The Mutant *HTT* mRNA-Protein Interactome

Implications in RNA Toxicity in Huntington's Disease

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Judith Schilling

aus

Mülheim an der Ruhr

Bonn, Mai 2017

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Dr. Ina Vorberg
2. Gutachter: Dr. Jörg Höhfeld
Tag der Promotion: 28. August 2017
Erscheinungsjahr: 2017

Summary

Huntington's disease (HD) is an autosomal dominant neurodegenerative disorder that is characterized by progressive motor, cognitive, and psychiatric symptoms. The mutant gene product contains an elongated stretch of CAG repeats that translates into an extended sequence of polyglutamines within the Huntingtin (HTT) protein. Many of the pathologic cellular mechanisms underlying HD are based on aberrant protein interactions of mutant HTT. Amongst others, affected processes include transcription, energy metabolism, axonal transport, synaptic transmission, and the proteostasis network. Additionally, mutant *HTT* RNA impacts cellular functions as well. A toxic gain-of-function of the mutant CAG repeat RNA can be explained by the sequestration of various RNA binding proteins, leading to deregulation of cellular mechanisms like RNA interference, alternative splicing, and gene expression. One specific example is the MID1 complex that enhances translation of mutant *HTT* exon 1 RNA, thereby increasing the production of toxic protein species.

This study shows that MID1 specifically regulates the translation of structured RNAs, revealing a requirement for the mechanism of MID1-dependent translation. Furthermore, MID1's close association with the translation initiation complex is confirmed. The identification of new and the verification of known binding partners locates MID1 within a large complex comprising eukaryotic initiation factors and ribosomal proteins. Together with the known CAG length-dependent binding of MID1 to *HTT* exon 1 RNA, this suggests that MID1 is an important factor directing the translational machinery to mutant CAG repeat RNAs, thereby possibly acting as a disease modifier. Moreover, in HD patients, MID1 expression is upregulated, corroborating this hypothesis.

The comparison of MID1 protein interactions with *HTT* exon 1 binding partners shows substantial overlap. 25% of the identified proteins are shared binding partners. Interestingly, gene ontology analysis of the *HTT* exon 1 interactome shows that 43% of the proteins can be assigned to the process of splicing and 53% of these preferentially bind to mutant *HTT* exon 1 RNA. Therefore, the analysis of splicing changes in a cellular model of inducible mutant *HTT* exon 1 RNA expression was conducted. Specific splicing events were identified that can be attributed to *HTT* exon 1 protein binding partners and this was confirmed in HD patient brain samples.

Together, this study extends the knowledge of MID1-dependent mechanisms of translation, characterizes the *HTT* RNA-protein network and identifies aberrant downstream effects that might contribute to HD pathogenesis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

µ	Micro (factor of 10^{-6})
4E-BP	4E-binding protein
AR	Androgen receptor
ATF-1	Acting transcription factor 1
ATXN8OS	ATXN8 Opposite Strand
BACE1	Beta-secretase 1
bDNA	Biotinylated DNA
BDNF	Brain-derived neurotrophic factor
CNS	Central nervous system
coIP	Co-immunoprecipitation
CPSF6	Cleavage and polyadenylation specificity factor subunit 6
CRE	cAMP-responsive element
CREB1	cAMP response element-binding protein
CREM	cAMP response element modulator
CUGBP1	CUG RNA-binding protein 1
DDX5	DEAD-Box helicase 5
DM1	Dystrophia myotonica
DMPK	Dystrophia myotonic protein kinase
DRPLA	Dentatorubral-pallidoluysian atrophy
EF-1α	Elongation factor 1 alpha
EGFP	Enhanced green fluorescent protein
eIF2α	Eukaryotic initiation factor 2
eIF4G	Eukaryotic translation initiation factor gamma 1
FISH	Fluorescence in situ hybridization
FMR1	Fragile x mental retardation 1
FMRP	Fragile X Mental Retardation Protein
FRDA	Friedreich ataxia
FXS	Fragile X syndrome
FXTAS	Fragile X tremor/ataxia syndrome
GO	Gene ontology
h	Hour(s)
HD	Huntington disease
HEAT	Huntingtin, elongation factor 3, protein phosphatase 2A and TOR1
HITS-CLIP	High-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation
HPA	Human protein atlas
HTT	Huntingtin
HTTAS	Huntingtin antisense
HTTex1(CAG)ex	HTT exon 1 RNA with an expanded CAG repeat tract
IHC	Immunohistochemistry
IP	Immunoprecipitation
L	Liter
m	Milli (factor of 10^{-3})
M	Molar (mol/m^3)

MAP2	Microtubule-associated protein 2
MAPT	Microtubule-associated protein tau
MBNL	Muscleblind-like
MID1	Midline-1
min	Minute(s)
miRNA	Micro RNA
mTOR	Mechanistic target of rapamycin
mTORC1	mTOR complex 1
n	Nano (factor of 10^{-9})
NCL	Nucleolin
NES	N-terminal nuclear export signal
NeuN	Neuronal nuclei
NFκB	Nuclear factor κB
NRSE	Neuron-restrictive silencer element
NUDT21	Nudix hydrolase 21
ORF	Open reading frame
PABP1	Poly(A) binding protein 1
PERK	(PKR)-like endoplasmic reticulum kinase
pH	Potential of hydrogen
PIC	preinitiation complex
p	Pico (factor 10^{-12})
PKC	Protein kinase C
PKR	Protein kinase R
polyQ	Polyglutamine
PP2A	Protein phosphatase 2A
PRD	Proline-rich domain
PRPF8	Pre-mRNA-processing-splicing factor 8
PSD95	Postsynaptic density protein 95
PSPC1	Paraspeckle component 1
qPCR	Quantitative polymerase chain reaction
RACK1	Receptor for activated protein C kinase 1
RAN translation	Repeat-associated non-ATG translation
RBPs	RNA-binding proteins
RING	Really Interesting New Gene
RNP	Ribonucleoprotein
RPL22	Ribosomal protein L22
RPL5	60S ribosomal protein L5
RPLP0	60S acidic ribosomal protein P0
RPS3	40S ribosomal protein S3
S6K	40S ribosomal S6 kinase
SBMA	Spinal and bulbar muscular atrophy
SCA	Spinocerebellar ataxia
sCAGs	Small CAG-repeated RNAs
SF3B2	Splicing factor 3B subunit 2
siRNA	Short interfering RNA
SNP	Single nucleotide polymorphism

snRNP	Small nuclear ribonucleoprotein complex
SNRNP40	U5 small nuclear ribonucleoprotein 40 kDa protein
SRSF6	Serine/arginine-rich splicing factor
STAU1	Double-stranded RNA-binding protein Staufen homolog 1
TF	Transcription factor
TNR	Trinucleotide repeat
TRAIL	TNF-related apoptosis inducing ligand
TRE	Tetracycline responsive element
tTA	Tetracycline transactivator protein
UTR	Untranslated region

1 Introduction

1.1 Trinucleotide repeat disorders

Trinucleotide repeat (TNR) disorders are a large group of both common and rare neurodegenerative and neuromuscular diseases that arise from microsatellite expansions of three base pairs. Generally, microsatellites denote di-, tri-, or tetra nucleotide tandem repeats in the genome and the repeat size is polymorphic within the alleles of an individual. Microsatellite repeats represent 3 % of the entire human genome [1], with TNRs being the most abundant ones.

Table 1 | Molecular and clinical characteristics of trinucleotide repeat disorders.

Disease	Gene	Repeat	Normal/ expanded repeat	location	Main clinical features
DM1	DMPK	CTG	5 - 37/ 50 – 10,000	3′ UTR	Myotonia, cardiac conduction defects, muscle weakness
DRPLA	ATN1	CAG	7 – 34/ 49 -88	coding	Ataxia, seizures, choreoathetosis, dementia
FRDA	FXN	GAA	6 – 32/ 200 -1,700	Intron	Sensory ataxia, cardiomyopathy, diabetes
FXS	FMR1	CGG	6 – 60/ > 200	5′ UTR	Mental retardation, facial dysmorphism, autism
FXTAS	FMR1	CGG	6 – 60/ 60 - 200	5′ UTR	Ataxia, tremor, Parkinsonism, dementia
HD	HTT	CAG	6 – 34/ 36 - 121	coding	Chorea, dystonia, cognitive decline, psychiatric problems
SBMA	AR	CAG	9 – 36/ 38 - 62	coding	Progressive motor weakness, gynecomastia, decreased fertility
SCA1	ATXN1	CAG	6 – 44/ 39 - 82	coding	Ataxia, dysarthria, spasticity, cognitive impairments
SCA2	ATXN2	CAG	15 – 24/ 32 - 200	coding	Ataxia, decreased reflexes, occasional parkinsonism
SCA3	ATXN3	CAG	13 – 36/ 61 - 84	coding	Ataxia, parkinsonism, spasticity
SCA7	ATXN7	CAG	4 – 35/ 37 - 306	coding	Ataxia, blindness, dysarthria
SCA8	ATXN8/ ATXNOS80	CTG/ CAG	16 – 34/ > 74	3′UTR/ coding	Ataxia, dysarthria, nystagmus

DM1, dystrophia myotonica; DRPLA, dentatorubral-pallidoluysian atrophy; FRDA, Friedreich ataxia; FXS, Fragile X syndrome; FXTAS, Fragile X tremor/ataxia syndrome; HD, Huntington disease; SBMA, spinal and bulbar muscular atrophy, SCA, spinocerebellar ataxia.

Repeat lengths within a physiologic range can be neutral or regulatory and will only become toxic above or below a specific threshold. Their frequent occurrence in coding regions is possible because variations in repeat lengths do not cause a frameshift within the encoded protein.

In TNR disorders, the repeat length varies in respect to the genetic locus and whether it is located in the untranslated region (UTR), intronic sequence or the coding region of the respective gene. A selection of the most common TNR disorders is presented in Table 1.

Since microsatellite expansions are dynamic, the length of the repeat is variable between individuals and the repeat transmission between generations is unstable [2,3]. This offers an explanation for the variability of the disease phenotypes and the concept of genetic anticipation. First, the larger the expansion the more severe is the disease phenotype and progression. Second, the germline instability can lead to earlier age of onset or more severe phenotypes between generations because longer repeats are correlated with earlier age of onset and increased severity of disease [4–6]. Additionally, in most TNR diseases a premutation range of repeat size can be defined where individuals rarely develop the disease but are likely to pass on a fully penetrating mutation to their offspring. An exception to this is the fragile x mental retardation 1 (*FMR1)* gene, where the length of the repeat expansion determines two distinct neurodegenerative diseases. The mutation is found in the 5' UTR, a regulatory region that is aberrantly methylated if the repeat tract exceeds 200 CGG repeats [7]. As a consequence, *FMR1* transcription is silenced and the gene product, Fragile X Mental Retardation Protein (FMRP), is not translated leading to Fragile X Syndrome (FXS). However, in case the repeat number lies between 60 to 200 units, affected individuals develop Fragile X-associated tremor ataxia syndrome (FXTAS, described in chapter 1.3.1, 1.3.3, 1.3.5).

1.2 Huntington's disease

Members of the largest group of TNR disorders share a CAG repeat expansion in the coding region of the respective gene. Because CAG translates into a glutamine amino acid, this subgroup is referred to as polyglutamine (polyQ) diseases. The most common polyQ disease is Huntington's disease (HD), an autosomal dominant neurological disorder where the pathologic mutation is found in the huntingtin (*HTT*) gene on chromosome 4p16.3 [8]. In healthy individuals, the CAG repeat is polymorphic with 6 to 35 units. In patients, as with other TNR disorders, the age of onset is inversely correlated to the number of CAG repeat units [9]. Repeats between 36 to 39 units show a variable penetrance and later onset of disease compared to individuals carrying 40 repeat units or more. This effect is aggravated in the case of very long repeat tracts: CAG repeat mutations with 60 or more repeats result in the juvenile onset form of HD (younger than 20 years) [10]. Interestingly, homozygous patients show a similar age of onset as heterozygotes, however the disease progression can be more severe [11].

Typically, the disease develops from a presymptomatic period without apparent clinical signs to a prodromal phase that is characterized by subtle changes in motor functions, behavior, and cognition. Eventually, patients enter the manifest stage at a mean age of 35 to 44 years with slow progression of symptoms. The appearance and sequence of motor, cognitive, and psychiatric symptoms is variable between individuals. However, the disease is invariably fatal with a median survival of 18 years from motor onset [12]. Motor symptoms like chorea (involuntary movements) usually appear early during the disease course, while bradykinesia (slow execution of voluntary movement) and rigidity appear in late stage patients. Both cognitive and psychiatric changes manifest in early disease stages and worsen over time: patients often show cognitive slowing, decreased attention, signs of apathy, irritability, impulsivity, and depression. Apart from the clinical features that can be attributed to the neuropathology, HD patients also show skeletal muscle wasting, weight loss, cardiac failure, testicular atrophy, and osteoporosis [13]. Activation of the immune system and an altered immune response is already evident during the premanifest stage of disease [14] including activated microglia in the brain [15]. Eventually, the activation of microglia correlates with the severity of disease progression [16]. Whether dysfunction of the immune system, and specifically neuroinflammation, is cause or consequence is still unclear.

While the worldwide prevalence and incidence of HD is unclear, the disease is apparent in all populations but with large variations in frequency. This is best exemplified in British Columbia, Canada, where people of European descent show a prevalence of 17.2 cases per 100,000 in contrast to the remaining population with 2.1 cases per 100,000 [17]. Underlying this different susceptibility to HD are specific haplotypes characterized by longer CAG repeat lengths, that in turn lead to a higher chance for *de novo* mutations [18].

1.2.1 Neuropathology

The most prominent neuropathologic feature of HD includes a selective degeneration of neurons in the striatum. GABAergic medium-sized spiny neurons are affected in the early stages of disease [19]. Other areas, like the globus pallidus, substantia nigra, and amygdala show variable degeneration and the loss of neurons is accompanied by an increase in astrocytes [20]. Medium spiny neurons of the indirect pathway of movement control in the basal ganglia are affected, explaining the etiology of chorea [21]. Massive striatal atrophy can be observed as early as 11 years prior to the clinical onset of HD [22]. Apart from striatal atrophy, a severe loss of cerebral cortex and subcortical white matter occurs in HD [23], which may account for the cognitive and neuropsychiatric impairments that often precede the onset of chorea.

Another feature of the disease is the presence of large intraneuronal inclusions containing huntingtin protein (HTT) [24,25]. Astrocytes develop HTT inclusions as well, however the frequency of HTT

inclusion-positive cells compared to neurons is much lower [26]. Not only full-length but also small N-terminal fragments of HTT have been visualized in inclusions [27]. The propensity of these fragments to aggregate depends on the length of the HTT fragments, the polyQ stretch, and their fibrillar or ribbon-like morphology resembles scrapie prion rods and beta-amyloid fibrils in Alzheimer's disease [28].

1.2.2 HTT structure and function

The *HTT* gene product with a normal polyQ length of 23 glutamines is a large protein of 3,144 amino acids and a molecular weight of 348 kDa. The polyQ stretch lies between an N-terminal nuclear export signal (NES) [29] and a proline-rich domain (PRD). Like the polyQ stretch, the PRD is polymorphic and probably involved in protein-protein interactions. The secondary structure of this region has been resolved for 17 glutamines (Q): the NES forms an α-helical structure, the 17Q can adopt various conformations including an α-helix, random coil, and extended loop [30]. The remaining parts of the protein are not as well characterized. Some structured domains could be identified as HEAT repeats (Huntingtin, elongation factor 3, protein phosphatase 2A and TOR1) that are important for protein-protein interactions [31,32]. HTT undergoes several posttranslational modifications like proteolysis, phosphorylation, acetylation, palmitoylation, ubiquitination, and sumoylation. Whether these modifications are relevant for the physiologic functions of HTT is largely unknown, but especially proteolysis plays an important role in the disease context of HD and will be discussed in chapter 1.2.3.
While *HTT* is ubiquitously expressed throughout the body, levels are highest in the brain and in testes [33,34]. Of note, HTT protein levels are relatively constant throughout different brain regions and thus cannot be correlated to neuronal susceptibility to cell death [35]. HTT can be found in the nucleus and cytoplasm, and in neurons in the soma, dendrites, and synaptic terminals [36,37].

A large body of work on HTT interaction partners has been compiled, ultimately identifying more than 350 HTT-interacting proteins [38–42]. This list underlines the diverse functions HTT performs throughout the cell, implicating it in pathways like cellular dynamics, metabolism, protein turnover, gene expression, and signal transduction. While many aspects remain unclear, HTT for example traffics various organelles in axons and dendrites like vesicles containing brain-derived neurotrophic factor (BDNF) [43], endosomes, lysosomes [44], and autophagosomes [45]. BDNF is a growth factor that supports the survival of existing neurons, assists the growth of new neurons and synapses, and is important in brain plasticity [46,47]

Concerning autophagy, HTT also facilitates recognition of ubiquitinylated proteins leading to cargo loading into autohagosomes [48]. Interestingly, deleting the wild-type polyQ stretch enhances autophagy and longevity in mice [49] suggesting a regulative role for the variations in polyQ length. Another pathway wild-type HTT is involved in is endocytosis: HTT interacts with proteins involved in clathrin-mediated endocytosis [50] and possibly takes part in several steps including membrane coating,

invagination, and dynamin 1 activation [51]. The association with several transcription factors (TF) [38,52,53] and the fact that a polyQ motif has been identified in many TFs to play a regulatory role, implies wild-type HTT in transcription. For instance, HTT inactivates the silencing activity of neuron-restrictive silencer element (NRSE), thereby stimulating the transcription of BDNF [54].

HTT expression starts early during development and its vital importance is highlighted by studies showing that *HTT* knockout in mice is lethal on embryonic day 7.5 [4,55,56]. Knockdown of *HTT* expression leads to cortical and striatal malformations and mice die shortly after birth [57] while conditionally targeting *HTT* in the adult forebrain results in a progressive degenerative neuronal phenotype [58]. Experiments investigating HTT function from the opposite angle show that overexpression has neuroprotective effects against excitotoxicity or ischemic injuries [59]. Supporting this function is the discovery of a non-coding single nucleotide polymorphism (SNP) in the *HTT* promoter affecting *HTT* expression [60]. Depending on the SNP variant, wild-type *HTT* is downregulated or mutant alleles are upregulated in HD patients associated with earlier or delayed age of onset, respectively [60].

1.2.3 Cellular mechanisms of protein pathobiology

Considering the multiple roles of wild-type HTT it is clear that a pathogenic mutation will affect various cellular pathways and indeed, all known biological functions HTT is involved in are disturbed by the polyQ expansion. The dominant inheritance of HD and HTT's vital role during embryogenesis points at mechanisms based on a protein gain-of-function. However, certain parts of the pathobiology can be attributed to a loss-of-function of wild-type HTT both through diminished expression levels and dominant-negative effects of the mutant allele [43,61].

The complexity of HTT itself in terms of structure and modifications, the existence of many different protein fragments and inclusions (which implies the existence of aggregated precursors) complicates the identification of toxic species. The formation of soluble N-terminal mutant HTT fragments is believed to be the main driver of protein pathology observed in HD (Figure 1). On the one hand, these toxic species arise from proteolysis of the full-length protein and result in differently sized protein fragments, depending on the protease [27]. On the other hand, aberrant alternative splicing of *HTT* mRNA generates an exon 1-containing transcript that is translated into the shortest known HTT fragment [62]. Toxicity of soluble N-terminal mutant HTT fragments has been demonstrated in various contexts [63–65]. Nuclear translocation of HTT fragments leads to cell death through transcriptional disruption [66]. Moreover, mutant HTT associates with mitochondria and impairs their function by disturbing mitochondrial distribution and transport rate [65]. Transport defects caused by mutant HTT can also be observed in axonal transport of GABA receptor- [67] and BDNF-containing [43] vesicles, resulting in synaptic dysfunction and neuronal death, respectively.

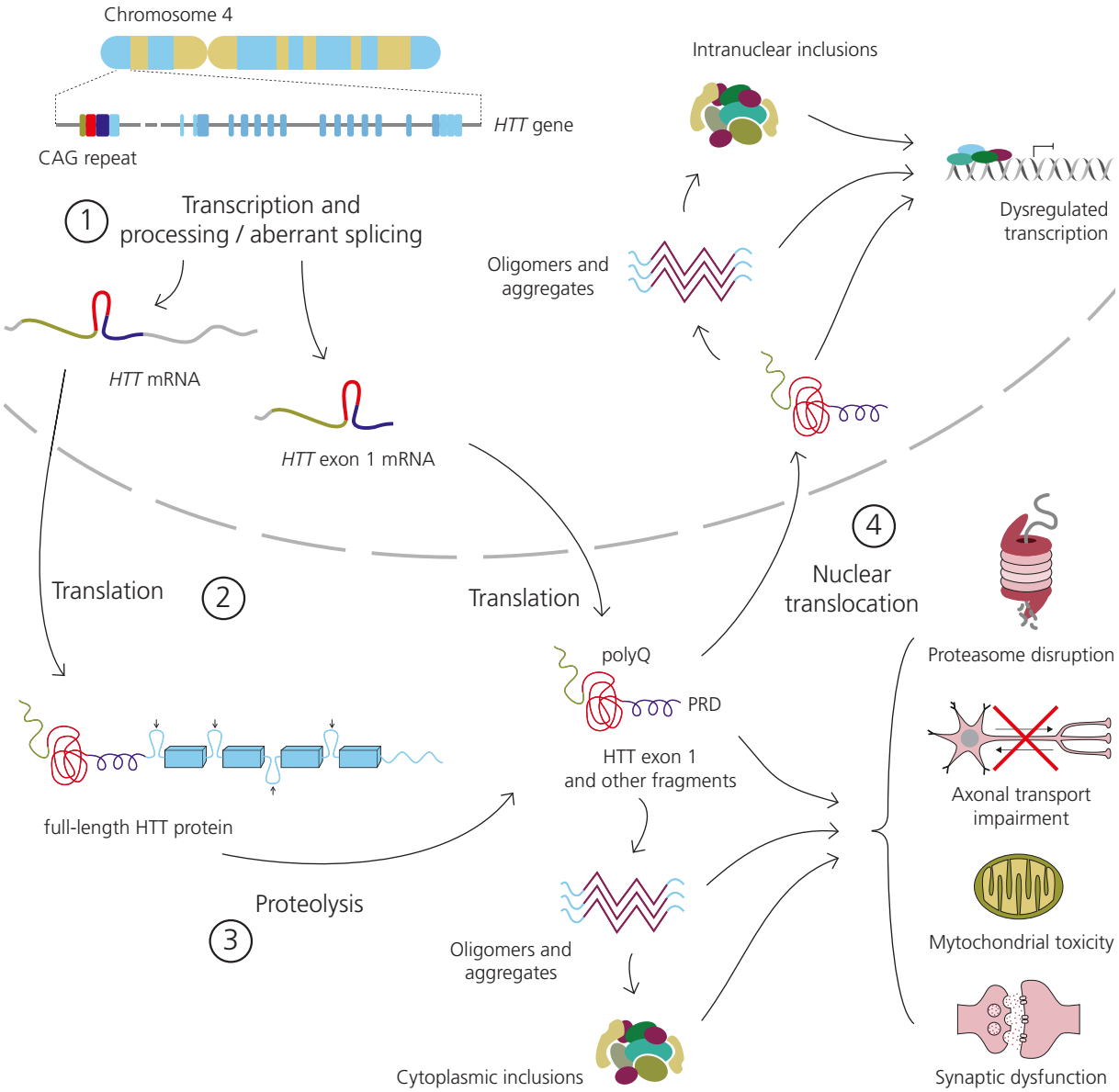

Figure 1 | Cellular mechanisms of protein pathology.
(1) Initially, transcription of the mutant *HTT* gene produces a full-length transcript that is both normally processed but also aberrantly spliced generating a short 5' UTR and exon 1 containing mRNA. (2) In the cytoplasm, mRNA is translated and the protein is differentially modified (compared to wild-type HTT). (3) Amongst other post-translational modification, full-length HTT is cleaved by proteases into various shorter fragments. These fragments have a high propensity to aggregate, form inclusions, and all species affect many cellular pathways. (4) Short fragments are also able to translocate back into the nucleus, disturbing transcription and again, forming aggregates and inclusions.

Since the striatum does not produce BDNF, survival of striatal neurons depends on the delivery of BDNF through the cortico-striatal synapse [68]. Therefore, the disruption of both BDNF transcription [69] and axonal transport by mutant HTT offers an explanation for the differential loss of striatal neurons. Apart from cell autonomous pathologic mechanisms and inter-neuronal dependency, interactions

between neurons and glial cells also play an important role in HD pathobiology. Mutant HTT reduces the expression of glutamate transporters in astrocytes, resulting in diminished uptake, overstimulation of adjoining neurons, and ultimately leading to excitotoxicity [70].

In addition to soluble HTT fragments, different species of aggregated mutant HTT have been shown to be cytotoxic [71]. For example, HTT aggregates can obstruct the proteasome system [72] impairing its overall cellular activity. The contribution of intracellular inclusions to cytotoxicity is not as clear. On the one hand, mutant HTT inclusions are able to sequester wild-type HTT leading to a loss-of-function of the normal protein [73]. On the other hand, the formation of inclusions is predictive for survival in neuronal cultures expressing mutant HTT [74].

The underlying cause of mutant HTT gain-of-function in respect to the formation of aggregates could be attributed to a change of protein conformation. Determining a mutation-dependent conformational change and its impact on HTT aggregation is difficult owing to the inherent flexibility of the protein structure and the lack of a direct assay identifying different conformations [75,76]. Apart from a conformational change in the mutant protein that could be causative for the oligomerization, the observation that nucleation of HTT amyloid fibrils is enhanced with increasing polyQ lengths [77] offers an explanation for the repeat length-dependent age-of-onset of the disease.

1.3 RNA-mediated toxicity

In addition to polyQ protein-mediated toxicity in HD, mutant CAG repeat RNA itself mediates pathologic mechanisms. The first indication for RNA being directly involved in a disease process came from the identification of the mutation causing another TNR disorder, myotonica dystrophy (DM1). The mutant CTG expansion lies in the 3' UTR of the dystrophia myotonic protein kinase (*DMPK)* gene and thus, is not translated [78]. Since the expression of DMPK is not affected, only mutant RNA remains as a possible toxic species [78]. Supporting RNA toxicity as a valid concept in polyQ diseases, Li et al. observed neurodegeneration in a *Drosophila* spinocerebellar ataxia type 3 (SCA3) model expressing untranslated and translated repeats of elongated CAG repeats. Interestingly, the insertion of CAA interruptions into the translated CAG repeat (CAA also codes for glutamine) markedly mitigated cell toxicity while preserving the polyQ tract [79] . These experiments unambiguously identify RNA structure as the cause for RNA toxicity and neurodegeneration. Since RNA structure is fundamental to its toxicity, it is worthwhile discussing RNA structure in the context of TNR disorders.

Inherent to the toxicity of mutant TNR RNA is a gain-of-function that can be attributed to a "gain-of-structure". Just like in proteins, RNA structure can be described as a hierarchical organization starting at the primary sequence. The secondary structure arises from Watson-Crick base pairing of complementary nucleotides leading to antiparallel double-helical structures varying in length. Long-

range interactions of secondary RNA motifs, stacking of helical structures, and metal ion stabilization between different motifs lead to the complex tertiary structure of RNA, one famous example being tRNA. Various RNA structural motifs have been identified. Hairpins or stem-loops are the most common, and variations of one or more mismatches within a duplex structure have been described (Figure 2A). The stability of secondary structures primarily depends on the nature of the TNR bases but also on the flanking sequences, i.e. the genetic context they are found in [80]. Apart from GAA, all isolated TNR RNAs associated with disease form hairpin structures with several possible alignments [81]. Detailed analysis of RNAs from TNR disease-causing genes demonstrated that the flanking sequences can influence the hairpin structure and that mutant repeat lengths cause longer stem-loop structures [80,82–84]. *HTT* RNA not only contains a CAG but also a polymorphic CCG repeat that translates into the PRD described above. RNA structure probing studies showed that the repeat stretches interact with each other, stabilizing a hairpin even in the healthy range of CAG repeats [84].

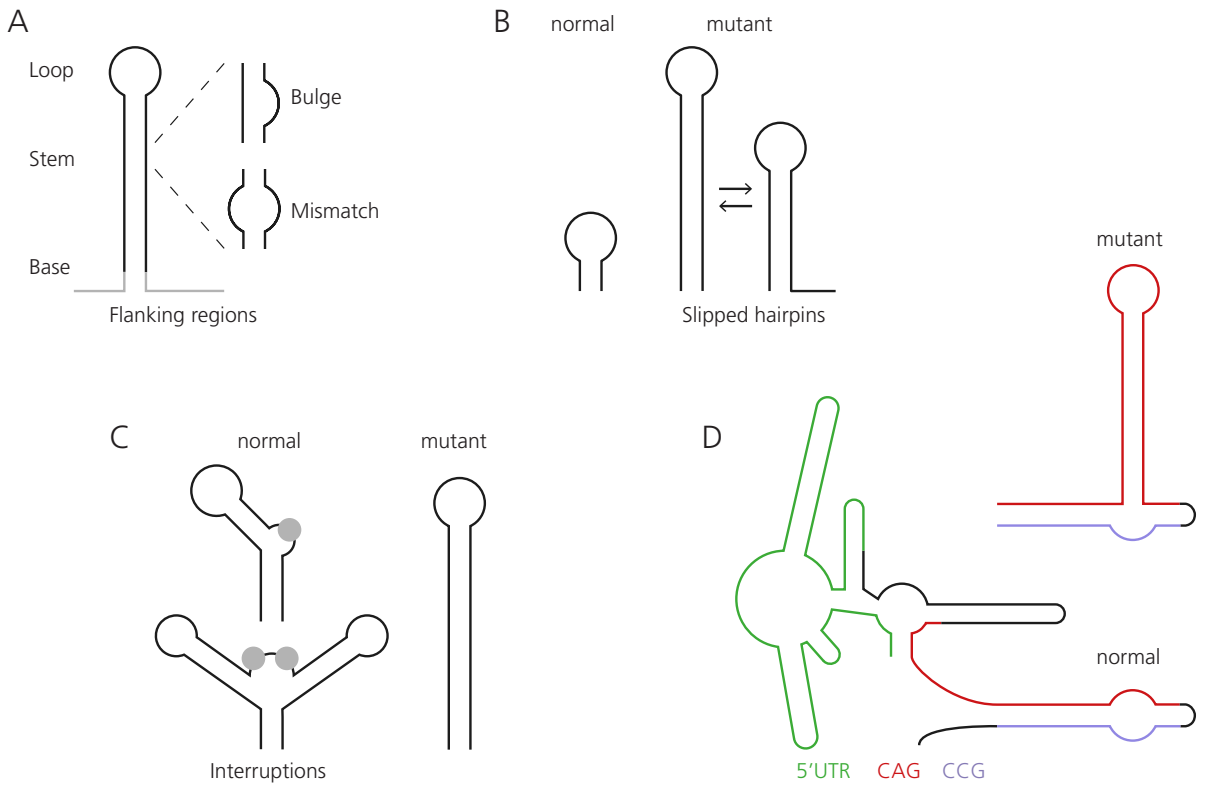

Figure 2 | TNR RNA structure and mutation-dependent effects.
(A) CNG repeats fold into hairpins that consist of a base region that can involve both the repeat region and the adjoining flanking regions. The stem can include different motifs and the terminal loop varies in size, both depending on the primary sequence. (B) *DMPK* and *ATXN3* RNA are examples for slipped hairpin structures, i.e. the RNA can fold into similar variants that align differently. (C) RNAs with a long but interrupted TNR that causes a mismatch adopt branched secondary structures. A mutation leading to a pure repeat tract can cause diseases as seen in SCA1, SCA2, and FXTAS. (D) A short CAG repeat within *HTT* RNA interacts both with the 5' UTR and polymorphic CCG repeat, while the pathogenic expansion produces a new structural motif [85].

In this model, the disease-causing mutation does not alter overall structure but only the stem length of the hairpin. Analysis of the full-length sequence of *HTT* exon 1 including the adjacent 5' UTR revealed more extensive interactions of the CAG repeat with flanking regions, resulting in structural differences for mutant CAG repeat lengths [85]. A transcript with 17 CAG repeats lacks the hairpin completely, while mutant repeats form a protruding CAG hairpin (Figure 2D). Even though the repeat units, their length, and the affected genes differ, most TNR diseases share common mechanisms regarding the molecular and cellular pathology of RNA-mediated toxicity. The following paragraphs give an overview of the underlying principles with a focus on HD.

1.3.1 Functional disruption of RNA-binding proteins

One RNA gain-of-function mechanism common to several TNR disorders involves aberrant interactions with RNA-binding proteins (RBPs) responsible for alternative splicing, transport, localization, stability, and translation of RNAs effectively impeding their normal function.

In DM1, the CUG repeat-containing 3' UTR leads to the retention of *DMPK* mRNA in the nucleus [86] and as a consequence sequesters the family of muscleblind-like (MBNL) proteins that are well known splicing factors. MBNL1 is recruited into stable *DMPK* RNA foci [87–89] and its retention leads to aberrant alternative splicing changes. The affected transcripts correlate with clinical phenotypes, e.g. the missplicing of chloride channel 1 can be linked to myotonia [90], sarcoplasmic/endoplasmic reticulum Ca^{2+} ATPase 1 to muscle wasting [91], and microtubule-associated protein tau (MAPT) to cognitive deficits [92]. These effects are likely reinforced through DEAD-Box helicase 5 (DDX5): by unwinding RNA secondary structure, DDX5 supports aberrant MBNL1 binding [93].

HD and FXTAS offer another example for this pathomechanism. Similar to mutant *HTT* mRNA, mutant *FMR1* mRNA with 60 to 600 CGG repeats adopts a pathogenic secondary structure, serving as a platform for RNA-binding proteins. Intranuclear inclusions were found in FXTAS patients [94] and these inclusions contain *FMR1* RNA and MBNL1 protein [95]. In HD patient-derived fibroblasts *HTT* RNA sequesters MBNL into nuclear foci [84] and subcellular fractionation of murine brains shows the age-dependent accumulation of expanded *HTT* RNA in the nucleus of an HD mouse model [96].

A direct link between neurodegeneration and mutant TNR RNA offers Nucleolin. Nucleolin is an important protein of the nucleolus, the location of ribosomal subunit assembly within the nucleus. Under physiologic conditions Nucleolin protects a control element of the rRNA promoter from CpG hypermethylation. Nucleolin dysfunction results in reduction of rRNA transcription, which in turn disturbs ribosome homeostasis. This leads to nucleolar stress that is linked to neurodegeneration and apoptosis [97–99]. In HD and SCA3, Nucleolin is sequestered by mutant CAG repeat RNA activating these downstream mechanisms [100,101]. In detail, free ribosomal proteins interact with the MDM3 E3 ubiquitin ligase resulting in downregulated p53 ubiquitination and increased accumulation. These events

activate mitochondrial cytochrome c release and the caspase cascade inducing apoptosis [102]. In line, p53 is upregulated in cell and animal models of HD, as well as in human brain of HD patients, emphasizing the relevance of nucleolar stress in HD.

Another protein linked to RNA toxicity is protein kinase R (PKR, also known as EIF2AK2). PKR has been shown to preferentially bind mutant *HTT* RNA and IHC staining of activated PKR is increased in brain tissue from HD patients [103]. PKR is activated by short stem-loop RNAs [104], possibly explaining how mutant *HTT* RNA can activate PKR. Activated PKR phosphorylates the α subunit of eukaryotic initiation factor 2 (eIF2α), thereby negatively regulating translation [105]. The eIF2α pathway is critically involved in local translation at synapses and memory formation [106] and is implicated in HD: the chemical inhibition of (PKR)-like endoplasmic reticulum kinase (PERK) in neurons, another eIF2α kinase, reduces Htt toxicity [64,107].

Finally, the RNA binding site of serine/arginine-rich splicing factor (SRSF6) is predicted to be a CAG or CAGCAA repeat motif and indeed, SRSF6 preferentially binds the 5' UTR of mutant *HTT* mRNA in mice [62]. Moreover, SRSF6 accumulates in inclusions in striatal tissue from HD patients and shows elevated expression levels [108]. Through an unknown mechanism, these interactions cause the mis-splicing of *HTT* exon 1 itself, *MAP2* and *MAPT*, leading to an imbalance of tau isoforms [62,108,109].

Table 2 presents an overview of validated direct TNR RNA-protein interactions with known downstream effects.

Table 2 | Proteins directly interacting with expanded TNR RNAs.

Repeat	Disease	Protein	Effect	Reference
CUG	DM1	MBNL1/ DDX5	Aberrant splicing	110,93
	DM1	hnRNP H	Impaired nuclear export	111
	SCA8	MBNL1	Aberrant splicing	112
CAG	HD	MID1	Enhanced translation	113
	HD	PKR	PKR activation	103
	HD	Dicer	Neurotoxicity	114,115
	HD	Nucleolin	Nucleolar stress	101
	SCA3	Nucleolin	Nucleolar stress, apoptosis	100
	SCA3	U2AF56, NXF1	Nuclear export	96
	SCA2,3,7	MID1	Enhanced translation	116
CGG	FXTAS	hnRNP A2/B1, Purα	Neurodegeneration	117
	FXTAS	Sam68	Aberrant splicing	118
	FXTAS	DROSHA, DGCR8	Reduced processing of miRNAs	119

1.3.2 CAG Repeat-dependent translational regulation by MID1

In HD, SCA2, SCA3, and SCA7 it has been shown that enhanced translation of the mutant CAG repeat RNA is involved in RNA-mediated toxicity and this process is mediated by the midline-1 (MID1) complex [113,116]. Since a large part of the work in this thesis focuses on MID1 and its involvement in repeat RNA translation, the following paragraph provides an introduction to the MID1 protein and its known functions.

MID1 belongs to the family of Really Interesting New Gene (RING) finger proteins [120] and contains six distinct domains: the N-terminal RING finger motif, two Bbox domains, a coiled-coil domain, a fibronectin type III, and a B30.2 domain. All domains are important for protein-protein interactions; for example, the C-terminal domain associates with microtubules [121] and the coiled-coil domain is responsible for homodimerization and heterodimerization with MID2, a close homologue [122]. Both proteins bind alpha 4, a regulatory subunit of protein phosphatase 2A (PP2A) phosphatase, through their Bboxes [122,123]. This interaction and dimerization are a prerequisite for association of MID1 and MID2 with microtubules [122]. Additionally, MID1 is an E3 ubiquitin ligase targeting the catalytic subunit of PP2A (PP2Ac) through the interaction with alpha 4 [124]. This MID1-dependent proteasomal degradation of PP2Ac downregulates mTOR complex 1 (mTORC1) signaling by disturbing the mechanistic target of rapamycin (mTOR)/Raptor complex [125].

Apart from its interaction with alpha 4, MID1 has been shown to associate with proteins involved in mRNA transport and translation, for example elongation factor 1 alpha (EF-1α), receptor for activated protein C kinase 1 (RACK1), Annexin A2, Nucleophosmin, 40S ribosomal protein SA, S3, and S8 [126]. Moreover, the MID1 complex is associated with G- and U-rich RNAs as part of a ribonucleoprotein (RNP) complex which plays a role in translation regulation [126].

In HD, MID1 binds *HTT* RNA in a length-dependent manner and induces translation by recruiting 40S ribosomal S6 kinase (S6K) preferentially to mutant *HTT* RNA, and simultaneously inhibits its binding partner PP2A and induces mTOR [113,124,125]. Since S6K is a target of PP2A and mTOR this leads to increased phosphorylation of S6K, which in turn activates S6K-dependent phosphorylation of its targets eukaryotic translation initiation factor 4B (eIF4B) and ribosomal protein S6. Subsequently, eIF4B promotes ribosome assembly and translation initiation. Interestingly, MID1-dependent translation is RNA structure-specific and particularly regulates repeat RNAs containing stem loops [116]. Besides *HTT* RNA this was also shown for mutant *ATXN2*, *ATXN3*, and *ATXN7* RNA [116].

1.3.3 Bi-directional Transcription

Bi-directional transcription produces sense and antisense transcripts from the same genetic locus and often plays regulatory roles in the expression of the sense transcript [127–129]. In the case of SCA8, this phenomenon results in the production of a sense transcript encoding the ataxin 8 protein harboring the mutant CAG repeat, while the antisense ATXN8 Opposite Strand (*ATXN8OS)* RNA contains a CUG repeat [130]. The translated mutant polyQ tract within ataxin 8 induces polyQ protein-mediated toxicity, while mutant *ATXN8OS* RNA localizes into RNA foci and sequesters MBNL1 [112].

Similarly, the huntingtin antisense (*HTTAS*) transcript is transcribed from the *HTT* locus and contains the disease-causing repeat [114]. Mutant CAG repeat lengths lead to downregulation of promoter activity in a reporter assay and the *HTTAS* transcript is downregulated in human HD frontal cortex supporting this analysis [114]. Experiments in cellular models confirm *HTTAS*-dependent regulation of *HTT* transcript levels: overexpression of *HTTAS* reduces endogenous *HTT* RNA. Consistently, siRNA-mediated knockdown of *HTTAS* increases *HTT* RNA levels and these effects are repeat length- and partially Dicer-dependent [114] (see next paragraph). Other diseases, where bi-directional transcription is implicated, include DM1 [131,132] and FXTAS [133].

1.3.4 Activation of siRNA-mediated Gene Silencing

Dicer is a type III endonuclease that recognizes and cleaves long double-stranded RNAs and precursor miRNAs [134]. The resulting products are 20 to 25 nucleotides long double-stranded short RNAs that enter the RNA interference pathway as micro (mi)RNAs or short interfering (si)RNAs, respectively. Binding of miRNAs by Argonaute proteins and the RNA-induced silencing complex (RISC) leads to the recognition of complementary mRNAs. The mRNAs are cleaved by RNases and consequently, mRNA translation is prevented. The elongated TNR stretch in mutant transcripts from *DMPK*, *HTT* and *ATXN1* mRNA serve as a substrate for Dicer-dependent production of 21 nucleotide long siRNAs [135]. The respective products and their downstream silencing effects could be detected in DM1, HD, and SCA1 patient-derived cells and post-mortem tissue [115,135]. Transfecting small RNAs isolated from brains of HD patients or cellular models is neurotoxic. Importantly, this effect is mediated by small CAG-repeated RNAs (sCAGs) as shown by co-transfection of complementary oligonucleotides [115].

The occurrence of bidirectional transcription generates another possible layer of regulation: antisense transcripts from disease-causing TNR mutations could act on additional targets through the RNA interference machinery.

1.3.5 RAN Translation

The translation of TNR transcripts can start at an arbitrary codon within the repeat without the need of an ATG start signal producing proteins of all three reading frames. This process is termed repeat-associated non-ATG (RAN) translation and was first described for SCA8. Investigating bi-directional transcription in SCA8, surprisingly, Zu et al. found homopolymeric proteins translated from an *ATXN8* minigene in the absence of an ATG start codon [136]. This type of translation depends on hairpin formation and the length of the repeat tract, it is independent of frameshifting, and even occurs when an ATG codon is present [136]. Keeping bi-directional transcription in mind, six possibly toxic proteins are translated from a single genetic locus: the CAG sense transcript may produce polyglutamine, polyserine, and polyalanine. Conversely, from the CUG antisense transcript polyleucine, polycysteine, and polyalanine may be RAN translated. Importantly, RAN proteins from CAG repeats enhance apoptosis in cultured cells and cerebellar Purkinje cells from human SCA8 patients are stained positive with a peptide antibody recognizing a putative polyalanine RAN protein [136]. Sense and anti-sense RAN proteins can be detected in human HD brains, these proteins are toxic to cells and their accumulation and aggregation is CAG length-dependent [137]. Moreover, RAN proteins can be identified in DM1 patient-derived cells [136], FXTAS patient brains and various FXTAS models [138].

1.3.6 Trans-dominant effects

Although many of the toxic downstream effects can be attributed to specific aberrant interactions of mutant TNR RNA with proteins, the case of CUG RNA-binding protein 1 (CUGBP1) is not as straightforward. CUGBP1 is part of the CELF protein family binding CUG repeat RNAs like *DMPK* [139] and is involved in splicing regulation and translation. Unlike MBNL1, it is not recruited to inclusions observed in DM1 [140,141], but protein levels are increased in various DM1 tissues [142,143] and a CUGBP1 overexpressing transgenic mouse model develops a muscle phenotype and shows characteristic DM1 splicing changes [144]. The underlying mechanism is a protein kinase C (PKC)-dependent hyperphosphorylation and stabilization of CUGBP1 that is induced by expanded CUG repeat RNA [145]. It is unclear how mutant *DMPK* RNA can trigger this signaling event. Nevertheless, downstream effects on RNA targets have been identified: mis-splicing of troponin T type 2 mRNA is linked to defective cardiac functions [146] and insulin receptor to insulin resistance [142].

1.4 RNA toxicity in HD: Aim of study

As part of the polyQ diseases, HD is caused by a TNR mutation in the coding region of *HTT* and many aspects of protein pathology have been elucidated. Research investigating the contribution of RNA toxicity to the HD phenotype clearly shows the detrimental role of the expanded CAG repeat on the transcript level (Figure 3).

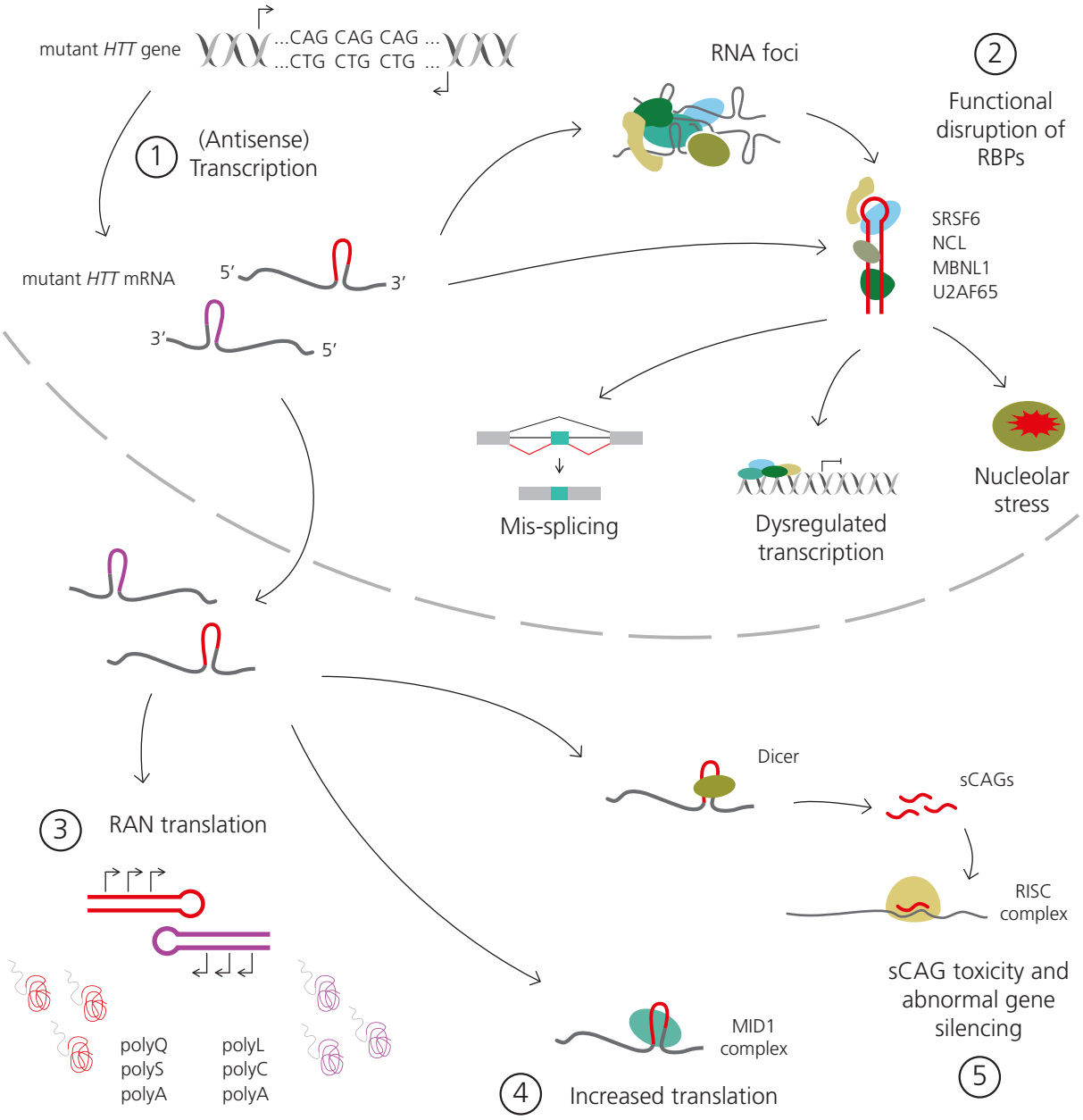

Figure 3 | Mechanisms of RNA toxicity in HD.
(1) Sense and antisense transcription produces two transcripts that both contain an elongated TNR repeat, possibly multiplying downstream mechanisms. (2) In the nucleus, mutant *HTT* RNA may disrupt RBPs through sequestration into stable RNA foci, impeding their physiologic functions. (3) In the cytoplasm, mutant CAG repeats induce RAN translation that leads to additional aberrant protein species. (4 and 5) Increased or abnormal interactions with RBPs upregulate mutant *HTT* RNA translation (MID1) and generate toxic sCAGs (Dicer), respectively.

The change in secondary structure leads to aberrant protein interactions, impeding their normal functions or even enhancing it, as in the case of MID1-dependent translational upregulation. Moreover, the repeat tract can induce RAN translation, producing more and possibly toxic protein species and is a substrate for Dicer-dependent gene silencing. These mechanisms underline the importance of the *HTT* RNA-protein interactome and specifically, how the mutation affects these interactions.

To better understand mutant *HTT* RNA mediated processes, this work analyzes *HTT* RNA-protein interactions from two angles. On the one hand, the known binding partner MID1 is analyzed in terms of its complex composition, determinants of repeat translation, and its expression in the context of HD. On the other hand, this work aims to systematically map *HTT* RNA-protein interactions and to analyze the implications of the disease causing mutation on RNA binding proteins and the associated downstream effects.

2 Methods

2.1 Chemicals

All standard chemicals were obtained from Sigma-Aldrich and Carl Roth.

2.2 Equipment

12 Tube Magnetic Separation Rack, New England Biolabs
7900 HT Fast Real-Time PCR System, Applied Biosystems
Biological Safety Cabinet Class 2 - Mars, ScanLaf
CASY® Cell Counter, Innovatis
DNA Engine® Dyad Peltier Thermal Cycler, Bio-Rad
FLUOstar® Omega Microplate Reader, BMG LABTECH
HERAcell 240i CO2 Incubator, Thermo Scientific
HERAEUS Fresco 21 Centrifuge, Thermo Scientific
HERAEUS Multifuge X3R Centrifuge, Thermo Scientific
Mini PROTEAN® Tetra Cell Electrophoresis System, Bio-Rad
NanoDrop 2000c Spectrophotometer, Thermo Scientific
PowerPac™ Basic and Universal Power Supply, Bio-Rad
Precellys® 24 Homogeniser (Peqlab)
Sonopuls HD 2070 Ultrasonic Homogenizer, Bandelin
Stella 3200, Raytest
Sub-Cell® GT Agarose Gel Electrophoresis System, Bio-Rad
Thermomixer Comfort, Eppendorf
Trans-Blot SD Semi-Dry Transfer Cell, Bio-Rad

2.3 Standard Procedures

2.3.1 Gel Electrophoresis & Immunoblotting

Agarose gel electrophoresis of nucleic acids, SDS-PAGE of proteins, and immunoblotting were performed as described in Ausubel Current Protocols [147]. To estimate DNA lengths, GeneRuler™ 100 bp or 1 kb DNA Ladder (Thermo Fisher Scientific) was used; for RNA the RiboRuler™ High Range RNA Ladder (Thermo Fisher Scientific) was used. 4x SDS Buffer (62.5 mM Tris-HCl (pH 6.8), 20 %

glycerol, 2 % SDS, 5 % β-mercaptoethanol, 0.025 % (w/v H_2O) bromophenol blue) was used for protein analysis.

Table 3 | List of antibodies

Name	Catalogue number
CPSF6	Abcam ab175237
CREB1	CST #9197S
eIF3A	CST #3411
eIF4A	CST #2013
eIF4B	CST #3592
eIF4G	CST #2498
FLAG-HRP	Sigma A8592
FMRP	Abcam ab17722
NUDT21	Abcam ab183660
PABP1	CST #4992
PRPF8	Abcam ab79237
RACK1	BD Trans. 610177
RALY	Abcam ab170105
RPL5	CST #51345
RPLP0	Abcam ab192866
RPS3	Abcam ab128995
SF3B2	Abnova H00010992-M01
SNRNP40	Abnova PAB21803
SON	Sigma HPA023535

2.4 Cell Culture

2.4.1 Cell lines and Cultivation

The human neuroblastoma cell lines SHSY-5Y and SHSY-5Y-EGFP-HTTex1Q68 were kept in cell+ flasks (Sarstedt) in DMEM-Glutamax™ (Invitrogen life technologies) supplemented with 15% filter-sterilized fetal bovine serum (FBS, PAN-Biotech GmbH). Human embryonic kidney cells 293T (HEK-T) cells were cultivated under the same conditions except for the amount of FBS added (10%).

The SHSY-5Y-EGFP-HTTex1Q68 cell line was generated by Yvonne Dürnberger (Deutsches Zentrum für neurodegenerative Erkrankungen, Bonn) using the Lenti-X Tet-On 3G Inducible Expression System (Clontech Laboratories) and the following protocol. The huntingtin exon 1 coding sequence was cloned in the inducible expression vector pLVX TRE3G. To generate lentiviral particles encoding the gene of interest, HEK-T cells were transfected with the expression vector and Lenti-X Packaging single shots (VSV-G). For all steps cells were grown under tetracycline-free conditions. Lentiviral particles were collected 48 h and 72 h post transfection, pooled and concentrated using PEG precipitation. After precipitation lentiviral particles were carefully resuspended in phosphate buffered saline (PBS) and stored at -80 °C. For transduction SHSY-5Y cells were incubated with lentiviral

particles in DMEM culturing medium containing 5 % FBS and 8 µg/ml polybrene. The well plate containing the cells and viral particles was briefly centrifuged and afterwards incubated for 24 h. The following day the medium was exchanged with standard culturing medium (DMEM, 10 % FBS, 5 % Penicillin/ Streptomycin) and the cells expanded. Aliquots of the generated stable cell line were frozen in liquid nitrogen for future experiments.

2.4.2 siRNA Transfection

Oligofectamine

Cells were seeded in a 24-well plate (10^5 cells per well in 500 µl growth medium) in DMEM Glutamax supplemented with 10 % FBS. On the following day, siRNA-mediated knockdown was performed. The respective siRNAs were combined equally to a final concentration of 20 µM and non-silencing siRNA served as control (sequences see Table 4). For each well, two mixes were prepared. Mix one contained 2.5 µl siRNA cocktail in 40 µl OptiMEM (Gibco) and mix two 1.5 µl Oligofectamine (Thermo Fisher Scientific) in 7.5 µl OptiMEM. Mix two was incubated 5 min at room temperature, then added to mix one and incubated for another 20 min at room temperature. The growth medium was aspirated from the cells and 200 µl OptiMEM was added carefully to each well. 50 µl transfection mix was added to the cells. After 4 h at 37 °C, 125 µl DMEM Glutamax containing 30 % FBS was added to each well and the plates were returned to the incubator.

Lipofectamine 2000

The respective siRNAs were combined equally to a final concentration of 20 µM and non-silencing siRNA served as control (sequences seeTable 4). For each well, two mixes were prepared. Mix one contained 1 µl Lipofectamine 2000 in 50 µl OptiMEM and mix two 2.5 µl siRNA cocktail in 50 µl OptiMEM. Mix two was added to mix one drop-wise and incubated for 5 min at room temperature. 300 µl growth medium was aspirated from each well, 100 µl transfection mix was added, and cells were incubated for 48 hrs at 37 °C.

Table 4 | siRNA sequences

Name	sequence	target
Non-silencing	AATTCTCCGAACGTGTCACGT	Non-silencing
Hs-MID1-3	CACCGCAUCCUAGUAUCACACTT	MID1
Hs-MID1-4	CAGGAUUACAACUUUUAGGAATT	MID1
Hs-MID1-8	TTGAGTGAGCGCTATGACAAA	MID1
Hs-MID1-9	AAGGTGATGAGGCTTCGCAAA	MID1
Hs-MID1-10	TAGAACGTGATGAGTCATCAT	MID1

2.4.3 Plasmid Transfection

Plasmid preparation

Bacterial transformation and culture were performed as described previously [147]. For DNA isolation the JETSTAR Plasmid Purification Maxi Kit (Genomed GmbH) was used according to the manufacturer's instructions.

Lipofectamine 2000

For each well, two mixes were prepared. Amounts listed are for cells grown in 24-well format in 400 µl medium. Mix one contained 50 ng plasmid DNA (s. appendix x) in 50 µl OptiMEM and mix two 0.25 µl Lipofectamine 2000 (Thermo Fisher Scientific) in 50 µl. Mix one was added to mix two drop-wise and was incubated for 20 min at room temperature. Then, 100 µl were added to each well and the cells were incubated for 24 h at 37 °C.

PolyFect

One day prior to transfection, 2 x 10^6 HEK-T cells were seeded in a 75 cm^2 cell culture flask. To 500 µl OptiMEM 30 µl PolyFect (Qiagen) and 10 µg pCMV-MID1-Tag2A was added and incubated for 10 min at room temperature. The transfection mix was added and cells were incubated for 48 h at 37 °C.

2.5 Gene Expression Analysis

2.5.1 RNA preparation & Reverse Transcription

Total mRNA was prepared using the RNeasy® Plus Mini Kit (Qiagen) with QIAshredder columns (Qiagen) for homogenization. cDNA synthesis reactions were prepared according to the manufacturers instructions (TaqMan Reverse Transcription Reagents, Roche Applied Biosystems) and the following temperature profile was used: 25 °C for 10 min, 48 °C for 1 h, 95 °C for 5 min, and cooling down to 4 °C.

2.5.2 Realtime PCR

cDNA was diluted five- to ten-fold including a standard series (five 1:2 dilutions starting at the same concentration as the samples). Per well, 9 µl SYBRGreen PCR Master Mix (Applied Biosystems, Nr. 4309155), 3 µl of a 2 µM primer pair mix (s. appendix Table 6), and 6 µl diluted cDNA were combined. Each sample was analyzed in quadruplets.

2.6 Luciferase Assays

HEK-T cells were seeded in a 24-well plate (10^5 cells per well in 500 µl growth medium) and the siRNA-mediated knockdowns were performed the following day using Oligofectamine (protocol s. chapter 2.4.2). Plasmid DNA was transfected the next day using Lipofectamine 2000 for plasmid transfection (protocol s. chapter 2.4.3). On day four, cells were harvested for luciferase activity measurements and samples for western blot analysis and qPCR were taken to analyze knockdown efficiency. For luciferase measurements, cells were washed once with 200 µl PBS, 100 µl 1 x Passive Lysis Buffer (Dual-Luciferase Reporter Assay System, Promega) was added to each well and the cells were either frozen at -20 °C or directly lysed for 15 minutes at room temperature with gentle shaking. Then, 400 µl H_2O was added and the lysate was homogenized by pipetting. Each plasmid DNA was transfected in triplicates with either MID1 or MID2 single knockdown, MID1 and MID2 double knockdown, or control siRNA. Firefly and Renilla luciferase activity was measured in individual wells of a 96-well white bottom plate. To this end, from each lysate ten microliters were transferred in triplicates for both measurements. 40 µl of Firefly or Renilla luciferase substrate buffer was added to the wells and the luciferase activity was measured.

To check for knockdown efficiency on RNA level, two wells of HEK-T cells transfected with the different siRNAs and psiCHECK-2 control vector were pooled and RNA was prepared. cDNA synthesis and qPCR analysis were performed as described in chapter 2.5. For the standard series, RNA from cells transfected with control siRNA was used.

2.7 Immunoprecipitation (IP)

TKM Buffer: 20mM Tris pH 7.4, 100mM KCl, 5mM $MgCl_2$, 0.5% NP40, 1mM DTT, protease inhibitor

Buffer D: 20mM Tris, 100mM KCl, 0.2mM EDTA, 20% glycerol, 0.5mM DTT, protease inhibitor

Unless indicated otherwise, agarose beads were washed with TKM buffer and pelleted for 1 min at 1,000 x g. The amounts of agarose beads described denote 50% bead slurry.

2.7.1 MID1 IP for Mass Spectrometry

Agarose beads were prepared as follows: 200 µl FLAG-beads per sample were washed twice. For untransfected control lysate FLAG-beads were coated with 1 mg/ml FLAG peptide in TKM buffer rotating for 1 h at 4 °C and again, beads were washed twice.

Two 75 cm^2 flasks of either HEK-T cells transfected with pCMV-MID1-Tag2A (PolyFect transfection protocol, chapter 2.4.3) or untransfected cells were harvested using a cell scraper and centrifuged for 10 min at 500 x g. The cell pellets were resuspended in 1 ml TKM Buffer, lysed with Precellys

(program: 6,000 – 1 x 10 – 005), and centrifuged at 21,000 x g for 10 min at 4 °C. For pre-clearing, 200 µl of IgG-agarose beads were added to the supernatants and incubated rotating for 30 min at 4 °C. The beads were pelleted for 5 min at 21,000 x g. Lysate from MID1-FLAG expressing HEK-T cells was added to 200 µl uncoated αFLAG beads and untransfected control lysate to 200 µl FLAG peptide-coated αFLAG agarose beads. After overnight incubation rotating at 4 °C, the beads were washed 6 times. Finally, the bead pellets were resuspended in 50 µl 1x SDS Buffer, boiled for 10 min at 95 °C and 30 µl supernatant were sent to external Mass Spectrometry analysis (chapter 2.9).

2.7.2 MID1 IP for Validation

A 75 cm^2 flask of pCMV-MID1-Tag2A transfected HEK-T cells was harvested with a cell scraper and the cell pellet was collected by centrifugation for 10 min at 500 x g. The cell pellet was resuspended in 1 ml TKM Buffer, sonicated (10 sec duration, 50 % cycle, 50 % amplitude) and centrifuged at 21,000 x g for 10 min at 4 °C. 45 µl lysate was aliquoted and frozen with 15 µl 4x SDS buffer at -20 °C. For pre-clearing, 100 µl mouse IgG-agarose beads were added to the supernatant, incubated for 30 min rotating at room temperature, and centrifuged for 5min at 21,000 x g. To prevent unspecific binding, 50 µl FLAG-beads and 50 µl IgG beads were washed and blocked with 2 mg/ml BSA in TKM buffer for 30 min at room temperature. Then, the lysate was divided in 500 µl aliquots and 50 µl αFLAG beads or 50 µl IgG beads were added in a final volume of 1 ml buffer. After overnight incubation at 4 °C and rotating, the beads were washed 6 times for 10 min and finally dissolved in 60 - 100 µl 1x SDS Buffer. Immunoprecipitations were incubated for 10 min at 95 °C and analyzed by western blot.

2.7.3 MID1 IP with Ribosome Disassembly

Immunoprecipitation was performed as described in chapter 2.7.2. Additionally, a second IP and control reaction were prepared with TKM buffer containing 40 mM EDTA during precipitation and washing.

2.8 RNA-protein pulldown

2.8.1 Pulldown for Mass Spectrometry

PCR

The coding sequence of HTT exon 1 with varying CAG repeat lengths was amplified from plasmid DNA (constructs described in Krauß et al. 2013 [113]) using GoTaq® Green Master Mix (Promega). A 50 µl reaction contained 1 µl of 10 µM forward and reverse primer mix (sequences see Table 6), 1 µl DMSO and 100 ng DNA template. The forward primer incorporates the T7 phage promotor sequence

upstream of HTT exon 1 to facilitate *in vitro* RNA synthesis by T7 RNA polymerase. The following temperature profile was used: 95 °C for 3 min, 25 cycles of 95 °C for 30 s, 58 °C for 30 s, and 72 °C for 3 min. The final elongation was carried out at 72 °C for 10 min. Correct PCR product length was validated on a 1.5 % agarose gel, DNA was purified by phenol/chloroform extraction, and precipitated using ethanol.

In vitro transcription of biotinylated RNA

The T7 RiboMAX™ Express Large Scale RNA Production System (Promega) was used according to the manufacturer's instructions supplemented with 0.5 mM biotin-UTP (Thermo Fisher Scientific, AM8450). DNA template was removed by DNase digestion; RNA was purified by phenol-chloroform extraction, and precipitated using ethanol. Correct RNA length was validated on a 1.5 % agarose gel.

RNA-protein pulldown protocol

10x RNA Structure Buffer: 100mM Tris pH7, 1M KCl, 100mM $MgCl_2$

rNTP stock: 25mM rATP/ rCTP/ rGTP, 12.5mM rUTP (Promega), 12.5mM bUTPs

TKM Buffer: 20mM Tris, 100mM KCl, 5mM $MgCl_2$, 1mM DTT, 1% NP40, protease inhibitor, 1µl /1ml RNase inhibitor

Biotinylated RNA was folded in RNA structure buffer by incubation at 72 °C for 10 min and subsequently cooling it down to room temperature slowly. For pulldowns, 40 µl Dynabeads® M-280 Streptavidin (Thermo Fisher Scientific) per sample were washed twice with TKM buffer and then incubated with 40 pmol of folded biotinylated RNA for 30 min at room temperature using gentle rotation. As control, beads were coated with a mix of rNTPs and bUTPs. The coated beads were washed twice with TKM buffer and resuspendend in 40 µl TKM Buffer.

For protein lysates, four 150 cm^2 cell culture flasks with 90 % confluent SH-SY5Y were harvested using a cell scraper, pooled and centrifuged at 500 x g for 10 min. The cell pellet was lysed in 1 ml TKM Buffer, homogenized using Precellys, and centrifuged for 10 min at 12,000 x g and 4 °C. The protein concentration was measured with Qubit® Protein Assay Kit (Thermo Fisher Scientific) and 1 mg of protein lysate was added to the control and biotinylated RNA coated magnetic beads in a final volume of 400 µl TKM Buffer. The samples were incubated overnight at 4 °C with gentle rotation. The beads were washed 3 times at 4 °C, the supernatant was completely removed, and the beads were resuspended in 20 µl 1x SDS Buffer. After incubation at 95 °C for 10 min, the eluted proteins were sent to external Mass Spectrometry analysis (chapter 2.9).

2.8.2 Pulldown for Validation

PCR

The synthesis of HTT exon 1 coding sequence was done essentially as described in chapter 2.8.1, however, a different forward primer was used for amplification (sequence see Table 6). This primer

incorporates additional bases that are complementary to a biotinylated DNA (bDNA) linker (Table 6), which can be used to coat the *in vitro* transcribed RNA onto beads.

In vitro transcription of biotinylated RNA

RNA was synthesized as described in chapter 2.8.1 without the addition of biotin-UTP.

RNA-protein pulldown protocol

10x Renaturation buffer: 100mM Tris (pH7.5), 1M KCl, 1mM EDTA

5x Folding Buffer: 200mM Tris pH7, 250mM $MgCl_2$, 100mM KCl, 2.5mM EDTA

40 pmol *in vitro* transcribed RNA and 40 pmol bDNA linker were combined in 1x Renaturation Buffer, in a final volume of 50 µl. For denaturation, the nucleic acids were heated for 1 min at 85°C, and subsequently cooled down slowly to room temperature. Then, 20 µl of 5x Folding Buffer and 30 µl H_2O were added and incubated for 1 hr at 37°C to facilitate re-folding and annealing. 20 µl Dynabeads® M-280 Streptavidin (Thermo Fisher Scientific) per sample were washed with 500 µl TKM Buffer and incubated with RNA / bDNA linker hybrids and only bDNA linker for 30min at room temperature, rotating. A confluent 75 cm^2 flask of SHSY-5Y cells was harvested using a cell scraper, pelleted for 5 min at 500 x g, and lysed in 1 ml TKM Buffer by sonication (10 sec duration, 50 % cycle, 50 % amplitude). The lysate was centrifuged for 10 min at 21,000 x g, 45 µl lysate were added to 15 µl 1x SDS loading buffer and frozen at -20 °C. The remaining lysate was pre-cleared on 80 µl Dynabeads in TKM Buffer for 5 hrs at 4 °C, rotating. The protein lysate was divided into equal aliquots and added to RNA coated magnetic beads and bDNA control beads in a final volume of 500 µl. The RNA-protein pulldown was incubated rotating at 4 °C overnight. Next, the beads were washed 3 times for 10 min with 500 µl TKM Buffer at room temperature and finally, 40 µl 1x SDS loading buffer was added to the beads, incubated at 95°C for 10 min and analyzed on western blot.

2.9 Mass Spectrometry

Protein samples were analyzed by Alina Dagane (Max Delbrück Center for Molecular Medicine, Berlin) as follows. The eluted proteins were concentrated into one band on an SDS-PAGE gel. The band was excised and the proteins contained were processed using an automated sample preparation setup [148]. The generated peptides were purified on StageTips [149]. The samples were measured on a Q-Exactive mass spectrometer (Thermo-Fisher) coupled to a Proxeon nano-LC system (Thermo-Fisher) in data-dependent acquisition mode, selecting the top 10 peaks for HCD fragmentation. A 1-h gradient (solvent A: 5 % acetonitrile, 0.1 % formic acid; solvent B: 80 % acetonitrile, 0.1 % formic acid) was applied for the samples using an in-house prepared nano-LC column (0.075 mm × 150 mm, 3 µm Reprosil C18, Dr. Maisch GmbH). A volume of 2 µl sample was injected and the peptides eluted with

3 h gradients of 5 to 75 % solvent B at flow rates of 0.25 µl/min. MS acquisition was performed at a resolution of 70,000 in the scan range from 300 to 1700 *m*/*z*. The normalized collision energy was set to 26 eV. The mass window for precursor ion selection was set to 2.0 *m*/*z*. The recorded spectra were analyzed using the MaxQuant software package (Version 1.3.0.5) [150] by matching the data to the Uniprot yeast database (downloaded on 06.05.2012) with a false discovery rate (FDR) of 1%.

2.10 Human Brain Tissue

Human brain tissue was acquired from W.M.C. van Roon-Mom, department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands. The tissue was obtained with the families' full consent and with the approval of the Leiden University Medical Center Institutional Ethics Committee. Detailed description of the tissue processing can be found in Waldvogel et al. 2008 [151].

Table 5 | Clinical features of brain tissue donors.

ID	Sex	Age	CAG repeat	Grade	PMD	Cause of death	qPCR	S/C	IHC
C1	M	42	--	--	14	--	+		+
C2	F	64	18 / 23	--	6	--			+
C3	F	59	15 / 17	--	21	--			+
C4	M	41	--	--	16	--	+		+
C5	M	64	17 / 18	--	7	--			+
C6	M	89	--	--	19	--	+		+
C7	M	48	--	--	--	myocard infarct	+	+/+	
C8	F	78	--	--	--	subdural hematoma	+	+/+	
C9	F	89	--	--	--	subdural hematoma	+	-/+	
HD1	M	41	19 / 39	1	11		+		+
HD2	M	40	18 / 51	3	15		+		+
HD3	F	67	15 / 42	1	9		+		+
HD4	M	75	19 / 43	3	3		+		+
HD5	F	53	21 / 47	--	12		+		+
HD6	M	57	17 / 43	--	--				+
HD7	F	57	--	--	--		+	+/+	
HD8	M	62	--	--	--		+	+/+	
HD9	M	48	--	--	--		+	+/+	

M, male; F, female; Grade: Vonsattel [152]; PMD, post-mortal delay; S/C, striatal/cerebellar samples analyzed by qPCR.

2.11 Immunohistochemistry

2.11.1 Coating slides

Coating solution: 0.5 % gelatin, 0.05 % chromium potassium sulphate in H_2O

Slides were rinsed in 75 % alcohol with 1% hydrochloric acid and dried for 30 min at 60 °C. After, they were dipped in gelatine solution once and dried overnight at 60 °C.

2.11.2 IHC procedure for tissue sections

PBS-T: 0.2 % Triton X-100 in PBS

Normal goat serum: 1 % normal goat serum, 0.04 % merthiolate in PBS-T

Methanol solution: 0.1 % H_2O_2 in 50% methanol

Sections were cut from fixed-frozen human tissue blocks and were used in a free floating immunohistochemical staining procedure [153,154]. Blocks were frozen onto the cutting stage of a freezing microtome using OCT compound as an adhesive. The sections were cut at either 30 – 50 µm and collected in 1 % sodium azide in PBS. Next, sections were washed 3 times in PBS-T for 5 min on a shaker at room temperature. They were incubated for 20 min at room temperature in methanol solution and again washed 3 times. Incubation in the primary antibody (MID1, 1:100) in normal goat serum was done overnight at 4 °C on a shaker. Following 3 washing steps, the sections were incubated in secondary antibody (goat anti-rabbit biotin, Chemicon AP132B, 1:1,000) in normal goat serum for 3 h at room temperature on a shaker. Again, the sections were washed 3 times, and then incubated in tertiary antibody (streptavidin-HRP, Southern Technology 7100-05, 1:1,000) in normal goat serum for 3 h at room temperature on a shaker. After washing 3 times, the sections were incubated with DAB solution (Sigma FAST Tablets D4293) for up to 20 minutes at room temperature and were washed 3 more times before they were mounted on a gelatine-coated glass slide and left for dehydration at room temperature overnight.

2.11.3 Nissl Staining with Cresyl Violet

Staining solution: 0.6 mM acetic acid, 2 % aqueous Cresyl Violet solution (filtered), 100 mM sodium acetate; combined in a ratio of 9:1:0.5

Dried tissue sections mounted on a glass slide were washed thoroughly in dH_2O and incubated in staining solution for 15 to 20 min. Then, slides were quickly washed in dH_2O and dehydrated in increasing amounts of ethanol for 5 min each (75 %, 85 %, 95 %), 2 times in 100 % ethanol for 10 min and 2 times in xylene for 10 min. Slides were covered with PERTEX and a coverslip.

2.11.4 Quantitative Analysis of MID1 positive Cells

For quantification of MID1 IHC staining in human cortical sections the Definiens Developer XD 2.3 was used. The region of interest (white matter) was selected manually, the thresholds for signal detection were adjusted (0.35 for cresyl violet and 0.4 for DAB staining), and nucleus size was set to 40 µm.

2.12 Preparation of Mouse Brain Regions

Female wild-type mice and transgenic mice of the Hdh(CAG)150 knock-in model of HD [155] were sacrificed by cervical dislocation, the whole brain was retrieved from the skull and placed on a -20 °C cold metal plate on ice. The brain regions were dissected as described in detail elsewhere [156], then placed in RNAlater® solution (Thermo Fisher Scientific) to preserve RNA integrity. For isolation of RNA, appropriate amounts of tissue were lysed in Buffer RLT using Precellys (program: 6,000 – 1 x 10 – 005, twice) and processed as described in chapter 2.5.1.

2.13 Transcriptome Profiling

Samples of doxycycline-treated and untreated SHSY-5Y-EGFP-HTTex1Q68 cells were used for RNA purification in triplicates. cDNA labeling, data acquisition and statistical analysis were performed by Jennifer Winter (Institute of Human Genetics, Mainz) using the Clariom™ D Assay, human (Thermo Fisher Scientific), and the Expression Console and Transcriptome Analysis Console (Affymetrix).

2.14 Online tools and Statistical Analysis

For Gene Ontology Analysis ToppGene Suite was used [157]. For pathway analysis data were analyzed through the use of IPA (Ingenuity® Systems, www.ingenuity.com). The RNAfold Server was used to predict RNA secondary structures [158]. Intersections of different lists of genes was analyzed using the Venn diagrams tool (http://bioinformatics.psb.ugent.be/webtools/Venn/).
Statistical analyses were conducted as appropriate and are detailed in the figure legends, with the help from the Image Data Analysis Facility, Deutsches Zentrum für Neurodegenerative Erkrankungen, Bonn.

3 Results

A fundamental mechanism underlying RNA-mediated toxicity in TNR disorders is the disturbed interaction with RBPs. However, to date not many direct protein binding partners of mutant RNA have been identified. Here, I will focus on HD as a prominent example for TNR disorders with two major goals. First, the characterization of one particular RBP that has been shown to enhance mutant HTT RNA translation: the MID1 complex. Second, the identification of novel RBPs binding mutant HTT RNA and the analysis of downstream effects.

3.1 The MID1 complex

3.1.1 MID1 regulates translation of RNA containing a CAG repeat in the 3′ UTR

Elongated CAG repeat RNA translation in the genetic context of *HTT* exon 1 has been shown to be MID1-dependent [113]. Moreover, the MID1 complex is also involved in translational regulation of two other RNAs, namely androgen receptor (*AR*) [159] and beta-secretase 1 (*BACE1*) [160] and has been shown to bind *AXTN2*, *ATXN3*, and *ATXN7* RNA [116]. To investigate MID1-dependent translation outside of a genetic context, an isolated CAG repeat was used in an *in vitro* luciferase reporter system.

First, the positional effect of the repeat was analyzed. HEK-T cells were transfected with a plasmid containing two luciferase coding sequences: a Renilla luciferase with a repeat of 50 CAG triplets either in its 5' or 3' UTR, and a Firefly luciferase as an internal transfection control used for normalization (Figure 4.A). Luciferase translation was quantified indirectly by measuring the catalytic activity: the oxidation of coelenterazine (Renilla luciferase) and luciferin (Firely luciferase), respectively, produces bioluminiscence that can be measured by a luminometer.

Figure 4.B shows that only a CAG repeat in the 3' UTR but not the 5' UTR enhances Renilla luciferase translation. To test if this effect depends on MID1, Renilla-3'UTR-$(CAG)_{50}$ activity was determined after siRNA-mediated knockdown. Since MID1 and MID2 show functional redundancy during avian embryogenesis [161], additionally, a MID2 and double knockdown was performed to determine if MID1 and MID2 are functionally redundant in this context, too. MID1 but not MID2 affects Renilla-3'UTR-$(CAG)_{50}$ translation (Figure 4.D). The effect of the double knockdown on Renilla-3'UTR-$(CAG)_{50}$ translation was not as pronounced in comparison to the MID1 single knockdown. Since the knockdowns were performed with equimolar amounts of siRNAs, the double knockdown contained half the amount of MID1 and MID2 siRNAs compared to single knockdowns, offering an explanation for the aforementioned effect. MID1 knockdown was validated using western blot and quantitative polymerase chain reaction (qPCR); since no MID2 antibody is available, knockdown efficiency of

MID2 was analyzed by qPCR only (Figure 4.C and E). In conclusion, these *in vitro* reporter assays show that MID1, but not MID2, regulates translation of RNAs that contain a 3' UTR-$(CAG)_{50}$ repeat, i.e. translational regulation is not necessarily specific for a certain genetic context.

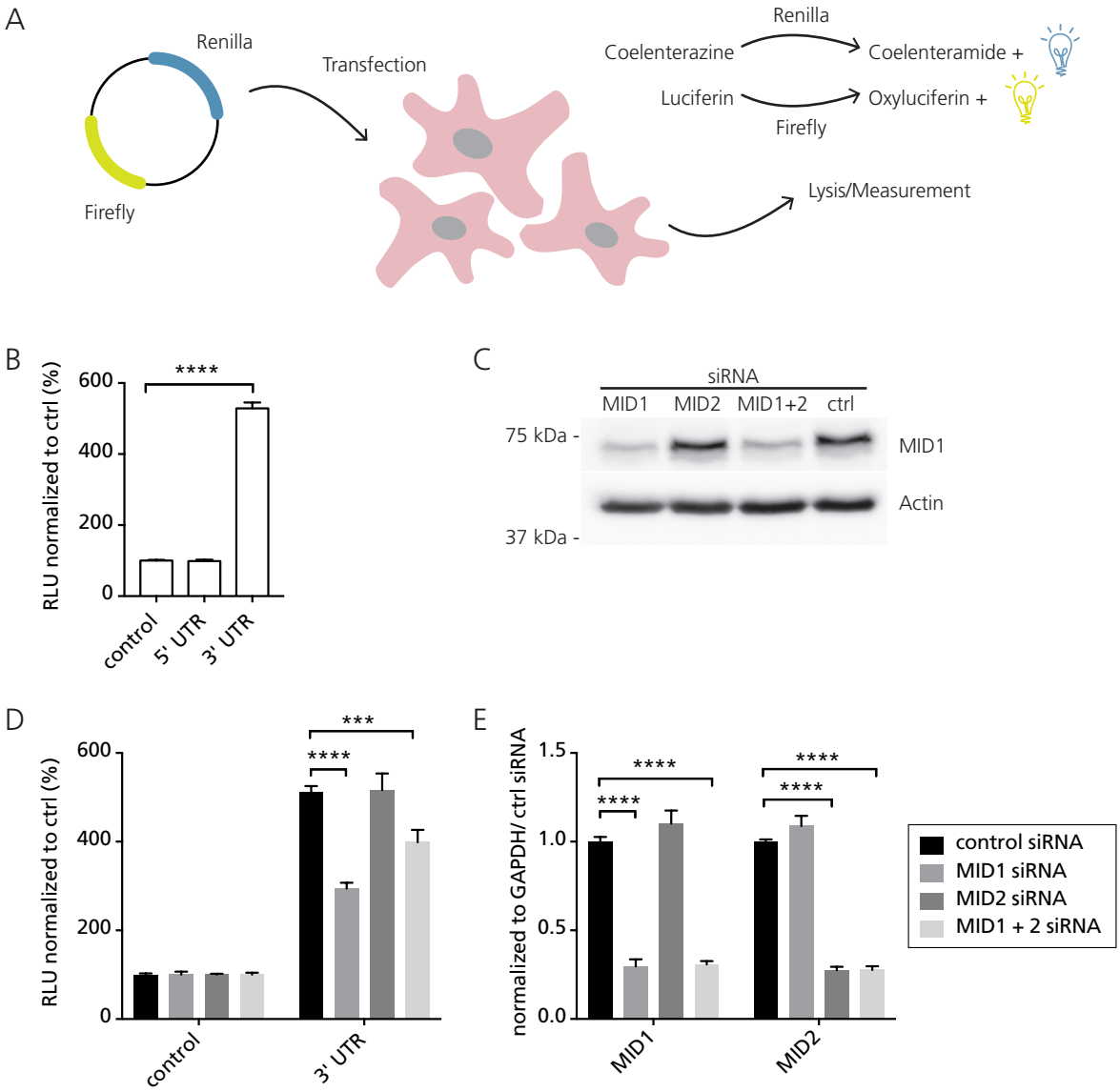

Figure 4 | Translation of RNA carrying a CAG repeat in the 3'UTR is regulated by MID1.
(A) Luciferase assay scheme illustrating transfection of the luciferase coding sequence-containing plasmids into HEK-T cells. Renilla luciferase contains the respective repeat tract, while Firefly luciferase was used for normalization of transfection efficiency. Experiments were repeated on separate days, for each condition cells were seeded in triplicates. Bioluminescence was measured after lysis and addition of appropriate substrates, each lysate was measured in triplicates. (B) Luciferase measurement comparing Renilla luciferase with a $(CAG)_{50}$ repeat either in the 5' or 3' UTR. Columns represent mean values +/- SE of RLU normalized to control (Renilla luciferase vector without repeats), p-values are the result of an unpaired Student's t-test ($p^{****}<0.0001$), n = 3. (C) Western blot detecting MID1 or Actin to show knockdown efficiency, corresponding to Lucferase experiment in (D). (D) Luciferase measurement of Renilla-3'UTR-$(CAG)_{50}$ after knockdown of MID1 and/ or MID2. Corresponding legend is depicted to the right. Columns represent mean values +/- SE of RLU normalized to control (Renilla luciferase vector without repeats), p-values are the result of an unpaired Student's t-test ($p^{***}<0.0005$, $p^{****}<0.0001$), n = 4. (E) qPCR showing MID1 and MID2 knockdown efficiency on transcript level. p-values are the result of an unpaired Student's t-test ($p^{****}<0.0001$), n = 4. SE, standard error; RLU, relative light units (Renilla/ Firefly luciferase); n = number of experiments on separate days.

3.1.2 RNA secondary structure influences MID1-dependent translation

To evaluate MID1 specificity for trinucleotide repeats and the contribution of RNA secondary structure, Renilla-3'UTR-$(CAG)_{50}$ translation was compared to constructs containing a CAG repeat with regular CAA interruptions and a pure CAA repeat, respectively. Figure 5.B to D shows secondary structure predictions [158] of Renilla luciferase RNA with the different repeats in the 3' UTR (the repeat sequence is indicated by an arrow). CAA interruptions prevent the formation of a hairpin structure compared to a pure CAG repeat, but still the RNA folds into a structured molecule with multiple short CAG hairpins. Expectedly, pure CAA repeats do not form stable secondary structures. Generally, the existence of a repeat sequence in the 3' UTR enhanced translation in the reporter assay, however MID1 affected translation of structured RNAs only (Figure 5.A) [116].

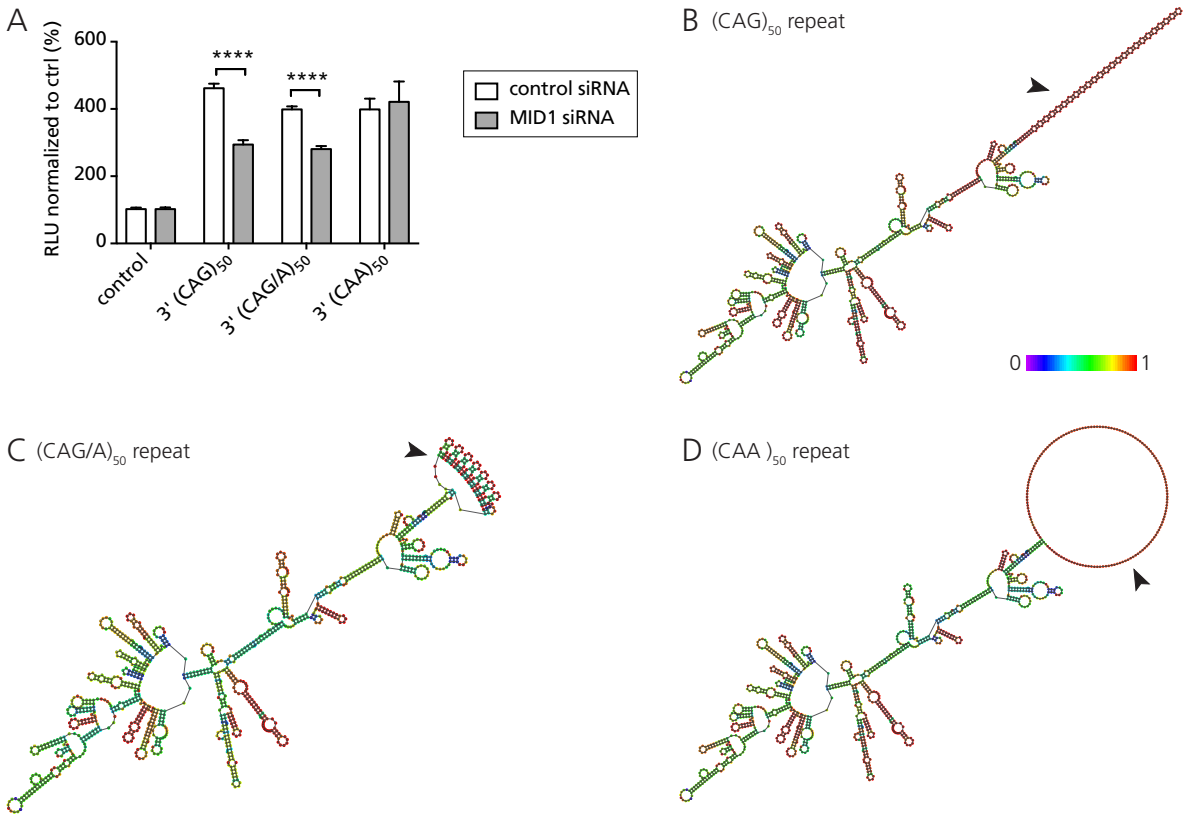

Figure 5 | MID1-dependent translation of different TNR RNAs.
(A) Luciferase measurements of Renilla luciferase fused to either $(CAG)_{50}$, $(CAG/A)_{50}$, or $(CAA)_{50}$ repeats in its 3' UTR. Columns represent mean values +/- SE of RLU normalized to control (Renilla luciferase vector without repeats). p-values are the result of an unpaired Student's t-test (p****<0.0001), n = 3. (B to D) Secondary structure predictions of Renilla luciferase RNA with indicated triplet repeats in its 3' UTR corresponding to (A). Arrowheads indicate TNR sequences and the structures are colored by base-pairing probabilities. For paired regions the color denotes the probability of being paired, while for unpaired regions the color denotes the probability of being unpaired. SE, standard error; RLU, relative light units (Renilla/ Firefly luciferase); n = number of experiments on separate days.

3.1.3 MID1 is part of the translational machinery

Since MID1 regulates translation of several RNAs in a genetic context but also generally structured RNAs with CAG repeats, the question remains how MID1 is directed to these specific RNAs. Other factors most likely play an important role in guiding MID1 and/ or vice versa. Therefore, understanding the MID1 complex and its protein interactions are fundamental to gain deeper insight into how MID1 regulates translation. To analyze the MID1 protein interactome quantitatively, FLAG-tagged MID1 was over-expressed in HEK-T cells and immunoprecipitated from the lysate using FLAG antibody-bound agarose beads (Figure 6.A). Immobilized MID1 complex and interacting proteins were extensively washed and eluted. To discriminate unspecific binding of proteins to the FLAG antibody agarose beads, untransfected lysate from HEK-T cells was used in a separate preparation. To obtain replicates, cells were seeded and transfected, on different days, so that triplicates of both immunoprecipitation (IP) and control could be analyzed. Mass spectrometric and statistical analyses were performed by A. Dagane (MDC, Berlin).

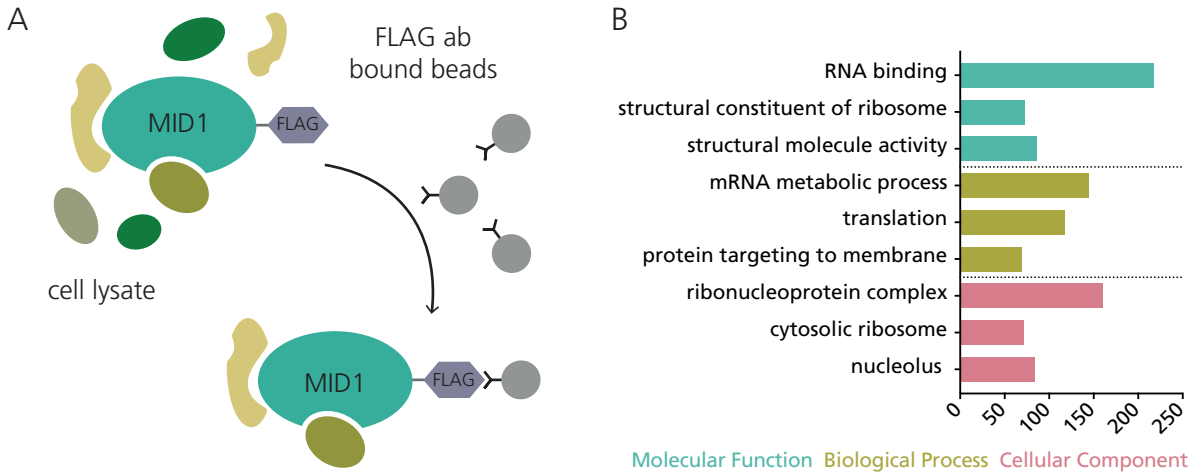

Figure 6 | MID1 binding partners are enriched in translation associated proteins.
(A) Immunoprecipitation scheme of over-expressed FLAG-tagged MID1. (B) GO analysis of 302 proteins identified as MID1 interaction partners showing the top three GO terms in each category.

302 MID1-specific interaction partners were identified (

Table 7). To put these proteins into context, a gene ontology analysis (GO) was conducted (Figure 6.B). This bioinformatic tool annotates genes with defined terms describing the gene products known function. GO terms are grouped into three classes, namely molecular function, biological process, and cellular compartment. GO terms of molecular function classify molecular activities, biological process terms define an event with several separate steps, and a cellular component term describes distinct structures within or outside the cell. Using a GO database, it is possible to find significant terms shared among a list of genes and thereby infer biological meaning.

Several known MID1 binding partners were identified: RACK1 (also known as GNB2L1), 40S ribosomal protein A (RPSA), 40S ribosomal protein 3 (RPS3), and 40S ribosomal protein 8 (RPS8) [162].
Two thirds of the MID1 interactome were RNA-binding proteins and more than 50 % of these proteins are involved in translation, supporting MID1's known function in translational regulation. MID1 is part of an RNP complex and MID1-dependent translation has been shown to act through the modulation of the PP2A-mTOR axis of translational control. However, its location in the context of the translational machinery has not been resolved. Interestingly, 73 proteins binding MID1 were annotated as a structural constituent of the ribosome, possibly placing MID1 in close proximity to the ribosome. Ingenuity pathway analysis, a functional analysis tool, clearly strengthend the importance of MID1 in translational control. The top three pathways identified were eIF2 signaling, regulation of eIF4 and p70S6K signaling, and mTOR signaling.
Of note, both cytosolic ribosomal proteins and proteins associated with the nucleolus were identified. Considering that MID1 is found in the cytoplasm and associates with microtubules [121,163], this observation is unexpected. However, all proteins are translated in the cytoplasm and since the IP was performed with crude cell lysate, the identification of nucleolar proteins could reflect unusual interactions with MID1. To take into account the cellular compartmentalization, the association with nucleolar proteins should be studied in detail in the intact cellular environment.

3.1.4 MID1 is located close to the ribosome

To validate the association of MID1 with the translational complex, the co-immunoprecipitation (coIP) of MID1 with different components was assessed by western blot. Figure 7 shows representative blots grouped according to distinct steps during translation.
Cleavage and polyadenylation specificity factor subunit 6 (CPSF6) and nudix hydrolase 21 (NUDT21) are polyA-binding proteins involved in processing of the 3' polyA of pre-mRNAs (Figure 7.A). Poly(A) binding protein 1 (PABP1) binds the 3' polyA tail and interacts with eukaryotic translation initiation factor 4 gamma 1 (eIF4G) to facilitate circularization of the mRNA, an important step during translation initiation. Besides eIF4G (Figure 7.B), other eukaryotic initiation factors (eIFs) were validated. eIF3A helps to assemble the 43S preinitiation complex and like eIF4A, eIF4B, and eIF4G recruits the preinitiation complex to the mRNA.
Interestingly, FMRP co-immunoprecipitates with the MID1 complex (Figure 7.C). As described above, this is the protein product of *FMR1*, where a TNR mutation in the 5' UTR leads to FXTAS or FXS, respectively. FMRP is involved in activity-dependent translation at the synapse [164,165] and directly binds to the ribosome [166]. RACK1, a known interaction partner, was detected as a positive control and the ribosomal proteins of the large subunit 60S acidic ribosomal protein P0 (RPLP0) and 60S ribosomal protein L5 (RPL5), as well as a part of the small subunit, namely 40S ribosomal protein S3 (RPS3),

were validated. Worth mentioning is the obvious higher enrichment of this group of proteins compared to the proteins depicted in Figure 7.A and B. In HEK-T, FMRP, RACK1, RPLP0, RPL5, and RPS3 are expressed at much lower levels than the polyA-binding proteins and eIFs. More than double the amount of input lysate compared to polyA-binding proteins and eIFs had to be loaded on western blot to visualize the ribosomal and ribosome-associated protein.

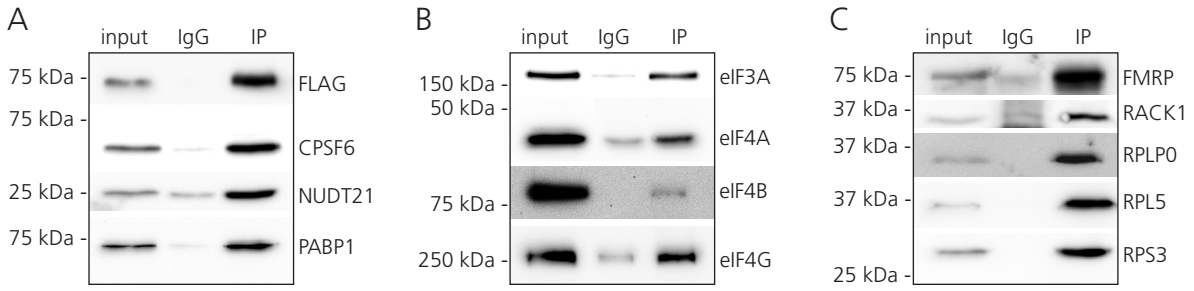

Figure 7 | Validation of MID1 IP Mass Spectrometry results.
MID1-FLAG was expressed in HEK-T cells and purified by IP. The presence of MID1-interacting proteins in precipitates (IP) was analyzed by western blot using specific antibodies. As negative control, unspecific IgG agarose beads were used (IgG). (A) Detection of the polyA-binding proteins CPSF6, NUDT21, and PABP1. (B) Detection of eIFs 3A, 4A, 4B, and 4G. (C) Detection of FMRP, RACK1, RPLP0, RPL5, and RPS3.

To analyze the MID1 function within this translation complex in more detail, coIPs were repeated in the presence of high concentrations of EDTA (40 mM). EDTA leads to the disassembly of the ribosomal subunits [167] thereby facilitating the differential mapping of MID1. Figure 8.A shows proteins that associate with MID1 depending on the integrity of the ribosome. To assure that EDTA treatment did not affect the efficiency of MID1 precipitation, the blots were incubated with FLAG antibody to detect MID1. In contrast to the polyA binding proteins and eIF3A, the binding of FMRP, RACK1, and RPLP0 to MID1 was stable even in the presence of EDTA (Figure 8.B). RPL5 and RPS3 binding was clearly reduced, arguing for an indirect interaction with MID1.

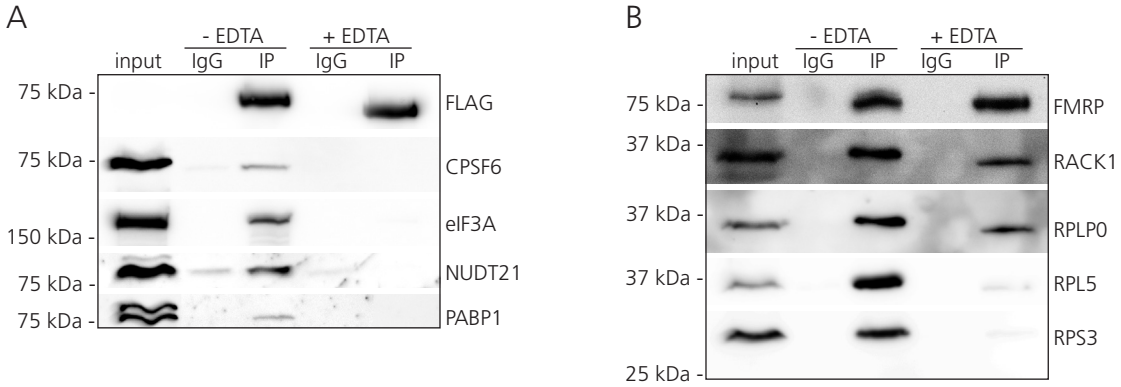

Figure 8 | Effect of ribosome disassembly on the composition of the MID1 complex.
MID1-FLAG was expressed in HEK-T cells and purified by IP either without EDTA or in the presence of high concentrations of EDTA. The presence of MID1-interacting proteins in precipitates (IP) was analyzed by western blot using specific antibodies. As negative control, unspecific IgG agarose beads were used (IgG). Detection of (A) CPSF6, eIF3A, NUDT21, PABP1, and (B) FMRP, RACK1, RPLP0, RPL5, and RPS3.

3.1.5 MID1 transcript and protein levels are upregulated in HD patient cortices

Reducing mutant HTT protein levels, for example by antisense oligonucleotides directed against the *HTT* transcript, is one possibly beneficial treatment of HD [168]. Doing so allele-specifically is difficult, however targeting the MID1 complex promises a solution to this problem [160] since it specifically modulates translation of the mutant transcript. Human data on MID1 expression exists but not in the context of HD. Therefore, human post mortem cortical sections from control subjects and HD patients were immunohistochemically (IHC) stained for the MID1 protein. MID1 positive cells were found throughout all cortical layers, however the intensity especially of the neuropil staining was much higher in HD patient samples compared to controls (Figure 9 and 10). Moreover the number of positively stained cells seems much lower in controls compared to HD patients.

Taking a closer look at the morphology of the cells, MID1 seems to be expressed in different cell types. In the cortical layers cells with long apical and basal processes were positively stained, indicating neuronal cells (Figure 11). The cresyl violet counter stain coloring Nissl substance in neurons can easily identify pyramidal neurons. The cell nucleus is relatively small while the cell body has a conical shape with one apical dendrite that extends vertically from the soma. This characteristic cell type of layers three and five in the motor cortex was clearly stained in an HD patient (Figure 11.F), however in other samples it could not be found but rather staining of smaller neuron-like cells. Additionally, cells with numerous highly branched fine processes can be identified in the cortical layers suggesting a glial cell type, reminiscent of astrocytes. MID1 staining in the white matter showed even greater differences between HD patients and controls: while in controls only few cells stained positive, the white matter of HD patients was densely populated. These cells have fewer processes compared to the glial cells in the cortical layers.

Figure 9 and 10.
IHC staining from C2 and HD4. Sections from the middle temporal gyrus were immunostained for MID1 and cresyl violet was used for Nissl substance staining. The overview image on the left shows all cortical layers from the meninges down to the white matter. Locations of the magnified images on the right are indicated with white boxes. Scale bar, 20 μm.

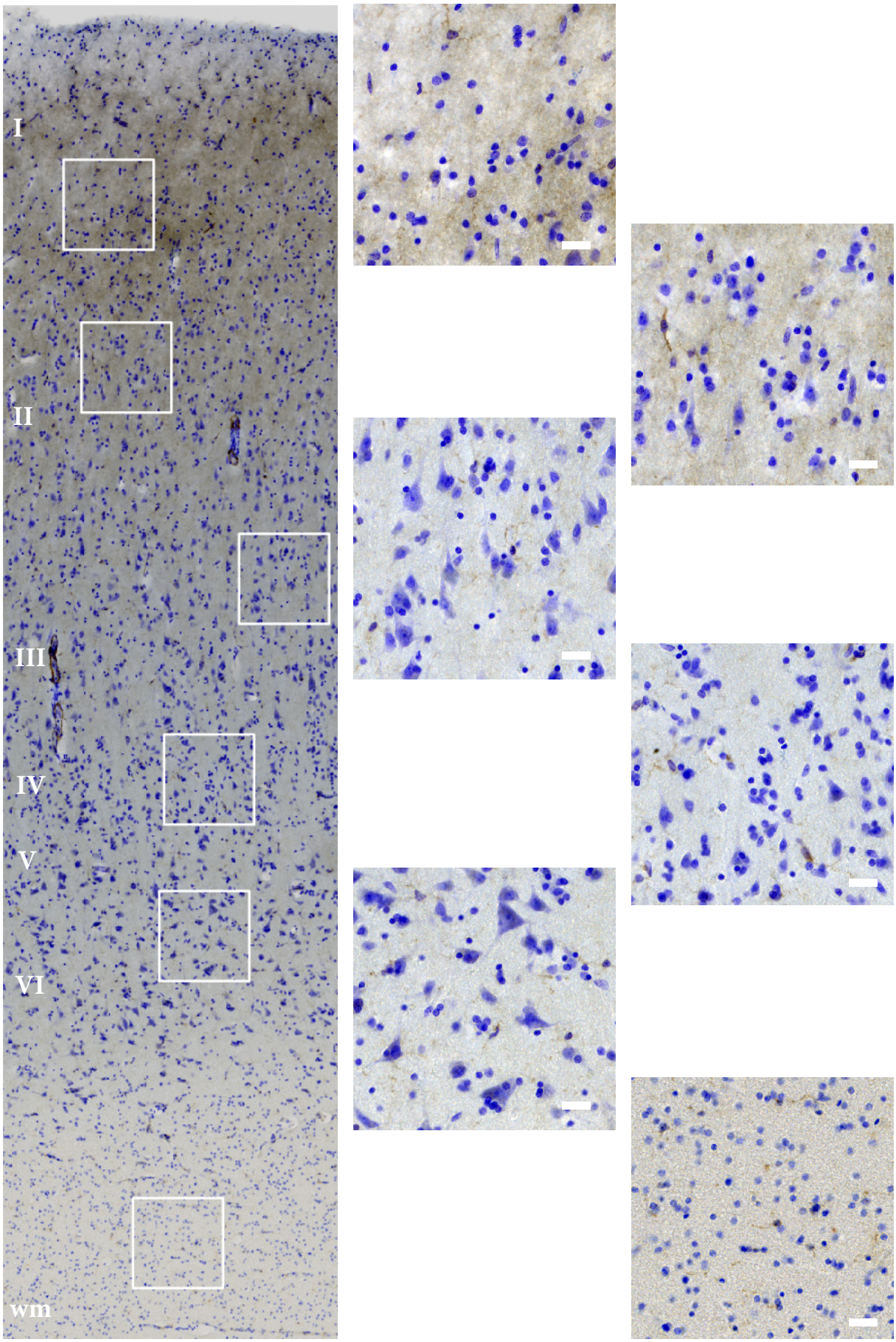

Figure 9 | MID1 IHC staining of human cortical layers and white matter of a control subject.

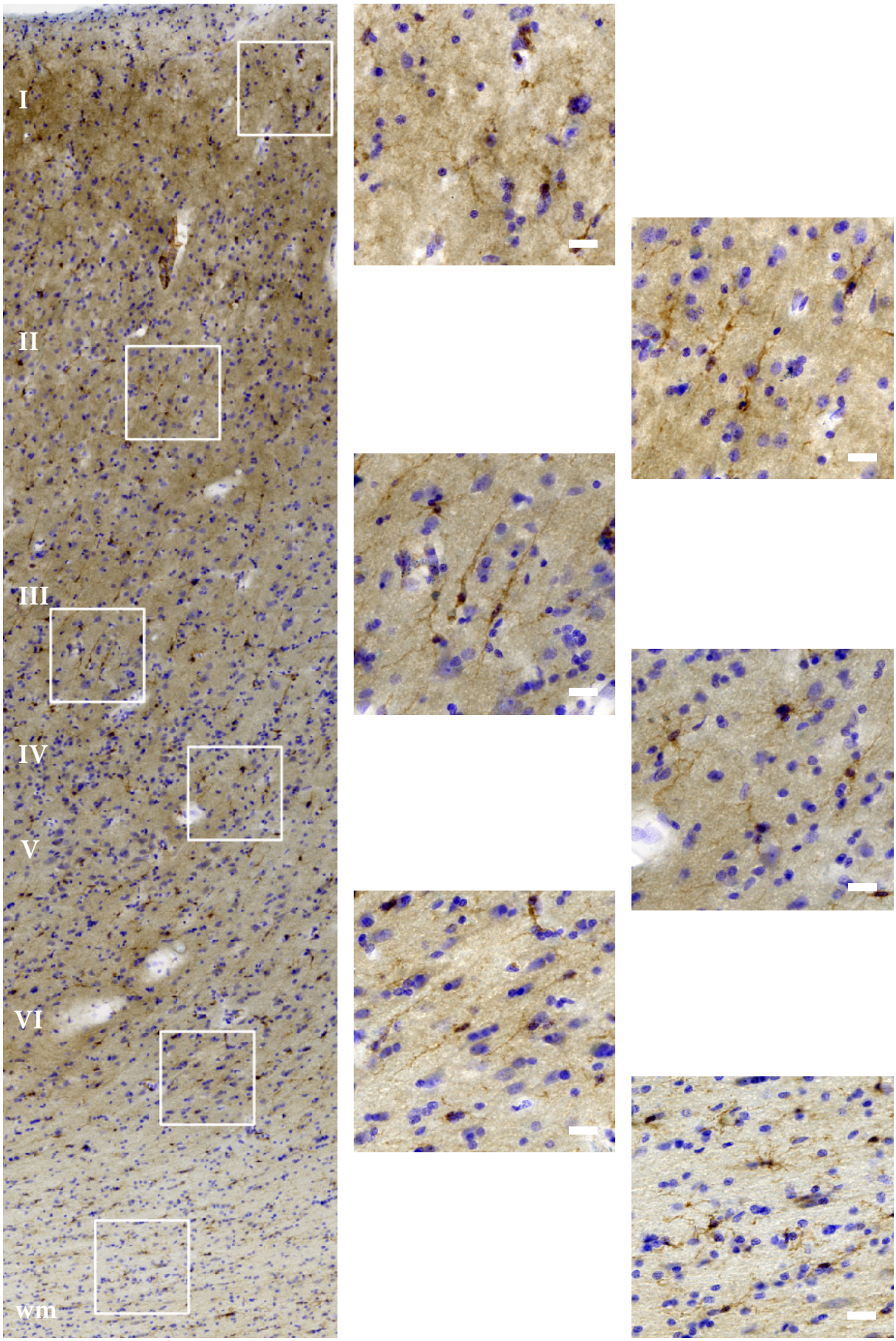

Figure 10 | MID1 IHC stainings of cortical layers and white matter of an HD patient.

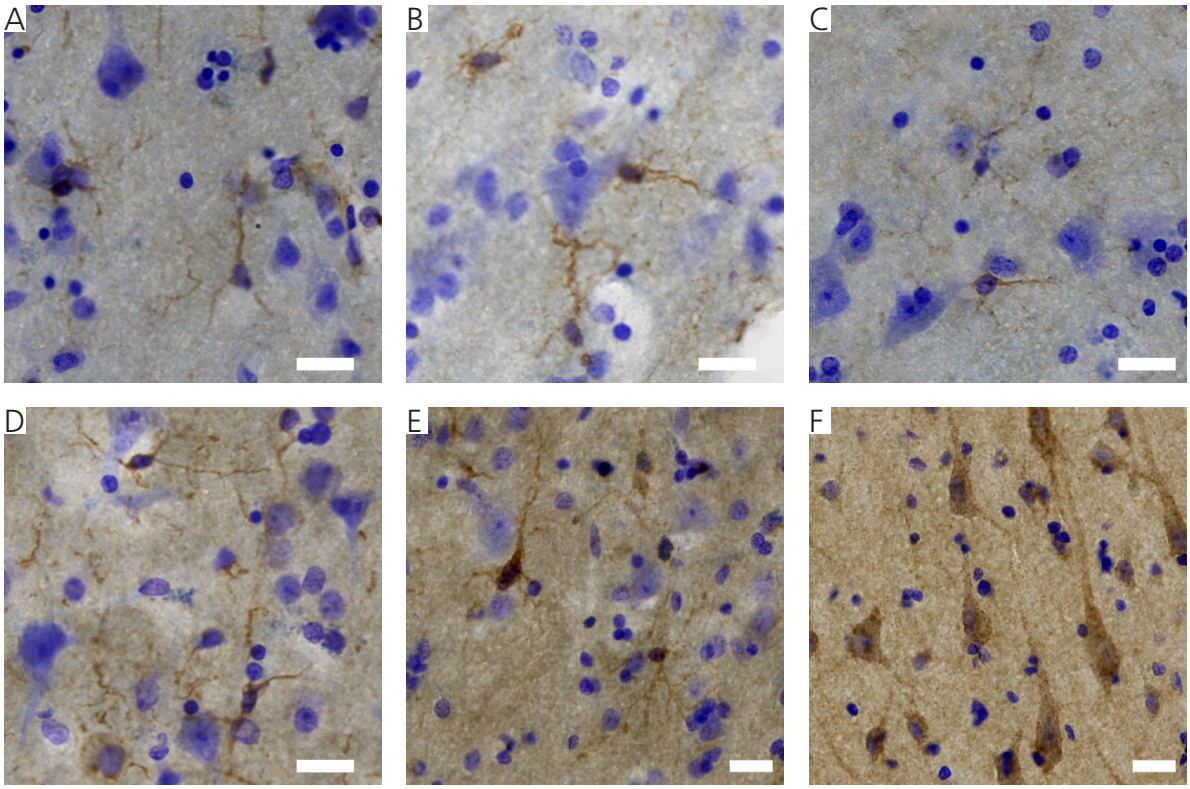

Figure 11 | Examples of MID1 IHC stainings in human cortical layers.
MID1 stainings of (A-C) controls subjects (C6, C3, and C2) and (D-F) HD patients (HD3, HD4, HD6). Scale bar, 20 µm.

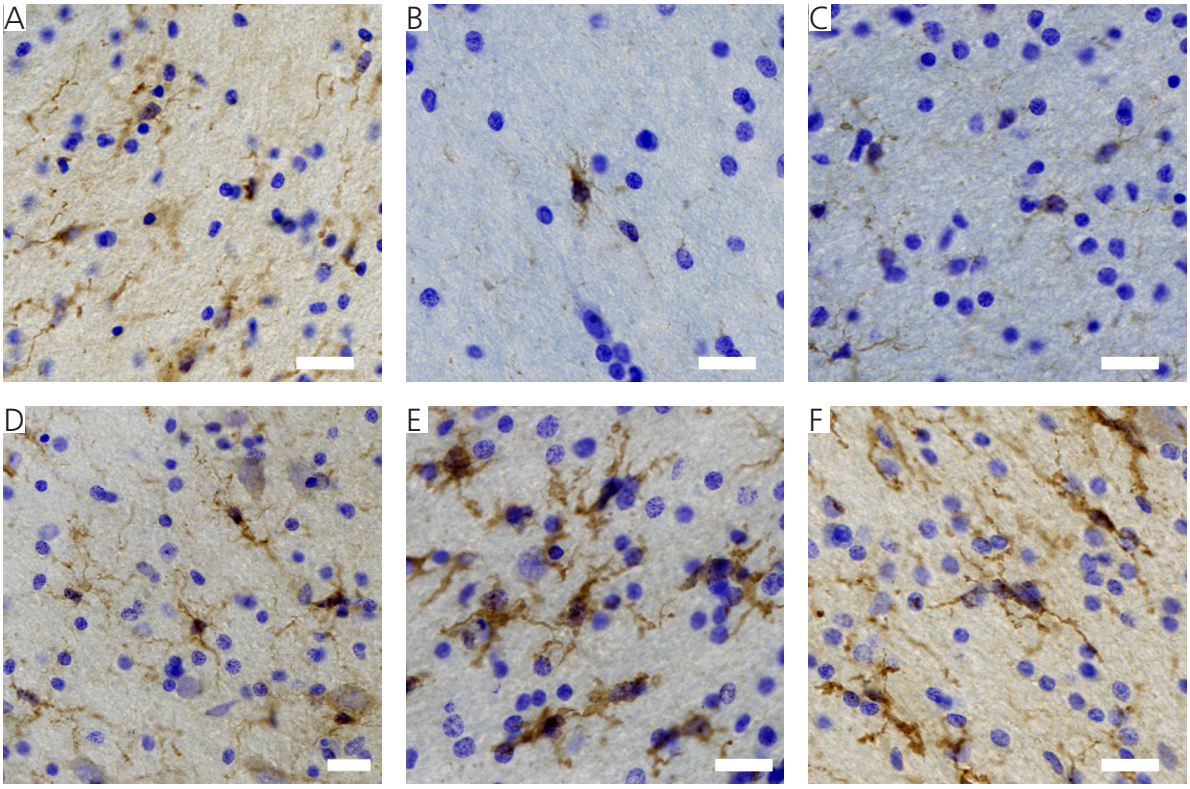

Figure 12 | Examples of MID1 IHC stainings in human cortical white matter.
MID1 stainings of (A-C) controls subjects (C2, C1, and C3) and (D-F) HD patients (HD4, HD5, HD1). Scale bar, 20 µm.

Since the neuropil staining of the cortical layers was very intense, quantification of positive cells was only possible in the white matter. Automated identification of positive cells showed significantly more MID1 positive cells in HD patients' brains compared to controls (Figure 13.A). To assess whether this upregulation of *MID1* is also seen on transcript level, qPCR analysis of unfixed post mortem material from HD patients and controls was conducted. Figure 13.B shows *MID1* mRNA levels normalized to the housekeeping gene ribosomal protein L22 (*RPL22*). *MID1* expression is significantly higher ($p < 0.001$) in the cortex of HD patients (n = 8) compared to control tissue (n = 6). Moreover, striatal and cerebellar tissue was analyzed for *MID1* expression. The cerebellum showed a similar trend as the cortex, while *MID1* expression in the striatum seems to be weaker in HD patients. It was refrained from doing statistical analysis because sample size was too small (striatum control n = 2, other tissues n = 3). Overall, these studies show that (i) MID1 is expressed in different cell types in the cortex, (ii) the number of positive cells in the cortical white matter is significantly upregulated in HD patient tissue, and (iii) also *MID1* mRNA levels are significantly upregulated in the cortex of HD patients.

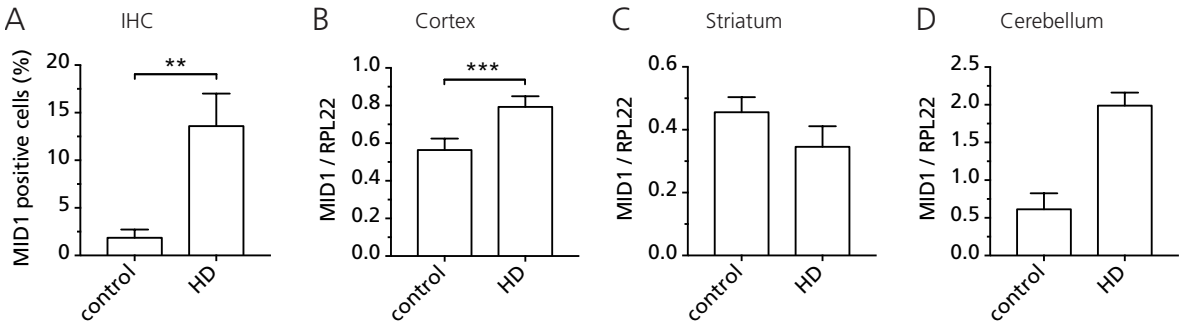

Figure 13 | Quantification of MID1 expression in brain regions of HD patients and controls.
(A) Quantification of MID1 positive cells in cortical white matter of HD patients and controls. n = 6, $p<0.005$, result of an unpaired t-test. (B-D) qPCR analysis of *MID1* expression normalized to *RPL22* in indicated brain areas of HD patients and controls. (B) $n_{control} = 6$, $n_{HD} = 8$, $p^{***}<0.001$, (C and D) $n_{control} = 2$, $n_{HD} = 3$. Columns represent mean values +/- SE, p-values are the result of a factorial ANOVA determining the genotype effect and correcting for *RPL22* expression and a confounding effect of the qPCR experiments. SE, standard error.

3.1.6 MID1 is expressed in the murine brain age- and genotype-dependently

Apart from the spatial component of gene expression, mouse models of HD permit the temporal analysis of MID1 expression. Therefore, we used the Hdh(CAG)150 (Q150) mouse model [155] to study *Mid1* expression in different brain areas and at a different age. This transgenic mouse model carries a CAG repeat of 150 units in the endogenous *Htt* gene.

Mice were sacrificed at either two months of age ("young") or between 11 to 14 months ("old"), the brain was dissected and the cortex, hippocampus, striatum, and cerebellum were analyzed by qPCR (Figure 14). *Mid1* expression depends on both age and genotype, but to different degrees in the

respective areas. In the young cortex (p-value < 0.05), striatum, and cerebellum (p-values < 0.01) Mid1 expression was significantly upregulated in transgenic mice compared to wild-type mice. In all transgenic brain tissues a downregulation of *Mid1* expression with age could be observed, with highest significance in the striatum (p-value < 0.001). In the wild-type mice, an age-dependent effect of *Mid1* expression could only be seen in striatal tissue (p-value < 0.01). *Mid1* expression in the hippocampus showed a similar trend but did not reach statistical significance. In contrast to the results from human tissue, there was no genotype-dependent effect in old animals. Overall, this shows that in young mice *Mid1* expression is higher in the Q150 animals, while a clear age-dependent effect is only seen in the HD mouse model.

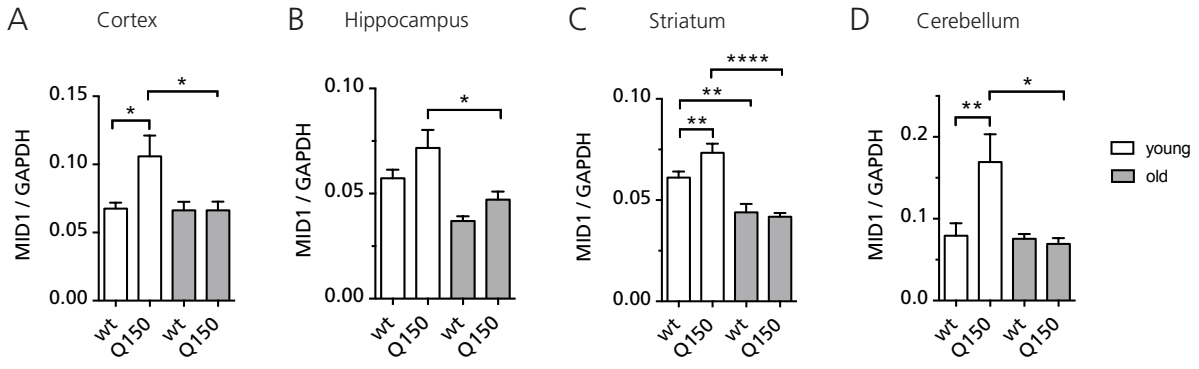

Figure 14 | Mid1 expression analysis in the HD Q150 mouse model.
Mid1 qPCR analysis of indicated tissues from wild-type (wt) and Q150 transgenic animals. MID1 levels are normalized to GAPDH. Young animals were two months old (n_{wt}=10, n_{tg}=9) and old animals 11 to 14 months (n_{wt}=7, n_{tg}= 6). Columns represent mean values +/- SE, p-values are the result of a one-way ANOVA with multiple comparisons. SE, standard error.

3.2 *HTT* RNA binding partners

In HD, the MID1 protein is one example for a protein binding partner of mutant *HTT* RNA that leads to abnormal processes. Other such proteins exist but mostly have been studied in the context of RNA foci [84]. The second part of this work aims to map mutant *HTT* RNA interactions in an unbiased approach.

3.2.1 Mutant *HTT* exon 1 RNA-binding proteins are enriched in splicing factors

To investigate a possible gain-of-function of mutant *HTT* exon 1 RNA with respect to its protein binding partners, an RNA pulldown using *in vitro*-transcribed RNA followed by quantitative mass spectrometry was performed (Figure 15.A). *HTT* exon 1 RNA containing 18, 40, or 70 CAG repeats, respectively, was amplified from vectors using primers incorporating a T7 sequence upstream of the open reading frame (ORF) and a linker sequence that allowed annealing of the RNA to a biotinylated DNA oligonucleotide. Via this binding the RNA was captured on streptavidin-coated magnetic beads. These RNA-coated magnetic beads were incubated with lysate from a human neuroblastoma cell line (SHSY-5Y), immobilized proteins were extensively washed to remove unbound proteins and finally, RNA-bound proteins were eluted. Five replicates with different passages of cells were prepared and each replicate consisted of four experiments: one control pulldown and three RNA pulldowns with *HTT* exon 1 and different CAG repeat lengths. Mass spectrometric and statistical analyses were performed by A. Dagane (MDC, Berlin).

Altogether, 1050 proteins were identified. For statistical analysis only proteins that (i) were identified in at least two out of five RNA pulldown replicates but not in controls and (ii) proteins bound to RNA with a mean intensity at least twice as high compared to controls were included. The intensity denotes all peak intensities from an eXtracted Ion Current (XIC) chromatogram of all isotopic clusters associated with the identified amino acid sequence. In this manner, 308 proteins were found to specifically bind *HTT* exon 1 RNA. Five published proteins that target *HTT* RNA, namely Dicer, SRSF6, Nucleolin, PKR and FMRP, were identified by mass spectrometry, underlining the strength of our approach [108,135,169]. Furthermore, FMRP, pre-mRNA-processing-splicing factor 8 (PRPF8), splicing factor 3B subunit 2 (SF3B2), U5 small nuclear ribonucleoprotein 40 kDa protein (SNRNP40), RPLP0, and protein SON (SON) were validated as *HTT* exon 1 protein binding partners on western blot (Figure 15.B). Interestingly, Paraspeckle component 1 (PSPC1), a nucleolar protein that can be found in paraspeckles (structures close to splicing speckles that are involved in gene expression) was also validated on western blot.

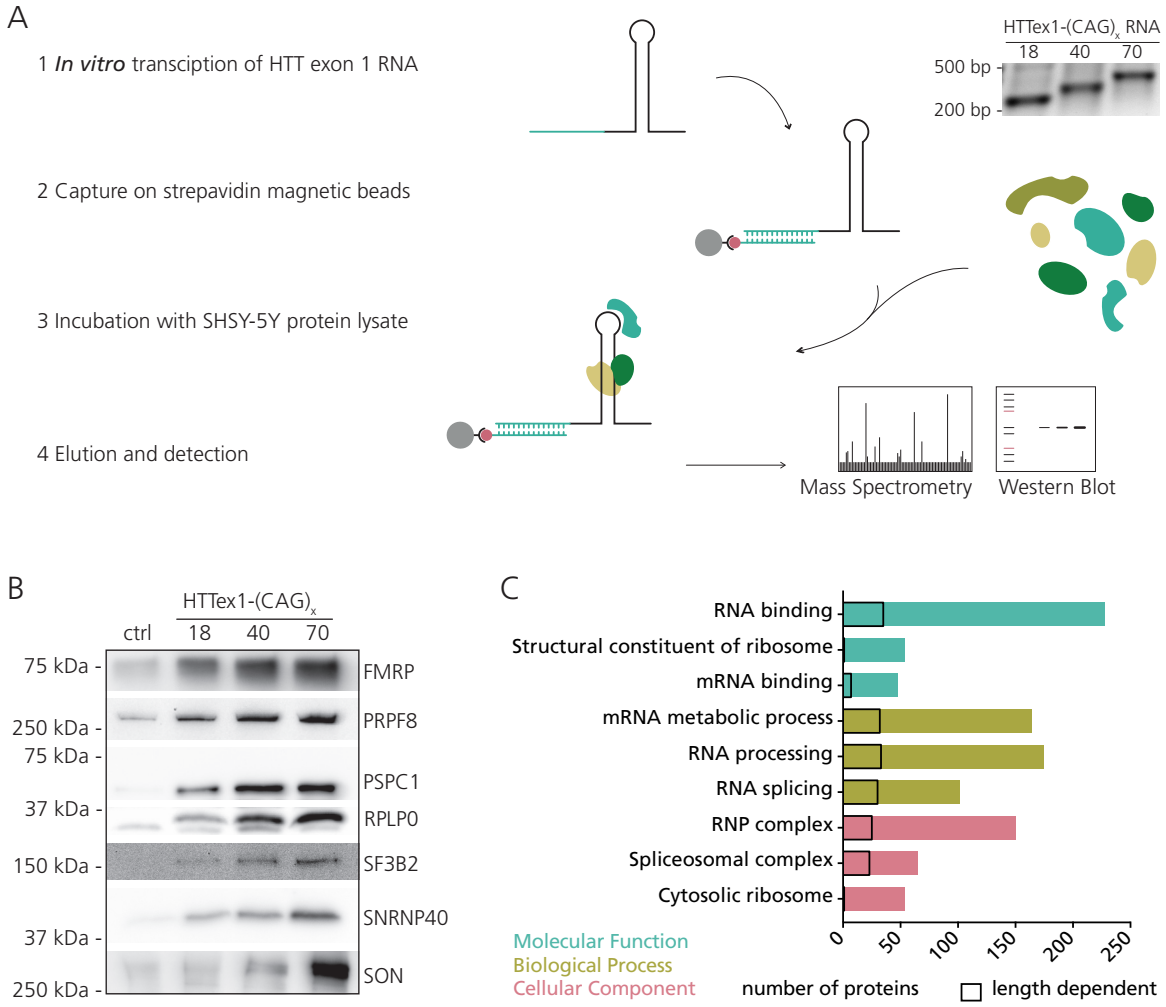

Figure 15 | Splicing factors bind HTT RNA CAG repeat length-dependently.
(A) RNA pulldown scheme depicting the experimental workflow. After *in vitro* transcription, HTT RNA was immobilized on magnetic beads through the interaction with a biotinylated DNA linker. A representative agarose gel picture of *HTT* RNA with different CAG repeat lengths is shown in the upper right corner. Next, RNA was incubated with SHSY-5Y cell lysate, and RNA bound proteins were analyzed by mass spectrometry and validated on western blot using specific antibodies. (B) Western blot validation of selected proteins in an RNA pulldown assay. (C) GO analysis of the 308 proteins binding to *HTT* exon 1 RNA. Black boxes indicate the number of proteins within this GO term that bind CAG length-dependently.

To identify proteins that bind in a CAG repeat-length dependent manner, an additional statistical analysis was conducted by I. Atanassov (Max Planck Institute for Biology of Aging, Cologne). Briefly, only proteins that were identified in at least four replicates in at least one experiment were used for analysis and missing values were added by Perseus imputation. Next, the RNA pulldown experiments were compared to the control: proteins with a positive log2 fold change (sample/ control) and a p-value < 0.01 were defined to specifically binding *HTT* exon 1. By doing a pairwise comparison, proteins that bind in a CAG length-dependent manner could be identified. This analysis identified 36 proteins that preferentially bind *HTT* exon 1 RNA with an expanded CAG repeat tract (*HTT*ex1(CAG)$_{ex}$, Table 8). To analyze the network of these proteins and infer biological meaning, a

GO term analysis was conducted. Not unexpectedly, many ribosomal proteins bound to the RNA and are involved in RNA processing (Figure 15.C). Interestingly, the majority of proteins specifically binding *HTT*ex1(CAG)$_{ex}$ can be mapped to GO terms connected to splicing in contrast to ribosome-associated terms (Figure 15.C, black boxes). This CAG length-dependent interaction was also seen on western blot. Altogether, these experiments validate published *HTT* RNA binding partners and identify new proteins specifically recruited to *HTT*ex1(CAG)$_{ex}$ RNA, implicating them as possible disease modifiers.

3.2.2 Conditional expression of *HTT* exon 1-(CAG)$_{68}$ leads to retention of introns with weak 5′ splice sites

Considering the evidence for mis-splicing in HD [170,171] and the aberrant binding of splicing factors to mutant *HTT* exon 1 RNA, the analysis of splicing changes in an HD model is rational. Therefore, an inducible model of SHSY-5Y cells expressing *HTT* exon 1 C-terminally tagged with enhanced green fluorescent protein (EGFP, provided by Y. Dürnberger, DZNE Bonn) was used for transcriptome profiling. Tetracycline-controlled transcriptional activation, termed TetON, is a method to reversibly induce transcription of a gene of interest by the addition of the antibiotic tetracycline or a derivative, for example doxycycline. *HTT* exon 1 is under the control of the tetracycline responsive element (TRE) promoter, which in turn is activated by tetracycline transactivator protein (tTA). However, tTA is only active in the presence of tetracycline (or doxycycline), linking addition of the antibiotic to the cell culture medium to transcriptional induction of *HTT* exon 1. Transcriptome profiling uses DNA microarray technology with probe sets covering the entire transcribed genome. This effectively measures the expression levels of all known coding and non-coding splice variants, because probes are designed to capture exonic sequences as well as sequences at intron-exon junctions.

RNA from induced SHSY-5Y cells expressing *HTT* exon 1 and non-induced cells as control was used to generate labeled cDNA for microarray hybridization (performed by J. Winter, Institute of Human Genetics, Mainz). Subsequent bioinformatic analysis identified 103 alternative splicing events upon *HTT* exon 1 expression events that could be mapped to defined splicing categories (Figure 16.B), with the highest average splicing score reached by retained introns. Since a GO analysis of this short list was inconclusive, a simple literature search was performed to put these genes into perspective. cAMP response element-binding protein (CREB1) has an integral role in memory formation and neuronal plasticity and is implicated in the pathology of HD [172,173]. Thus, the *CREB1* intron retention event was chosen for further validation experiments. To this end, primers lying within the adjoining exons were designed to detect the successful splicing event and an alternative reverse primer complementary to the intron sequence to amplify from transcripts containing the retained intron.

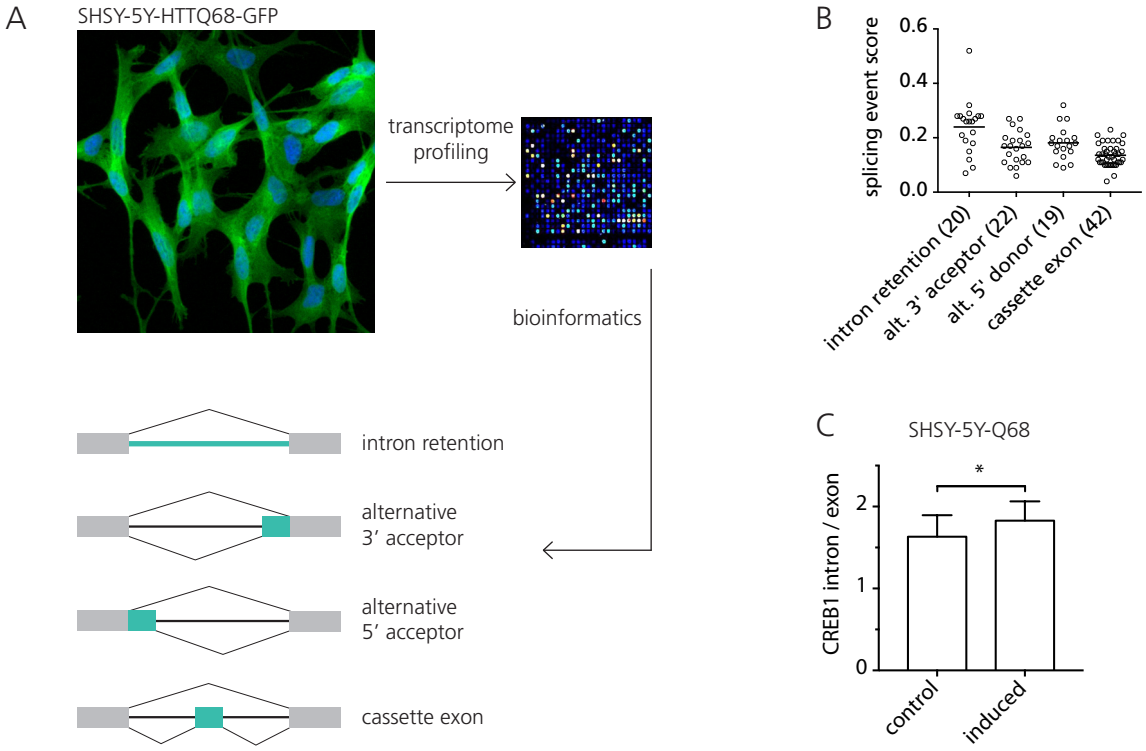

Figure 16 | Splicing events detected by transcriptome profiling and target validation.
(A) Experimental set-up and schematic depiction of splicing events. RNA from SHSY-5Y-HTTQ68-EGFP cells was used to detect alternative splicing events upon HTTQ68 expression. Microarray with complete transcriptome probes were used. (B) Results of the bioinformatic analysis quantifying splicing events. Amounts of different splicing events are plotted against their splicing event score. (C) qPCR validation of the *CREB1* intron retention event in SHSY-5Y-Q58-EGFP cells. Columns represent mean values +/- SE, p*<0.05, n = 7 replicate experiments using different passages of cells. p-value is the result of a factorial ANOVA determining the effect of HTTQ68 expression and correcting for *CREB1* exon expression and a confounding effect of the qPCR experiments. SE, standard error.

CREB1 intron retention upon *HTT* exon 1 expression in SHSY-5Y-HTTQ68-GFP cells is significantly upregulated detected by qPCR ($p < 0.05$, Figure 16.C).

Considering the identification of a specific splicing event and association of various splicing factors with *HTT* exon 1 RNA, it is feasible to examine the role of the validated splicing factors within the complex process of splicing. This in turn revealed that PRPF8, an integral part of the U5 small nuclear ribonucleoprotein complexes (snRNPs), interacts with the 5' splice site [174] and specifically mediates the splicing of weak 5' splice sites [175]. To investigate whether this is also true for transcripts differentially spliced depending on *HTT* exon 1 expression, bioinformatic analysis using the MaxEntScan algorithm was performed by J. Winter. Indeed, the 5' splice sites in retained introns were significantly weaker compared to non-retained introns (Figure 17.A) and a prediction of 5' splice site sequence motifs was in line with reported motif enrichments (Figure 17.B and C) [175].

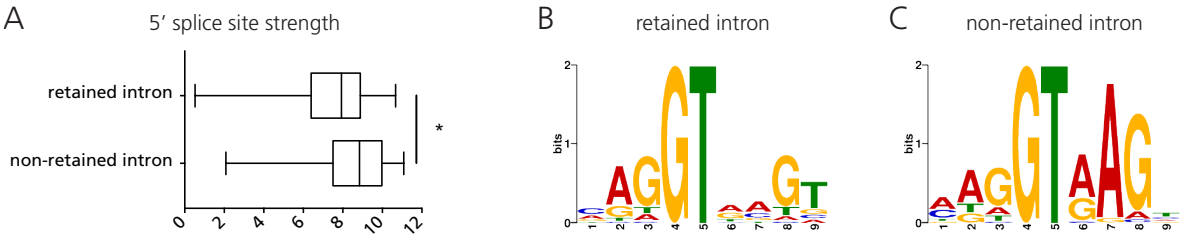

Figure 17 | 5' splice site characterization of intron retention events in an HD cell model.
(A) 5' splice site strength of retained introns is significantly weaker compared to non-retained introns. (B and C) Motif predictions of 5' splice sites for indicated introns.

3.2.3 *CREB1* intron retention is PRPF8-dependent in a cellular model of HD and upregulated in human HD cortex

To analyze if the *CREB1* intron retention event is mediated by PRPF8, a siRNA-mediated knockdown of PRPF8 in non-induced SHSY-5Y-HTTQ58-GFP was performed. The intron retention event was measured by qPCR as described above. Figure 18.A shows that, indeed, this splicing event depends on PRPF8 since the amount of unprocessed transcript increased upon PRPF8 knockdown. Knockdown efficiency was validated on protein level (Figure 18.B). Next, *CREB1* expression and the intron retention event were analyzed in human cortical tissue of HD patients and control subjects. Both expression of normally processed *CREB1* transcript as well as the intron retained transcript was upregulated in HD patient material (Figure 18.C and D).

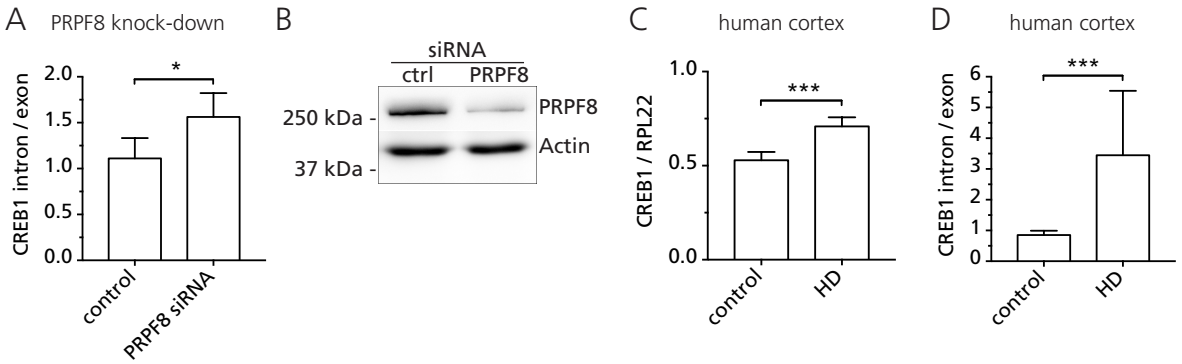

Figure 18 | *CREB1* intron retention is PRPF8 dependent and upregulated in HD patient cortex.
(A) *CREB1* intron retention event measured by qPCR after PRPF8 knockdown in uninduced SHSY-5Y-HTTQ68-EGFP cells, p*<0.05, n = 4 replicate experiments using different passages of cells. (B) Western blot of cell lysates from (A) verifying PRPF8 protein knockdown. qPCR analysis of (C) *CREB1* expression in cortical tissue of control subjects and HD patients and (D) *CREB1* intron retention event in cortical tissue of control subjects and HD patients, p***<0.001. Columns represent mean values +/- SE, p-values are the result of a factorial ANOVA determining the knockdown or genotype effect and correcting for *CREB1* exon or *RPL22* expression and a confounding effect of the qPCR experiments, $n_{control} = 6$, $n_{HD} = 8$. SE, standard error.

4 Discussion

HD is caused by a CAG repeat expansion mutation translating into a neurotoxic polyQ protein. In addition to polyQ-mediated pathogenesis, the mutant CAG repeat RNA elicits toxic mechanisms as well. One such mechanism underlying the toxic gain-of-function of mutant *HTT* RNA is the aberrant recruitment of proteins. The aim of this thesis was to identify and characterize these RNA-protein interactions. First, a previously identified protein, MID1, that binds to and regulates *HTT* RNA in a CAG length-dependent manner, was investigated. The results show a MID1-dependency on RNA secondary structure and a, so far unknown, binding of MID1 to the ribosome. Second, an unbiased approach to systematically map mutant *HTT* exon 1 RNA interactions with proteins demonstrates that deregulated splicing is a major mechanism underlying RNA-mediated toxicity in HD.

4.1 Mechanism of MID1-dependent translation

The MID1 protein is part of an RNP complex that regulates translation of specific RNAs. In the context of HD, MID1 binds *HTT* mRNA in a CAG repeat length-dependent manner, up-regulating translation of the mutant allele. This implicates MID1 as a modifier of HD pathology. Specifically, MID1 indirectly modulates translation by targeting PP2Ac for proteasomal degradation [124]. PP2A negatively affects the assembly of mTORC1, placing MID1 upstream of mTOR signaling and identifying it as a positive regulator of mTORC1 [125]. PP2A and its target S6K also localize to *HTT*ex1$(CAG)_{ex}$ RNA, however, how MID1 exactly encounters its substrates is unclear. Therefore, we characterized the MID1 complex by quantitative mass spectrometry to identify protein interaction partners that could be crucial in this process. The identified protein interactome of MID1 supports its involvement in mTOR signaling and additionally suggests eIF2 signaling and regulation of eIF4 and p70S6K signaling. Here, I concentrate on mTOR signaling since these pathways converge on the same effector molecules.

mTORC1 is a kinase complex comprising mTOR, raptor and LST8 that is directly involved in the control of translation initiation. Translational initiation is a rate-limiting step in protein synthesis and therefore has the potential for control. During cap-dependent translation initiation, the assembly of the elongation-competent 80S ribosome depends on two processes: on the one hand, the formation of the 43S preinitiation complex (PIC), and on the other hand mRNA activation [176]. The 43S PIC consists of the eIF2 ternary complex, the 40S ribosome, eIF3, and eIF5. mRNA activation includes the sequential assembly of several eIFs on the 5' cap structure of mRNA. In detail, eIF4E binds the 5' cap,

the two DEAD-box RNA helicases eIF4A and eIF4G start unwinding the mRNA's 5' secondary structure and serve as a scaffold for eIF4B, eIF3, and polyA-binding protein (PABP). eIF4B promotes eIF4A's helicase activity, while PABP facilitates mRNA circularization by interacting with eIF4G. This specific tertiary structure of mRNA and interacting proteins is thought to improve translation efficiency by facilitating the recycling of 40S ribosomes. Finally, eIF3 recruits the 43S PIC, 5' to 3' scanning begins, eventually leading to initiation codon recognition and binding of the 60S ribosomal subunit. In this context, eIF3 is a dynamic scaffold for the interaction of mTOR/raptor with its targets S6K and 4E-binding protein (4E-BP) [177]. Activated mTOR/raptor is recruited to eIF3 that is bound by inactivated S6K [177]. Upon mTOR-mediated phosphorylation, S6K dissociates and phosphorylates eIF4B and S6. Phosphorylated eIF4B now enhances eIF4A helicase activity [177], promoting translation initiation. The mTOR/eIF3 complex is now in close proximity to the 5' cap complex, which is bound by hypophosphorylated 4E-BP that prevents the interaction of eIF4G and eIF4A to eIF4E, thereby inhibiting translation [178]. mTOR-mediated phosphorylation of 4E-BP leads to its dissociation, permitting translation initiation to continue.

In this study, all 13 subunits of the eIF3 complex were identified by mass spectrometry as binding partners of MID1. Their involvement places MID1 close to the 43S PIC. In line with this, RACK1, a known binding partner of MID1 and a scaffolding protein that can be mapped close to eIF3 and the 40S ribosomal subunit in the 43S PIC [179], was found to be part of the MID1 complex. Considering the above-mentioned dynamic model of mTOR translation initiation control, MID1 actively promotes the assembly of mTORC1 and its subsequent recruitment to the eIF3 complex by targeting its opposing phosphatase PP2A for degradation (Figure 19). The fact that PP2A was not identified in this study underlines the dynamic nature of the interaction with MID1. To elucidate the exact steps of translation initiation that MID1 is crucial for, it is feasible to map MID1 in polysome gradients under conditions that affect different steps of this dynamic process.

Considering MID1's role in *HTT*ex1$(CAG)_{ex}$ RNA translation, the effects on S6K phosphorylation are likely to be crucial for TNR RNA translation. Phospho-S6K targets eIF4B that in turn increases the helicase activity of eIF4A, which is especially important for the unwinding of structured RNAs. Here we show that both eIF4B and eIF4A are part of the MID1 complex. Usually, RNA structures controlling translation initiation are located in the 5' UTR und significantly suppress their translation efficiency. Interestingly, this study shows that a pure $(CAG)_{50}$ repeat in the 5' UTR is not capable of inhibiting the translation of a reporter, however if it is located in the 3' UTR, translation is increased.

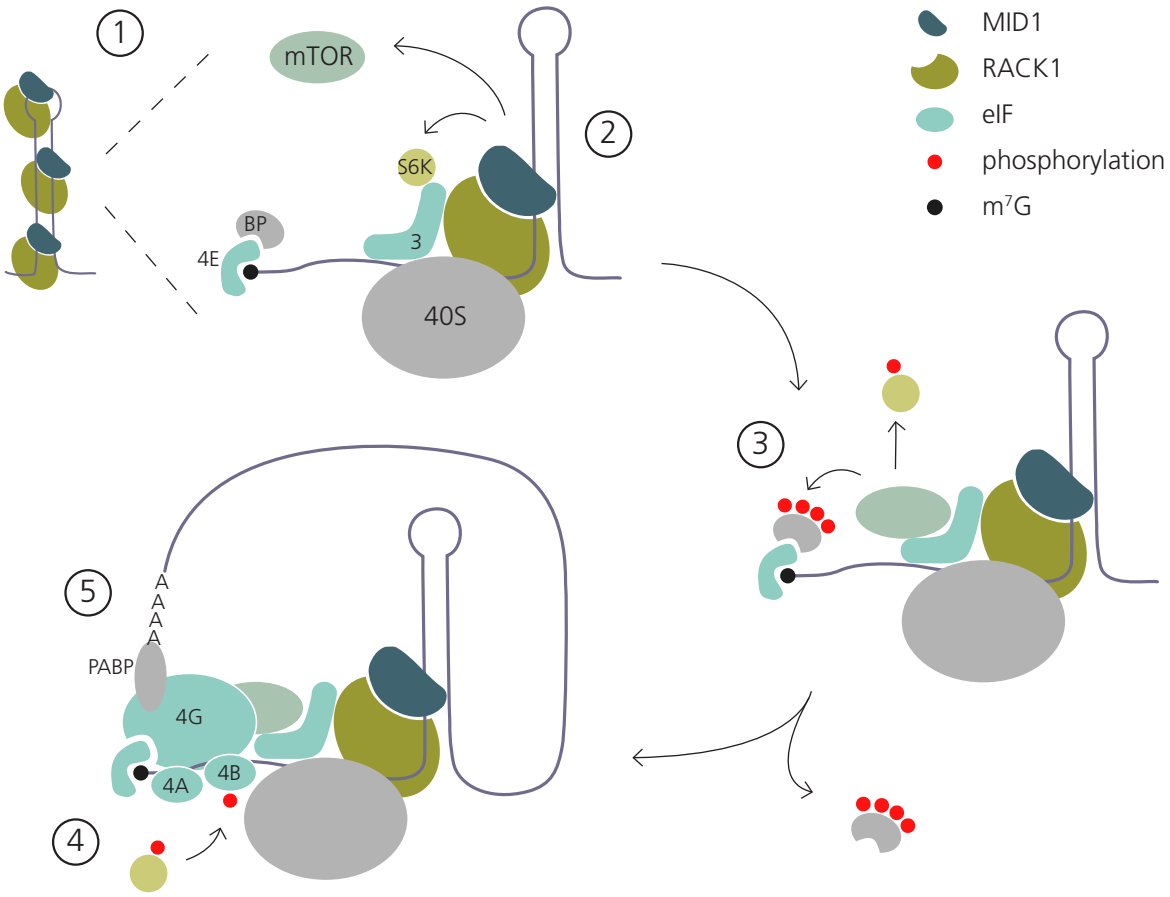

Figure 19 | Model mechanism of MID1-dependent translation.
(1) MID1 and RACK1 (and probably additional binding partners) define the repeat RNA-specific RNP that excessively binds to structured RNAs. (2) RACK1 serves as a scaffold for eIF3 and the 40S ribosomal subunit and possibly other factors of the 43S PIC. MID1 signaling leads to increased activation and thereby recruitment of mTOR. (3) Activated mTOR binds to eIF3 and phosphorylates S6K that dissociates from eIF3. Now mTOR is in close proximity to its other target 4E-BP and phosphorylates it. This leads to the dissociation of 4E-BP, canceling its inhibitory effect on eIF4G and eIF4A binding to eIF3E. (4) Activated S6K phosphorylates eIF4B, enhancing its helicase activity that unwinds the structured RNA. (5) eIF4G is recruited to eIF4E and, together with PABP, facilitates circularization of the transcript.

This phenomenon is dependent on MID1, similarly to the induction of *HTT*ex1(CAG)$_{ex}$ RNA translation. This, together with experiments employing differently structured and unstructured TNR repeats in the 3' UTR, shows that MID1 influences the translation of RNAs with structured 3' UTRs.

To see if this effect really depends on eIF4A activity, the effect of a MID1 knockdown on the eIF4A phosphorylation status should give some insight.

To test the dependency of MID1-protein interactions on ribosome integrity, the MID1 IPs were repeated with high EDTA concentrations. Of note, the disassembly of the ribosome during MID1 IP had a negative effect on eIF3A and PABP1 binding, while RACK1 interaction was preserved. This argues for a direct binding to MID1 although additional experiments are necessary to prove this. Several approaches are possible. First, a yeast two-hybrid assay is a protein-fragment complementation

assay that identifies direct protein-protein interaction. Second, IPs could be repeated including UV cross-linking of proteins and increasing the stringency of the washing steps. This has several advantages. Only proteins that interact directly are covalently cross-linked if photo-reactive amino acid analogs are taken up by cells and incorporated into nascent proteins [180]. Moreover, this can be performed in living cells and the modified proteins can be detected by western blot. This helps to resolve protein-protein interactions in the cellular space and time.

RACK1 and FMRP binding to MID1 were maintained independently of ribosome disassembly. Interestingly, RACK1 controls internal ribosomal entry site (IRES)-mediated translation of viruses [181], i.e. RACK1 seems to be involved in translation of mRNAs with specific structures. Similarly, FMRP is known to preferentially bind its targets throughout their ORF as well as along the 3' UTR of a subset of target mRNAs [182]. Whether MID1 directly binds mRNA is unknown. Alternatively, the interaction with specific mRNAs may be mediated by interacting proteins like RACK1 and FMRP. In this model, MID1 directly interacts with RACK1, FMRP, and RPLP0, while the association with eIF3 and S6K1 is RNA-dependent. Through the increased binding to structured RNAs, this complex excessively recruits mTORC1. High-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation (HITS-CLIP) identified HTT mRNA as a FMRP target [183] and indeed, we also show that FMRP binds *HTT*ex1$(CAG)_{ex}$ RNA in a repeat length-dependent manner. However, FMRP is mainly known to repress translation of its mRNA targets and this process is modulated by phosphorylation. FMRP's function and how this could play a role in MID1-dependent translation will be further discussed in chapter 4.6.

4.2 MID1 expression in HD

Initially, MID1 was identified in the context of Opitz syndrome, a monogenic disorder that is characterized by body axis patterning and midline formation defects [120]. Therefore, MID1 expression patterns have been widely studied during embryogenesis and development, showing that in human fetal organs MID1 expression is highest in the kidney, followed by brain and lung [120]. In the adult brain, heart, and placenta MID1 is expressed most abundantly. In mice, MID1 is expressed throughout the whole body starting at embryonic day 10.5, with the exception of the heart [120]. Several large-scale projects have mapped gene expression on RNA as well as protein level in a variety of human tissues, cell types, and in model organisms. Figure 20 summarizes the results from three projects regarding MID1 expression in human brain tissues and cell types: the human protein atlas (HPA) [184], FANTOM5 [185], and GTEx [186]. In line with the RNAseq data on cerebellar expression (Figure 20.A), our results show highest MID1 transcript levels in the human cerebellum.

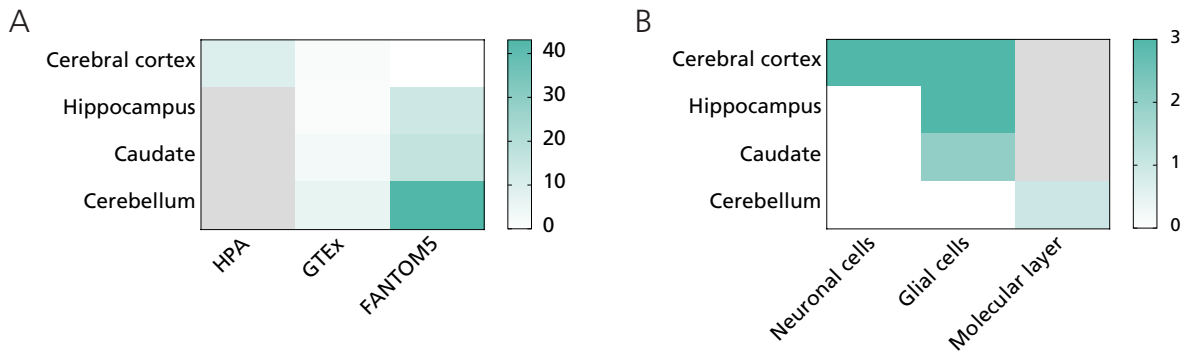

Figure 20 | Published MID1 expression patterns in human brain tissues.
(A) RNAseq data. Expression levels from human protein atlas, HPA and FANTOM5 are reported in TPM, transcripts per million; data from GTEx in RPKM, reads per kilobase million. (B) Protein expression data from HPA determined by IHC staining. Grey denotes no available data. Data was obtained from http://www.proteinatlas.org/ENSG00000101871-MID1/tissue [184].

Detailed analysis of protein expression in cellular subtypes reveals equal MID1 levels in neurons and glial cells according to HPA (Figure 20.B). The HPA project also uses morphology for cell type identification and cerebral cortex stainings are in agreement with our data. IHC stainings on cortical sections, specifically the middle temporal gyrus, show that MID1 protein seems to be expressed cell type-specific in neurons and glia.

Considering MID1's potential in the treatment of HD, knowledge about its expression pattern, especially in the human brain and in the context of HD is important, but was missing. Here, we show that both in mice and humans, *MID1* mRNA expression is genotype-dependent, with higher expression in young Q150 mice and HD patients compared to controls. However, *Mid1* mRNA levels show no difference between genotypes in old mice, which contradicts the human data. Possibly, this particular mouse model of HD does not reproduce the *MID1* expression changes seen in humans. The mutant CAG repeat is much longer compared to human HD cases and is actually in the range of juvenile HD. A longer CAG repeat might lead to earlier upregulation of MID1 expression in humans as well though this is difficult to test owing to the rarity of juvenile HD cases [187].

The analysis of MID1 protein expression determined by IHC staining and automated quantitative analysis reveals that MID1 protein levels are consistent with transcript levels: the amounts of positively stained cells in the white matter is significantly higher in HD patients.

The cortical layers representing grey matter and the underlying white matter differ considerably in regard to their cell type composition. Grey matter mainly consists of cell bodies and neuropil, i.e. areas of unmyelinated axons, dendrites, and glial processes. In contrast, the white matter mostly contains myelinated sheaths of neuronal axons. Neuronal cell types in the grey matter are distributed in discrete layers, encompassing structurally and functionally diverse neurons like pyramidal and stellate neurons, and many subtypes [188,189,190]. Clearly, MID1 is differentially expressed in

various cell types and unambiguously identifying these especially in the context of HD could help to understand MID1-dependent pathomechanims. Fluorescent double staining with antibodies directed against neuronal cell-type markers like neuronal nuclei (NeuN), microtubule-associated protein 2 (MAP2), neurofilament proteins, or postsynaptic density protein 95 (PSD95) should give some insight into the intercellular and intracellular distribution of MID1.

4.3 MID1 and the immune system

The most striking difference between HD patients and controls was seen in the number of MID1 positively stained cells in the white matter. These cells show glial morphology, however, distinguishing astrocytes and microglia unequivocally without cellular markers is challenging. Interestingly, changes in both astrocytes and microglia accompany HD [20], underlining the involvement of the immune system. These non-neuronal cells of the central nervous system (CNS) have specific roles in supporting neurons and controlling the microenvironment of the brain.

Astrocytes are heterogeneous, for example their morphology differs considerable depending on their location: in white matter mainly fibrous astrocytes can be found. They are smaller and less branched than the protoplasmic astrocytes of the grey matter. Astrogliosis, the activation of astrocytes upon various insults, is neuroprotective under normal conditions. In HD, multiple lines of evidence suggest a detrimental role and point at astrocytic dysfunction caused by mutant HTT [191,192]. Mutant HTT accumulates in glial nuclei in HD brains, of note in mouse brains Htt aggregates are more frequent in neurons compared to glia [70]. Expressing mutant Htt selectively in astrocytes in a transgenic mouse model causes age-dependent neurological phenotypes, even though it is expressed at lower levels than endogenous wild-type Htt [191]. Furthermore, mutant Htt in astrocytes decreases the expression of glutamate transporter offering an explanation for the observed excitotoxicity in HD [191,193].

In the brains of HD patients, the morphological changes of astrocytes that accompany astrogliosis increase with severity of disease [192]. Like astrocytes, microglia can exert both positive and negative effects, and are the primary mediators of neuroinflammation. Under physiological conditions they contain branched cytoplasmic processes and function as the macrophages of the central nervous system playing an important phagocytic role. In HD, the severity of disease progression is accompanied by microglial activation and can be detected even before onset of symptoms [15,16].

The TNF-related apoptosis inducing ligand (TRAIL) possibly bridges the gap between the immune system and the observed up-regulation of MID1 in neuroglia. TRAIL has central functions in the regulation of the immune system and is implicated in many diseases, notably, in neurodegenerative disorders like Alzheimer's disease and multiple sclerosis [194–196]. Interestingly, in Eosinophilic esophagitis (an inflammatory disorder of the esophagus) TRAIL upregulates MID1

expression. MID1-dependent downregulation of PP2A increases nuclear factor κB (NFκB) activation and thereby promotes inflammation [197]. In HD, the TRAIL-dependent activation of MID1 expression would additionally lead to increased mutant HTT translation, further promoting neurodegeneration. To test this hypothesis, several experiments are rational. First, analyzing the co-expression of MID1 and Iba1 (marker for microglia), or GFAP (marker for astrocytes), respectively in double immune staining gives insight into the exact immune cell type MID1 is expressed in. Furthermore, the status of the cells, i.e. whether they are resting, activated, or reactive is important to assess possible effects on HD pathomechanisms. Second, modulating TRAIL activity in HD models and analyzing the effect on HTT expression could elucidate TRAIL's role in HD and possibly provide a druggable target to treat HD. Finally, one should test if MID1 silencing in the respective cell types inhibits NFκB signaling.

4.4 *HTT*ex1(CAG)$_{ex}$ RNA gain-of-function: aberrant protein interactions

RBPs are endowed with a central role in cellular processes owing to their ability to bind RNA, the mediator of genetic information. Depending on their various functions, RBPs orchestrate splicing, transport, localization, stability, and/ or translation and consequently, their disruption impacts disease processes. In TNR disorders it is accepted that the aberrant binding of RBPs to mutant RNA plays an integral role in the pathobiology [198]. The downstream effects are widespread and depend on the physiologic RBP function and in several disorders they can explain the particular phenotype [199,200]. Since the aberrant interaction lies upstream of the observed toxicity, understanding mutant RNA-protein interactions could help to discover new targets for the treatment of TNR disorders.

This study reports an unbiased *in vitro* approach that identifies novel protein interaction partners of *HTT*ex1 RNA. Moreover, we can confirm the binding of several known *HTT*ex1 RNA binding partners, namely Dicer, SRSF6, Nucleolin, PKR and FMRP, emphasizing the validity of our approach. These proteins are known to contribute to RNA toxicity: Dicer activates the RNA interference pathway by producing sCAGs [115], SRSF6 promotes the mis-splicing of *HTT* itself, *MAPT* and *MAP2* isoforms in HD [108], aberrant interactions with Nucleolin elicit nucleolar stress [100], while PKR modulates eIF2 signaling and is disturbed in HD [103]. FMRP will be discussed in more detail below.

Additionally, proteins that are annotated binding partners of other mutant TNR RNAs were found. *HTT*ex1 RNA binds DDX5 that previously has been shown to bind *DMPK* RNA[93]. The TNR tracts within these transcripts resemble each other structurally and functionally [135,201], indicating that shared binding partners could account for similar symptoms between these diseases. In DM1, DDX5 mediates the unwinding of RNA secondary structure, thereby supporting aberrant MBNL1 binding [93]. Given

that we identified nine additional RNA helicases that specifically bind *HTT*ex1 RNA, three of them in a CAG repeat length-dependent manner (DDX46, DHX15, and DHX36), it seems likely that a similar mechanism is involved in HD. Performing specific knockdowns in HD models could give some insight into their role as modifiers of *HTT*ex1(CAG)$_{ex}$ RNA toxicity.

From the quantitative identification of protein interactions with specific CAG repeat lengths we can draw conclusions regarding the specific interactions of *HTT*ex1(CAG)$_{ex}$ RNA. This analysis highlights the importance of splicing factors in the protein network of mutant *HTT* RNA. Out of 36 proteins that bind CAG repeat-dependently, 30 are annotated with the GO term 'RNA splicing'. Two such proteins are PRPF8 and SNRNP40, both integral components of the spliceosome [202]. The splicing of precursor mRNA (pre-mRNA) involves more than 300 proteins that assemble into small nuclear ribonucleoprotein complexes (snRNPs). Distinct snRNPs consist of specific noncoding small nuclear RNAs (snRNAs) and core proteins. They can be divided into U1, U2, U4, U5, and U6 snRNPs. The major spliceosomal pathway is characterized by the step-wise assembly of snRNPs that mediate distinct functions (Figure 21). The U1 and U2 snRNPs recognize the 5' splice site and the branch point, respectively, on the pre-mRNA. The U4/U6.U5 tri-snRNP then associates with the pre-mRNA, effectively rearranging the bound snRNPs. This ultimately leads to catalytic activation of the spliceosomal complex, which removes the intronic sequence and joins the 5' and 3' exons. PRPF8 and SNRNP40 are core components of the U5 snRNP. PRPF8 directly contacts the 5' splice site, the branch point and 3' splice site, and engages the U5 and U6 snRNAs [174,203,204].

PRPF8 depletion causes widespread mis-splicing preferentially of transcripts that contain weak (non-consensus) 5' splice sites [175]. This study shows that PRPF8 and SNRNP40 bind *HTT* exon 1 RNA CAG repeat length-dependently. Moreover, the conditional expression of *HTT*ex1(CAG)$_{ex}$ RNA induces retention of introns with weak 5' splice sites and this effect is mediated by PRPF8. Together these results suggest that in HD, PRPF8 is sequestered by mutant RNA leading to suboptimal spliceosome assembly and activation, resulting in the usage of strong 5' splice sites. Strikingly, the mis-splicing of one identified transcript, *CREB1*, is significantly increased in cortical tissue of HD patients, emphasizing the biological relevance of our results.

The question how PRPF8 may loose its function in HD remains. One possibility is the sequestration into RNA foci. However, MBNL1, a well-defined protein that is captured by stable RNA foci in multiple TNR disorders [84,87,112], was absent in the list of *HTT* exon 1 protein binding partners. This suggests that aberrant binding of proteins to *HTT*ex1(CAG)$_{ex}$ RNA and the sequestration into stable

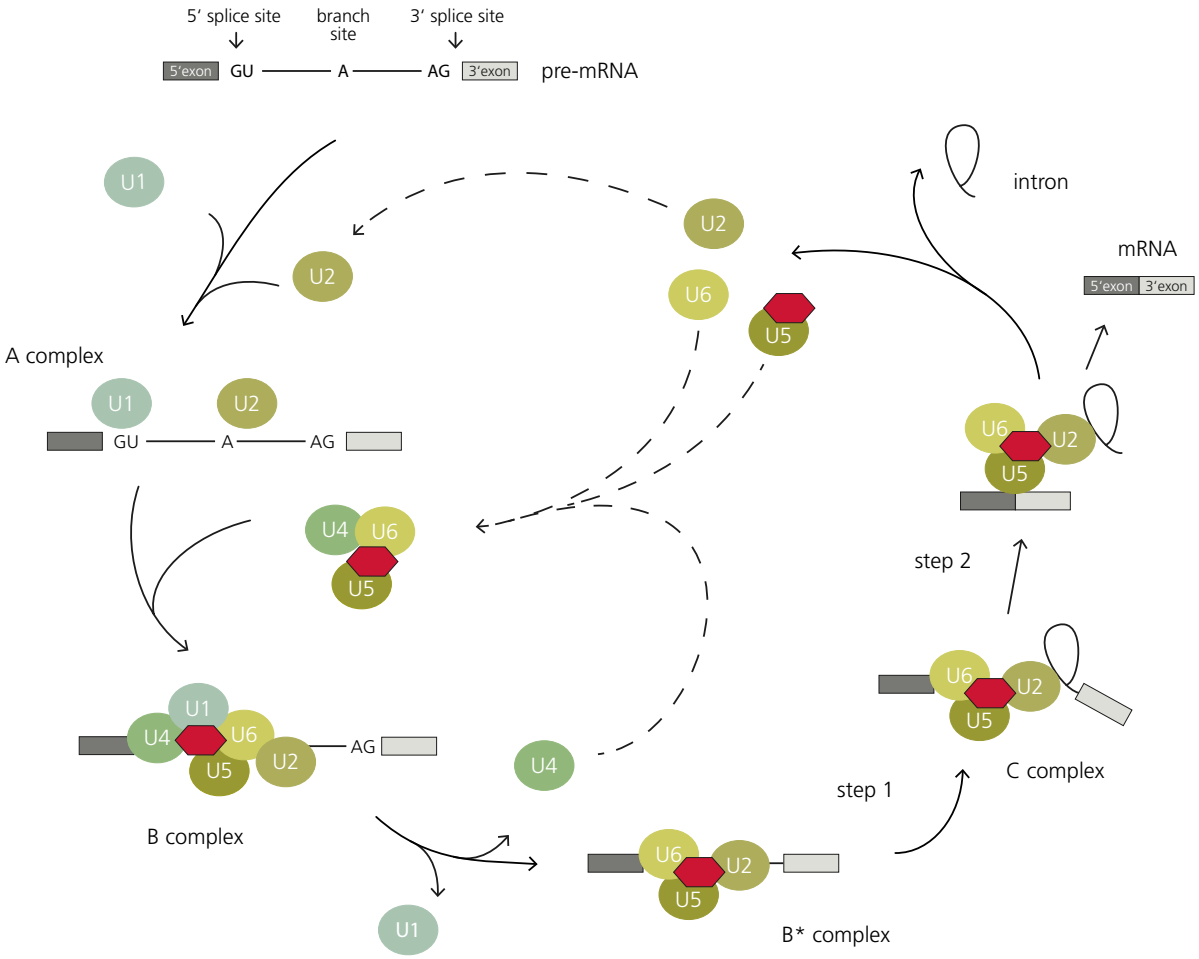

Figure 21 | The splicing cycle.
The pre-mRNA consists of consensus sites (invariant nucleotides at these sites are indicated) in the 5' splice site, branch point, and 3' splice site. The steps of spliceosome assembly are shown advancing counter-clockwise. The A complex comprises U1 and U2 snRNPs: U1 recognizes the 5' splice site and U2 the branch site. The tri-snRNP (U4/U6.U5 snRNPs) joins to form the B complex. The subsequent remodeling leads to the dissociation of U1 and U4 snRNPs, and the formation of the activated B* complex. B* is competent to perform the first catalytic reaction that results in the formation of a lariat structure in the intron. The second reaction is performed by the C complex leading to the removal of the intronic sequence and joining of the 5' and 3' exons. PRPF8, depicted in red, is a component of the U5 snRNP.

RNA foci are distinct mechanisms. Whether PRPF8 localization to RNA foci remains to be established. RNA foci of TNR RNAs can be visualized in fixed cells and tissues by RNA fluorescence in situ hybridization (FISH) [205]. This has guided the study of aberrant RNA foci interactions and sequestration of RBPs, however temporal resolution is not possible with this method. Considering that aberrant protein interactions of soluble mutant RNA can be detrimental, the study of RNA dynamics in living cells is feasible. Visualizing a pure CGG repeat of 60 units has been achieved using a genetically encoded system: the RNA of interest is tagged with an apatamer that binds a small molecule mimic of the GFP fluorophore [206]. Using this technology several open questions concerning *HTT*ex1 RNA could be addressed. First, one could study the effect of mutant CAG repeat length on *HTT*ex1

RNA dynamics. Second, using appropriately tagged fluorescent proteins, the interactions with *HTT*ex1(CAG)$_{ex}$ RNA could be examined, providing information on dynamic processes such as nuclear export, translation, and transport. Lastly, small molecules targeting *HTT*ex1(CAG)$_{ex}$ RNA can be tested for their ability to disassemble RNA foci or disrupt detrimental RNA-protein interactions.

4.5 CREB1 in HD

Our discovery of increased transcription and subsequent mis-splicing of *CREB1* in the cortex of HD patients support its role in HD. CREB1 is a widely studied transcription factor of the bZIP superfamily that binds to a cAMP-responsive element (CRE) located upstream from the transcriptional initiation site [207] and influences transcription together with other members of its family, namely cAMP response element modulator (CREM) and the acting transcription factor 1 (ATF-1) [208]. Its transcriptional activity is regulated by multiple stimuli like synaptic activity and the resulting Ca^{2+} influx, neurotransmitters, and growth factors [209,210]. The phosphorylation status of CREB1 is important for the transcription of multiple genes [211] and can be mediated by several kinases [212]. Distinct splicing isoforms of CREB1 activate or repress transcription, adding another layer of regulation [213]. CREB1 signaling seems to be neuroprotective: overexpression of constitutively active CREB1 prevents cell death, while expression of a dominant negative form of CREB1 leads to apoptosis in both sympathetic neurons and cerebellar granule cells [214,215]. CREB1 is also implicated in HD: in a cell model [172] and in the striatum of transgenic mice expressing HTT-Q111 [173] CRE-mediated transcription is downregulated. Moreover, CREB1 loss-of-function in an HD mouse model accelerates motor dysfunction [216].

Of note, the above-mentioned studies investigating CREB1 function in neuronal survival and HD were performed in cellular cultures and animal models, not necessarily reflecting the pathogenic changes in HD patients. Supporting our finding that CREB1 transcripts are upregulated in cortical tissues of HD patients, adipose tissue from HD patients shows elevated levels of *CREB1* expression [217]. Interestingly, BDNF transcription during neuronal activity mostly depends on CREB1 [218] and cAMP levels are reduced in the cortex of HD patients [173]. Thus, reduced BDNF delivery at the cortico-striatal synapse that promotes vulnerability of striatal neurons in HD [68] could also be mediated by CREB1-dependent gene expression. Evaluating the effects of the intron retention event on protein levels gives more insight. Three CREB1 protein isoforms are annotated that are produced by alternative splicing [219]. The intron retention event affects intron 2-3 and incorporates a stop codon nine bases downstream of exon two. Translation would produce a truncated protein and the intron retention effectively disturbs all protein isoforms since the physiologic alternative splicing events are downstream of intron 2-3. Our observed up-regulation of *CREB1* could be a compensatory effect to increase levels of correctly spliced *CREB1* mRNA.

4.6 The MID1 complex and *HTT* exon 1 RNA share many protein binding partners

To understand the specific mechanisms of *HTT*ex1(CAG)$_{ex}$ RNA translation mediated by the MID1 complex, comparing the MID1 and *HTT* exon 1 RNA protein interactome provides some insight. The overlap is quite substantial: more than one third of each mass spectrometry set binds to both MID1 protein and *HTT*ex1(CAG)$_{ex}$ RNA. One shared binding partner is FMRP. Just like MID1, FMRP is an RBP, regulating both translation and transport of specific RNAs, many of which are involved in neuronal plasticity [183,220–222]. Interestingly, it regulates local translation at the synapse activity-dependently [223] and directly associates with the ribosome [166]. Substantial effort has been directed towards the identification of FMRP consensus binding motifs, identifying both sequence-dependent and structural motifs as FMRP binding determinants [169,182,220,221,224]. One such study by Darnell and colleagues used high-throughput sequencing of RNAs isolated by HITS-CLIP to identify FMRP interactions with mouse brain polyribosomal mRNA [183]. Given that FMRP function might be disturbed in HD, we compared this data set with a large study analyzing proteomic changes in HD mouse models [225]. Langfelder et al. analyzed tissue samples from the striatum, cortex, and cerebellum of 2-, 6-, and 10-month-old knock-in mice with polyQ lengths of 20, 50, 80, 92, 111, 140, 175 as well as littermate control wild-type animals. FMRP target RNAs show 13 % overlap with misregulated proteins in the HD models, suggesting that translational misregulation of these specific transcripts in HD might be mediated by FMRP. *HTT* transcript and protein can be found in this overlapping list of FMRP RNA targets and misregulated proteins in HD. Indeed, we identified FMRP in two replicates in our study binding to *HTT* exon 1 RNA with 40 CAG repeats. Even though it is not included in the list of proteins preferentially binding *HTT*ex1(CAG)$_{ex}$ RNA, we could show on western blot that FMRP binding to *HTT*ex1(CAG)$_{ex}$ increases with longer CAG repeats. Moreover, several known FMRP interaction partners were identified in both mass spectrometry analyses: RPLP0, RPL5, RPL8, PABP1, Staufen (*STAU1*), Nucleolin, CAPRIN1, DDX5, GEMIN4, and YBX1, strengthening the hypothesis that MID1 together with FMRP, at least of *HTT*ex1(CAG)$_{ex}$ RNA, regulates translation. The results of MID1-dependent TNR repeat translation underline the theory of specific RNP complexes: a CAG repeat in the 3' UTR increases translation. While this is MID1-dependent, the MID1 knockdown did not completely reverse the effect, showing that other RBPs must be involved.

These results have several implications. MID1 together with FMRP might regulate a certain subset of RNAs in terms of localization and translation. Interestingly, mutations in the *MID1* and *FMR1* genes leading to protein loss-of-function both lead to mental retardation [120,226] potentially reflecting the effects from shared RNA targets. The FMRP phosphorylation status controls its usually inhibitory effect on translation [227] and this modification depends on several factors, for example synaptic

activity [223] and ubiquitination [228]. However, the exact determinants of FMRP phosphorylation are under debate, some studies identifying the mTOR pathway, PP2A, and S6K [223,229,230], while others suggest different kinases [231,232]. Still, it is conceivable that MID1 may play a role in FMRP phosphorylation regulation through acting on kinases targeting FMRP. Alternatively, considering that MID1 is an E3 ubiquitin ligase, MID1 might be targeting FMRP for proteasomal degradation. These hypotheses could be tested by siRNA-mediated knockdown of MID1 and subsequent phospho-FMRP detection by western blot and *in vitro* ubiquitination assays, respectively.

Taken together, this study provides new evidence for MID1's contribution to HD pathogenesis. MID1 induces translation of a subset of RNAs with specific secondary structures. Moreover, *MID1* expression is upregulated in cortical tissue from HD patients, most likely exacerbating this detrimental process. Apart from MID1, mutant *HTT* RNA interacts with multiple proteins disturbing their function. Mapping this RNA-protein interactome shows that the interaction of mutant *HTT* RNA with PRPF8 changes the alternative splicing pattern of *CREB1* in cell models as well as cortical tissue from HD patients. These data support the crucial role of RNA toxicity in HD and suggest several strategies for therapeutic interventions. First, targeting the MID1 protein is promising considering that it specifically upregulates translation of the mutant *HTT* allele. Second, affecting other RBPs, for example splice factors, could help to alleviate associated phenotypes. Lastly, directly targeting the mutant *HTT* RNA could help to suppress RNA-mediated toxicity.

5 References

1. Subramanian, S. *et al.* Triplet repeats in human genome: distribution and their association with genes and other genomic regions. *Bioinformatics* **19,** 549–52 (2003).
2. Fu, Y. H. *et al.* Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* **67,** 1047–58 (1991).
3. La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352,** 77–9 (1991).
4. Duyao, M. P. *et al.* Inactivation of the mouse Huntington's disease gene homolog Hdh. *Science* **269,** 407–10 (1995).
5. Tezenas du Montcel, S. *et al.* Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain* **137,** 2444–2455 (2014).
6. Yum, K., Wang, E. T. & Kalsotra, A. Myotonic dystrophy: disease repeat range, penetrance, age of onset, and relationship between repeat size and phenotypes. *Curr. Opin. Genet. Dev.* **44,** 30–37 (2017).
7. Verkerk, A. J. *et al.* Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65,** 905–14 (1991).
8. Group, T. H. D. C. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72,** 971–983 (1993).
9. Craufurd, D. & Dodge, A. Mutation size and age at onset in Huntington's disease. *J. Med. Genet.* **30,** 1008–11 (1993).
10. Andresen, J. M. *et al.* The relationship between CAG repeat length and age of onset differs for Huntington's disease patients with juvenile onset or adult onset. *Ann. Hum. Genet.* **71,** 295–301 (2007).
11. Lee, J.-M. *et al.* CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. *Neurology* **78,** 690–5 (2012).
12. Ross, C. a *et al.* Huntington disease: natural history, biomarkers and prospects for therapeutics. *Nat. Rev. Neurol.* **10,** 204–16 (2014).
13. van der Burg, J. M. M., Björkqvist, M. & Brundin, P. Beyond the brain: widespread pathology in Huntington's disease. *Lancet. Neurol.* **8,** 765–74 (2009).
14. Björkqvist, M. *et al.* A novel pathogenic pathway of immune activation detectable before clinical onset in Huntington's disease. *J. Exp. Med.* **205,** 1869–1877 (2008).
15. Tai, Y. F. *et al.* Microglial activation in presymptomatic Huntington's disease gene carriers. *Brain* **130,** 1759–1766 (2007).
16. Pavese, N. *et al.* Microglial activation correlates with severity in Huntington disease: A clinical and PET study. *Neurology* **66,** 1638–1643 (2006).
17. Fisher, E. R. & Hayden, M. R. Multisource ascertainment of Huntington disease in Canada: Prevalence and population at risk. *Mov. Disord.* **29,** 105–114 (2014).
18. Squitieri, F. *et al.* DNA haplotype analysis of Huntington disease reveals clues to the origins and

mechanisms of CAG expansion and reasons for geographic variations of prevalence. *Hum. Mol. Genet.* **3,** 2103–14 (1994).

19. Graveland, G. A., Williams, R. S. & DiFiglia, M. Evidence for degenerative and regenerative changes in neostriatal spiny neurons in Huntington's disease. *Science* **227,** 770–3 (1985).
20. Vonsattel, J.-P. *et al.* Neuropathological Classification of Huntington's Disease. *J. Neuropathol. Exp. Neurol.* **44,** (1985).
21. Mitchell, I. J., Cooper, A. J. & Griffiths, M. R. The selective vulnerability of striatopallidal neurons. *Prog. Neurobiol.* **59,** 691–719 (1999).
22. Aylward, E. H. *et al.* Onset and rate of striatal atrophy in preclinical Huntington disease. *Neurology* **63,** 66–72 (2004).
23. de la Monte, S. M., Vonsattel, J. P. & Richardson, E. P. Morphometric demonstration of atrophic changes in the cerebral cortex, white matter, and neostriatum in Huntington's disease. *J. Neuropathol. Exp. Neurol.* **47,** 516–25 (1988).
24. DiFiglia, M. *et al.* Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* **277,** 1990–3 (1997).
25. Sahl, S. J., Weiss, L. E., Duim, W. C., Frydman, J. & Moerner, W. E. Cellular inclusion bodies of mutant huntingtin exon 1 obscure small fibrillar aggregate species. *Sci. Rep.* **2,** 895 (2012).
26. Jansen, A. H. P. *et al.* Frequency of nuclear mutant huntingtin inclusion formation in neurons and glia is cell-type-specific. *Glia* **65,** 50–61 (2017).
27. Landles, C. *et al.* Proteolysis of mutant huntingtin produces an exon 1 fragment that accumulates as an aggregated protein in neuronal nuclei in Huntington disease. *J. Biol. Chem.* **285,** 8808–23 (2010).
28. Scherzinger, E. *et al.* Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 4604–9 (1999).
29. Atwal, R. S. *et al.* Huntingtin has a membrane association signal that can modulate huntingtin aggregation, nuclear entry and toxicity. *Hum. Mol. Genet.* **16,** 2600–2615 (2007).
30. Kim, M. W., Chelliah, Y., Kim, S. W., Otwinowski, Z. & Bezprozvanny, I. Secondary Structure of Huntingtin Amino-Terminal Region. *Structure* **17,** 1205–1212 (2009).
31. Andrade, M. A. & Bork, P. HEAT repeats in the Huntington's disease protein. *Nat. Genet.* **11,** 115–116 (1995).
32. Palidwor, G. A. *et al.* Detection of alpha-rod protein repeats using a neural network and application to huntingtin. *PLoS Comput. Biol.* **5,** e1000304 (2009).
33. Sharp, A. H. *et al.* Widespread expression of Huntington's disease gene (IT15) protein product. *Neuron* **14,** 1065–74 (1995).
34. Sapp, E. *et al.* Huntingtin localization in brains of normal and Huntington's disease patients. *Ann. Neurol.* **42,** 604–612 (1997).
35. Aronin, N. *et al.* CAG expansion affects the expression of mutant Huntingtin in the Huntington's disease brain. *Neuron* **15,** 1193–201 (1995).

36. DiFiglia, M. *et al.* Huntingtin is a cytoplasmic protein associated with vesicles in human and rat brain neurons. *Neuron* **14,** 1075–81 (1995).
37. Trottier, Y. *et al.* Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. *Nature* **378,** 403–6 (1995).
38. Steffan, J. S. *et al.* The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6763–8 (2000).
39. Kegel, K. B. *et al.* Huntingtin is present in the nucleus, interacts with the transcriptional corepressor C-terminal binding protein, and represses transcription. *J. Biol. Chem.* **277,** 7466–76 (2002).
40. Goehler, H. *et al.* A Protein Interaction Network Links GIT1, an Enhancer of Huntingtin Aggregation, to Huntington's Disease. *Mol. Cell* **15,** 853–865 (2004).
41. Kaltenbach, L. S. *et al.* Huntingtin Interacting Proteins Are Genetic Modifiers of Neurodegeneration. *PLoS Genet.* **3,** e82 (2007).
42. Ratovitski, T. *et al.* Huntingtin protein interactions altered by polyglutamine expansion as determined by quantitative proteomic analysis. *Cell Cycle* **11,** 2006–2021 (2012).
43. Gauthier, L. R. *et al.* Huntingtin controls neurotrophic support and survival of neurons by enhancing BDNF vesicular transport along microtubules. *Cell* **118,** 127–38 (2004).
44. Caviston, J. P., Zajac, A. L., Tokito, M. & Holzbaur, E. L. F. Huntingtin coordinates the dynein-mediated dynamic positioning of endosomes and lysosomes. *Mol. Biol. Cell* **22,** 478–92 (2011).
45. Wong, Y. C. & Holzbaur, E. L. F. The Regulation of Autophagosome Dynamics by Huntingtin and HAP1 Is Disrupted by Expression of Mutant Huntingtin, Leading to Defective Cargo Degradation. *J. Neurosci.* **34,** 1293–1305 (2014).
46. Acheson, A. *et al.* A BDNF autocrine loop in adult sensory neurons prevents cell death. *Nature* **374,** 450–453 (1995).
47. Huang, E. J. & Reichardt, L. F. Neurotrophins: Roles in Neuronal Development and Function. *Annu. Rev. Neurosci.* **24,** 677–736 (2001).
48. Rui, Y.-N. *et al.* Huntingtin functions as a scaffold for selective macroautophagy. *Nat. Cell Biol.* **17,** 262–275 (2015).
49. Zheng, S. *et al.* Deletion of the Huntingtin Polyglutamine Stretch Enhances Neuronal Autophagy and Longevity in Mice. *PLoS Genet.* **6,** e1000838 (2010).
50. Velier, J. *et al.* Wild-type and mutant huntingtins function in vesicle trafficking in the secretory and endocytic pathways. *Exp. Neurol.* **152,** 34–40 (1998).
51. El-Daher, M.-T. *et al.* Huntingtin proteolysis releases non-polyQ fragments that cause toxicity through dynamin 1 dysregulation. *EMBO J.* **34,** 2255–2271 (2015).
52. Marcora, E., Gowan, K. & Lee, J. E. Stimulation of NeuroD activity by huntingtin and huntingtin-associated proteins HAP1 and MLK2. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 9578–83 (2003).
53. Dunah, A. W. *et al.* Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. *Science* **296,** 2238–43 (2002).
54. Zuccato, C. *et al.* Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-

controlled neuronal genes. *Nat. Genet.* **35,** 76–83 (2003).

55. Nasir, J. *et al.* Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygotes. *Cell* **81,** 811–23 (1995).
56. Zeitlin, S., Liu, J. P., Chapman, D. L., Papaioannou, V. E. & Efstratiadis, A. Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. *Nat. Genet.* **11,** 155–63 (1995).
57. White, J. K. *et al.* Huntingtin is required for neurogenesis and is not impaired by the Huntington's disease CAG expansion. *Nat. Genet.* **17,** 404–410 (1997).
58. Dragatsis, I., Levine, M. S. & Zeitlin, S. Inactivation of Hdh in the brain and testis results in progressive neurodegeneration and sterility in mice. *Nat. Genet.* **26,** 300–6 (2000).
59. Leavitt, B. R. *et al.* Wild-type huntingtin protects neurons from excitotoxicity. *J. Neurochem.* **96,** 1121–1129 (2006).
60. Bečanović, K. *et al.* A SNP in the HTT promoter alters NF-κB binding and is a bidirectional genetic modifier of Huntington disease. *Nat. Neurosci.* **18,** 807–16 (2015).
61. Drouet, V. *et al.* Allele-specific silencing of mutant huntingtin in rodent brain and human stem cells. *PLoS One* **9,** e99341 (2014).
62. Sathasivam, K. *et al.* Aberrant splicing of HTT generates the pathogenic exon 1 protein in Huntington disease. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 2366–70 (2013).
63. Mangiarini, L. *et al.* Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* **87,** 493–506 (1996).
64. Reijonen, S., Putkonen, N., Nørremølle, A., Lindholm, D. & Korhonen, L. Inhibition of endoplasmic reticulum stress counteracts neuronal cell death and protein aggregation caused by N-terminal mutant huntingtin proteins. *Exp. Cell Res.* **314,** 950–60 (2008).
65. Orr, A. L. *et al.* N-terminal mutant huntingtin associates with mitochondria and impairs mitochondrial trafficking. *J. Neurosci.* **28,** 2783–92 (2008).
66. Shimohata, T. *et al.* Expanded polyglutamine stretches interact with TAFII130, interfering with CREB-dependent transcription. *Nat. Genet.* **26,** 29–36 (2000).
67. Twelvetrees, A. E. *et al.* Delivery of GABAARs to Synapses Is Mediated by HAP1-KIF5 and Disrupted by Mutant Huntingtin. *Neuron* **65,** 53–65 (2010).
68. Baquet, Z. C., Gorski, J. A. & Jones, K. R. Early striatal dendrite deficits followed by neuron loss with advanced age in the absence of anterograde cortical brain-derived neurotrophic factor. *J. Neurosci.* **24,** 4250–8 (2004).
69. Zuccato, C. *et al.* Loss of Huntingtin-Mediated BDNF Gene Transcription in Huntington's Disease. *Science (80-.).* **293,** 493–498 (2001).
70. Shin, J.-Y. *et al.* Expression of mutant huntingtin in glial cells contributes to neuronal excitotoxicity. *J. Cell Biol.* **171,** 1001–12 (2005).
71. Takahashi, T. *et al.* Soluble polyglutamine oligomers formed prior to inclusion body formation are cytotoxic. *Hum. Mol. Genet.* **17,** 345–56 (2008).
72. Bence, N. F., Sampat, R. M. & Kopito, R. R. Impairment of the Ubiquitin-Proteasome System by

Protein Aggregation. *Science (80-.).* **292,** 1552–1555 (2001).

73. Busch, A. *et al.* Mutant huntingtin promotes the fibrillogenesis of wild-type huntingtin: a potential mechanism for loss of huntingtin function in Huntington's disease. *J. Biol. Chem.* **278,** 41452–61 (2003).
74. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R. & Finkbeiner, S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431,** 805–10 (2004).
75. Jayaraman, M. *et al.* Slow amyloid nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. *J. Mol. Biol.* **415,** 881–99 (2012).
76. Ossato, G. *et al.* A two-step path to inclusion formation of huntingtin peptides revealed by number and brightness analysis. *Biophys. J.* **98,** 3078–85 (2010).
77. Chen, S., Ferrone, F. A. & Wetzel, R. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 11884–9 (2002).
78. Mahadevan, M. *et al.* Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. *Science* **255,** 1253–5 (1992).
79. Li, L.-B., Yu, Z., Teng, X. & Bonini, N. M. RNA toxicity is a component of ataxin-3 degeneration in Drosophila. *Nature* **453,** 1107–11 (2008).
80. Michlewski, G. & Krzyzosiak, W. J. Molecular architecture of CAG repeats in human disease related transcripts. *J. Mol. Biol.* **340,** 665–679 (2004).
81. Sobczak, K., de Mezer, M., Michlewski, G., Krol, J. & Krzyzosiak, W. J. RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res.* **31,** 5469–82 (2003).
82. Napierała, M. & Krzyzosiak, W. J. CUG repeats present in myotonin kinase RNA form metastable 'slippery' hairpins. *J. Biol. Chem.* **272,** 31079–85 (1997).
83. Broda, M., Kierzek, E., Gdaniec, Z., Kulinski, T. & Kierzek, R. Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure. *Biochemistry* **44,** 10873–82 (2005).
84. de Mezer, M., Wojciechowska, M., Napierala, M., Sobczak, K. & Krzyzosiak, W. J. Mutant CAG repeats of Huntingtin transcript fold into hairpins, form nuclear foci and are targets for RNA interference. *Nucleic Acids Res.* **39,** 3852–63 (2011).
85. Busan, S. & Weeks, K. M. Role of context in RNA structure: flanking sequences reconfigure CAG motif folding in huntingtin exon 1 transcripts. *Biochemistry* **52,** 8219–25 (2013).
86. Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H. & Housman, D. E. Expansion of a CUG trinucleotide repeat in the 3′ untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 7388–93 (1997).
87. Fardaei, M., Larkin, K., Brook, J. D. & Hamshere, M. G. In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res.* **29,** 2766–71 (2001).
88. Mankodi, A. *et al.* Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet.* **10,** 2165–70 (2001).
89. Cardani, R., Mancinelli, E., Rotondo, G., Sansone, V. & Meola, G. Muscleblind-like protein 1 nuclear sequestration is a molecular pathology marker of DM1 and DM2. *Eur. J. Histochem.* **50,**

177–82 (2006).

90. Charlet-B, N. *et al.* Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol. Cell* **10,** 45–53 (2002).
91. Kimura, T. *et al.* Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca2+-ATPase in myotonic dystrophy type 1. *Hum. Mol. Genet.* **14,** 2189–2200 (2005).
92. Jiang, H., Mankodi, A., Swanson, M. S., Moxley, R. T. & Thornton, C. A. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum. Mol. Genet.* **13,** 3079–3088 (2004).
93. Laurent, F.-X. *et al.* New function for the RNA helicase p68/DDX5 as a modifier of MBNL1 activity on expanded CUG repeats. *Nucleic Acids Res.* **40,** 3159–71 (2012).
94. Greco, C. M. *et al.* Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain* **125,** 1760–71 (2002).
95. Tassone, F., Iwahashi, C. & Hagerman, P. J. FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). *RNA Biol.* **1,** 103–5 (2004).
96. Tsoi, H., Lau, C. K., Lau, K. F. & Chan, H. Y. E. Perturbation of U2AF65/NXF1-mediated RNA nuclear export enhances RNA toxicity in polyQ diseases. *Hum. Mol. Genet.* **20,** 3787–3797 (2011).
97. Kalita, K., Makonchuk, D., Gomes, C., Zheng, J.-J. & Hetman, M. Inhibition of nucleolar transcription as a trigger for neuronal apoptosis. *J. Neurochem.* **105,** 2286–99 (2008).
98. Parlato, R. *et al.* Activation of an endogenous suicide response after perturbation of rRNA synthesis leads to neurodegeneration in mice. *J. Neurosci.* **28,** 12759–64 (2008).
99. Rieker, C. *et al.* Nucleolar disruption in dopaminergic neurons leads to oxidative damage and parkinsonism through repression of mammalian target of rapamycin signaling. *J. Neurosci.* **31,** 453–60 (2011).
100. Tsoi, H. *et al.* CAG expansion induces nucleolar stress in polyglutamine diseases. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 13428–13433 (2012).
101. Tsoi, H. & Chan, H. Y. E. Expression of Expanded CAG Transcripts Triggers Nucleolar Stress in Huntington's Disease. *The Cerebellum* **12,** 310–312 (2013).
102. Zhang, Y. & Lu, H. Signaling to p53: ribosomal proteins find their way. *Cancer Cell* **16,** 369–77 (2009).
103. Peel, A. L. *et al.* Double-stranded RNA-dependent protein kinase, PKR, binds preferentially to Huntington's disease (HD) transcripts and is activated in HD tissue. *Hum. Mol. Genet.* **10,** 1531–8 (2001).
104. Mayo, C. B., Wong, C. J., Lopez, P. E., Lary, J. W. & Cole, J. L. Activation of PKR by short stem–loop RNAs containing single-stranded arms. *RNA* **22,** 1065–1075 (2016).
105. Sadler, A. J. & Williams, B. R. G. Structure and function of the protein kinase R. *Curr. Top. Microbiol. Immunol.* **316,** 253–92 (2007).
106. Klann, E. & Dever, T. E. Biochemical mechanisms for translational regulation in synaptic plasticity. *Nat. Rev. Neurosci.* **5,** 931–942 (2004).
107. Leitman, J. *et al.* ER stress-induced eIF2-alpha phosphorylation underlies sensitivity of striatal

neurons to pathogenic huntingtin. *PLoS One* **9,** e90803 (2014).

108. Fernández-Nogales, M., Santos-Galindo, M., Hernández, I. H., Cabrera, J. R. & Lucas, J. J. Faulty splicing and cytoskeleton abnormalities in Huntington's disease. *Brain Pathol.* **26,** 772–778 (2016).
109. Fernández-Nogales, M. *et al.* Huntington's disease is a four-repeat tauopathy with tau nuclear rods. *Nat. Med.* **20,** 881–5 (2014).
110. Yuan, Y. *et al.* Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.* **35,** 5474–86 (2007).
111. Kim, D.-H. *et al.* HnRNP H inhibits nuclear export of mRNA containing expanded CUG repeats and a distal branch point sequence. *Nucleic Acids Res.* **33,** 3866–74 (2005).
112. Daughters, R. S. *et al.* RNA Gain-of-Function in Spinocerebellar Ataxia Type 8. *PLoS Genet.* **5,** e1000600 (2009).
113. Krauß, S. *et al.* Translation of HTT mRNA with expanded CAG repeats is regulated by the MID1–PP2A protein complex. *Nat. Commun.* **4,** 1511 (2013).
114. Chung, D. W., Rudnicki, D. D., Yu, L. & Margolis, R. L. A natural antisense transcript at the Huntington's disease repeat locus regulates HTT expression. *Hum. Mol. Genet.* **20,** 3467–77 (2011).
115. Bañez-Coronel, M. *et al.* A pathogenic mechanism in Huntington's disease involves small CAG-repeated RNAs with neurotoxic activity. *PLoS Genet.* **8,** e1002481 (2012).
116. Griesche, N. *et al.* Regulation of mRNA Translation by MID1: A Common Mechanism of Expanded CAG Repeat RNAs. *Front. Cell. Neurosci.* **10,** (2016).
117. Jin, P. *et al.* Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. *Neuron* **55,** 556–64 (2007).
118. Sellier, C. *et al.* Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J.* **29,** 1248–61 (2010).
119. Sellier, C. *et al.* Sequestration of DROSHA and DGCR8 by Expanded CGG RNA Repeats Alters MicroRNA Processing in Fragile X-Associated Tremor/Ataxia Syndrome. *Cell Rep.* **3,** 869–880 (2013).
120. Quaderi, N. A. *et al.* Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22. *Nat. Genet.* **17,** 285–91 (1997).
121. Schweiger, S. *et al.* The Opitz syndrome gene product, MID1, associates with microtubules. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 2794–9 (1999).
122. Short, K. M., Hopwood, B., Yi, Z. & Cox, T. C. MID1 and MID2 homo- and heterodimerise to tether the rapamycin-sensitive PP2A regulatory subunit, alpha 4, to microtubules: implications for the clinical variability of X-linked Opitz GBBB syndrome and other developmental disorders. *BMC Cell Biol.* **3,** 1 (2002).
123. Liu, J., Prickett, T. D., Elliott, E., Meroni, G. & Brautigan, D. L. Phosphorylation and microtubule association of the Opitz syndrome protein mid-1 is regulated by protein phosphatase 2A via binding to the regulatory subunit alpha 4. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 6650–5 (2001).
124. Trockenbacher, A. *et al.* MID1, mutated in Opitz syndrome, encodes an ubiquitin ligase that

targets phosphatase 2A for degradation. *Nat. Genet.* **29,** 287–94 (2001).

125. Liu, E., Knutzen, C. A., Krauss, S., Schweiger, S. & Chiang, G. G. Control of mTORC1 signaling by the Opitz syndrome protein MID1. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 8680–5 (2011).
126. Aranda-Orgillés, B. *et al.* Active transport of the ubiquitin ligase MID1 along the microtubules is regulated by protein phosphatase 2A. *PLoS One* **3,** e3507 (2008).
127. Merino, E., Balbás, P., Puente, J. L. & Bolívar, F. Antisense overlapping open reading frames in genes from bacteria to humans. *Nucleic Acids Res.* **22,** 1903–8 (1994).
128. Chen, J. *et al.* Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Res.* **32,** 4812–4820 (2004).
129. Morris, K. V, Santoso, S., Turner, A.-M., Pastori, C. & Hawkins, P. G. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet.* **4,** e1000258 (2008).
130. Moseley, M. L. *et al.* Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat. Genet.* **38,** 758–769 (2006).
131. Cho, D. H. *et al.* Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol. Cell* **20,** 483–9 (2005).
132. Gudde, A. E. E. G. *et al.* Antisense transcription of the myotonic dystrophy locus yields low-abundant RNAs with and without (CAG)n repeat. *RNA Biol.* **0,** 00–00 (2017).
133. Ladd, P. D. *et al.* An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. *Hum. Mol. Genet.* **16,** 3174–87 (2007).
134. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409,** 363–366 (2001).
135. Krol, J. *et al.* Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. *Mol. Cell* **25,** 575–86 (2007).
136. Zu, T. *et al.* Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 260–5 (2011).
137. Bañez-Coronel, M. *et al.* RAN Translation in Huntington Disease. *Neuron* **88,** 667–677 (2015).
138. Todd, P. K. *et al.* CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **78,** 440–55 (2013).
139. Timchenko, L. T. *et al.* Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res.* **24,** 4407–14 (1996).
140. Jiang, H., Mankodi, A., Swanson, M. S., Moxley, R. T. & Thornton, C. A. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum. Mol. Genet.* **13,** 3079–88 (2004).
141. Mankodi, A., Lin, X., Blaxall, B. C., Swanson, M. S. & Thornton, C. A. Nuclear RNA Foci in the Heart in Myotonic Dystrophy. *Circ. Res.* **97,** 1152–1155 (2005).
142. Savkur, R. S., Philips, A. V & Cooper, T. A. Aberrant regulation of insulin receptor alternative

splicing is associated with insulin resistance in myotonic dystrophy. *Nat. Genet.* **29,** 40–7 (2001).

143. Timchenko, N. A. *et al.* RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. *J. Biol. Chem.* **276,** 7820–6 (2001).
144. Ho, T. H., Bundman, D., Armstrong, D. L. & Cooper, T. A. Transgenic mice expressing CUG-BP1 reproduce splicing mis-regulation observed in myotonic dystrophy. *Hum. Mol. Genet.* **14,** 1539–47 (2005).
145. Kuyumcu-Martinez, N. M., Wang, G.-S. & Cooper, T. A. Increased Steady-State Levels of CUGBP1 in Myotonic Dystrophy 1 Are Due to PKC-Mediated Hyperphosphorylation. *Mol. Cell* **28,** 68–78 (2007).
146. Philips, A. V, Timchenko, L. T. & Cooper, T. A. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280,** 737–41 (1998).
147. Ausubel, F. M. *et al. Current Protocols in Molecular Biology*. *Molecular Biology* **1,** (2003).
148. Kanashova, T. *et al.* Differential proteomic analysis of mouse macrophages exposed to adsorbate-loaded heavy fuel oil derived combustion particles using an automated sample-preparation workflow. *Anal. Bioanal. Chem.* **407,** 5965–76 (2015).
149. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2,** 1896–906 (2007).
150. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372 (2008).
151. Waldvogel, H. J. *et al.* The collection and processing of human brain tissue for research. *Cell Tissue Bank.* **9,** 169–179 (2008).
152. Vonsattel, J. P. *et al.* Neuropathological classification of Huntington's disease. *J. Neuropathol. Exp. Neurol.* **44,** 559–77 (1985).
153. Waldvogel, H. J., Kubota, Y., Fritschy, J., Mohler, H. & Faull, R. L. Regional and cellular localisation of GABA(A) receptor subunits in the human basal ganglia: An autoradiographic and immunohistochemical study. *J. Comp. Neurol.* **415,** 313–40 (1999).
154. Waldvogel, H. J., Billinton, A., White, J. H., Emson, P. C. & Faull, R. L. M. Comparative cellular distribution of GABAA and GABAB receptors in the human basal ganglia: immunohistochemical colocalization of the alpha 1 subunit of the GABAA receptor, and the GABABR1 and GABABR2 receptor subunits. *J. Comp. Neurol.* **470,** 339–56 (2004).
155. Heng, M. Y., Tallaksen-Greene, S. J., Detloff, P. J. & Albin, R. L. Longitudinal evaluation of the Hdh(CAG)150 knock-in murine model of Huntington's disease. *J. Neurosci.* **27,** 8989–98 (2007).
156. Bétemps, D. *et al.* Detection of Disease-associated alpha-synuclein by Enhanced ELISA in the Brain of Transgenic Mice Overexpressing Human A53T Mutated alpha-synuclein. *J. Vis. Exp.* e52752 (2015). doi:10.3791/52752
157. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37,** W305–W311 (2009).
158. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6,** 26 (2011).

159. Köhler, A. *et al.* A hormone-dependent feedback-loop controls androgen receptor levels by limiting MID1, a novel translation enhancer and promoter of oncogenic signaling. *Mol. Cancer* **13,** 146 (2014).
160. Hettich, M. M. *et al.* The Anti-Diabetic Drug Metformin Reduces BACE1 Protein Level by Interfering with the MID1 Complex. *PLoS One* **9,** e102420 (2014).
161. Granata, A. *et al.* Evidence of functional redundancy between MID proteins: implications for the presentation of Opitz syndrome. *Dev. Biol.* **277,** 417–24 (2005).
162. Aranda-Orgillés, B. *et al.* The Opitz syndrome gene product MID1 assembles a microtubule-associated ribonucleoprotein complex. *Hum. Genet.* **123,** 163–76 (2008).
163. Buchner, G. *et al.* MID2, a homologue of the Opitz syndrome gene MID1: similarities in subcellular localization and differences in expression during development. *Hum. Mol. Genet.* **8,** 1397–407 (1999).
164. Weiler, I. J. *et al.* Fragile X mental retardation protein is translated near synapses in response to neurotransmitter activation. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 5395–400 (1997).
165. Feng, Y. *et al.* FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association. *Mol. Cell* **1,** 109–18 (1997).
166. Chen, E., Sharma, M. R., Shi, X., Agrawal, R. K. & Joseph, S. Fragile X mental retardation protein regulates translation by binding directly to the ribosome. *Mol. Cell* **54,** 407–17 (2014).
167. Kruppa, J. & Sabatini, D. Release of poly a(+) messenger RNA from rat liver rough microsomes upon disassembly of bound polysomes. *J. Cell Biol.* **74,** (1977).
168. Aronin, N. & DiFiglia, M. Huntingtin-lowering strategies in Huntington's disease: Antisense oligonucleotides, small RNAs, and gene editing. *Mov. Disord.* **29,** 1455–1461 (2014).
169. Darnell, J. C. & Klann, E. The translation of translational control by FMRP: therapeutic targets for FXS. *Nat. Neurosci.* **16,** 1530–1536 (2013).
170. Lin, L. *et al.* Transcriptome sequencing reveals aberrant alternative splicing in Huntington's disease. *Hum. Mol. Genet.* **25,** 3454–3466 (2016).
171. Labadorf, A. T. *et al.* Evidence of Extensive Alternative Splicing in Post Mortem Human Brain HTT Transcription by mRNA Sequencing. *PLoS One* **10,** e0141298 (2015).
172. Wyttenbach, A. *et al.* Polyglutamine expansions cause decreased CRE-mediated transcription and early gene expression changes prior to cell death in an inducible cell model of Huntington's disease. *Hum. Mol. Genet.* **10,** 1829–45 (2001).
173. Gines, S. *et al.* Specific progressive cAMP reduction implicates energy deficit in presymptomatic Huntington's disease knock-in mice. *Hum. Mol. Genet.* **12,** 497–508 (2003).
174. Teigelkamp, S., Newman, A. J. & Beggs, J. D. Extensive interactions of PRP8 protein with the 5′ and 3′ splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J.* **14,** 2602–12 (1995).
175. Wickramasinghe, V. O. *et al.* Regulation of constitutive and alternative mRNA splicing across the human transcriptome by PRPF8 is determined by 5′ splice site strength. *Genome Biol.* **16,** 201 (2015).

176. Pestova, T. V. *et al.* Molecular mechanisms of translation initiation in eukaryotes. *Proc. Natl. Acad. Sci.* **98,** 7029–7036 (2001).
177. Holz, M. K., Ballif, B. A., Gygi, S. P. & Blenis, J. mTOR and S6K1 Mediate Assembly of the Translation Preinitiation Complex through Dynamic Protein Interchange and Ordered Phosphorylation Events. *Cell* **123,** 569–580 (2005).
178. Gingras, A. C., Raught, B. & Sonenberg, N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* **68,** 913–63 (1999).
179. des Georges, A. *et al.* Structure of mammalian eIF3 in the context of the 43S preinitiation complex. *Nature* **525,** 491–5 (2015).
180. Suchanek, M., Radzikowska, A. & Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat. Methods* **2,** 261–268 (2005).
181. Majzoub, K. *et al.* RACK1 controls IRES-mediated translation of viruses. *Cell* **159,** 1086–1095 (2014).
182. Anderson, B. R., Chopra, P., Suhl, J. A., Warren, S. T. & Bassell, G. J. Identification of consensus binding sites clarifies FMRP binding determinants. *Nucleic Acids Res.* **44,** 6649–6659 (2016).
183. Darnell, J. C. *et al.* FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. *Cell* **146,** 247–261 (2011).
184. Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science (80-.).* **347,** 1260419–1260419 (2015).
185. Lizio, M. *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* **16,** 22 (2015).
186. Carithers, L. J. *et al.* A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv. Biobank.* **13,** 311–319 (2015).
187. Quarrell, O., O'Donovan, K. L., Bandmann, O. & Strong, M. The Prevalence of Juvenile Huntington's Disease: A Review of the Literature and Meta-Analysis. *PLoS Curr.* **4,** e4f8606b742ef3 (2012).
188. Sugino, K. *et al.* Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nat. Neurosci.* **9,** 99–107 (2006).
189. Molyneaux, B. J., Arlotta, P., Menezes, J. R. L. & Macklis, J. D. Neuronal subtype specification in the cerebral cortex. *Nat. Rev. Neurosci.* **8,** 427–437 (2007).
190. Zeisel, a. *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science (80-.).* **347,** 1138–42 (2015).
191. Bradford, J. *et al.* Expression of mutant huntingtin in mouse brain astrocytes causes age-dependent neurological symptoms. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 22480–5 (2009).
192. Faideau, M. *et al.* In vivo expression of polyglutamine-expanded huntingtin by mouse striatal astrocytes impairs glutamate transport: a correlation with Huntington's disease subjects. *Hum. Mol. Genet.* **19,** 3053–67 (2010).
193. Cepeda, C. *et al.* NMDA receptor function in mouse models of Huntington disease. *J. Neurosci. Res.* **66,** 525–539 (2001).

194. López-Gómez, C. *et al.* TRAIL/TRAIL Receptor System and Susceptibility to Multiple Sclerosis. *PLoS One* **6,** e21766 (2011).
195. Cantarella, G. *et al.* Neutralization of TNFSF10 ameliorates functional outcome in a murine model of Alzheimer's disease. *Brain* **138,** 203–216 (2015).
196. Wu, Y.-Y. *et al.* Alterations of the Neuroinflammatory Markers IL-6 and TRAIL in Alzheimer's Disease. *Dement. Geriatr. Cogn. Dis. Extra* **5,** 424–434 (2015).
197. Collison, A. M. *et al.* TNF-related apoptosis-inducing ligand (TRAIL) regulates midline-1, thymic stromal lymphopoietin, inflammation, and remodeling in experimental eosinophilic esophagitis. *J. Allergy Clin. Immunol.* **136,** 971–982 (2015).
198. Ciesiolka, A., Jazurek, M., Drazkowska, K. & Krzyzosiak, W. J. Structural Characteristics of Simple RNA Repeats Associated with Disease and their Deleterious Protein Interactions. *Front. Cell. Neurosci.* **11,** 1–19 (2017).
199. Schilling, J., Griesche, N. & Krauß, S. Mechanisms of RNA-Induced Toxicity in Diseases Characterised by CAG Repeat Expansions. *eLS* 1–8 (2016). doi:10.1002/9780470015902.a0026464
200. Mohan, A., Goodwin, M. & Swanson, M. S. RNA-protein interactions in unstable microsatellite diseases. *Brain Res.* **1584C,** 3–14 (2014).
201. Mykowska, A., Sobczak, K., Wojciechowska, M., Kozlowski, P. & Krzyzosiak, W. J. CAG repeats mimic CUG repeats in the misregulation of alternative splicing. *Nucleic Acids Res.* **39,** 8938–51 (2011).
202. Grainger, R. J. Prp8 protein: At the heart of the spliceosome. *Rna* **11,** 533–557 (2005).
203. Reyes, J. L., Gustafson, E. H., Luo, H. R., Moore, M. J. & Konarska, M. M. The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5′ splice site. *RNA* **5,** 167–79 (1999).
204. Chiara, M. D. *et al.* Identification of proteins that interact with exon sequences, splice sites, and the branchpoint sequence during each stage of spliceosome assembly. *Mol. Cell. Biol.* **16,** 3317–26 (1996).
205. Urbanek, M. O. & Krzyzosiak, W. J. RNA FISH for detecting expanded repeats in human diseases. *Methods* **98,** 115–123 (2016).
206. Strack, R. L., Disney, M. D. & Jaffrey, S. R. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat–containing RNA. *Nat. Methods* **10,** 1219–1224 (2013).
207. Montminy, M. R., Sevarino, K. A., Wagner, J. A., Mandel, G. & Goodman, R. H. Identification of a cyclic-AMP-responsive element within the rat somatostatin gene. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 6682–6 (1986).
208. Lonze, B. E. & Ginty, D. D. Function and Regulation of CREB Family Transcription Factors in the Nervous System. *Neuron* **35,** 605–623 (2002).
209. Bonni, A., Ginty, D. D., Dudek, H. & Greenberg, M. E. Serine 133-Phosphorylated CREB Induces Transcription via a Cooperative Mechanism That May Confer Specificity to Neurotrophin Signals. *Mol. Cell. Neurosci.* **6,** 168–183 (1995).
210. Deisseroth, K., Bito, H. & Tsien, R. W. Signaling from synapse to nucleus: postsynaptic CREB phosphorylation during multiple forms of hippocampal synaptic plasticity. *Neuron* **16,** 89–101

(1996).

211. Gonzalez, G. A. & Montminy, M. R. Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. *Cell* **59,** 675–80 (1989).
212. Deisseroth, K. & Tsien, R. W. Dynamic multiphosphorylation passwords for activity-dependent gene expression. *Neuron* **34,** 179–82 (2002).
213. Walker, W. H., Girardet, C. & Habener, J. F. Alternative exon splicing controls a translational switch from activator to repressor isoforms of transcription factor CREB during spermatogenesis. *J. Biol. Chem.* **271,** 20145–1050 (1996).
214. Riccio, A., Ahn, S., Davenport, C. M., Blendy, J. A. & Ginty, D. D. Mediation by a CREB family transcription factor of NGF-dependent survival of sympathetic neurons. *Science* **286,** 2358–61 (1999).
215. Bonni, A. *et al.* Cell survival promoted by the Ras-MAPK signaling pathway by transcription-dependent and -independent mechanisms. *Science* **286,** 1358–62 (1999).
216. Choi, Y.-S. *et al.* CREB is a key regulator of striatal vulnerability in chemical and genetic models of Huntington's disease. *Neurobiol. Dis.* **36,** 259–268 (2009).
217. McCourt, A. C. *et al.* Analysis of White Adipose Tissue Gene Expression Reveals CREB1 Pathway Altered in Huntington's Disease. *J. Huntingtons. Dis.* **4,** 371–382 (2015).
218. Tao, X., Finkbeiner, S., Arnold, D. B., Shaywitz, A. J. & Greenberg, M. E. Ca2+ influx regulates BDNF transcription by a CREB family transcription factor-dependent mechanism. *Neuron* **20,** 709–26 (1998).
219. Consortium, U. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45,** D158–D169 (2017).
220. Brown, V. *et al.* Microarray Identification of FMRP-Associated Brain mRNAs and Altered mRNA Translational Profiles in Fragile X Syndrome. *Cell* **107,** 477–487 (2001).
221. Ascano, M. *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492,** 382–386 (2012).
222. Wang, H. *et al.* Dynamic association of the fragile X mental retardation protein as a messenger ribonucleoprotein between microtubules and polyribosomes. *Mol. Biol. Cell* **19,** 105–14 (2008).
223. Narayanan, U. *et al.* FMRP phosphorylation reveals an immediate-early signaling pathway triggered by group I mGluR and mediated by PP2A. *J. Neurosci.* (2007). doi:10.1523/JNEUROSCI.2969-07.2007
224. Suhl, J. A., Chopra, P., Anderson, B. R., Bassell, G. J. & Warren, S. T. Analysis of FMRP mRNA target datasets reveals highly associated mRNAs mediated by G-quadruplex structures formed via clustered WGGA sequences. *Hum. Mol. Genet.* **23,** 5479–5491 (2014).
225. Langfelder, P. *et al.* Integrated genomics and proteomics define huntingtin CAG length–dependent networks in mice. *Nat. Neurosci.* **19,** 623–633 (2016).
226. Martin, J. P. & Bell, J. A pedigree of mental defect showing sex-linkage. *J. Neurol. Psychiatry* **6,** 154–7 (1943).
227. Ceman, S. *et al.* Phosphorylation influences the translation state of FMRP-associated polyribosomes. *Hum. Mol. Genet.* **12,** 3295–305 (2003).

228. Nalavadi, V. C., Muddashetty, R. S., Gross, C. & Bassell, G. J. Dephosphorylation-Induced Ubiquitination and Degradation of FMRP in Dendrites: A Role in Immediate Early mGluR-Stimulated Translation. *J. Neurosci.* (2012). doi:10.1523/JNEUROSCI.5057-11.2012

229. Narayanan, U. *et al.* S6K1 phosphorylates and regulates fragile X mental retardation protein (FMRP) with the neuronal protein synthesis-dependent mammalian target of rapamycin (mTOR) signaling cascade. *J. Biol. Chem.* **283,** 18478–82 (2008).

230. Coffee, R. L. *et al.* In vivo neuronal function of the fragile X mental retardation protein is regulated by phosphorylation. *Hum. Mol. Genet.* **21,** 900–915 (2012).

231. Bartley, C. M. *et al.* FMRP S499 Is Phosphorylated Independent of mTORC1-S6K1 Activity. *PLoS One* **9,** e96956 (2014).

232. Bartley, C. M. *et al.* Mammalian FMRP S499 Is Phosphorylated by CK2 and Promotes Secondary Phosphorylation of FMRP. *eNeuro* **3,** (2016).

Acknowledgements

Ich möchte mich zu allererst bei Sybille bedanken: dafür, dass ich meine Doktorarbeit in ihrem Labor anfertigen durfte! Vor allem auch für die Unterstützung im Laboralltag und die vielen Konferenzen und externen Laboraufenthalte, die sie mir ermöglicht hat. Ich habe viel gelernt und es hat viel Spaß gemacht! Dank gilt auch Ina, die sich für mein Projekt begeistern lies und die Arbeit als meine erste Gutachterin auf sich genommen hat bzw. Prof. Höhfeld als mein Zweitgutachter.

Glücklicherweise hatte ich in den letzten Jahren tolle Kollegen: vor allem Frank und Stephie, Felix, Katrin, Manuel, Nora, und Lars! Meiner Mama und Ralf gilt besonderer Dank weil sie mich immer und bei allem unterstützen und obwohl sie nicht viel Ahnung von meiner Arbeit haben, es auch noch toll finden! Simon hat wie immer toll gekocht und mich auch im Gurkenzustand ertragen.
Vielen Dank für den emotionalen support!

Zu guter Letzt, ein riesen Dankeschön an alle (last minute) Korrekturleser: Saskia, Frank, Willeke und Nina!

Appendix

Table 6 | List of primers
f, forward; r, reverse; h, human; m, mouse.

Name	Sequence
hCREB exon_r	TTCGCTTTTGGGAATCAGTT
hCREB intron_r	GTTCTCTCCAAATCTAGGACC
hCREB_f	CAATGGGCAGACAGTTCAAG
hGAPDH_f	CCACCCATGGCAAATTCC
hGAPDH_r	TGGGATTTCCATTGATGACAAG
hMID1_f	CTGCCAGGTCTGGTGTCATG
hMID1_r	AATCAGGCTTAGGGCCCTTCT
hMID2_f	GGGAAAGGACTTACAGGCCC
hMID2_r	GTTTCTTGTTGGGGTGCGTG
hRPL22_f	TGACATCCGAGGTGCCTTTC
hRPL22_r	GTTAGCAACTACGCGCAACC
mGAPDH_f	GCACAGTCAAGGCCGAGAAT
mGAPDH_r	GCCTTCTCCATGGTGGTGAA
mMID1_f	CAAAGTGGCACCAAGTATATCTTCA
mMID1_r	TCCGGGCTCGCTGCTA
T7 HTTex1_f	CCAAGCTTCTAATACGACTCACTATAGGGAGAATGGCGACCCTGGAAAAGCT
Linker T7 Httex1_f	GAATTAATACGACTCACTATAGGGAGAATAGATAGTATGGCGACCCTGGAAAAGCT
HTTex1_r	GGTCGGTGCAGCGGCTCCTCAGC
Biotin DNA Linker	ACTATCTATTCTCCC (BtnTg)

Table 7 | Statistical analysis of proteins identified in MID1 immunoprecipitation.
Log2 ratio and p-values were calculated using measured protein intensities, i.e. eXtracted Ion Current (XIC) of all isotopic clusters associated with the identified amino acid sequence. Log2 ratio was calculated from the intensity sum of samples/ controls. p-values are the result of a two-sided t-test, samples versus control. In cases where intensities had been measured in 2 (out of 3) replicates, the third intensity value was added through imputation. If no intensity could be measured in all 3 replicates, the intensities were set from 0 to 1 in order to still be able to calculate a ratio (same applies to cases where only 1 intensity could be measured). Statistical analysis was performed by A. Dagane (MDC Berlin).

Protein name	Gene ID	log2 ratio	p-value
ATP-binding cassette sub-family D member 3	ABCD3	2.75E+01	2.04E-03
ATP-binding cassette sub-family F member 2	ABCF2	2.88E+01	5.96E-04
Apoptotic chromatin condensation inducer in the nucleus	ACIN1	2.72E+01	1.39E-02
Aldehyde dehydrogenase X, mitochondrial	ALDH1B1	2.62E+01	3.61E-04
Mitochondrial 10-formyltetrahydrofolate dehydrogenase	ALDH1L2	2.66E+01	1.97E-02
THO complex subunit 4	ALYREF	2.99E+01	7.55E-04
Serine/threonine-protein phosphatase 6 reg. ankyrin repeat subunit A	ANKRD28	2.63E+01	1.85E-02
Coatomer subunit delta	ARCN1	2.65E+01	2.36E-02
Activating signal cointegrator 1 complex subunit 3	ASCC3	2.59E+01	7.96E-03
ATPase family AAA domain-containing protein 3A	ATAD3A	1.77E+00	3.98E-02
Sodium/potassium-transporting ATPase subunit alpha-1	ATP1A1	2.72E+01	4.39E-04
Ribosome biogenesis protein BMS1 homolog	BMS1	2.90E+01	8.92E-03
Ribosome biogenesis protein BRX1 homolog	BRIX1	2.68E+01	1.41E-02
Caprin-1	CAPRIN1	2.80E+01	5.88E-03
Coiled-coil domain-containing protein 124	CCDC124	2.94E+01	1.91E-04
T-complex protein 1 subunit gamma	CCT3	2.63E+01	4.86E-02
T-complex protein 1 subunit epsilon	CCT5	2.64E+01	1.42E-03
Cell division cycle 5-like protein	CDC5L	3.04E+01	3.51E-03
Centrosomal protein of 170 kDa	CEP170	2.71E+01	2.53E-03
Chromatin target of PRMT1 protein	CHTOP	2.81E+01	3.45E-02
CLIP-associating protein 2	CLASP2	2.74E+01	4.83E-03
Methylosome subunit pICln	CLNS1A	2.97E+01	2.16E-03
Coatomer subunit gamma-2	COPG2	2.59E+01	2.98E-04
Coronin-1C	CORO1C	3.03E+01	2.00E-02
Cleavage and polyadenylation specificity factor subunit 6	CPSF6	2.93E+01	4.13E-03
Cleavage and polyadenylation specificity factor subunit 7	CPSF7	2.73E+01	2.20E-03
Probable ATP-dependent RNA helicase DDX17	DDX17	3.22E+01	1.28E-02
Probable ATP-dependent RNA helicase DDX20	DDX20	2.73E+01	2.81E-02

Nucleolar RNA helicase 2	DDX21	3.01E+01	8.42E-03
Probable ATP-dependent RNA helicase DDX23	DDX23	2.93E+01	4.58E-03
ATP-dependent RNA helicase DDX3X	DDX3X	2.66E+01	5.42E-04
Probable ATP-dependent RNA helicase DDX41	DDX41	2.73E+01	4.88E-03
Probable ATP-dependent RNA helicase DDX5	DDX5	2.97E+01	1.38E-02
ATP-dependent RNA helicase DDX50	DDX50	2.65E+01	5.75E-03
Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	DHX15	2.76E+01	6.67E-03
Putative ATP-dependent RNA helicase DHX30	DHX30	2.99E+01	1.03E-02
ATP-dependent RNA helicase A	DHX9	7.63E+00	4.50E-03
Elongation factor 2	EEF2	4.68E+00	8.14E-03
116 kDa U5 small nuclear ribonucleoprotein component	EFTUD2	3.04E+01	6.54E-03
Eukaryotic translation initiation factor 3 subunit A	EIF3A	3.34E+01	2.89E-04
Eukaryotic translation initiation factor 3 subunit B	EIF3B	3.19E+01	6.22E-03
Eukaryotic translation initiation factor 3 subunit C	EIF3C	3.23E+01	1.43E-03
Eukaryotic translation initiation factor 3 subunit D	EIF3D	3.01E+01	1.72E-03
Eukaryotic translation initiation factor 3 subunit E	EIF3E	3.15E+01	2.60E-03
Eukaryotic translation initiation factor 3 subunit F	EIF3F	3.08E+01	9.00E-03
Eukaryotic translation initiation factor 3 subunit G	EIF3G	2.99E+01	1.64E-04
Eukaryotic translation initiation factor 3 subunit I	EIF3I	3.07E+01	4.95E-03
Eukaryotic translation initiation factor 3 subunit J	EIF3J	2.75E+01	2.97E-04
Eukaryotic translation initiation factor 3 subunit K	EIF3K	2.75E+01	4.35E-02
Eukaryotic translation initiation factor 3 subunit L	EIF3L	3.20E+01	1.62E-03
Eukaryotic translation initiation factor 3 subunit M	EIF3M	3.06E+01	2.52E-02
Eukaryotic translation initiation factor 3 subunit H	EIF3S3	2.98E+01	4.34E-03
Eukaryotic initiation factor 4A-I	EIF4A1	2.82E+01	1.28E-02
Eukaryotic initiation factor 4A-III	EIF4A3	2.66E+01	2.37E-02
Eukaryotic translation initiation factor 4B	EIF4B	3.07E+01	1.05E-02
Eukaryotic translation initiation factor 6	EIF6	2.71E+01	1.74E-02
Emerin	EMD	2.73E+01	1.99E-02
Erlin-2	ERLIN2	2.81E+01	7.54E-03
Exosome component 10	EXOSC10	2.76E+01	3.45E-03
Exosome complex component RRP45	EXOSC9	2.59E+01	2.53E-02
Constitutive coactivator of PPAR-gamma-like protein 1	FAM120A	2.74E+01	1.36E-02
Phenylalanine--tRNA ligase alpha subunit	FARSA	2.78E+01	7.54E-03
Phenylalanine--tRNA ligase beta subunit	FARSB	2.78E+01	1.17E-02
40S ribosomal protein S30	FAU	2.85E+01	2.31E-03
Protein furry homolog-like	FRYL	3.02E+01	2.54E-02
Gem-associated protein 4	GEMIN4	2.71E+01	2.53E-03
Guanine nucleotide-binding protein subunit beta-2-like 1	GNB2L1	3.25E+01	3.47E-04
Nucleolar GTP-binding protein 2	GNL2	2.75E+01	7.11E-03
Guanine nucleotide-binding protein-like 3	GNL3	2.69E+01	4.03E-04
Golgin subfamily A member 3	GOLGA3	2.93E+01	3.96E-03
General transcription factor 3C polypeptide 2	GTF3C2	2.75E+01	3.97E-03
General transcription factor 3C polypeptide 3	GTF3C3	2.57E+01	3.56E-02
General transcription factor 3C polypeptide 4	GTF3C4	2.70E+01	2.46E-03
Nucleolar GTP-binding protein 1	GTPBP4	2.87E+01	1.26E-02
Histone H2B	HIST1H2BN	2.99E+01	1.62E-02
Heterogeneous nuclear ribonucleoproteins C1/C2	HNRNPC	3.08E+01	3.22E-04
Heterogeneous nuclear ribonucleoprotein D0	HNRNPD	2.60E+01	2.60E-02
Heterogeneous nuclear ribonucleoprotein F	HNRNPF	2.74E+01	4.67E-03
Heterogeneous nuclear ribonucleoprotein K	HNRNPK	2.96E+01	9.43E-04
Heterogeneous nuclear ribonucleoprotein M	HNRNPM	6.56E+00	1.98E-02
Heterogeneous nuclear ribonucleoprotein R	HNRNPR	2.99E+01	4.29E-03
Heterogeneous nuclear ribonucleoprotein U	HNRNPU	3.23E+01	2.34E-03
Isoleucine--tRNA ligase, cytoplasmic	IARS	2.72E+01	3.12E-03
Insulin-like growth factor 2 mRNA-binding protein 1	IGF2BP1	3.18E+01	1.09E-03
Insulin-like growth factor 2 mRNA-binding protein 3	IGF2BP3	2.81E+01	1.22E-02
Interleukin enhancer-binding factor 2	ILF2	3.10E+01	1.08E-02
Interleukin enhancer-binding factor 3	ILF3	3.27E+01	2.99E-03
Importin-8	IPO8	2.74E+01	1.30E-02
Insulin receptor substrate 4	IRS4	1.33E+00	3.38E-03
Influenza virus NS1A-binding protein	IVNS1ABP	3.33E+01	3.05E-03
Tyrosine-protein kinase JAK1	JAK1	2.88E+01	4.64E-03
BTB/POZ domain-containing protein KCTD17	KCTD17	2.96E+01	1.23E-02
BTB/POZ domain-containing protein KCTD5	KCTD5	2.96E+01	3.67E-04
Kinesin-like protein KIF11	KIF11	1.48E+00	1.25E-03
La-related protein 1	LARP1	3.17E+01	9.37E-04
La-related protein 4	LARP4	2.86E+01	7.22E-03
La-related protein 4B	LARP4B	2.65E+01	8.21E-03
LIM domain and actin-binding protein 1	LIMA1	2.98E+01	7.34E-03
Leucine-rich PPR motif-containing protein, mitochondrial	LRPPRC	2.63E+01	8.82E-03
Putative RNA-binding protein Luc7-like 2	LUC7L2	3.00E+01	3.39E-03
Luc7-like protein 3	LUC7L3	2.81E+01	5.20E-03
Microtubule-associated protein 1B	MAP1B	3.01E+01	3.76E-03
Serine/threonine-protein kinase MARK2	MARK2	2.60E+01	3.76E-03
Methionine--tRNA ligase, cytoplasmic	MARS	2.62E+01	1.62E-03
Matrin-3	MATR3	2.87E+01	1.32E-03
DNA replication licensing factor MCM7	MCM7	2.83E+01	8.19E-03
E3 ubiquitin-protein ligase Midline-1	MID1	3.76E+01	3.26E-04
Putative helicase MOV-10	MOV10	2.77E+01	2.49E-02
28S ribosomal protein S17, mitochondrial	MRPS17	2.91E+01	6.96E-03
28S ribosomal protein S22, mitochondrial	MRPS22	2.74E+01	4.69E-03

28S ribosomal protein S25, mitochondrial	MRPS25	2.71E+01	1.14E-02
28S ribosomal protein S27, mitochondrial	MRPS27	2.64E+01	3.48E-02
Protein LYRIC	MTDH	2.72E+01	1.03E-02
Myb-binding protein 1A	MYBBP1A	2.90E+01	1.44E-04
Myosin-10	MYH10	1.13E+00	2.28E-02
Myosin-9	MYH9	2.92E+01	4.78E-03
Unconventional myosin-Ib	MYO1B	2.71E+01	2.88E-02
Nicotinamide phosphoribosyltransferase	NAMPT	2.75E+01	8.55E-03
Nucleosome assembly protein 1-like 1	NAP1L1	2.57E+01	7.18E-03
Nuclear cap-binding protein subunit 1	NCBP1	2.81E+01	4.89E-04
Nucleolin	NCL	2.86E+01	6.74E-03
Nucleolar complex protein 4 homolog	NOC4L	2.80E+01	8.27E-03
Probable 28S rRNA (cytosine(4447)-C(5))-methyltransferase	NOP2	2.62E+01	5.08E-03
Cleavage and polyadenylation specificity factor subunit 5	NUDT21	2.96E+01	6.21E-04
OTU domain-containing protein 4	OTUD4	2.74E+01	1.45E-03
Prolyl 4-hydroxylase subunit alpha-1	P4HA1	2.79E+01	4.44E-03
Proliferation-associated protein 2G4	PA2G4	3.08E+01	7.55E-03
Polyadenylate-binding protein 1	PABPC1	3.23E+01	2.17E-03
Polyadenylate-binding protein 4	PABPC4	3.22E+01	2.78E-03
Programmed cell death protein 4	PDCD4	2.94E+01	8.37E-03
Proline-, glutamic acid- and leucine-rich protein 1	PELP1	2.66E+01	9.09E-03
Serine/threonine-protein phosphatase PGAM5, mitochondrial	PGAM5	2.86E+01	3.52E-03
Protein arginine N-methyltransferase 5	PRMT5	3.54E+01	6.16E-04
Pre-mRNA-processing factor 19	PRPF19	3.02E+01	3.67E-03
U4/U6 small nuclear ribonucleoprotein Prp3	PRPF3	2.70E+01	1.69E-02
U4/U6 small nuclear ribonucleoprotein Prp31	PRPF31	2.97E+01	8.19E-04
U4/U6 small nuclear ribonucleoprotein Prp4	PRPF4	2.64E+01	3.83E-03
Pre-mRNA-processing factor 6	PRPF6	2.98E+01	6.12E-03
Pre-mRNA-processing-splicing factor 8	PRPF8	3.11E+01	3.06E-03
Ribose-phosphate pyrophosphokinase 1	PRPS1	5.58E+00	3.60E-03
Ribose-phosphate pyrophosphokinase 2	PRPS2	2.93E+01	1.56E-02
Phosphoribosyl pyrophosphate synthase-associated protein 1	PRPSAP1	3.05E+01	8.50E-03
Phosphoribosyl pyrophosphate synthase-associated protein 2	PRPSAP2	3.25E+01	2.15E-04
Protein PRRC2A	PRRC2A	2.85E+01	8.43E-03
Protein PRRC2C	PRRC2C	2.97E+01	2.05E-02
26S protease regulatory subunit 4	PSMC1	2.97E+01	5.78E-03
26S protease regulatory subunit 7	PSMC2	3.07E+01	8.06E-04
26S protease regulatory subunit 6A	PSMC3	2.89E+01	5.50E-03
26S protease regulatory subunit 6B	PSMC4	1.19E+00	1.54E-02
26S protease regulatory subunit 8	PSMC5	4.55E+00	3.71E-04
26S protease regulatory subunit 10B	PSMC6	2.84E+01	1.52E-03
26S proteasome non-ATPase regulatory subunit 1	PSMD1	2.97E+01	2.12E-03
26S proteasome non-ATPase regulatory subunit 10	PSMD10	2.85E+01	2.02E-02
26S proteasome non-ATPase regulatory subunit 11	PSMD11	3.02E+01	1.94E-02
26S proteasome non-ATPase regulatory subunit 12	PSMD12	2.90E+01	4.18E-03
26S proteasome non-ATPase regulatory subunit 13	PSMD13	2.95E+01	3.11E-03
26S proteasome non-ATPase regulatory subunit 14	PSMD14	2.80E+01	1.85E-03
26S proteasome non-ATPase regulatory subunit 2	PSMD2	2.15E+00	4.38E-03
26S proteasome non-ATPase regulatory subunit 3	PSMD3	3.04E+01	1.48E-04
26S proteasome non-ATPase regulatory subunit 4	PSMD4	2.79E+01	1.70E-04
26S proteasome non-ATPase regulatory subunit 6	PSMD6	2.96E+01	9.65E-03
26S proteasome non-ATPase regulatory subunit 7	PSMD7	3.75E+00	1.13E-02
26S proteasome non-ATPase regulatory subunit 8	PSMD8	2.75E+01	1.56E-03
Poly(U)-binding-splicing factor PUF60	PUF60	2.75E+01	9.95E-04
Pyrroline-5-carboxylate reductase	PYCR1	2.55E+01	3.49E-03
RNA-binding protein 10	RBM10	3.30E+01	4.88E-04
RNA-binding protein 14	RBM14	2.94E+01	6.57E-03
RNA-binding protein 25	RBM25	2.75E+01	7.71E-03
RNA-binding protein 26	RBM26	2.63E+01	3.79E-02
RNA-binding protein 27	RBM27	2.70E+01	1.20E-02
RNA-binding protein 28	RBM28	2.63E+01	5.02E-03
RNA-binding motif protein, X chromosome	RBMX	3.01E+01	8.90E-03
RNA 3-terminal phosphate cyclase-like protein	RCL1	2.65E+01	1.64E-03
Reticulocalbin-2	RCN2	2.73E+01	2.67E-03
Replication factor C subunit 3	RFC3	2.62E+01	2.34E-02
Telomere-associated protein RIF1	RIF1	3.04E+01	9.16E-03
Serine/threonine-protein kinase RIO1	RIOK1	3.02E+01	3.05E-03
RING finger protein 219	RNF219	2.94E+01	2.50E-03
RNA-binding protein 39	RNPC2	2.86E+01	8.92E-03
60S ribosomal protein L10	RPL10	3.26E+01	7.84E-03
60S ribosomal protein L10a	RPL10A	3.22E+01	2.24E-02
60S ribosomal protein L11	RPL11	3.19E+01	1.04E-03
60S ribosomal protein L12	RPL12	3.18E+01	6.43E-03
60S ribosomal protein L13	RPL13	3.30E+01	4.30E-04
60S ribosomal protein L13a	RPL13A	3.17E+01	2.25E-03
60S ribosomal protein L14	RPL14	3.07E+01	7.88E-04
60S ribosomal protein L15	RPL15	5.67E+00	8.90E-03
60S ribosomal protein L17	RPL17	3.16E+01	4.69E-04
60S ribosomal protein L18	RPL18	3.28E+01	1.68E-03
60S ribosomal protein L18a	RPL18A	3.24E+01	2.10E-03
Ribosomal protein L19	RPL19	3.26E+01	4.36E-03
60S ribosomal protein L21	RPL21	3.11E+01	1.99E-03

60S ribosomal protein L22	RPL22	3.05E+01	1.47E-03
60S ribosomal protein L22-like 1	RPL22L1	2.70E+01	6.60E-03
60S ribosomal protein L23	RPL23	3.09E+01	4.70E-03
60S ribosomal protein L23a	RPL23A	3.24E+01	1.26E-03
60S ribosomal protein L24	RPL24	3.03E+01	1.24E-03
60S ribosomal protein L26	RPL26	3.24E+01	1.20E-03
60S ribosomal protein L27	RPL27	3.18E+01	3.06E-03
60S ribosomal protein L27a	RPL27A	3.06E+01	5.33E-04
60S ribosomal protein L28	RPL28	3.22E+01	6.23E-04
60S ribosomal protein L29	RPL29	3.19E+01	3.76E-04
60S ribosomal protein L3	RPL3	7.15E+00	6.72E-03
60S ribosomal protein L30	RPL30	3.04E+01	2.06E-03
60S ribosomal protein L31	RPL31	3.15E+01	3.12E-04
60S ribosomal protein L32	RPL32	3.18E+01	3.51E-04
60S ribosomal protein L34	RPL34	2.90E+01	4.73E-03
60S ribosomal protein L35	RPL35	3.13E+01	1.43E-02
60S ribosomal protein L35a	RPL35A	3.14E+01	2.01E-02
60S ribosomal protein L36	RPL36	3.02E+01	1.49E-02
60S ribosomal protein L36a	RPL36A	2.97E+01	5.51E-04
60S ribosomal protein L37a	RPL37A	3.07E+01	1.05E-04
60S ribosomal protein L38	RPL38	2.90E+01	3.49E-02
60S ribosomal protein L4	RPL4	3.36E+01	1.82E-03
60S ribosomal protein L5	RPL5	3.26E+01	6.62E-03
60S ribosomal protein L6	RPL6	6.39E+00	9.35E-03
60S ribosomal protein L7	RPL7	3.33E+01	2.21E-03
60S ribosomal protein L7a	RPL7A	8.22E+00	1.74E-03
60S ribosomal protein L8	RPL8	3.33E+01	9.73E-04
60S ribosomal protein L9	RPL9	3.06E+01	1.10E-04
60S acidic ribosomal protein P0	RPLP0	3.23E+01	2.18E-03
60S acidic ribosomal protein P2	RPLP2	2.87E+01	4.32E-03
40S ribosomal protein S10	RPS10	3.21E+01	1.86E-03
40S ribosomal protein S11	RPS11	3.24E+01	1.88E-03
40S ribosomal protein S12	RPS12	3.14E+01	4.05E-04
40S ribosomal protein S13	RPS13	3.22E+01	1.60E-03
40S ribosomal protein S14	RPS14	3.15E+01	4.92E-04
40S ribosomal protein S15	RPS15	3.15E+01	3.33E-02
40S ribosomal protein S15a	RPS15A	3.19E+01	1.70E-03
40S ribosomal protein S16	RPS16	3.24E+01	1.08E-03
40S ribosomal protein S17	RPS17	3.19E+01	1.78E-03
40S ribosomal protein S18	RPS18	8.03E+00	7.90E-09
40S ribosomal protein S19	RPS19	3.22E+01	1.65E-03
40S ribosomal protein S2	RPS2	3.28E+01	1.71E-03
40S ribosomal protein S20	RPS20	3.21E+01	4.89E-04
40S ribosomal protein S21	RPS21	2.88E+01	2.47E-03
40S ribosomal protein S23	RPS23	3.25E+01	1.31E-03
40S ribosomal protein S24	RPS24	3.00E+01	1.10E-04
40S ribosomal protein S25	RPS25	3.14E+01	5.19E-03
40S ribosomal protein S26	RPS26	3.08E+01	2.54E-02
40S ribosomal protein S27	RPS27	3.00E+01	3.65E-03
40S ribosomal protein S3	RPS3	3.28E+01	3.26E-03
40S ribosomal protein S3a	RPS3A	3.30E+01	4.75E-04
40S ribosomal protein S4, X isoform	RPS4X	6.65E+00	2.28E-03
40S ribosomal protein S6	RPS6	3.18E+01	2.28E-03
40S ribosomal protein S7	RPS7	3.28E+01	1.99E-02
40S ribosomal protein S8	RPS8	3.24E+01	3.63E-03
40S ribosomal protein S9	RPS9	3.31E+01	3.94E-03
40S ribosomal protein SA	RPSA	3.32E+01	1.29E-03
Ribosome-binding protein 1	RRBP1	2.84E+01	3.20E-02
RRP12-like protein	RRP12	2.58E+01	1.13E-02
Ribosomal L1 domain-containing protein 1	RSL1D1	2.69E+01	1.56E-02
U4/U6.U5 tri-snRNP-associated protein 1	SART1	2.91E+01	1.18E-04
Splicing factor, arginine/serine-rich 15	SCAF4	2.74E+01	1.68E-02
Protein SDA1 homolog	SDAD1	2.64E+01	1.69E-02
Plasminogen activator inhibitor 1 RNA-binding protein	SERBP1	3.28E+01	2.19E-02
Splicing factor 3B subunit 1	SF3B1	2.81E+01	8.87E-03
Splicing factor 3B subunit 3	SF3B3	2.72E+01	1.14E-02
Superkiller viralicidic activity 2-like 2	SKIV2L2	2.75E+01	8.05E-03
U5 small nuclear ribonucleoprotein 200 kDa helicase	SNRNP200	3.06E+01	5.98E-03
U5 small nuclear ribonucleoprotein 40 kDa protein	SNRNP40	2.72E+01	1.41E-02
Small nuclear ribonucleoprotein Sm D1	SNRPD1	3.02E+01	1.48E-03
Small nuclear ribonucleoprotein Sm D2	SNRPD2	2.92E+01	2.94E-04
Small nuclear ribonucleoprotein Sm D3	SNRPD3	2.95E+01	1.93E-02
Small nuclear ribonucleoprotein-associated proteins B and B	SNRPN	3.06E+01	4.62E-04
Spectrin alpha chain, non-erythrocytic 1	SPTAN1	3.41E+01	9.63E-04
Spectrin beta chain, non-erythrocytic 1	SPTBN1	3.41E+01	9.41E-05
SRSF protein kinase 1	SRPK1	2.94E+01	5.81E-03
SRSF protein kinase 2	SRPK2	2.60E+01	2.88E-02
Serine/arginine repetitive matrix protein 1	SRRM1	2.85E+01	1.46E-02
Serrate RNA effector molecule homolog	SRRT	2.57E+01	2.04E-03
Serine/arginine-rich splicing factor 1	SRSF1	2.74E+01	1.86E-04
Serine/arginine-rich splicing factor 2	SRSF2	2.69E+01	4.59E-02
Serine/arginine-rich splicing factor 3	SRSF3	2.87E+01	1.42E-03

Double-stranded RNA-binding protein Staufen homolog 1	STAU1	2.88E+01	1.18E-02
Serine/threonine-protein kinase 38	STK38	2.78E+01	2.59E-02
SUN domain-containing protein 2	SUN2	2.97E+01	6.35E-04
Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	2.82E+01	1.08E-02
Very-long-chain enoyl-CoA reductase	TECR	2.70E+01	4.59E-03
Testis-expressed sequence 10 protein	TEX10	2.64E+01	1.64E-03
THO complex subunit 2	THOC2	2.53E+01	1.37E-02
Tight junction protein ZO-2	TJP2	2.65E+01	6.69E-03
Transmembrane protein 33	TMEM33	2.74E+01	1.11E-02
Tropomodulin-3	TMOD3	2.63E+01	4.62E-03
TRMT1-like protein	TRMT1L	2.82E+01	9.34E-03
Tubulin beta-3 chain	TUBB3	2.57E+01	5.04E-04
Tubulin beta-4A chain	TUBB4A	2.61E+01	2.90E-02
Splicing factor U2AF 35 kDa subunit	U2AF1	2.90E+01	7.02E-03
Splicing factor U2AF 65 kDa subunit	U2AF2	2.98E+01	4.30E-03
U2 snRNP-associated SURP motif-containing protein	U2SURP	2.78E+01	1.78E-05
E3 ubiquitin-protein ligase UBR5	UBR5	2.78E+01	2.75E-02
U4/U6.U5 tri-snRNP-associated protein 2	USP39	2.88E+01	2.14E-02
Transitional endoplasmic reticulum ATPase	VCP	2.80E+01	7.61E-05
Vimentin	VIM	1.35E+00	2.08E-02
Methylosome protein 50	WDR77	3.30E+01	2.27E-02
Exportin-T	XPOT	2.64E+01	7.27E-03
Nuclease-sensitive element-binding protein 1	YBX1	3.07E+01	3.00E-03
YTH domain-containing protein 1	YTHDC1	2.78E+01	8.82E-03
YTH domain-containing family protein 2	YTHDF2	1.60E+00	3.42E-02
Zinc finger CCCH domain-containing protein 18	ZC3H18	2.79E+01	1.14E-02
Zinc finger CCCH-type antiviral protein 1	ZC3HAV1	2.79E+01	1.27E-02
Zinc finger protein 622	ZNF622	2.75E+01	1.14E-02

Table 8 | Statistical analysis of proteins indentified in HTT RNA pulldown.

Log2 ratio and p-values were calculated using measured protein intensities, i.e. eXtracted Ion Current (XIC) of all isotopic clusters associated with the identified amino acid sequence. Log2 ratio was calculated from the intensity sum of samples/ controls. p-values were calculated using the programming language 'R'. Log2 ratio is the mean ratio of RNA pulldown/ control. If the protein was absent in the control, the intensity was set from 0 to 1 to be able to calculate a ratio. Statistical analysis was performed by A. Dagane (MDC Berlin). The first column indicates proteins specifically binding mutant HTT exon 1 RNA based on the statistical analysis conducted by I. Atanassov (MPI, Cologne).

			18CAG		40CAG		70CAG	
HTT	Protein name	Gene. names	log2 ratio	p-value	log2 ratio	p-value	log2 ratio	p-value
+	Bcl-2-associated transcription factor 1	BCLAF1	#N/A	#N/A	2.76E+01	1.05E-02	2.92E+01	5.32E-03
+	Probable ATP-dependent RNA helicase DDX46	DDX46	#N/A	#N/A	2.77E+01	6.08E-02	2.95E+01	4.36E-02
+	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	DHX15	4.27E-01	4.40E-02	1.49E+00	3.16E-03	2.26E+00	9.26E-04
+	Probable ATP-dependent RNA helicase DHX36	DHX36	#N/A	#N/A	2.68E+01	2.58E-03	2.64E+01	9.41E-04
+	116 kDa U5 small nuclear ribonucleoprotein component	EFTUD2	1.64E+01	8.46E-01	1.80E+01	6.60E-03	1.90E+01	3.53E-03
+	Eukaryotic initiation factor 4A-III	EIF4A3	7.73E-01	5.04E-02	1.29E+00	5.26E-03	2.09E+00	1.26E-02
+	Heterogeneous nuclear ribonucleoproteins C1/C2	HNRNPC	8.22E-01	2.01E-02	1.63E+00	2.03E-03	2.24E+00	9.75E-04
+	Heterogeneous nuclear ribonucleoprotein U-like protein 1	HNRNPUL1	1.01E+01	4.76E-02	1.03E+01	4.06E-03	1.00E+01	1.35E-01
+	Pinin	PNN	#N/A	#N/A	2.78E+01	8.02E-02	2.91E+01	9.62E-03
+	Pre-mRNA-processing factor 19	PRPF19	2.28E+01	5.60E-01	2.33E+01	1.50E-03	2.40E+01	6.02E-05
+	U4/U6 small nuclear ribonucleoprotein Prp31	PRPF31	#N/A	#N/A	2.59E+01	1.65E-02	2.60E+01	2.38E-05
+	Pre-mRNA-processing factor 40 homolog A	PRPF40A	#N/A	#N/A	2.85E+01	3.58E-04	2.97E+01	2.08E-04
+	Pre-mRNA-processing-splicing factor 8	PRPF8	3.63E-01	2.11E-01	2.53E+00	1.87E-04	3.92E+00	7.91E-03
+	Poly(U)-binding-splicing factor PUF60	PUF60	#N/A	#N/A	2.63E+01	1.06E-04	2.69E+01	2.34E-03
+	RNA-binding protein Raly	RALY	#N/A	#N/A	2.80E+01	1.50E-02	2.86E+01	7.49E-03
+	RNA-binding protein 25	RBM25	#N/A	#N/A	2.80E+01	6.80E-04	2.86E+01	2.19E-04
+	RNA-binding protein 39	RBM39	2.67E+01	1.58E-03	2.79E+01	2.61E-03	2.98E+01	3.12E-03
+	40S ribosomal protein S27	RPS27	#N/A	#N/A	2.76E+01	3.74E-03	2.74E+01	1.50E-04
+	Splicing factor 3A subunit 1	SF3A1	#N/A	#N/A	2.73E+01	1.25E-04	2.83E+01	2.47E-04
+	Splicing factor 3A subunit 3	SF3A3	#N/A	#N/A	2.58E+01	3.75E-04	2.69E+01	9.21E-04
+	Splicing factor 3B subunit 1	SF3B1	2.81E+01	8.02E-02	2.90E+01	8.00E-04	3.00E+01	1.69E-03
+	Splicing factor 3B subunit 2	SF3B2	#N/A	#N/A	2.80E+01	6.73E-03	2.87E+01	3.61E-03
+	Splicing factor 3B subunit 3	SF3B3	2.28E+01	3.93E-02	2.40E+01	1.13E-03	2.49E+01	2.43E-04
+	Splicing factor 3B subunit 4	SF3B4	#N/A	#N/A	#N/A	#N/A	2.70E+01	7.80E-04
+	U5 small nuclear ribonucleoprotein 200 kDa helicase	SNRNP200	4.06E-01	1.14E-01	1.64E+00	2.48E-03	2.52E+00	8.24E-04
+	U5 small nuclear ribonucleoprotein 40 kDa protein	SNRNP40	#N/A	#N/A	#N/A	#N/A	2.77E+01	4.19E-04
+	Small nuclear ribonucleoprotein Sm D1	SNRPD1	#N/A	#N/A	2.81E+01	7.23E-02	2.87E+01	2.92E-04
+	Protein SON	SON	#N/A	#N/A	2.90E+01	1.21E-02	3.10E+01	7.88E-03
+	Serine/arginine repetitive matrix protein 2	SRRM2	#N/A	#N/A	2.85E+01	4.95E-02	3.02E+01	1.75E-02
+	Serine/arginine-rich splicing factor 1	SRSF1	2.87E+01	4.73E-02	2.90E+01	4.86E-03	3.06E+01	1.01E-03
+	Serine/arginine-rich splicing factor 3	SRSF3	#N/A	#N/A	2.92E+01	1.76E-03	3.01E+01	3.95E-04
+	Serine/arginine-rich splicing factor 6	SRSF6	#N/A	#N/A	2.99E+01	3.74E-01	3.07E+01	2.16E-03
+	Transformer-2 protein homolog beta	TRA2B	1.10E+01	3.75E-01	1.17E+01	5.44E-03	1.26E+01	1.29E-04
+	Splicing factor U2AF 65 kDa subunit	U2AF2	1.01E+00	2.69E-02	1.86E+00	1.10E-02	2.37E+00	1.43E-02
+	U2 snRNP-associated SURP motif-containing protein	U2SURP	#N/A	#N/A	2.59E+01	1.11E-03	2.64E+01	1.13E-03
+	Zinc finger CCCH-type antiviral protein 1	ZC3HAV1	#N/A	#N/A	2.71E+01	3.94E-04	2.71E+01	8.88E-02
	ATP-binding cassette sub-family F member 2	ABCF2	#N/A	#N/A	#N/A	#N/A	2.59E+01	1.78E-01
	Apoptotic chromatin condensation inducer in the nucleus	ACIN1	#N/A	#N/A	2.84E+01	2.04E-01	2.89E+01	1.35E-04

Cytosolic acyl coenzyme A thioester hydrolase	ACOT7	2.59E+01	1.85E-02	2.57E+01	5.60E-04	2.58E+01	1.80E-04
Double-stranded RNA-specific adenosine deaminase	ADAR	2.15E+01	2.26E-02	2.14E+01	1.84E-02	2.12E+01	6.79E-03
Activator of 90 kDa heat shock protein ATPase homolog 1	AHSA1	2.98E+01	1.06E-01	2.98E+01	1.96E-01	2.96E+01	1.78E-01
A-kinase anchor protein 8	AKAP8	#N/A	#N/A	#N/A	#N/A	2.32E+01	8.70E-02
Delta-1-pyrroline-5-carboxylate synthase	ALDH18A1	#N/A	#N/A	2.57E+01	2.01E-01	#N/A	#N/A
THO complex subunit 4	ALYREF	1.71E+01	4.58E-02	1.71E+01	4.35E-02	1.66E+01	2.10E-01
AP-2 complex subunit alpha-1	AP2A1	#N/A	#N/A	2.44E+01	6.93E-06	#N/A	#N/A
AP-3 complex subunit delta-1	AP3D1	#N/A	#N/A	#N/A	#N/A	2.64E+01	2.24E-02
Intron-binding protein aquarius	AQR	#N/A	#N/A	#N/A	#N/A	2.59E+01	1.67E-02
ADP-ribosylation factor 5	ARF3	#N/A	#N/A	#N/A	#N/A	2.57E+01	1.89E-01
ATPase family AAA domain-containing protein 3A	ATAD3A	#N/A	#N/A	2.76E+01	1.78E-01	#N/A	#N/A
ATP synthase subunit b, mitochondrial	ATP5F1	#N/A	#N/A	2.31E+01	2.84E-02	#N/A	#N/A
Pre-mRNA-splicing factor SPF27	BCAS2	#N/A	#N/A	#N/A	#N/A	2.52E+01	7.12E-02
UPF0568 protein C14orf166	C14orf166	#N/A	#N/A	2.53E+01	7.59E-02	#N/A	#N/A
UPF0468 protein C16orf80	C16orf80	#N/A	#N/A	2.60E+01	1.05E-02	2.59E+01	7.00E-04
Complement component 1 Q subcomponent-binding protein, mitochondrial	C1QBP	#N/A	#N/A	#N/A	#N/A	2.57E+01	7.65E-02
tRNA-splicing ligase RtcB homolog	C22orf28	2.24E+01	1.10E-03	2.25E+01	1.48E-03	2.23E+01	2.04E-03
Caprin-1	CAPRIN1	2.71E-01	4.71E-01	1.14E+00	6.98E-02	2.64E-01	5.40E-01
Cell division cycle 5-like protein	CDC5L	2.12E+01	1.30E-01	2.21E+01	1.33E-03	2.24E+01	1.80E-03
Coiled-coil-helix-coiled-coil-helix domain-containing protein 3, mitochondrial	CHCHD3	2.13E+01	5.69E-02	#N/A	3.74E-01	2.12E+01	5.50E-01
Chromodomain-helicase-DNA-binding protein 4	CHD4	2.48E+01	9.91E-02	#N/A	#N/A	#N/A	#N/A
Cytoskeleton-associated protein 4	CKAP4	5.91E+00	9.62E-02	5.70E+00	2.61E-01	5.75E+00	2.70E-01
Ciliary neurotrophic factor receptor subunit alpha	CNTFR	#N/A	#N/A	#N/A	#N/A	2.59E+01	2.14E-01
Cleavage and polyadenylation specificity factor subunit 1	CPSF1	#N/A	#N/A	2.56E+01	2.42E-03	#N/A	#N/A
Cleavage and polyadenylation specificity factor subunit 6	CPSF6	#N/A	2.70E-03	8.98E-01	3.23E-01	1.18E+00	1.94E-01
Cleavage and polyadenylation specificity factor subunit 7	CPSF7	2.37E+01	7.89E-02	2.38E+01	7.10E-02	2.37E+01	2.45E-03
Cellular retinoic acid-binding protein 1	CRABP1	#N/A	#N/A	#N/A	#N/A	2.59E+01	2.67E-02
Crooked neck-like protein 1	CRNKL1	#N/A	#N/A	#N/A	#N/A	2.62E+01	4.91E-03
Pre-mRNA-splicing factor CWC22 homolog	CWC22	#N/A	#N/A	#N/A	#N/A	2.43E+01	2.32E-04
Death-associated protein kinase 3	DAPK3	2.78E+01	9.76E-04	#N/A	#N/A	2.76E+01	2.14E-03
Aspartate--tRNA ligase, cytoplasmic	DARS	#N/A	#N/A	2.57E+01	2.58E-02	#N/A	#N/A
Dopamine beta-hydroxylase	DBH	2.64E+01	7.29E-02	#N/A	#N/A	2.64E+01	1.80E-01
ATP-dependent RNA helicase DDX1	DDX1	8.24E-01	3.45E-01	1.23E+00	5.23E-05	9.21E-01	2.89E-04
Probable ATP-dependent RNA helicase DDX17	DDX17	2.07E+00	1.85E-03	2.50E+00	5.71E-04	2.56E+00	5.24E-06
ATP-dependent RNA helicase DDX18	DDX18	2.51E+01	3.74E-01	#N/A	#N/A	2.56E+01	7.18E-02
Probable ATP-dependent RNA helicase DDX20	DDX20	1.04E+01	3.43E-01	#N/A	#N/A	1.08E+01	2.74E-02
Nucleolar RNA helicase 2	DDX21	8.21E-01	7.07E-03	1.05E+00	8.38E-03	4.78E-01	7.02E-02
ATP-dependent RNA helicase DDX24	DDX24	#N/A	#N/A	#N/A	#N/A	2.40E+01	1.78E-01
ATP-dependent RNA helicase DDX3X	DDX3X	1.22E+00	6.73E-03	1.52E+00	2.21E-03	1.12E+00	1.48E-02
Probable ATP-dependent RNA helicase DDX47	DDX47	1.96E+01	1.27E-01	1.98E+01	1.55E-02	1.98E+01	2.10E-02
Probable ATP-dependent RNA helicase DDX5	DDX5	1.05E+00	1.65E-02	1.32E+00	7.76E-04	1.16E+00	7.64E-03
ATP-dependent RNA helicase DDX50	DDX50	2.47E+01	1.55E-01	2.55E+01	7.16E-02	2.54E+01	2.28E-03
Putative ATP-dependent RNA helicase DHX30	DHX30	2.17E+00	9.42E-03	2.95E+00	6.63E-03	2.47E+00	5.79E-03
Pre-mRNA-splicing factor ATP-dependent RNA helicase PRP16	DHX38	#N/A	#N/A	2.37E+01	3.74E-01	2.41E+01	1.79E-01
ATP-dependent RNA helicase DHX8	DHX8	#N/A	#N/A	#N/A	#N/A	2.62E+01	2.41E-03
ATP-dependent RNA helicase A	DHX9	3.93E+00	1.95E-02	3.97E+00	4.96E-03	3.49E+00	2.94E-02
Endoribonuclease Dicer	DICER1	2.96E+01	8.10E-03	2.96E+01	6.79E-03	2.98E+01	9.79E-03
DnaJ homolog subfamily A member 3, mitochondrial	DNAJA3	#N/A	#N/A	2.51E+01	1.62E-02	#N/A	#N/A
Dihydropyrimidinase-related protein 3	DPYSL3	2.69E+01	8.07E-03	#N/A	#N/A	#N/A	#N/A
E3 ubiquitin-protein ligase DZIP3	DZIP3	#N/A	#N/A	2.61E+01	1.79E-01	2.59E+01	7.37E-02
Enhancer of mRNA-decapping protein 4	EDC4	2.82E+01	1.00E-01	2.76E+01	1.51E-01	2.70E+01	1.51E-01
Interferon-induced, double-stranded RNA-activated protein kinase	EIF2AK2	6.69E+00	5.03E-02	6.71E+00	8.72E-03	6.23E+00	3.74E-03
Eukaryotic translation initiation factor 4 gamma 1	EIF4G1	#N/A	#N/A	2.54E+01	2.63E-05	#N/A	#N/A
RNA-binding protein EWS	EWSR1	1.19E+01	4.32E-02	1.20E+01	1.94E-02	1.18E+01	4.27E-02
FAS-associated factor 2	FAF2	#N/A	#N/A	2.52E+01	1.79E-01	#N/A	#N/A
Constitutive coactivator of PPAR-gamma-like protein 1	FAM120A	1.61E+01	1.78E-02	1.65E+01	3.01E-03	1.62E+01	1.13E-02
Protein FAM91A1	FAM91A1	2.55E+01	7.69E-02	#N/A	#N/A	#N/A	#N/A
Protein FAM98B	FAM98B	#N/A	#N/A	2.37E+01	7.90E-02	#N/A	#N/A
Fermitin family homolog 2	FERMT2	#N/A	#N/A	#N/A	#N/A	2.45E+01	2.05E-01
FH1/FH2 domain-containing protein 1	FHOD1	2.47E+01	2.76E-02	2.48E+01	4.73E-05	#N/A	#N/A
Protein flightless-1 homolog	FLII	#N/A	#N/A	#N/A	#N/A	2.69E+01	7.07E-02
Fragile X mental retardation protein 1	FMR1	#N/A	#N/A	2.52E+01	1.83E-01	#N/A	#N/A
RNA-binding protein FUS	FUS	8.95E-01	4.30E-02	1.23E+00	5.98E-03	6.16E-01	6.83E-02
Ras GTPase-activating protein-binding protein 1	G3BP1	2.66E+01	9.77E-02	2.70E+01	9.84E-03	2.66E+01	1.20E-03
Ras GTPase-activating protein-binding protein 2	G3BP2	2.76E+01	3.80E-02	2.86E+01	2.77E-02	#N/A	#N/A
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	#N/A	#N/A	2.63E+01	6.56E-02	2.63E+01	1.29E-01
Trifunctional purine biosynthetic protein adenosine-3	GART	#N/A	#N/A	#N/A	#N/A	2.54E+01	1.98E-01
GTP cyclohydrolase 1	GCH1	#N/A	#N/A	2.44E+01	9.14E-02	#N/A	#N/A
Gem-associated protein 4	GEMIN4	2.94E+01	2.16E-02	2.97E+01	3.74E-01	3.01E+01	2.65E-03
Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2	GNB2	5.95E+00	9.05E-01	6.25E+00	3.81E-02	5.61E+00	3.91E-01
Guanine nucleotide-binding protein subunit beta-2-like 1	GNB2L1	8.07E-01	4.86E-04	9.49E-01	1.93E-02	5.06E-01	5.69E-03
Guanine nucleotide-binding protein-like 3	GNL3	#N/A	#N/A	2.60E+01	1.71E-02	#N/A	#N/A
G-rich sequence factor 1	GRSF1	#N/A	#N/A	#N/A	#N/A	2.48E+01	1.78E-01
Histone H1x	H1FX	#N/A	#N/A	2.74E+01	1.78E-01	#N/A	#N/A
Core histone macro-H2A.1	H2AFY	#N/A	#N/A	2.63E+01	7.76E-02	2.62E+01	1.79E-01
Trifunctional enzyme subunit beta, mitochondrial	HADHB	#N/A	#N/A	#N/A	#N/A	2.61E+01	8.19E-02
Histone deacetylase	HDAC2	#N/A	#N/A	2.59E+01	3.74E-01	2.60E+01	1.78E-01
Histone H1.2	HIST1H1C	#N/A	#N/A	2.76E+01	1.92E-01	#N/A	#N/A
Helicase-like transcription factor	HLTF	2.51E+01	1.87E-01	2.47E+01	2.66E-02	2.47E+01	7.51E-02
Heterogeneous nuclear ribonucleoprotein A0	HNRNPA0	6.11E+00	7.01E-02	6.65E+00	3.33E-03	6.69E+00	2.49E-02
Heterogeneous nuclear ribonucleoprotein A1	HNRNPA1	5.92E+00	1.21E-01	6.59E+00	1.95E-02	6.57E+00	1.14E-02
Heterogeneous nuclear ribonucleoproteins A2/B1	HNRNPA2B1	5.65E-01	1.01E-01	1.21E+00	3.70E-02	1.46E+00	6.95E-03

Heterogeneous nuclear ribonucleoprotein A3	HNRNPA3	9.09E-01	4.40E-02	1.24E+00	7.03E-04	1.46E+00	2.85E-03
Heterogeneous nuclear ribonucleoprotein F	HNRNPF	2.37E+01	4.97E-04	2.37E+01	3.94E-04	2.38E+01	7.70E-04
Heterogeneous nuclear ribonucleoprotein H	HNRNPH1	9.73E-01	9.50E-02	7.37E-01	2.40E-02	7.80E-01	4.60E-02
Heterogeneous nuclear ribonucleoprotein H3	HNRNPH3	#N/A	#N/A	#N/A	#N/A	2.66E+01	2.00E-02
Heterogeneous nuclear ribonucleoprotein K	HNRNPK	1.16E+00	3.01E-02	1.31E+00	1.80E-04	1.31E+00	5.15E-03
Heterogeneous nuclear ribonucleoprotein M	HNRNPM	6.91E-01	1.21E-01	7.01E-01	5.66E-03	1.05E+00	1.12E-02
Heterogeneous nuclear ribonucleoprotein R	HNRNPR	1.62E+00	4.62E-03	2.45E+00	3.26E-03	2.05E+00	1.61E-04
Heterogeneous nuclear ribonucleoprotein U	HNRNPU	8.19E-01	6.39E-03	1.24E+00	4.45E-04	1.12E+00	2.56E-03
Heterogeneous nuclear ribonucleoprotein U-like protein 2	HNRNPUL2	#N/A	#N/A	2.70E+01	1.54E-05	2.70E+01	2.32E-04
Heterogeneous nuclear ribonucleoprotein D-like	HNRPDL	5.88E-01	6.01E-01	8.71E-01	4.07E-03	1.02E+00	2.61E-02
Estradiol 17-beta-dehydrogenase 11	HSD17B11	#N/A	#N/A	2.52E+01	1.87E-01	2.52E+01	1.79E-01
Peroxisomal multifunctional enzyme type 2	HSD17B4	#N/A	#N/A	2.62E+01	1.73E-02	2.60E+01	8.18E-02
Heat shock protein HSP 90-alpha	HSP90AA1	#N/A	#N/A	#N/A	#N/A	2.76E+01	1.87E-01
Endoplasmin	HSP90B1	#N/A	#N/A	#N/A	#N/A	2.65E+01	2.51E-01
Heat shock 70 kDa protein 1A/1B	HSPA1A	2.57E+01	8.09E-04	#N/A	#N/A	#N/A	#N/A
Insulin-like growth factor 2 mRNA-binding protein 3	IGF2BP3	1.88E+00	2.17E-03	2.40E+00	5.54E-03	2.06E+00	1.26E-02
Interleukin enhancer-binding factor 2	ILF2	2.50E+00	1.66E-04	2.75E+00	1.35E-03	2.40E+00	2.66E-03
Interleukin enhancer-binding factor 3	ILF3	3.62E+00	1.30E-02	3.72E+00	5.68E-03	3.32E+00	6.59E-03
Importin-4	IPO4	#N/A	#N/A	#N/A	#N/A	2.34E+01	1.52E-01
Importin-7	IPO7	#N/A	#N/A	#N/A	#N/A	2.52E+01	8.39E-02
Junction plakoglobin	JUP	#N/A	#N/A	#N/A	#N/A	2.39E+01	3.39E-02
KH domain-containing, RNA-binding, signal transduction-associated protein 1	KHDRBS1	1.15E+00	5.62E-02	1.39E+00	3.97E-06	1.06E+00	4.42E-03
Protein virilizer homolog	KIAA1429	#N/A	#N/A	2.27E+01	7.46E-02	2.27E+01	1.70E-02
DBIRD complex subunit KIAA1967	KIAA1967	#N/A	#N/A	#N/A	#N/A	2.57E+01	1.90E-01
Kinesin light chain 1	KLC1	#N/A	#N/A	2.45E+01	1.81E-01	2.48E+01	7.33E-02
Importin subunit alpha-4	KPNA4	#N/A	#N/A	2.18E+01	1.86E-01	#N/A	#N/A
La-related protein 1	LARP1	2.56E+01	1.82E-01	2.60E+01	7.59E-02	#N/A	#N/A
L-lactate dehydrogenase A chain	LDHA	#N/A	#N/A	#N/A	#N/A	2.56E+01	1.70E-01
LEM domain-containing protein 2	LEMD2	#N/A	#N/A	2.62E+01	7.73E-02	#N/A	#N/A
Luc7-like protein 3	LUC7L3	#N/A	#N/A	#N/A	#N/A	2.66E+01	2.04E-03
Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5	MACF1	#N/A	#N/A	#N/A	#N/A	2.59E+01	1.91E-02
Protein mago nashi homolog	MAGOHB	#N/A	#N/A	#N/A	#N/A	2.77E+01	3.23E-04
Microtubule-associated protein 4	MAP4	#N/A	#N/A	2.67E+01	1.70E-02	2.67E+01	1.86E-02
MAP7 domain-containing protein 1	MAP7D1	2.49E+01	3.74E-01	2.50E+01	2.40E-04	#N/A	#N/A
Matrin-3	MATR3	-5.47E-02	7.86E-01	1.23E+00	9.63E-03	1.31E+00	1.88E-03
MMS19 nucleotide excision repair protein homolog	MMS19	#N/A	#N/A	#N/A	#N/A	2.34E+01	8.81E-02
Mannosyl-oligosaccharide glucosidase	MOGS	#N/A	#N/A	#N/A	#N/A	2.52E+01	8.47E-02
RNA-binding protein Musashi homolog 1	MSI1	#N/A	#N/A	#N/A	#N/A	2.47E+01	7.41E-02
C-1-tetrahydrofolate synthase, cytoplasmicsynthetase	MTHFD1	#N/A	#N/A	2.55E+01	6.44E-03	2.55E+01	1.38E-03
Myb-binding protein 1A	MYBBP1A	1.10E+01	9.26E-02	1.09E+01	1.60E-01	1.09E+01	1.60E-01
Myelin expression factor 2	MYEF2	#N/A	#N/A	#N/A	#N/A	2.60E+01	8.64E-05
Myosin light chain 1/3, skeletal muscle isoform	MYL1	2.87E+01	7.72E-03	2.87E+01	3.39E-02	2.90E+01	2.46E-02
Myosin light chain kinase, smooth muscle	MYLK	#N/A	#N/A	#N/A	#N/A	2.41E+01	1.78E-01
Asparagine--tRNA ligase, cytoplasmic	NARS	#N/A	#N/A	2.57E+01	2.12E-02	2.56E+01	1.42E-03
Nucleolin	NCL	8.27E-01	7.49E-02	1.68E+00	1.17E-03	7.51E-01	1.39E-01
NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial	NDUFS3	#N/A	#N/A	#N/A	#N/A	2.55E+01	1.80E-01
Non-POU domain-containing octamer-binding protein	NONO	9.79E-01	3.99E-02	1.48E+00	5.16E-02	1.86E-01	6.36E-01
Nuclear pore complex protein Nup133	NUP133	2.47E+01	3.74E-01	#N/A	#N/A	2.41E+01	1.71E-02
Nuclear pore membrane glycoprotein 210	NUP210	2.45E+01	7.39E-02	#N/A	#N/A	#N/A	#N/A
Polyadenylate-binding protein 1	PABPC1	6.86E-01	5.55E-02	1.36E+00	2.47E-04	9.94E-01	1.11E-04
Polyadenylate-binding protein 4	PABPC4	5.69E-01	2.21E-01	1.06E+00	6.54E-02	8.17E-01	7.76E-02
Poly(rC)-binding protein 1	PCBP1	5.94E+00	1.67E-01	6.00E+00	3.47E-02	6.11E+00	1.87E-02
Proliferating cell nuclear antigen	PCNA	#N/A	#N/A	2.61E+01	1.22E-01	#N/A	#N/A
Polymerase delta-interacting protein 3	PDIP46	#N/A	#N/A	2.58E+01	6.77E-05	#N/A	#N/A
6-phosphofructokinase, muscle type	PFKM	2.10E+01	7.74E-02	2.08E+01	4.36E-02	2.11E+01	6.67E-02
Prohibitin-2	PHB2	#N/A	#N/A	#N/A	#N/A	2.52E+01	1.36E-01
D-3-phosphoglycerate dehydrogenase	PHGDH	#N/A	#N/A	2.55E+01	3.44E-02	#N/A	#N/A
Peptidyl-prolyl cis-trans isomerase	PPIA	#N/A	#N/A	#N/A	#N/A	2.61E+01	2.00E-01
Serine/threonine-protein phosphatase 2A 65 kDa reg. subunit A alpha	PPP2R1A	2.44E+01	1.95E-02	#N/A	#N/A	2.47E+01	2.43E-01
Protein regulator of cytokinesis 1	PRC1	#N/A	#N/A	2.55E+01	1.85E-01	#N/A	#N/A
Peroxiredoxin-1	PRDX1	2.74E+01	1.83E-01	#N/A	#N/A	#N/A	#N/A
Peroxiredoxin-6	PRDX6	#N/A	#N/A	#N/A	#N/A	2.42E+01	2.75E-01
Interferon-inducible double stranded RNA-dependent protein kinase activator A	PRKRA	2.76E+01	5.50E-02	2.77E+01	1.06E-02	2.77E+01	2.95E-04
Protein arginine N-methyltransferase 1	PRMT1	2.48E+01	8.22E-02	#N/A	#N/A	2.55E+01	5.28E-03
Pre-mRNA-processing factor 6	PRPF6	#N/A	#N/A	#N/A	#N/A	2.68E+01	3.15E-03
Protein PRRC2A	PRRC2A	#N/A	#N/A	2.64E+01	4.59E-04	#N/A	#N/A
26S protease regulatory subunit 6A	PSMC3	#N/A	#N/A	2.48E+01	2.21E-02	#N/A	#N/A
26S protease regulatory subunit 8	PSMC5	#N/A	#N/A	#N/A	#N/A	2.60E+01	1.95E-01
Paraspeckle component 1	PSPC1	2.65E+01	7.49E-02	2.71E+01	2.58E-02	2.67E+01	1.78E-01
Polypyrimidine tract-binding protein 1	PTBP1	2.59E-01	5.05E-01	1.34E+00	1.40E-02	1.18E+00	1.04E-02
Peroxidasin homolog	PXDN	#N/A	#N/A	#N/A	#N/A	2.59E+01	8.74E-02
Glutamine--tRNA ligase	QARS	2.41E+01	1.78E-01	#N/A	#N/A	#N/A	#N/A
Rac GTPase-activating protein 1	RACGAP1	#N/A	1.78E-01	#N/A	1.78E-01	1.58E+01	4.02E-02
Arginine--tRNA ligase, cytoplasmic	RARS	#N/A	#N/A	#N/A	#N/A	2.46E+01	8.60E-02
RNA-binding protein 14	RBM14	8.56E-01	3.96E-02	9.23E-01	1.52E-03	5.36E-01	8.04E-02
RNA-binding protein 6	RBM6	#N/A	#N/A	2.45E+01	8.81E-02	#N/A	#N/A
RNA-binding protein 8A	RBM8A	#N/A	#N/A	#N/A	#N/A	2.68E+01	1.92E-01
RNA-binding motif protein, X chromosome	RBMX	1.45E+00	3.14E-02	1.54E+00	4.70E-04	1.31E+00	2.29E-02
Protein RCC2	RCC2	#N/A	#N/A	#N/A	#N/A	2.55E+01	7.30E-02
Replication factor C subunit 1	RFC1	2.05E+01	7.91E-02	2.07E+01	4.95E-03	#N/A	3.74E-01
Replication factor C subunit 2	RFC2	#N/A	#N/A	#N/A	#N/A	2.53E+01	1.78E-01
Ribonuclease inhibitor	RNH1	3.05E+01	4.55E-02	2.99E+01	4.57E-02	3.08E+01	8.87E-02

RNA-binding protein with serine-rich domain 1	RNPS1	#N/A	#N/A	2.57E+01	8.87E-02	2.66E+01	7.23E-02
60S ribosomal protein L10	RPL10	7.91E-01	4.98E-03	1.14E+00	1.71E-02	2.34E-01	1.98E-01
60S ribosomal protein L10a	RPL10A	1.79E+01	5.05E-03	1.76E+01	2.51E-02	1.75E+01	2.50E-02
60S ribosomal protein L11	RPL11	1.09E+00	1.14E-01	8.79E-01	2.12E-02	2.69E-01	4.41E-01
60S ribosomal protein L12	RPL12	9.22E-01	8.28E-03	1.02E+00	3.09E-02	5.63E-01	9.80E-02
60S ribosomal protein L13	RPL13	7.95E-01	2.18E-02	1.12E+00	7.31E-03	5.42E-01	7.91E-02
60S ribosomal protein L13a	RPL13A	1.26E+01	3.22E-02	1.25E+01	9.70E-03	1.19E+01	8.99E-03
60S ribosomal protein L14	RPL14	2.89E+01	3.46E-02	2.92E+01	1.12E-02	2.82E+01	1.19E-01
60S ribosomal protein L15	RPL15	1.28E+01	6.64E-02	1.30E+01	5.80E-02	1.27E+01	2.71E-02
60S ribosomal protein L17	RPL17	1.16E+01	2.38E-01	1.20E+01	5.06E-02	1.18E+01	1.90E-02
60S ribosomal protein L18	RPL18	1.59E+00	1.09E-01	1.90E+00	5.37E-02	1.55E+00	4.45E-02
60S ribosomal protein L18a	RPL18A	1.23E+01	2.57E-03	1.24E+01	1.46E-03	1.19E+01	1.06E-02
Ribosomal protein L19	RPL19	3.10E+01	7.03E-02	3.12E+01	3.45E-02	3.16E+01	1.67E-03
60S ribosomal protein L21	RPL21	9.63E-01	1.85E-02	1.40E+00	3.32E-03	1.07E+00	3.78E-03
60S ribosomal protein L23a	RPL23A	1.18E+01	8.93E-01	1.24E+01	2.23E-03	1.21E+01	9.92E-02
60S ribosomal protein L26	RPL26	2.91E+01	2.27E-02	2.91E+01	2.94E-02	2.87E+01	3.74E-01
60S ribosomal protein L27	RPL27	1.12E+00	3.41E-03	1.28E+00	5.28E-03	8.19E-01	1.78E-02
60S ribosomal protein L28	RPL28	6.62E+00	1.13E-02	6.68E+00	1.90E-02	6.52E+00	1.52E-01
60S ribosomal protein L3	RPL3	1.05E+00	7.69E-02	9.20E-01	9.76E-03	5.22E-01	9.59E-02
60S ribosomal protein L31	RPL31	2.34E+01	2.32E-03	2.39E+01	9.81E-07	2.32E+01	3.75E-05
60S ribosomal protein L34	RPL34	#N/A	#N/A	2.90E+01	8.04E-02	#N/A	#N/A
60S ribosomal protein L35	RPL35	2.82E+01	1.01E-01	2.81E+01	6.39E-03	2.83E+01	3.74E-01
60S ribosomal protein L35a	RPL35A	1.27E+00	3.70E-02	1.29E+00	1.92E-02	6.83E-01	8.45E-01
60S ribosomal protein L36	RPL36	2.77E+01	7.39E-02	1.02E+01	1.58E-01	2.68E+01	3.74E-01
60S ribosomal protein L4	RPL4	1.13E+00	5.32E-02	1.76E+00	1.22E-03	1.27E+00	2.96E-03
60S ribosomal protein L5	RPL5	6.56E-01	1.55E-01	1.33E+00	5.33E-02	1.02E+00	1.04E-02
60S ribosomal protein L6	RPL6	1.15E+00	3.10E-01	2.18E+00	7.55E-02	1.87E+00	5.97E-02
60S ribosomal protein L7	RPL7	9.68E-01	1.02E-02	9.05E-01	8.38E-03	4.91E-01	2.89E-02
60S ribosomal protein L7a	RPL7A	1.15E+00	1.79E-02	1.47E+00	4.34E-03	1.01E+00	1.85E-04
60S ribosomal protein L8	RPL8	3.00E+01	2.16E-03	3.03E+01	4.95E-04	2.93E+01	4.77E-03
60S ribosomal protein L9	RPL9	6.38E-01	7.69E-03	9.60E-01	1.74E-02	5.14E-01	1.28E-02
60S acidic ribosomal protein P0	RPLP0	5.28E-01	3.04E-02	1.03E+00	2.46E-02	4.54E-01	2.59E-02
60S acidic ribosomal protein P1	RPLP1	#N/A	#N/A	2.69E+01	1.91E-01	#N/A	#N/A
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	RPN2	#N/A	#N/A	#N/A	#N/A	2.53E+01	1.84E-01
40S ribosomal protein S10	RPS10	1.15E+00	7.79E-03	1.53E+00	1.31E-02	8.36E-01	5.58E-02
40S ribosomal protein S11	RPS11	1.04E+00	5.16E-03	1.04E+00	9.28E-05	8.12E-01	2.26E-02
40S ribosomal protein S13	RPS13	5.69E-01	1.19E-01	1.41E+00	1.63E-02	8.61E-01	3.16E-03
40S ribosomal protein S14	RPS14	1.26E+01	6.79E-02	1.21E+01	9.16E-03	1.19E+01	3.05E-02
40S ribosomal protein S15	RPS15	2.57E+01	7.29E-02	2.62E+01	4.54E-02	2.58E+01	1.33E-04
40S ribosomal protein S15a	RPS15A	1.71E+01	4.02E-02	1.74E+01	1.83E-02	1.70E+01	6.03E-02
40S ribosomal protein S16	RPS16	2.90E+01	1.15E-05	2.92E+01	1.86E-03	2.87E+01	2.92E-02
40S ribosomal protein S17-like	RPS17L	2.66E+01	2.12E-02	2.68E+01	3.46E-03	2.64E+01	1.81E-02
40S ribosomal protein S18	RPS18	9.72E-01	5.13E-03	9.25E-01	1.44E-02	4.71E-01	6.32E-02
40S ribosomal protein S19	RPS19	2.66E+01	4.98E-02	2.71E+01	1.40E-03	#N/A	#N/A
40S ribosomal protein S2	RPS2	7.88E-01	8.47E-02	1.41E+00	3.73E-02	1.03E+00	2.34E-03
40S ribosomal protein S20	RPS20	7.07E-01	5.10E-02	1.05E+00	3.76E-02	5.78E-01	7.05E-02
40S ribosomal protein S23	RPS23	3.01E+01	1.06E-02	3.00E+01	1.32E-03	2.96E+01	4.90E-04
40S ribosomal protein S24	RPS24	3.93E-01	2.44E-01	1.11E+00	2.62E-02	8.00E-01	8.41E-03
40S ribosomal protein S25	RPS25	6.59E+00	1.65E-03	6.68E+00	1.96E-03	5.83E+00	1.25E-01
40S ribosomal protein S3	RPS3	7.77E-01	4.04E-03	8.81E-01	8.94E-03	4.75E-01	1.57E-04
40S ribosomal protein S4, X isoform	RPS4X	7.86E-01	7.53E-03	9.45E-01	4.77E-03	6.71E-01	4.48E-02
40S ribosomal protein S5	RPS5	#N/A	1.78E-01	1.66E+01	3.65E-01	1.65E+01	2.38E-02
40S ribosomal protein S6	RPS6	1.26E+01	4.02E-03	1.28E+01	1.56E-03	1.23E+01	1.47E-02
40S ribosomal protein S8	RPS8	8.62E-01	1.73E-02	1.26E+00	8.72E-03	7.32E-01	1.71E-03
40S ribosomal protein S9	RPS9	1.67E+00	2.60E-02	1.58E+00	2.12E-03	1.06E+00	6.28E-03
Sacsin	SACS	#N/A	#N/A	2.14E+01	1.86E-01	2.13E+01	7.71E-02
Histone deacetylase complex subunit SAP18	SAP18	#N/A	#N/A	#N/A	#N/A	2.85E+01	9.09E-02
Splicing factor 1	SF1	#N/A	#N/A	#N/A	#N/A	2.33E+01	2.25E-01
Splicing factor, proline- and glutamine-rich	SFPQ	6.88E-01	1.00E-01	1.49E+00	4.83E-02	2.30E-01	5.69E-01
Sphingosine-1-phosphate lyase 1	SGPL1	2.47E+01	1.79E-01	#N/A	#N/A	#N/A	#N/A
Superkiller viralicidic activity 2-like 2	SKIV2L2	#N/A	2.33E-02	5.36E+00	3.30E-02	5.53E+00	1.48E-02
Mitochondrial dicarboxylate carrier	SLC25A10	2.55E+01	3.73E-02	#N/A	#N/A	2.57E+01	1.84E-01
Structural maintenance of chromosomes protein 1A	SMC1A	#N/A	#N/A	#N/A	#N/A	2.45E+01	7.85E-02
Structural maintenance of chromosomes protein 3	SMC3	2.60E+01	1.79E-01	#N/A	#N/A	#N/A	#N/A
WD40 repeat-containing protein SMU1	SMU1	#N/A	#N/A	#N/A	#N/A	2.66E+01	7.27E-02
Staphylococcal nuclease domain-containing protein 1	SND1	#N/A	1.87E-02	5.58E+00	5.09E-02	5.23E+00	7.91E-01
U2 small nuclear ribonucleoprotein B	SNRPB2	#N/A	#N/A	#N/A	#N/A	2.73E+01	7.47E-05
Small nuclear ribonucleoprotein Sm D2	SNRPD2	#N/A	#N/A	#N/A	#N/A	2.73E+01	1.88E-02
Small nuclear ribonucleoprotein Sm D3	SNRPD3	2.70E+01	1.37E-04	2.74E+01	1.67E-02	2.77E+01	7.09E-02
Small nuclear ribonucleoprotein-associated protein N	SNRPN	#N/A	#N/A	#N/A	#N/A	2.75E+01	1.95E-01
Spermatogenesis-associated protein 5	SPATA5	2.34E+01	2.26E-01	2.34E+01	1.03E-01	#N/A	#N/A
Serine/arginine repetitive matrix protein 1	SRRM1	#N/A	#N/A	#N/A	#N/A	2.60E+01	1.29E-01
Serrate RNA effector molecule homolog	SRRT	#N/A	#N/A	2.60E+01	4.01E-02	#N/A	#N/A
Serine/arginine-rich splicing factor 10	SRSF10	#N/A	#N/A	#N/A	#N/A	2.74E+01	7.30E-02
Serine/arginine-rich splicing factor 2	SRSF2	#N/A	#N/A	#N/A	#N/A	2.81E+01	1.63E-02
Serine/arginine-rich splicing factor 4	SRSF4	#N/A	#N/A	#N/A	#N/A	2.58E+01	1.10E-01
Serine/arginine-rich splicing factor 7	SRSF7	1.14E+01	9.90E-02	1.29E+01	1.74E-02	1.36E+01	5.78E-03
Lupus La protein	SSB	2.58E+01	1.07E-01	2.66E+01	3.71E-02	2.60E+01	3.84E-02
FACT complex subunit SSRP1	SSRP1	2.52E+01	2.00E-01	#N/A	#N/A	#N/A	#N/A
Double-stranded RNA-binding protein Staufen homolog 1	STAU1	3.22E+01	1.43E-02	3.16E+01	5.58E-03	3.15E+01	1.27E-02
Double-stranded RNA-binding protein Staufen homolog 2	STAU2	2.79E+01	3.90E-03	2.75E+01	3.75E-03	2.77E+01	7.24E-02
Stomatin-like protein 2, mitochondrial	STOML2	#N/A	#N/A	2.54E+01	6.26E-05	2.55E+01	2.76E-02

Pre-mRNA-splicing factor SYF2	SYF2	#N/A	#N/A	#N/A	#N/A	2.38E+01	1.83E-01
Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	2.27E+01	3.09E-03	2.38E+01	1.52E-03	2.36E+01	1.07E-03
TATA-binding protein-associated factor 2N	TAF15	#N/A	#N/A	2.72E+01	9.00E-02	#N/A	#N/A
RISC-loading complex subunit TARBP2	TARBP2	#N/A	#N/A	#N/A	#N/A	2.77E+01	1.89E-01
Transducin beta-like protein 2	TBL2	#N/A	#N/A	2.59E+01	1.78E-01	#N/A	#N/A
THO complex subunit 2	THOC2	#N/A	#N/A	#N/A	#N/A	2.64E+01	2.01E-01
Thyroid hormone receptor-associated protein 3	THRAP3	#N/A	#N/A	2.74E+01	1.83E-01	2.81E+01	3.83E-03
Transportin-3	TNPO3	#N/A	#N/A	#N/A	#N/A	2.33E+01	1.91E-01
Target of Myb protein 1	TOM1	2.42E+01	3.74E-01	#N/A	#N/A	2.41E+01	1.28E-01
Nucleoprotein TPR	TPR	#N/A	#N/A	2.33E+01	5.46E-03	#N/A	#N/A
TNF receptor-associated factor 2	TRAF2	2.47E+01	1.78E-01	#N/A	#N/A	#N/A	#N/A
TRMT1-like protein	TRMT1L	#N/A	#N/A	#N/A	#N/A	2.39E+01	2.09E-01
Splicing factor U2AF 35 kDa subunit	U2AF1	1.20E+00	1.90E-01	1.52E+00	6.68E-04	2.46E+00	1.32E-03
Ubiquitin-like modifier-activating enzyme 1	UBA1	#N/A	#N/A	#N/A	#N/A	2.55E+01	3.21E-01
Regulator of nonsense transcripts 1	UPF1	2.72E+01	4.78E-04	2.76E+01	2.95E-05	2.69E+01	3.74E-01
Cytochrome b-c1 complex subunit 2, mitochondrial	UQCRC2	#N/A	#N/A	2.59E+01	8.58E-06	#N/A	#N/A
Ubiquitin carboxyl-terminal hydrolase 10	USP10	#N/A	#N/A	2.54E+01	3.74E-01	2.46E+01	1.99E-02
U4/U6.U5 tri-snRNP-associated protein 2	USP39	#N/A	#N/A	2.45E+01	7.66E-02	#N/A	#N/A
Vacuolar protein sorting-associated protein 35	VPS35	2.47E+01	3.74E-01	#N/A	#N/A	2.50E+01	3.74E-01
5-3 exoribonuclease 2	XRN2	2.67E+01	3.77E-03	2.68E+01	5.04E-03	2.68E+01	4.87E-02
Nuclease-sensitive element-binding protein 1	YBX1	3.14E+01	6.90E-02	3.11E+01	6.79E-03	3.05E+01	2.95E-03
YLP motif-containing protein 1	YLPM1	2.71E+01	1.75E-02	2.76E+01	5.41E-04	2.69E+01	2.41E-05
Zinc finger RNA-binding protein	ZFR	2.75E+01	2.41E-03	2.76E+01	3.23E-03	2.73E+01	5.61E-04
Zinc finger protein 318	ZNF318	2.52E+01	1.78E-01	#N/A	#N/A	#N/A	#N/A

Table 9 | Affymetrix splicing array

Protein name	Gene ID	Splicing Event Estimate	Splicing Event Score
mitochondrial translational initiation factor 2	MTIF2	Alternative 3' Acceptor Site	0.27
Dmx-like 1	DMXL1	Alternative 3' Acceptor Site	0.27
phosphatidylserine decarboxylase; microRNA 7109	PISD; MIR7109	Alternative 3' Acceptor Site	0.25
NonCoding		Alternative 3' Acceptor Site	0.23
NonCoding		Alternative 3' Acceptor Site	0.16
ZNRD1 antisense RNA 1	ZNRD1-AS1	Alternative 3' Acceptor Site	0.21
putative homeodomain transcription factor 1	PHTF1	Alternative 3' Acceptor Site	0.2
NonCoding		Alternative 3' Acceptor Site	0.19
phosphatidylinositol glycan anchor biosynthesis class N	PIGN	Alternative 3' Acceptor Site	0.19
acylphosphatase 2, muscle type	ACYP2	Alternative 3' Acceptor Site	0.06
Ca++-dependent secretion activator	CADPS	Alternative 3' Acceptor Site	0.18
Transcript Identified by AceView	pleefarbo	Alternative 3' Acceptor Site	0.17
NonCoding		Alternative 3' Acceptor Site	0.17
kinesin family member 16B	KIF16B	Alternative 3' Acceptor Site	0.17
bromodomain and WD repeat domain containing 1	BRWD1	Alternative 3' Acceptor Site	0.09
signal peptidase complex subunit 2	SPCS2	Alternative 3' Acceptor Site	0.14
discoidin domain receptor tyrosine kinase 1; microRNA 4640	DDR1; MIR4640	Alternative 3' Acceptor Site	0.13
acyl-CoA synthetase short-chain family member 2	ACSS2	Alternative 3' Acceptor Site	0.12
DEAD/H (Asp-Glu-Ala-Asp/His) box helicase 11 like 5	DDX11L5	Alternative 3' Acceptor Site	0.11
novel transcript	RP5-1085F17.3	Alternative 3' Acceptor Site	0.11
novel transcript	RP5-1085F17.3	Alternative 3' Acceptor Site	0.11
aspartic peptidase, retroviral-like 1; PCBP1 antisense RNA 1	ASPRV1; PCBP1-AS1	Alternative 3' Acceptor Site	0.09
NonCoding		Alternative 5' Donor Site	0.32
DPY19L2 pseudogene 3	DPY19L2P3	Alternative 5' Donor Site	0.27
CCCTC-binding factor (zinc finger protein)	CTCF	Alternative 5' Donor Site	0.27
mitochondrial translational release factor 1-like	MTRF1L	Alternative 5' Donor Site	0.22
tripartite motif containing 2	TRIM2	Alternative 5' Donor Site	0.2
nuclear RNA export factor 1	NXF1	Alternative 5' Donor Site	0.2
cell division cycle associated 2	CDCA2	Alternative 5' Donor Site	0.19
aldo-keto reductase family 1, member C2	AKR1C2	Alternative 5' Donor Site	0.19
T-cell leukemia translocation altered	TCTA	Alternative 5' Donor Site	0.18
eukaryotic translation initiation factor 4E nuclear import factor 1	EIF4ENIF1	Alternative 5' Donor Site	0.18
transmembrane protein 135	TMEM135	Alternative 5' Donor Site	0.17
rotatin	RTTN	Alternative 5' Donor Site	0.17
bromodomain and WD repeat domain containing 1	BRWD1	Alternative 5' Donor Site	0.16
transmembrane protein 67	TMEM67	Alternative 5' Donor Site	0.15
family with sequence similarity 83, member A	FAM83A	Alternative 5' Donor Site	0.15
zinc fingers and homeoboxes 3	ZHX3	Alternative 5' Donor Site	0.13
NonCoding		Alternative 5' Donor Site	0.1
adaptor-related protein complex 4, sigma 1 subunit	AP4S1	Alternative 5' Donor Site	0.1
PIF1 5-to-3 DNA helicase	PIF1	Alternative 5' Donor Site	0.09
pogo transposable element with ZNF domain	POGZ	Cassette Exon	0.23
zinc finger protein 519	ZNF519	Cassette Exon	0.21
chemokine (C-X-C motif) ligand 17	CXCL17	Cassette Exon	0.21
EF-hand calcium binding domain 7; deleted in lymphocytic leukemia 2-like	EFCAB7; DLEU2L	Cassette Exon	0.19
NonCoding		Cassette Exon	0.19
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1, 7.5kDa	NDUFA1	Cassette Exon	0.19
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1, 7.5kDa	NDUFA1	Cassette Exon	0.19
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1, 7.5kDa	NDUFA1	Cassette Exon	0.16
acylphosphatase 2, muscle type	ACYP2	Cassette Exon	0.18
acylphosphatase 2, muscle type	ACYP2	Cassette Exon	0.16

acylphosphatase 2, muscle type	ACYP2	Cassette Exon	0.15
polymerase (DNA-directed), delta 4, accessory subunit	POLD4	Cassette Exon	0.18
long intergenic non-protein coding RNA 1533	LINC01533	Cassette Exon	0.16
transmembrane protein 67	TMEM67	Cassette Exon	0.15
Transcript Identified by AceView	fyweyby	Cassette Exon	0.15
zinc finger protein 789	ZNF789	Cassette Exon	0.14
pregnancy up-regulated nonubiquitous CaM kinase	PNCK	Cassette Exon	0.14
pregnancy up-regulated nonubiquitous CaM kinase	PNCK	Cassette Exon	0.14
NonCoding		Cassette Exon	0.14
aldehyde dehydrogenase 3 family, member B2	ALDH3B2	Cassette Exon	0.14
notch 2 N-terminal like	NOTCH2NL	Cassette Exon	0.13
chromosome 19 open reading frame 12	C19orf12	Cassette Exon	0.13
dual serine/threonine and tyrosine protein kinase	DSTYK	Cassette Exon	0.12
mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase	MGAT1	Cassette Exon	0.12
NonCoding		Cassette Exon	0.12
chromosome 22 open reading frame 15	C22orf15	Cassette Exon	0.12
S100 calcium binding protein A13	S100A13	Cassette Exon	0.11
microphthalmia-associated transcription factor	MITF	Cassette Exon	0.11
NonCoding	RP11-1396O13.13	Cassette Exon	0.11
Transcript Identified by AceView	rerame	Cassette Exon	0.11
solute carrier family 7 (amino acid transporter light chain, y+L system), member 6	SLC7A6	Cassette Exon	0.11
NonCoding		Cassette Exon	0.11
NonCoding	RP11-451O13.1	Cassette Exon	0.1
novel putative transcript	AC093627.12	Cassette Exon	0.1
cytochrome c oxidase assembly factor 1 homolog	COA1	Cassette Exon	0.1
N-terminal EF-hand calcium binding protein 1	NECAB1	Cassette Exon	0.1
coiled-coil domain containing 171	CCDC171	Cassette Exon	0.1
ankyrin repeat and SOCS box containing 8	ASB8	Cassette Exon	0.1
NonCoding		Cassette Exon	0.1
zinc finger protein 20	ZNF20	Cassette Exon	0.1
ZRANB2 antisense RNA 2 (head to head)	ZRANB2-AS2	Cassette Exon	0.06
IZUMO family member 4	IZUMO4	Cassette Exon	0.04
polymerase (DNA directed) iota	POLI	Intron Retention	0.52
tubulin tyrosine ligase-like family member 5	TTLL5	Intron Retention	0.32
microtubule crosslinking factor 1	MTCL1	Intron Retention	0.29
cAMP responsive element binding protein 1	CREB1	Intron Retention	0.28
mucin 5AC, oligomeric mucus/gel-forming	MUC5AC	Intron Retention	0.28
tyrosyl-DNA phosphodiesterase 1	TDP1	Intron Retention	0.28
scaffold attachment factor B2	SAFB2	Intron Retention	0.28
long intergenic non-protein coding RNA 116	LINC00116	Intron Retention	0.27
major facilitator superfamily domain containing 8	MFSD8	Intron Retention	0.27
coiled-coil domain containing 13	CCDC13	Intron Retention	0.26
growth factor receptor bound protein 10	GRB10	Intron Retention	0.26
aryl hydrocarbon receptor nuclear translocator-like 2	ARNTL2	Intron Retention	0.26
ubiquitin specific peptidase 37	USP37	Intron Retention	0.22
speedy/RINGO cell cycle regulator family member E3	SPDYE3	Intron Retention	0.21
cyclin-dependent kinase-like 1 (CDC2-related kinase)	CDKL1	Intron Retention	0.19
acylphosphatase 2, muscle type	ACYP2	Intron Retention	0.07
RALY heterogeneous nuclear ribonucleoprotein	RALY	Intron Retention	0.18
pre-mRNA processing factor 4B	PRPF4B	Intron Retention	0.15
Transcript Identified by AceView	blosleybu; yukamu	Intron Retention	0.12
suppression of tumorigenicity 5	ST5	Intron Retention	0.09