75	4.27%
Satellite	374	0.36%
Retroposon	3	0.00%
DNA?	0	0.00%
LTR?	0	0.00%
RC	0	0.00%
RC?	0	0.00%
SINE?	0	0.00%
Unknown	0	0.00%
Total	**104066.31**	**100%**

Supplementary table 16 Distribution of significant DHX40 CRAC reads mapping to TE classes and families. *The table shows the number and the relative distribution of reads assigned to TE families. Reads were annotated using hg38_rmsk_TE.gtf. A gene ID list is available in the supplementary zip file. The CRAC bioinformatic analysis was performed in collaboration with Nicolás Lemus, MD PhD*

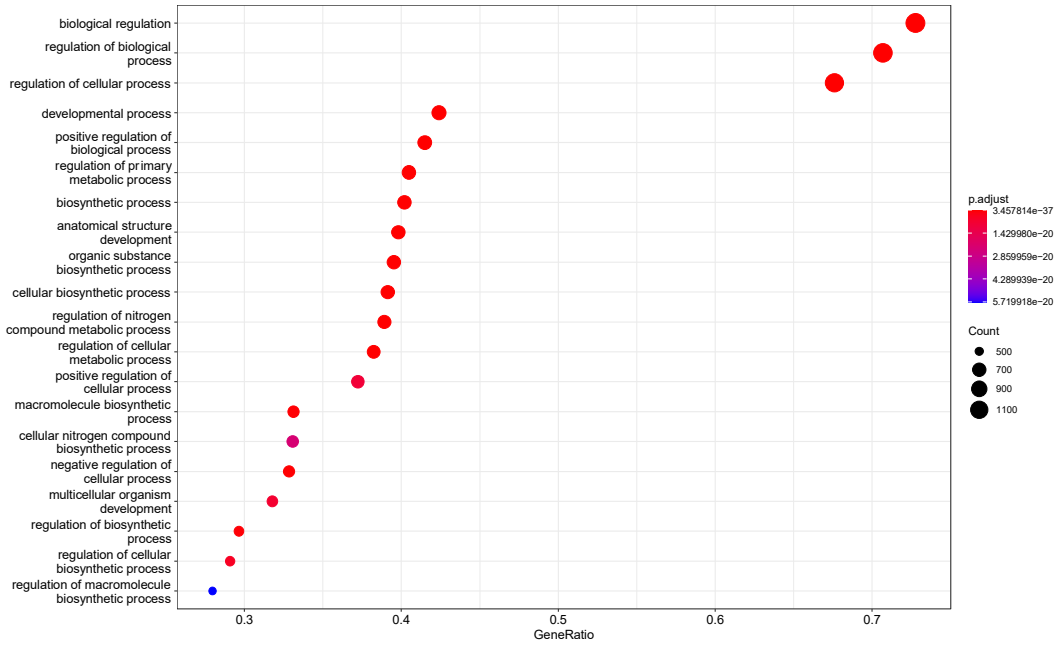
TE - Family	Normalised reads	Normalised reads (%)
Alu	32171.82	30.91
L1	24496.08	23.54
MIR	10570.49	10.16
L2	7800.95	7.50
ERV1	7056.57	6.78
hAT-Charlie	5433.7	5.22
RNA	4443.75	4.27
CR1	2440.1	2.34
ERVL	2177.37	2.09
ERVL-MaLR	2169.9	2.09
TcMar-Tigger	2152.16	2.07
TcMar-Mariner	1704.72	1.64
ERVK	377.5	0.36
Satellite	374	0.36
hAT-Tip100	318.07	0.31
RTE-X	261	0.25
Gypsy	76	0.07
TcMar-Tc2	34.95	0.03
hAT	3.36	0.00
SVA	3	0.00
hAT-Blackjack	0.33	0.00
Total	**104065.82**	**100**



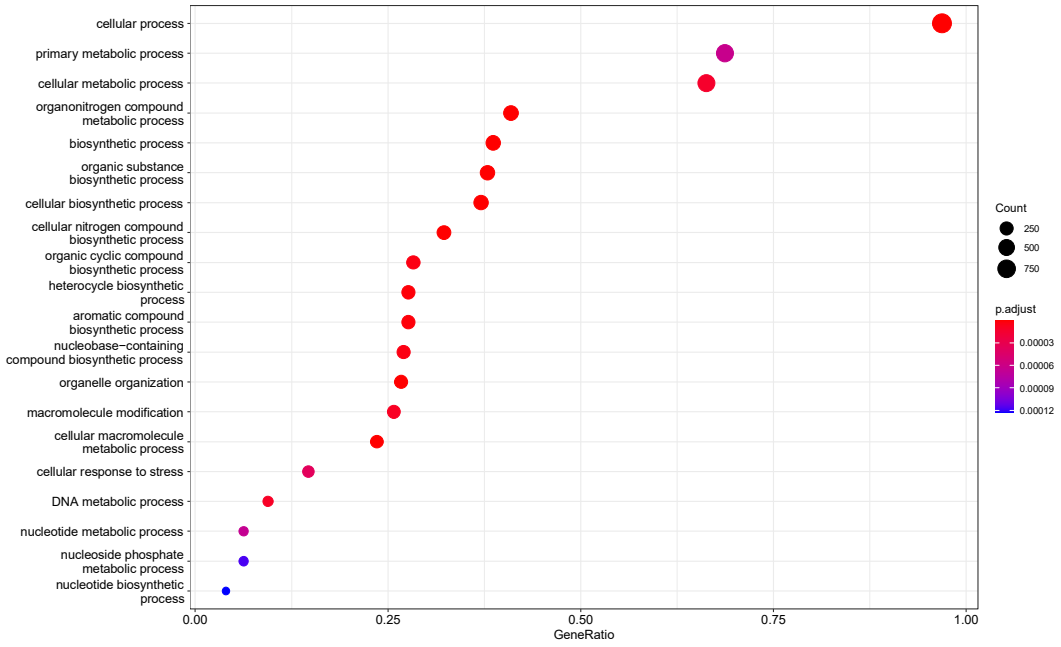
Supplementary figure 1 Functional enrichment analysis of protein-coding transcripts downregulated in siDHX40 sample compared to siNT control *Functional analysis of significantly (padj < 0.005) downregulated (FC<-2) protein-coding transcripts in the siDHX40 sample. Functional analysis was performed in R using the enrichplot package. The bioinformatic analysis was performed in collaboration with Nicolás Lemus, MD PhD*



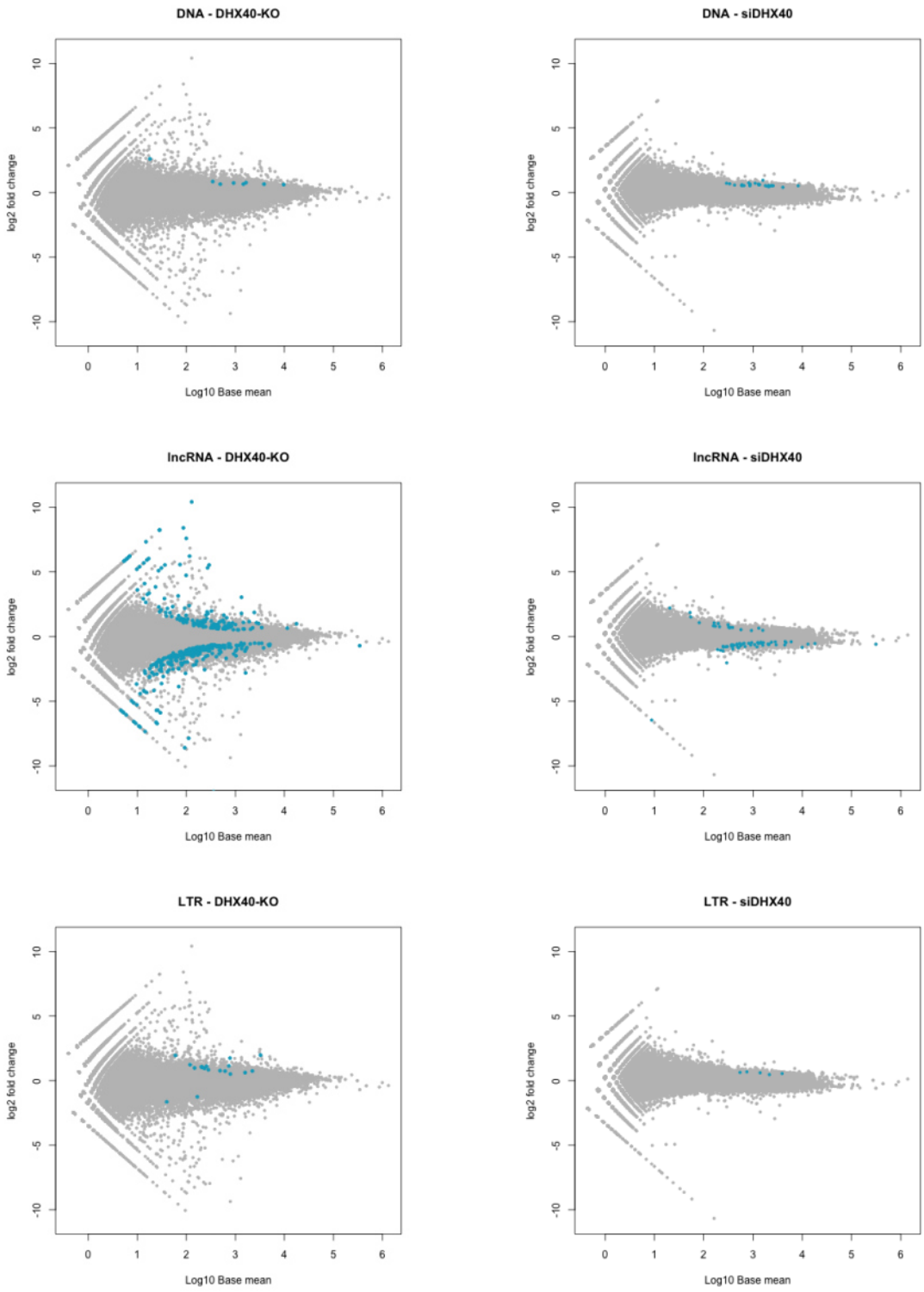
Supplementary figure 2 Functional enrichment analysis of protein-coding transcripts upregulated in siDHX40 sample compared to siNT control *Functional analysis of significantly (padj < 0.005) upregulated (FC>2) protein-coding transcripts in the siDHX40 sample. Functional analysis was performed in R using the enrichplot package. The bioinformatic analysis was performed in collaboration with Nicolás Lemus, MD PhD*



Supplementary figure 3 Functional enrichment analysis of protein-coding transcripts downregulated in ΔDHX40-1 sample compared to control *Functional analysis of significantly (padj < 0.005) downregulated (FC<-2) protein-coding transcripts in the ΔDHX40 sample. Functional analysis was performed in R using the enrichplot package. The bioinformatic analysis was performed in collaboration with Nicolás Lemus, MD PhD*

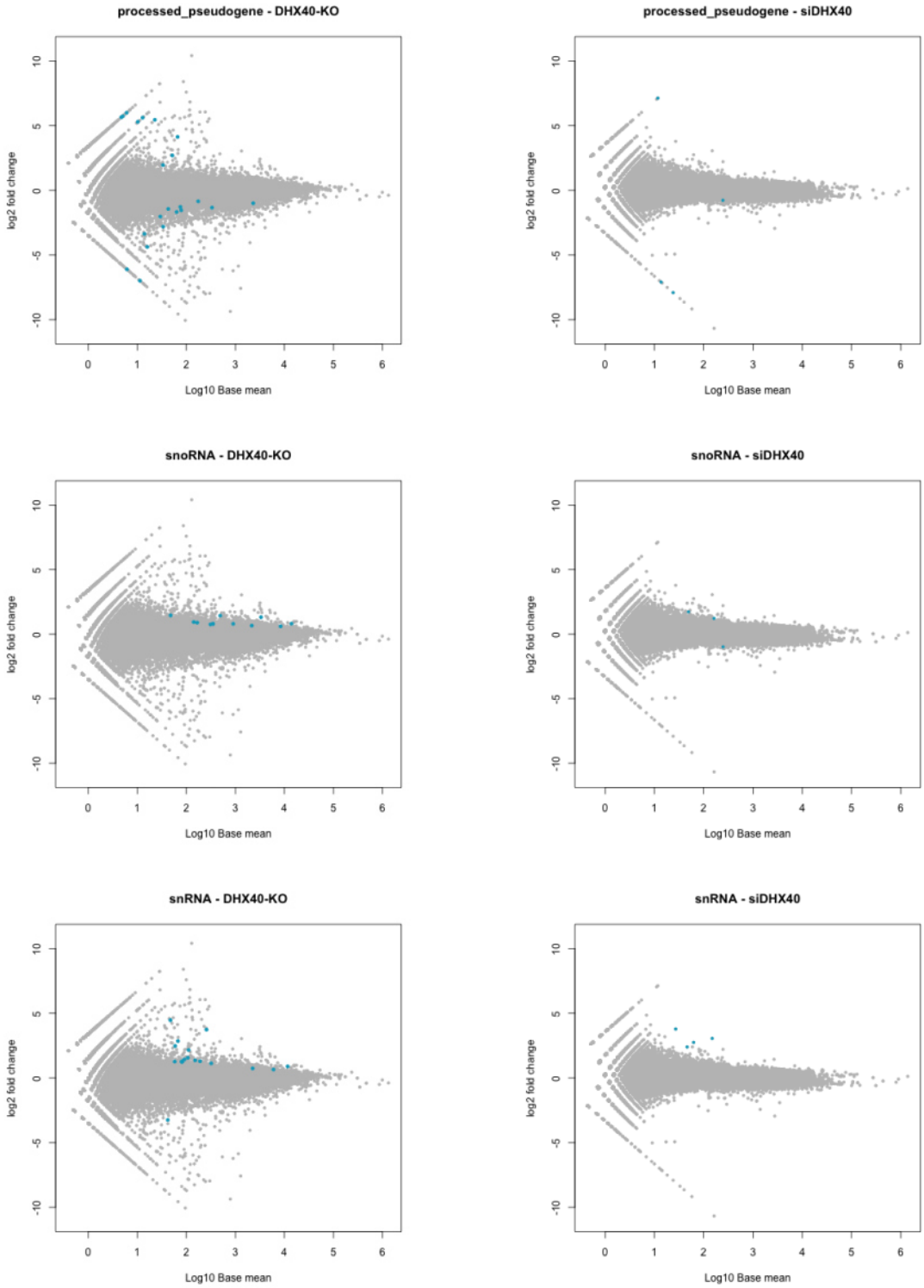


Supplementary figure 4 Functional enrichment analysis of protein-coding transcripts upregulated in ΔDHX40-1 sample compared to control *Functional analysis of significantly (padj < 0.005) upregulated (FC>2) protein-coding transcripts in the ΔDHX40-1 sample. Functional analysis was performed in R using the enrichplot package. The bioinformatic analysis was performed in collaboration with Nicolás Lemus, MD PhD*

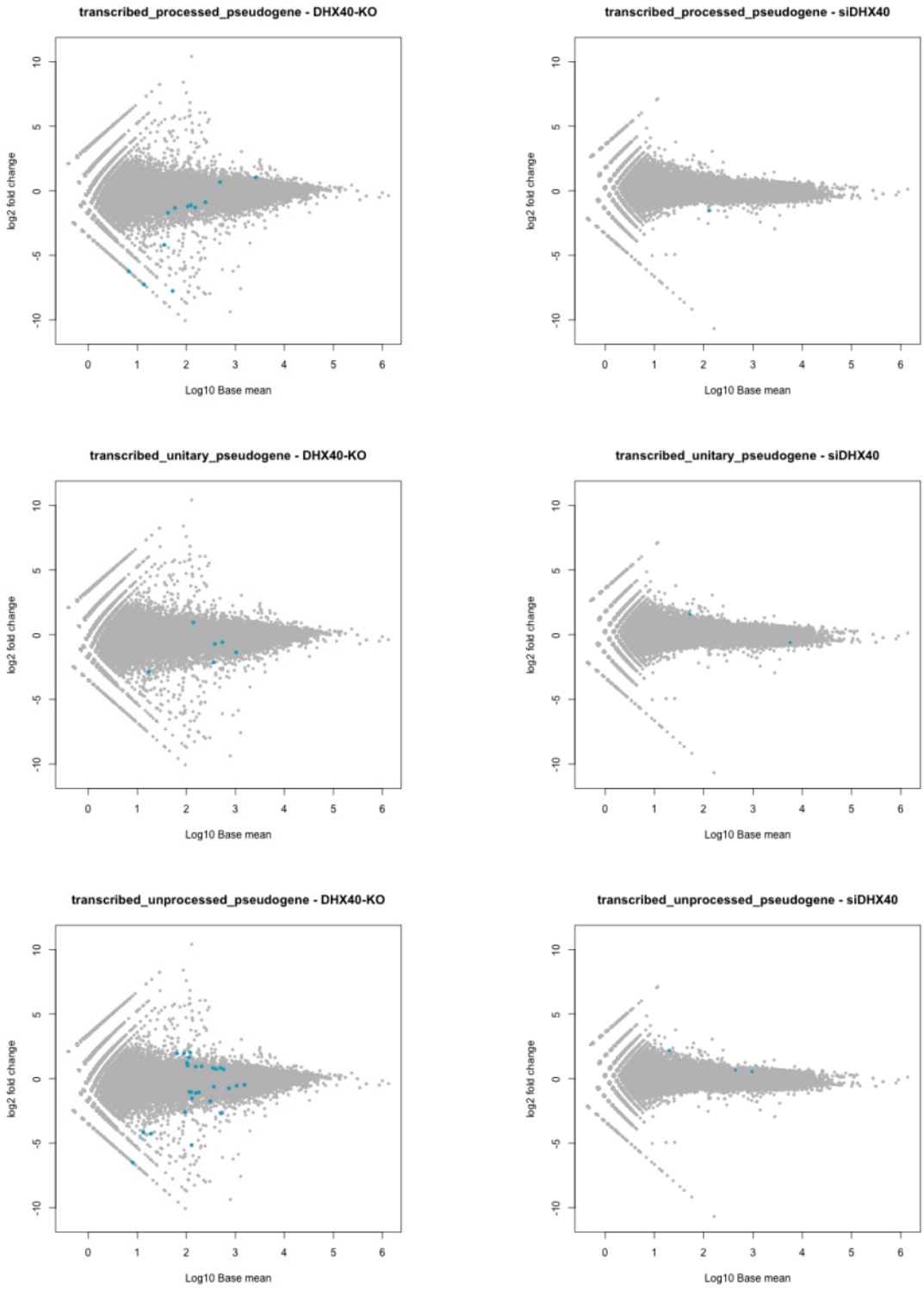


Supplementary figure 5 MA plots for significantly up- and downregulated transcripts in the siDHX40 and ΔDHX40-1 compared to controls. *Significantly (padj < 0.005) upregulated (FC>2) or downregulated (FC<-2) transcripts in ΔDHX40-1 (DHX40-KO) or siDHX40 samples relative to control. Differential expression analysis was performed in TEtranscript. Feature count and gene id tables are available in the supplementary zip file.*

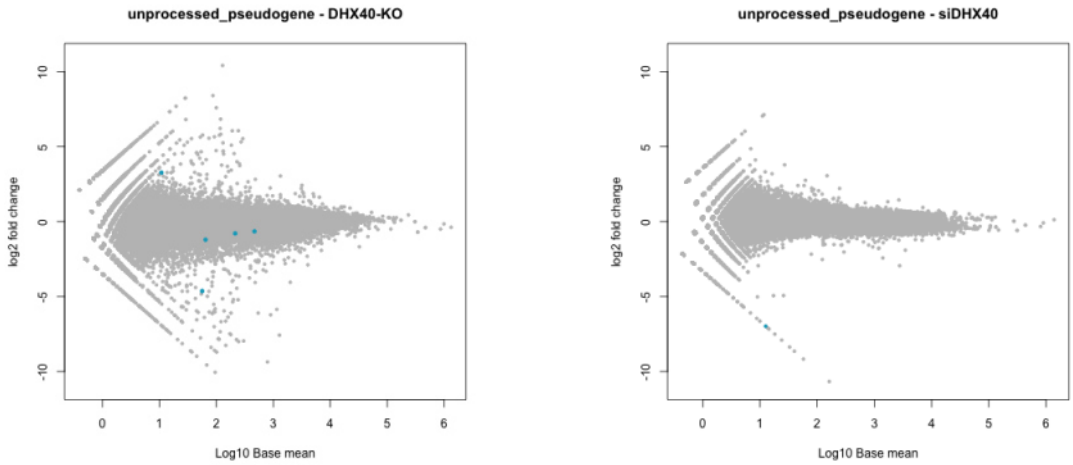
Abbreviations: DNA transposons (DNA), long terminal repeat transposable elements (LTR). The bioinformatic analysis was performed in collaboration with Nicolás Lemus, MD PhD



Supplementary figure 5 (continued)



Supplementary figure 5 (continued)



Supplementary figure 5 (continued)

Supplementary table 17 Number of GFP-positive cells in L1 retrotransposition assays *Number of GFP-positive cells in the L1 retrotransposition assay were measured by FACS. Number of positive cells were scaled to the Mock sample to allow comparison of different experiment. Numbers are from at least biological triplicates and technical duplicates of each biological replicate.*

Positive cells per 10,000 recorded events						
Non Transfected	**Mock**	**siNT**	**siABH5**	**siMETTL3**	**siPPIL4**	**siDHX40**
49.8	719.2	862.4	976.1	690.6	1429.5	1246.2
48.3	702.7	895.9	853.6	660.4	1218.2	1106.2
268.5	755.6	823.6	1140.0	433.7	1398.1	1160.0
279.4	691.2	861.2	1171.0	419.4	1310.5	998.8
257.5	681.5	764.1	1076.2	438.4	1137.5	1200.3
241.7	715.5	813.9	1002.1	362.6	1445.5	1061.6
173.0	801.0	771.3	1301.0		1392.2	
41.5	721.6	833.0	1118.8		1098.4	
45.0	652.9	733.4	1056.8			
55.7	784.4	735.8	1062.9			
36.7	657.6	752.4	915.9			
28.4	648.1	705.0	974.2			
			1162.5			
			943.8			
			1065.2			
			899.3			
			727.5			
			692.0			
			1034.4			
			821.1			
			690.8			
			707.4			

Supplementary table 18 Ordinary one-way ANOVA analysis for the L1 retrotransposition assay. *Summary of ordinary one-way ANOVA analysis of the data in Supplementary table 17 Analysis was perfomed in GraphPad Prism.*

F	104.1				
P value	<0,0001				
P value summary	****				
Significant diff. among means (P < 0.05)?	Yes				
R squared	0.8979				
Brown-Forsythe test					
F (DFn, DFd)	2,664 (6, 71)				
P value	0.0218				
P value summary	*				
Are SDs significantly different (P < 0.05)?	Yes				
Bartlett's test					
Bartlett's statistic (corrected)	23.81				
P value	0.0006				
P value summary	***				
Are SDs significantly different (P < 0.05)?	Yes				
ANOVA table	**SS**	**DF**	**MS**	**F (DFn, DFd)**	**P value**
Treatment (between columns)	9379891	6	1563315	F (6, 71) = 104,1	P<0,0001
Residual (within columns)	1066589	71	15022		
Total	10446480	77			
Data summary					
Number of treatments (columns)	7				
Number of values (total)	78				

Supplementary table 19 Tukey's multiple comparisons test for the L1 retrotransposition assay. *Analysis of the data in Supplementary table 17 and Supplementary table 18. Analysis was perfomed in GraphPad Prism.*

Tukey's multiple comparisons test	Mean Diff,	95,00% CI of diff,	Below threshold?	Sum-mary	Adjusted P Value
Non Transfected vs. Mock	-583.8	-735,6 to -432,0	Yes	****	<0,0001
Non Transfected vs. siNT	-668.9	-820,7 to -517,0	Yes	****	<0,0001
Non Transfected vs. siABH5	-845.3	-978,7 to -711,8	Yes	****	<0,0001
Non Transfected vs. siMETTL3	-373.7	-559,7 to -187,8	Yes	****	<0,0001
Non Transfected vs. siPPIL4	-1177	-1346 to -1007	Yes	****	<0,0001
Non Transfected vs. siDHX40	-1002	-1188 to -815,8	Yes	****	<0,0001
Mock vs. siNT	-85.08	-236,9 to 66,75	No	ns	0.618
Mock vs. siABH5	-261.5	-394,9 to -128,0	Yes	****	<0,0001
Mock vs. siMETTL3	210.1	24,11 to 396,0	Yes	*	0.0168
Mock vs. siPPIL4	-592.8	-762,6 to -423,1	Yes	****	<0,0001
Mock vs. siDHX40	-417.9	-603,9 to -232,0	Yes	****	<0,0001
siNT vs. siABH5	-176.4	-309,9 to -42,92	Yes	**	0.0027
siNT vs. siMETTL3	295.2	109,2 to 481,1	Yes	***	0.0002
siNT vs. siPPIL4	-507.7	-677,5 to -338,0	Yes	****	<0,0001
siNT vs. siDHX40	-332.8	-518,8 to -146,9	Yes	****	<0,0001
siABH5 vs. siMETTL3	471.5	300,2 to 642,8	Yes	****	<0,0001
siABH5 vs. siPPIL4	-331.3	-484,9 to -177,8	Yes	****	<0,0001
siABH5 vs. siDHX40	-156.5	-327,7 to 14,84	No	ns	0.096
siMETTL3 vs. siPPIL4	-802.9	-1004 to -602,0	Yes	****	<0,0001
siMETTL3 vs. siDHX40	-628	-842,7 to -413,3	Yes	****	<0,0001
siPPIL4 vs. siDHX40	174.9	-25,97 to 375,8	No	ns	0.1291

List of abbreviations

3'SA	3' splice acceptor
43S PIC	43S preinitiation complex
48S IC	48S initiation complex
5'SD	splice doner
amine	NH_2
AP-MS	Affinity purifications mass spectrometry
ATP	Adenosine triphosphate
AvgSpc	Average spectral counts
bp	Base pair
BP	Branch point
CARD	Caspase recruitment domains
CBP	CREB-binding protein
CC	Coiled-coil
cDNA	Complementary DNA
CFD	Cutting frequency determination
CRAC	Crosslinking and analysis of cDNA
CRISPR	Clustered regularly interspaced short palindromic repeats
Cry	Cryptic domain
cryo-EM	Cryogenic electron microscopy
Ct	Quantification cycle value
CTD	C-terminal domain
C-terminal	Carboxyl-terminal
CV	Column volumes
DE	Differential expression
DECID	Decay-inducing complex
DHX40-FLAG	DHX40-His6Prc2xFLAG
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dsRBD	Double-stranded RNA-binding domain
dsRNA	Double-stranded RNA
EGFP	Enhanced green fluorescent protein
eIF	Eukaryotic initiation factor
EJC	Exon junction complex
EMCV	Encephalomyocarditis virus
EN	Endonuclease

eRF	Eukaryotic release factor
ERV	Endogenous retrovirus
ESCRT	Endosomal sorting complexes required for transport
FACS	Fluorescence-activated Cell Sorting
FBS	Foetal bovine serum
FC	Fold change
FRT	FLP recombinase target
G4	G-quadruplex
GAIT	γ-interferon-activated inhibitor of the translation
gDNA	Genomic DNA
GOI	Gene of interest
GFP	Green fluorescent protein
GO	Gene ontology
GTP	Guanosine triphosphate
HB	Helix bundle
HDR	Homology-directed repair
His_n	Polyhistidine tag
hnRNP	Heterogeneous nuclear ribonucleoprotein
IFN	Type I interferon
IgG	Immunoglobulin G
ILS	Intron-lariat spliceosome
IMAC	Immobilised metal affinity chromatography
IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KAP	KRAB-associated protein
KPNA	Karyopherin Subunit Alpha
L1	LINE-1
L2	LINE-2
LB	Luria Broth according to Miller
LC-MS/MS	Liquid chromatography with tandem mass spectrometry
LDH	Lactate dehydrogenase
LINEs	Long interspersed elements
lncRNA	Long non-coding
LSU	Large subunit
LTR	Long terminal repeat
m^6A	N^6-methyladenosine
MBP	Maltose binding protein

miRNA	Micro RNA
misc. RNA	Miscellaneous RNAs
mRNA	Messenger RNA
mtDNA	Mitochondrial DNA
mt-rRNA	Mitochondrial ribosomal RNA
mt-tRNA	Mitochondrial tRNA
MWCO	Molecular weight cut-off
NADH	Nicotinamide adenine dinucleotide
ncRNA	Non-coding RNA
NDP	Nucleoside diphosphate
NGS	Next-generation-sequencing
NHEJ	Non-homologous end repair
NMD	Nonsense-mediated RNA decay
nt	Nucleotides
NTC	Prp19-associated complex
NTD	N-terminal domain
N-terminal	Amino-terminal
NTP	Nucleoside triphosphate
NTR	Prp19-related complex
OB	Oligosaccharide-binding fold
OH	Hydroxyl
ORF	Open reading frames
PABPC1	Polyadenylate-binding protein
padj	Adjusted p-value
PAM	Protospacer adjacent motif
PARP1	Poly(ADP-ribose) polymerase
PBS	Phosphate buffered saline
PBS-T	PBS supplemented with 0.1% Triton™ X-100
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
PK	Pyruvate kinase
PMSF	Phenylmethylsulfonyl fluoride
Pol II	RNA polymerase II
polyA	Polyadenylation
PPIase	Propyl isomerase
PTCs	Premature termination codons
PVDF	Polyvinylidene difluoride

PAA	Polyacrylamide
qPCR	Quantitative PCR
RBF	Ribosome biogenesis factor
RBP	RNA binding protein
RC	Rolling-circle
RIPA	Radioimmunoprecipitation assay
RLR	RIG-I-like receptor
RMG	Mitochondrial RNA granules
RNA	Ribonucleic acids
RNAi	RNA interference
RNP	Ribonucleoprotein
RP	Ribosomal protein
RPA	Replication protein A
RRM	RNA recognition motif
rRNA	Ribosomal RNA
RT	Reverse transcriptase
SAINT	Significance analysis of interactome
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SF1	Superfamilies
sgRNA	Single guide RNA
sgRNA	Single guide RNA
shRNA	Short hairpin RNA
SINE	Short interspersed element
siNT	Non-target siRNA
siRNA	Small interfering RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear
snRNP	Small nuclear ribonucleoprotein
ssRNA	Single-stranded RNA
SSU	Small subunit
SURF	Surveillance complex
SVA	SINE-R/VNTR/Alu-like
TAE	Tris-acetate-EDTA
TBS	Tris-buffered Saline
TBS-T	TBS supplemented with 0.1% Tween-20
TCA	Trichloroacetic acid

TE	Transposable elements
TF	Transcription factor
TIR	Terminal inverted repeat
TPRT	Target-primed reverse transcription
TRIM	Tripartite motif
tRNA	Transfer RNA
$tRNA_i^{Met}$	Initiator methionine transfer RNA
TSD	Target site duplication
UTR	Untranslated region
VL	Variable loop
WH	Winged-helix
YY1	Yin Yang 1

List of figures

List of tables