Advances in Experimental Medicine and Biology 1157

Luísa Romão *Editor*

The mRNA Metabolism in Human Disease

Springer

Advances in Experimental Medicine and Biology

Volume 1157

More information about this series at http://www.springer.com/series/5584

Luísa Romão
Editor

The mRNA Metabolism in Human Disease

Springer

Editor
Luísa Romão
Department of Human Genetics
National Institute of Health Doutor Ricardo Jorge
Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems
& Integrative Sciences Institute
University of Lisbon
Lisboa, Portugal

ISSN 0065-2598 ISSN 2214-8019 (electronic)
Advances in Experimental Medicine and Biology
ISBN 978-3-030-19965-4 ISBN 978-3-030-19966-1 (eBook)
https://doi.org/10.1007/978-3-030-19966-1

Contents

Contributors

Sandra Alves Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Porto, Portugal

Cecília M. Arraiano Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal

Marco M. Candeias Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

MaRCU – Molecular and RNA Cancer Unit, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Maria Francisca Coutinho Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Porto, Portugal

Paulo J. da Costa Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

Rafael Fernandes Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

Sara Fernandes Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Lisboa, Portugal

Department of Chemistry and Biochemistry, Faculty of Sciences, University of Lisbon, Lisboa, Portugal

Margarida Gama-Carvalho Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

Department of Chemistry and Biochemistry, Faculty of Sciences, University of Lisbon, Lisboa, Portugal

Vânia Gonçalves Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

Greco Hernández Laboratory of Translation and Cancer, Unit of Biomedical Research on Cancer, Instituto Nacional de Cancerología (INCan), Mexico City, Mexico

Peter Jordan Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

Adam Kosti Greehey Children's Cancer Research Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

Rafaela Lacerda Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

Tânia Marques Department of Chemistry and Biochemistry, Faculty of Sciences, University of Lisbon, Lisboa, Portugal

Liliana Matos Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Porto, Portugal

Juliane Menezes Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

Gonçalo Nogueira Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

Luiz O. F. Penalva Greehey Children's Cancer Research Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

Marcelo Pereira Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

Francisco Pinto Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

Luísa Romão Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

Juliana Inês Santos Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Porto, Portugal

Margarida Saramago Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal

Joana Silva Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

Mitzli X. Velasco Laboratory of Translation and Cancer, Unit of Biomedical Research on Cancer, Instituto Nacional de Cancerología (INCan), Mexico City, Mexico

Biomedical Sciences Ph.D. Program, National Autonomous University of Mexico, UNAM, Mexico City, Mexico

Greehey Children's Cancer Research Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

Sandra C. Viegas Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal

Networks of mRNA Processing and Alternative Splicing Regulation in Health and Disease

1

Peter Jordan, Vânia Gonçalves, Sara Fernandes, Tânia Marques, Marcelo Pereira, and Margarida Gama-Carvalho

Abstract

mRNA processing events introduce an intricate layer of complexity into gene expression processes, supporting a tremendous level of diversification of the genome's coding and regulatory potential, particularly in vertebrate species. The recent development of massive parallel sequencing methods and their adaptation to the identification and quantification of different RNA species and the dynamics of mRNA metabolism and processing has generated an unprecedented view over the regulatory networks that are established at this level, which contribute to sustain developmental, tissue specific or disease specific gene expression programs. In this chapter, we provide an overview of the recent evolution of transcriptome profiling methods and the surprising insights that have emerged in recent years regarding distinct mRNA processing events – from the 5′ end to the 3′ end of the molecule.

Keywords

Cancer · Regulatory networks · RNA processing · Splicing · Transcriptomics

1.1 Introduction

Human cells depend on the continuous expression of genes that encode the production of proteins and non-coding RNAs. Whereas some genes are continuously expressed, others require specific physiologic or developmental stimuli [1, 2]. The

P. Jordan (✉) · V. Gonçalves
Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal
e-mail: Peter.jordan@insa.min-saude.pt

S. Fernandes
Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Lisboa, Portugal

Department of Chemistry and Biochemistry, Faculty of Sciences, University of Lisbon, Lisboa, Portugal

T. Marques
Department of Chemistry and Biochemistry, Faculty of Sciences, University of Lisbon, Lisboa, Portugal

M. Pereira
Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

M. Gama-Carvalho (✉)
Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

Department of Chemistry and Biochemistry, Faculty of Sciences, University of Lisbon, Lisboa, Portugal
e-mail: mhcarvalho@fc.ul.pt

L. Romão (ed.), *The mRNA Metabolism in Human Disease*, Advances in Experimental Medicine and Biology 1157, https://doi.org/10.1007/978-3-030-19966-1_1

underlying mechanisms that regulate gene expression patterns are complex and operate in various layers along the molecular pathway leading from gene to gene product. One key step is the transcription of the gene by RNA polymerases into a primary RNA transcript, a process regulated at the levels of chromatin conformation at the gene locus and the binding of transcription factors to the gene promoter [3, 4]. This transcript then needs to be further processed in order to produce a mature RNA product, e.g. messenger RNA (mRNA), ribosomal RNA (rRNA), microRNA (miRNA or miR), small nuclear RNA (snRNA) or long non-coding RNA (lncRNA). This can involve various processing steps including RNA cleavage at specific sites, intron removal or splicing, or modifications at the RNA extremities, like the addition of a poly-adenosine tail to the 3′ end and a 7-methyl-guanosine 'cap' modification to the 5′ end of RNA polymerase II transcripts (Fig. 1.1). The final RNA product can undergo further post-transcriptional modifications, like the editing of selected nucleotides. All of these modification steps can have profound influence on the RNA life cycle, for example influencing mRNA properties such as half-life, nuclear export, or rate of translation at the ribosome [5, 6].

Recently developed massive parallel sequencing technologies have transformed our ability to identify gene regulatory networks by allowing for the genome-scale characterization of chromatin modifications, regulatory sequence elements, and transcriptome diversity and dynamics [7–10]. In this review we will discuss in more detail the current understanding of how post-transcriptional mRNA processing events contribute to transcriptome complexity and lead to functional networks of genes, both in normal physiology and in diseases like cancer.

1.2 New Technologies – New Insights: Contributions of Large-Scale Methods to Understanding Gene Expression Networks

The concept of transcriptome was used for the first time in a 1997 publication describing the application of the SAGE – Serial Analysis of Gene Expression – technology [11] to perform a detailed characterization of the universe of genes expressed in yeast [12]. Although the general features of gene expression patterns had begun to be explored almost two decades before, with the demonstration of the existence of different classes of transcripts regarding expression levels and stability, the detailed and systematic characterization of transcript abundance, structure and modifications required the much more recent development of highly parallelized methods. Following the generalization of gene expression profiling using microarrays, including splicing-sensitive arrays [13, 14], setting the stage for the first comprehensive characterization of transcriptome level regulatory networks, the advent of third generation sequencing methods and their application to transcriptome sequencing – or RNA-Seq – has had a tremendous impact on our current understanding of gene expression processes [7, 15]. In the wake of the new transcriptomic approaches, large international consortia aiming at the functional characterization of the human genome, like ENCODE (ENCyclopedia Of DNA Elements) [16], FANTOM (Functional ANoTation Of Mammalian genomes) [17], GTEx [18] and The Cancer Genome Atlas – TCGA [19], started providing an unprecedented view of the organization of gene expression processes and even re-defining our concept of gene. The currently accepted definitions have moved beyond the reference to genomic regions with transcriptional activity to focus on the biological function of the transcripts, either in the form of protein or RNA, establishing the determination of functionality as the critical challenge to overcome [20]. In spite of all advances, the accurate characterization of transcript structure and diversity is an ongoing pursuit that still requires the effective implementation of methods for the sensitive sequencing of full-length transcripts. Single Molecule Real Time Sequencing (SMRT) approaches such as the ones pioneered by PacBio [21] and the more recent solutions developed by Oxford Nanopore, supporting the high throughput direct sequencing of intact RNA molecules including the identification of base modifications [22] are promising to change the rules of the field. Coupled to the consolidation of single cell

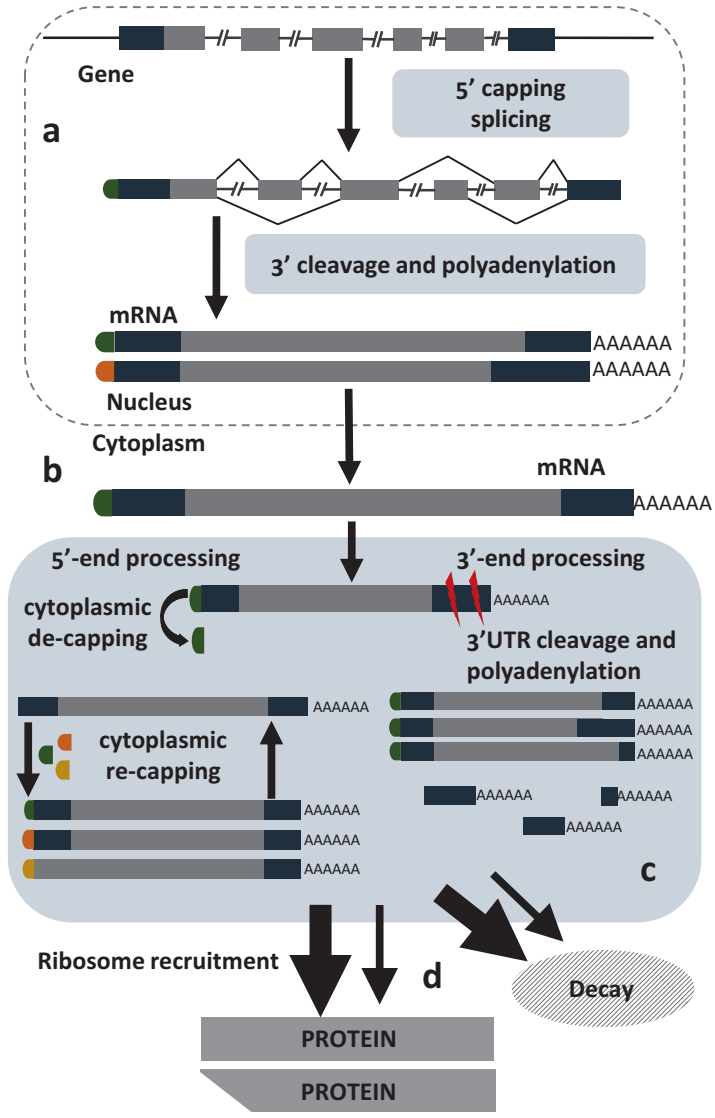

Fig. 1.1 Global view of differential co-transcriptional and post-transcriptional mRNA processing events discussed in this chapter. (**a**) During transcription, the pre-RNA undergoes co-transcriptional modifications such as the addition of the 5′-cap, alternative splicing and 3′-end cleavage and polyadenylation, all of which can be differentially regulated to define functional networks of transcripts. In the cytoplasm, several types of post-transcriptional modifications can also take place: the 5′-end can be processed through a cycle of de-capping and recapping, with the possibility of differential cap usage (**b**) and the 3′-end can be processed through cleavage and shortening of the 3′ untranslated region (UTR) associated with alternative polyadenylation, and the generation of stable, uncapped mRNA tails (**c**). These modifications contribute to the diversification of molecules present in the cell. The different sequence features arising from post-transcriptional modifications can dictate the efficiency of either protein translation or decay for each transcript or population of transcripts (**d**)

transcriptomics [23], these advances hallmark a new and exciting era for this expanding field.

1.2.1 The Basics: Characterizing the Transcriptome with RNA-Seq

The advent of highly parallelized sequencing methods with the ability to perform the simultaneous identification of the nucleotide sequence of millions of molecules in a quantitative way opened for the first time the possibility for the unbiased mapping of the transcriptome of different organisms, tissues or cell types [7]. However, the specific technical requirements for the implementation of these methods have a significant impact on their ability to achieve an accurate description of the RNA molecules present in each system. Taken to the letter, this would call for the ability to quantitatively identify full-length transcripts with differing structural arrangements (for example, with and without poly-A tails), varying from ten to hundred thousand nucleotides in length, with abundances that differ in the same orders of magnitude. Currently, there is no method that will support such a challenging application in a single assay. In fact, even the less ambitious aim of mapping the protein coding transcriptome is still faced with significant challenges due to technical limitations. Most of these limitations are related to the critical step called 'library preparation', whose complexity and impact on the final sequencing profile is often under-appreciated [24, 25]. In fact, given the high robustness and low error rate of the predominant massive parallel sequencing technologies, this is indeed the most critical step, during which significant biases can be introduced that may alter the relative representation of different RNA species in the sample, thus interfering with the accurate quantification of the target transcriptome [26]. A first problem to consider is the need to enrich the sample for the desired target in order to achieve appropriate sampling or "sequencing depth" for robust quantitative analysis [27].

Compared to the previous generation of hybridization-based methods for transcriptome analysis, sequencing based methods bring a novel concept – quantification is performed in a digital fashion, involving the detection (1) or absence of detection (0) of a given transcript. This implies that these methods are highly sensitive to the depth of sampling, which is reflected in the concept of "sequencing depth", i.e. the number of sequences or "reads" acquired per sample. Considering that ribosomal RNA makes for over 90% of the cellular transcriptome, its removal is imperative to achieve proper sequencing depth for other transcript families, either coding or non-coding [25]. Second, the sequencing-by-synthesis approach used by the predominant technologies requires the conversion of the target RNA molecules to their cDNA counterparts, i.e., a reverse-transcription step. Furthermore, since these methods can only sequence relatively short segments of DNA – usually between 50 and 300 base pairs (bp) and thus usually called "fragments" or "reads" – sample fragmentation is necessary to ensure an appropriate coverage across all of the transcriptome. Several approaches have been developed to try to achieve the ideal aim of a random fragmentation of the sample, generating a uniform distribution of molecular sizes, each of them linked to different problems and biases [25, 28, 29].

The need for fragmentation further leads to another set of problems that must be addressed. On the one hand, because of fragmentation, the likelihood of detecting a given transcript or isoform will not only be proportional to its abundance in the sample, but also to its size, as a larger molecule will generate a higher number of signature fragments. On the other hand, the reconstruction of the actual transcript isoforms present in the sample and the mapping of variations in splice site usage become very significant data analysis challenges. Finally, the actual sequencing reactions require that the target molecules are tagged on both ends through the addition of adaptors. Again, different technical solutions, each introducing different types of bias, have been developed [25]. The limited efficiency of all the

steps involved in library preparation requires the use of PCR-based amplification steps, which can lead to changes in the relative representation of the RNA species present in the original sample [30, 31]. Depending on the specificities of the application, the design of the adaptors can provide additional value, such as supporting strand-specific sequencing, which distinguishes the orientation of the original RNA molecule present in the sample; paired-end sequencing, which supports sequencing from both ends of the fragment, increasing the sample coverage per assay and providing additional information on the structure of the transcript; and the ability to identify fragments that have been duplicated during the library amplification steps [25].

The requirement for all these steps to produce a library that is appropriate for sequencing can obviously lead to multiple deviations from the composition of the original transcriptome. After the sequencing step, these problems will be compounded by the difficulties faced in the data analysis. In the simpler cases, the data analysis step aims to generate an image of the relative abundance of each transcript in the transcriptome, typically represented by an 'expression count table', where each known gene, transcript or exon ID is associated to the number of sequenced reads. This process generally relies on (1) the mapping of the sequenced reads to a known genome (although it is possible to perform direct transcriptome assembly); (2) the identification of uniquely annotated regions corresponding to the reads (thus creating problems in the case of transcripts or partial sequence elements that are encoded in multiple loci); and (3) the tentative reconstruction of the transcript structure from the assembly and parsimonious distribution of reads, in particular those mapping on exon-exon junctions [32]. Multiple data analysis pipelines exist that try to find accurate and efficient solutions for these problems, many of them focusing on the characterization of differential expression states between samples – for example, differences in expression at the gene, exon or splice-junction level, thus foregoing the need to generate an accurate representation of the transcriptome [33].

In spite of all the challenges highlighted above, RNA-seq studies are effectively generating an unprecedented detailed view of the mammalian transcriptome, with multiple adaptations of the library preparation methods that allow researchers to move beyond the characterization of expression steady states and provide detailed insights into gene expression dynamics and the underlying regulatory mechanisms and networks (Fig. 1.2).

1.2.2 Mapping Interactions, Dynamic Processes and Base Modifications

Standard transcriptomic approaches, either using RNA-seq or the previous microarray-based technologies, only allow for the steady-state characterization of transcript levels. However, since the advent of these methodologies, multiple adaptations to the sample preparation method have been implemented to allow researchers to obtain relevant information regarding both the regulation and dynamics of RNA processing events. These approaches generally involve the use of selective transcript enrichment, transcript labeling or transcript modification approaches. There are currently well over 50 different variations of RNA-seq based methods, for which we provide a few examples covering the different types of approach and discuss their application to the study of specific RNA processing events.

Selective transcript enrichment methods can be used for focusing RNA sequencing libraries on a target population or region of interest, gaining bigger sampling depths. In fact, most RNA-seq assays for gene expression analysis actually rely in the enrichment for mRNA molecules based on the use of oligo-dT molecules coupled to beads to capture poly-adenylated RNA. Alternatively, a similar process can be used to perform depletion of ribosomal RNA and thus perform a "negative" enrichment for other molecules [25].

More elaborate enrichment methods have been used to specifically capture and enrich samples for the starts and ends of transcripts in order

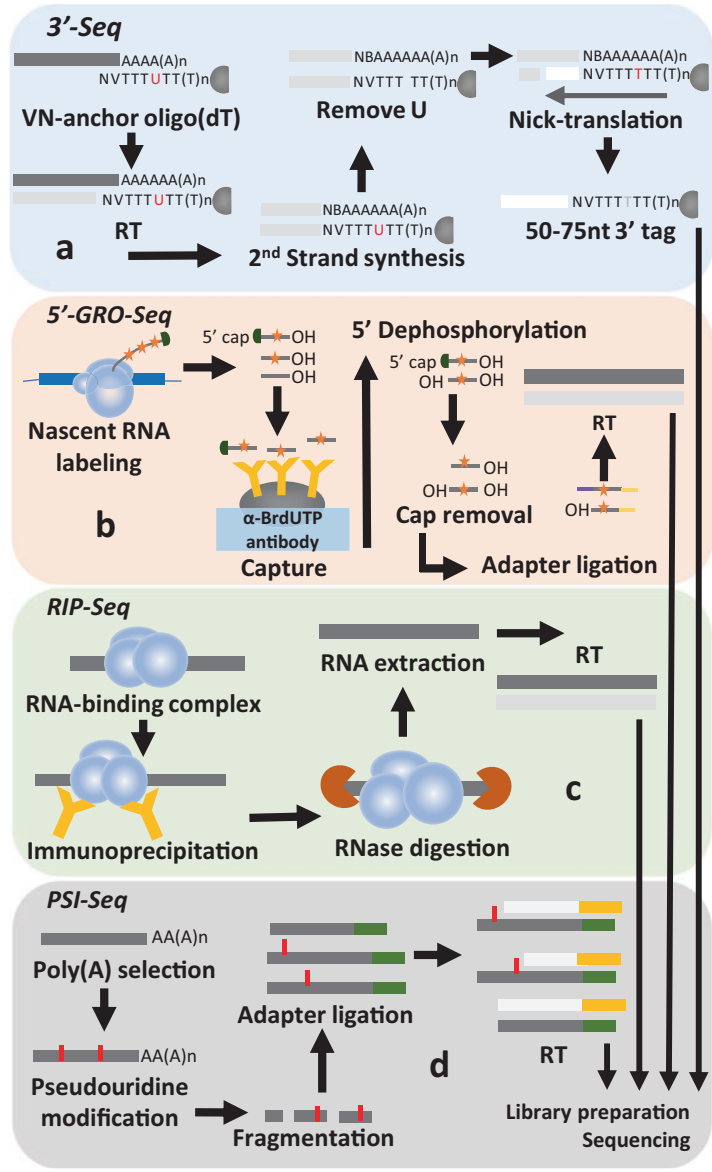

Fig. 1.2 RNA-seq derived methods for global analysis of mRNA processing. Four methods representative of the different types of strategy behind RNA-seq based approaches to characterize transcriptome diversity, dynamics and regulation. (**a**) 3′-Seq uses a poly-A based capture and controlled cleavage approach to focus on the identification of poly-A sites; (**b**) 5′-GRO-Seq uses pulse labeling and 5′-end capture to look at transcription start sites; (**c**) RIP-Seq relies on affinity purification of RNA-protein complexes to map RBP targets and binding sites; (**d**) PSI-Seq uses chemically based modification of pseudouridines to support the mapping of their positions on mRNA molecules. A more detailed description of the methods is presented in the main text. *RT* reverse transcription

to generate focused data on these regions. Among this family of applications, methods that specifically target the 3′UTRs have been critically important in the global characterization of 3′-end processing. The analysis of 3′UTRs typically relies on the capture of RNA molecules containing poly-A tails and identification of the immediate upstream region. However, due to the frequent presence of internal poly-A stretches in mRNA molecules, more complex approaches than just sequencing after poly-A based reverse transcription have been developed. For example, 3′-Seq allows for highly quantitative profiling of 3′UTR isoform expression by using a nick-translation step after reverse transcription to specifically capture the terminal 50–70 nt upstream of the poly-A tail [34] (Fig. 1.2a).

Capture approaches can also be applied to provide increased focus on transcript 5′ ends, thereby allowing a precise mapping of transcription start sites (TSS). The first high-throughput approaches for characterizing transcription

initiation at the genome level were derived from the CAGE method (Cap-Analysis of Gene Expression) [35–37], which uses a chemical-based oxidation method to promote the biotinylation of RNAs at the 5′-cap site, trapping associated first-strand cDNAs in an ensuing purification step [38]. Alternative approaches rely on the replacement of the cap structure with oligoribonucleotides [39–41]. This method consists of removing the 5′ cap with tobacco acid pyrophosphatase followed by ligation of oligos to the decapped mRNAs. The prior removal of 5′-phosphates of non-capped RNAs with alkaline phosphatase makes this oligo-capture specific to the 5′ cap. Following cDNA synthesis, libraries compatible with massive parallel-sequencing methods are generated to profile TSS with nucleotide level resolution. The combined use of these methods on the same sample has been applied to the detection and relative quantification of capped and uncapped mRNAs in different experimental conditions (see below).

By conjugating capture methods with the labeling of transcripts that are undergoing active transcription, it becomes possible to focus on dynamic processes while improving the ability to detect unstable transcripts. This approach was pioneered by Core and co-workers, who developed the GRO-Seq (Global Run-On) method, using the ribonucleotide analog 5-bromouridine 5′-triphosphate to label and purify newly synthesized RNA before preparing sequencing libraries [42]. This method was then refined to focus on the mapping of TSS by coupling it to 5′ end oligo-capture: the GRO-Cap [40] or 5′GRO-Seq [41, 43] methods (Fig. 1.2b). Global measurements of mRNA stability can be achieved by similar techniques, using for example 4-thiouridine (4sU) to pulse label newly synthesized RNAs, followed by a chase period. Newly synthesized and pre-existing mRNA fractions can then be separated on the basis of incorporated 4sU and processed according to an mRNA sequencing protocol for comparison with the total mRNA fraction in a time-course experiment [44]. Profiling of mRNAs engaged in active translation can be performed using ribosome profiling or Ribo-Seq [45], an adaptation of the older concept of polysome profiling [46] that allows for precise mapping of ribosome position. Together, these approaches have highlighted the importance of translational control by revealing significant differences in the dynamics of translating and steady-state mRNA levels (see below).

Alternative RNA processing is generally determined by the recognition of regulatory sequence elements by RNA binding proteins (RBPs). Methods to identify the interactions between these proteins and their RNA targets are based on the immunoprecipitation of the protein under conditions that preserve its interactions with RNA molecules (RNA Immuno-Precipitation or RIP), followed by the identification of the bound molecules. This was originally performed using microarrays – RIP-Chip [47, 48] – and was later adapted to RNA-seq methods – RIP-Seq [49]. RIP-Seq may be conjugated with RNAse digestion to provide positional information regarding the RBP binding site (Fig. 1.2c). However, this type of mapping is better suited for situations where the target RNA and protein are stably bound, for example using UV Cross-Linking Immunoprecipitation coupled to sequencing, as in the CLIP-Seq method [50].

The systematic mapping of RNA-protein interactions using genome wide approaches has become a central approach for the study of post-transcriptional regulatory mechanisms in health and disease [51, 52]. The functional annotation of mRNA targets and analysis of enriched sequence motifs has supported the post-transcriptional operon hypothesis [53], which proposed a role for RBPs in the establishment of functional gene expression networks, as well as the development of regulatory network models that describe the mechanisms underlying the coordinated regulation of pre-mRNA splicing [54, 55].

The prevalence and relevance of RNA base modifications on the post-transcriptional control of gene expression has become increasingly clear in the past few years [56]. RNA-seq methods have been adapted to support the mapping of these modifications in different cellular RNAs. These methods either rely on the use of antibodies that recognize the modified RNA bases like N6-methyladenosine (m6A) [57], or on the

chemical treatment of RNA molecules to specifically target the modified nucleotide. For example, Pseudouridine Site Identification sequencing (PSI-seq) uses N-Cyclohexyl-N′-(2-morpholinoethyl) carbodiimide to selectively modify pseudouridine residues, effectively halting reverse transcription during library preparation [58]. The positions of pseudouridines can then be mapped by sequencing of the truncated cDNAs and comparison with the annotated transcriptome (Fig. 1.2d). Using a very similar approach, pseudouridylation was detected in >200 yeast and human mRNAs involved in the response to nutrient deprivation [59]. These methods are opening the doors to the world of epi-transcriptomics, whose relevance and impact in mRNA metabolism have just began to be uncovered. RNA modifications are the object of a recent review [60] and, together with RNA editing, will not be explored in-depth in this chapter.

1.2.3 Emerging Views of Transcriptome Complexity

The first profiling studies of mammalian transcriptomes using RNA-seq began generating an unprecedented characterization of the abundance and diversity of RNA species, opening the doors to a deeper understanding of the underlying gene regulatory networks [7, 15, 61, 62].

Given the historical dependence of transcriptome profiling methods on 3′-end capture, systematic mapping of TSSs at single nucleotide resolution proved to be a challenging task in the early days of genomics. The first steps towards the genome wide unravelling of mammalian promotor regions were pioneered by the developers of the CAGE method [63], and later by the FANTOM consortium, who performed a systematic characterization of TSSs across the human and mouse genomes, generating the first comprehensive promoter landscape of a mammalian genome [35] and revealing for the first time the existence of generalized anti-sense transcription [64]. The methodology was then adapted for massive parallel sequencing approaches [17], providing a detailed view of transcriptional activity across a very large number of human and mouse tissues and cell types [65]. In a complimentary approach, the ENCODE consortium focused on generating systematic maps of the underlying regulatory events on selected cell lines, including chromatin modification and transcription factor binding sites [16]. The ENCODE project's high-depth analysis revealed that, in contrasts with the prevailing view of junk-DNA littered genomes, approximately 76% of human DNA sequences are actually transcribed to RNA, further reporting the identification of 8800 and 9600 genes encoding for small and long noncoding RNA molecules, respectively [16].

The conjugation of such RNA-seq based methods for genome wide mapping of TSSs using nucleotide level resolution with transcript labelling approaches like Gro-Seq, led to increased sensitivity for the detection of highly unstable molecules [42]. The results where surprising: in addition to revealing that transcription extends beyond the pre-mRNA 3′ cleavage site and confirming the prevalence of antisense transcription, these studies showed that most human promoters have an engaged polymerase upstream, moving in the opposite orientation to the annotated gene. Later work confirmed that bidirectional transcription occurs at these locations, with the productive sense being defined with the help of the pre-mRNA processing machinery. Indeed, the U1 snRNP was shown to engage the sense transcript in productive splicing versus systematic cleavage of the newly synthesized products at interspersed polyadenylation signals on the antisense direction [66]. Analysis of RNA-seq data across mammalian transcriptomes suggests that alternative transcription start and termination sites may play a greater role in transcript diversification than alternative splicing [67, 68]. However, by looking at steady state data, these analyses may fail to capture the dynamic processes that generate such variations, assuming their transcriptional, rather than post-transcriptional nature (see below).

1.3 Differential 5′-End Processing

All RNA polymerase II transcripts, both coding and non-coding, are subjected to a modification of their 5′-end termed "capping", which consists on the addition of a 7-methylguanosine nucleotide through an unusual 5′-PPP-5′ inverted bond. The basic cap structure, or "Cap0" is added co-transcriptionally in the nucleus by a complex of two proteins that performs three successive enzymatic activities: the removal the γ-phosphate from the 5′ triphosphate end of the transcript, the transfer of a GMP group to the resulting 5′ diphosphate, and the addition of a methyl group to the N7 amine of the guanine cap (reviewed by [69]). The capping enzyme (RNA guanylyltransferase and 5′ phosphatase; CE or RNGTT), catalyses the first two steps, while the third is performed by the RNA guanine-7 methyltransferase (RNMT), with the help of the activating subunit RNMT Activating Miniprotein (RAM). The subsequent 2′O methylation of the first nucleotide or of the first two nucleotides of the transcript leads to the formation of the "Cap1" or "Cap2" structure, performed by the Cap Methyltransferase enzymes, CMTR1 and 2, respectively. While the Cap0 structure has been known for a long time to be central to the protection of mRNA from exonucleases and in the recognition of mRNA molecules for export and translation [70], the 2′O ribose methylations play a critical role in the discrimination between self and viral RNA molecules by the innate immune system [71, 72]. Given the central role of the 5′-cap structure in mRNA metabolism, this RNA processing event was traditionally considered as a core and irreversible step in the mRNA biogenesis pathway. Recent results, however, have started to suggest that there is a hidden layer of differential regulation linked to 5′-capping that can not only have a profound impact on transcriptome diversity but also contribute to the definition of networks of co-regulated RNAs with significant impact in health and disease.

1.3.1 Differential Cap Modification

RNA capping is a powerful processing event with profound impact on the life of mRNA molecules, namely regarding their recruitment to the ribosome under standard translation conditions. Interestingly, the formation of the Cap0 structure has been shown to be regulated by the Myc and E2F1 transcription factors, being a critical requirement for cell transformation [70]. Early studies characterizing c-myc transcriptional control led to the surprising observation that its transactivation domain was able to promote cellular proliferation through a transcription independent increase in CDK7/9 levels [73]. This was shown to occur through increased cap methylation, which in turn led to enhanced recruitment to the ribosome and ensuing increase in translation efficiency. This behaviour was shared by a subset of other Myc transcriptional targets, thus demonstrating for the first time the differential impact of regulated 5′ cap methylation events. This observation was in agreement with previous results showing a Myc-dependent enhancement of VEGF mRNA translation [74]. Later studies have consolidated these observations, showing a similar pattern for the E2F1 transcription factor [75]. Thus, these transcription factors regulate CTD-phosphorylation leading to enhanced recruitment of active RNGTT enzyme and, consequently mRNA capping of specific mRNA targets critical for cell transformation [70, 76].

In addition to RNGTT, the RNMT enzyme function is known to be regulated by the CDK1B cyclin, allowing for the coordination of cap methylation with G1 phase transcription [77]. This regulatory event does not have a significant impact on steady state mRNA levels but controls the recruitment of specific cellular mRNAs to the translating ribosome, including molecules encoding proteins involved in cell cycle control and apoptosis, regulating cell proliferation. The RNMT Activating Miniprotein RAM has further been shown to be the target of ERK-dependent phosphorylation, leading to enhanced ubiquitination

and degradation. This regulatory event has a downstream effect on specific networks of transcripts involved in the regulation of pluripotency and neuronal differentiation [78]. Thus, high levels of active RAM were shown to be required for the expression of pluripotency genes in embryonic stem cells, including Sox2, Oct4 and Klf4, whereas RAM inhibition promotes the upregulation of neuronal-specific genes required for neural differentiation. Interestingly, although RAM was shown to have a global positive impact on mRNA translation levels, the gene specific effects were only observed at the level of transcript abundance. This could occur either as a consequence of reduced transcript stability or enhanced RNA Pol II transcription. Thus, the core mRNA capping machinery is subject to differential regulation of its activity, with an impact on specific transcript networks that are linked to proliferation, differentiation and transformation events [70].

In addition to Cap0 formation, the events leading to Cap1 and Cap2 structures are also regulated, resulting in the differential usage of these modifications. Although 2′O methylation of the first RNA nucleotide that defines the Cap 1 structure is prevalent among mRNAs, being critical for the immune recognition of self-RNAs as well as for translation, recent work suggests that the DNMT1 enzyme is regulated by DHX15 [79]. The DNMT1-DHX15 interaction controls the ribosomal recruitment and translation of a subset of cellular mRNAs critical for proliferation, thus providing the first example of a Cap 1-dependent gene regulatory network. In contrast to the generalized presence of the 2′O methylation on the first nucleotide, the modification of the second nucleotide leading to the Cap2 structure was only detected in about half of the cellular mRNAs in human cells [80, 81]. These studies, dating back to the 1970's, suggest that Cap2 formation is a primarily cytoplasmic event, leading to enhanced recruitment of the modified mRNAs to active ribosomes. The actual mRNA populations involved in this process have yet to be identified, but these observations clearly suggest another potential regulatory layer for cap-dependent differential expression [82].

In addition to these nucleotide modifications, first nucleotide adenosine 6-methylation is present in 20–30% of the mRNAs containing the Cap 1 structure (m^6Am) [83]. The functional consequences of this modification are unclear but given that different cap binding proteins and RBPs display different affinities for specific nucleotides and structures, it is possible that it has a differential impact on mRNA translation efficiency and may support the establishment of distinct regulatory networks [70]. In agreement with this hypothesis, the abundance of first nucleotide m^6Am on specific mRNAs was shown to vary across different mouse tissues [84]. Furthermore, given that this modification is nucleotide dependent, these observations imply that the choice of the first nucleotide of the transcript, either defined by the TSS or by post-transcriptional processing (see below) can have a significant impact on mRNA expression. In fact, a recent study has shown that cap-proximal nucleotides mediate a translational response program to cellular stress via alternative promotor usage and differential binding to the eukaryotic translation initiation factor eIF4E [85].

The development of RNA mass-spectrometry methods is supporting the identification of a previously unsuspected diversity of RNA 5′ cap structures and modifications [86, 87]. Among these is the recent identification of a 5′-end Nicotinamide Adenine dinucleotide cap in human cells, which was shown to act to destabilize mRNA molecules [88–90]. These discoveries are in line with the recent explosion of results regarding the presence of RNA nucleotide modifications in the body of mRNA and ncRNA molecules and their impact on gene expression processes, which have been recently reviewed in [56, 60].

1.3.2 Cytoplasmic Re-capping

Removal of the mRNA cap structure is one of the critical events leading to XRN1-dependent mRNA degradation [91]. However, several studies aiming at the identification of mRNA molecules targeted by endonucleolytic cleavage started hinting at the fact that stable un-capped

mRNA molecules are present in eukaryotic cells [92–94]. These studies relied on the capture of polyadenylated mRNAs with monophosphorylated 5′ ends coupled to either microarray analysis or RNA-Seq. As a consequence, highly regulated populations of uncapped mRNA molecules where identified, both in plants and mammals [92, 94]. In the former case, the mRNA molecules were found to display significant functional enrichment for regulatory proteins as well as characteristic sequence elements in their 5′-ends. These observations led to the suggestion that regulated post-transcriptional mRNA cleavage might present as a novel mechanism to diversify the eukaryotic transcriptome and, in particular, as an energy-efficient alternative to transcription initiation regarding the ability to generate varying 5′ mRNA ends [94]. However, such model requires the demonstration that uncapped mRNAs can become translationally active rather than just representing intermediary degradation products. The identification of a cytoplasmic mRNA capping complex [95], later shown to be composed by the RNGTT capping enzyme, the NCK adapter protein 1 NCK1 and an as yet unidentified 5′-end kinase was a critical observation that supported the concept of mRNA re-capping as a mechanism for regulating gene expression [96, 97]. Ensuing genome wide studies by the same lab identified specific mRNA populations that are the target of cytoplasmic capping and showed that this was critical for their association with the translation machinery [98]. In this study, uncapped cytoplasmic mRNAs were identified through susceptibility to *in vitro* degradation by the 5′-3′ exonuclease XRN1 coupled to microarray profiling of treated/untreated samples. By comparing the transcript populations between cells expressing a cytoplasmic dominant negative mutant of the capping enzyme and control cells, the authors not only identified a relatively large number of uncapped transcripts that are present at steady state levels in mammalian cells but, more importantly, a population of uncapped molecules that only accumulates upon inhibition of the cytoplasmic capping complex [98]. The fact that these mRNA targets only appear without a cap upon inactivation of the cytoplasmic capping complex implies that they are regulated by de-capping and re-capping events. These authors further showed that cytoplasmic capping was required to maintain the transcript's association with actively translating ribosomes. Interestingly, functional enrichment analysis of these "recapping targets" revealed a significant association to the mitotic cell cycle and RNA localization processes, including a significant number of RNA metabolic proteins that are linked to motor neuron diseases (FUS and GLE1) and Fragile X syndrome (FMR1), several of the CNOT proteins, HNRNPD (Auf1), and both exosome subunits (EXOSC2, EXOSC4, and EXOSC9) and exosome-associated 30 exonucleases (EXOSC10 and DIS3). Furthermore, a relatively large population of mRNAs, with enriched GO term annotations for "nucleotide binding" and "protein localization", whose stability depends on the presence of an active cytoplasmic capping complex was also identified in this study. These correspond to the roughly 2500 genes whose uncapped transcripts are present at high levels in control cells and disappear when cytoplasmic capping is inhibited. This observation suggests that there is a population of "natively uncapped" cytoplasmic transcripts, as previously reported by [94, 99], which undergoes cycles of de-capping and re-capping that are critical for the stability of the transcripts, establishing a cyclic process that the author's termed "cap homeostasis" [98]. The Schoenberg lab went on to demonstrate that this process occurs independently of polyA tail shortening or lengthening, thus firmly establishing a novel regulatory mechanism whereby the translation of a subset of cellular mRNAs is regulated by active de-capping and re-capping cycles [100]. This process may be conjugated with differential RNA stability and, depending on the precise site of cleavage that leads to the generation of an uncapped mRNA, can have profound impact on transcriptome and proteome diversity. Possibilities include the generation of transcripts encoding the same protein but with shortened 5′UTRs influencing their translation efficiency or subcellular localization, N-terminal truncated proteins, or novel non-coding RNAs. The mechanisms that define these

alternative 5′ ends of the mRNAs still remain to be characterized in detail but include Drosha or Ago2 dependent cleavage or other forms of endonucleolytic cleavage and pausing sites for XRN1 5′-to-3′ degradation [97]. Since identified re-capping mRNAs display specific sequence characteristics, like increased usage of alternative 3′-end processing (see below), enrichment of AU-rich elements in their 3′UTR and increased poly-A tail modifications [98, 100], it is possible that cap homeostasis is differentially regulated across tissues and in response to specific stimuli. In fact, results from a recent study link cytoplasmic re-capping to Hedgehog signaling, a pathway that is dysregulated in a wide range of human diseases, including cancer and neurodevelopmental disorders [101]. Future work will be crucial to determine the extent of interaction between regulated mRNA 5′ capping and altered gene regulatory networks in development and disease.

1.4 Differential Exon Usage

The unexpected discovery that the coding information in eukaryotic genes is discontinuous, containing intervening non-coding segments in the primary RNA transcript [102], has not only led to the discovery of mRNA splicing, but also provided the foundation for our current understanding that, through a regulated process known as alternative splicing, one gene can give rise to multiple, functionally distinct transcripts. The potential to generate final mRNA products with different exon combinations can be observed in over 95% of human genes [15, 62, 103].

Constitutive and alternative splicing of pre-mRNA follow the same basic rules. Both rely on the spliceosome, a multi-protein ribonucleoprotein (RNP) complex that assembles in an ordered and stepwise fashion to complete the splicing reaction [104, 105]. Continuous efforts to purify and characterize the spliceosome protein composition have revealed that it contains five core small nuclear RNA containing RNPs and several additional regulatory protein factors [106]. The stepwise assembly of this complex confers a dynamic cyclic nature to the splicing process. One important aspect to ensure the specificity of the splicing reaction is the initial step of exon definition, in which consensus sequences at the exon-intron borders are recognized, as recently reviewed [107]. In addition, the efficiency with which the snRNPs are recruited to a given exon is also defined. This is achieved by virtue of short cis-regulatory sequence elements in the pre-mRNA, designated as splicing enhancers or silencers, to which regulatory splicing factors bind, either promoting or repressing exon recognition by snRNPs. Besides aiding in the correct choice of splice sites during constitutive splicing, the interplay between active transcription and the available sequence-specific splicing factors at a given moment in the cell will determine which exon-exon junctions are created during the splicing process and, thus, which mature mRNA is made by alternative splicing [108–110]. The actual combination of enhancing and silencing splicing factors, in addition to splice sites strength, local RNA secondary structures, chromatin modifications, splicing enhancers or silencers availability, exon/intron architecture and the speed of RNA polymerase II transcription, is what controls alternative splicing [111], as described in more detail below.

Several different types of alternative exon usage can occur. The most frequently observed type of event in human transcripts is the inclusion or skipping of individual (cassette) exons, accounting for about 45% of all alternative splicing events. A variation of this type of event can occur at the 5′ or 3′ end of the transcript when either alternative promoters determine distinct transcription initiation sites or when alternative terminal exons determine the use of different polyadenylation sites. Another event type found in 5–20% of alternative transcripts occurs in exons carrying more than one splice site so that their length, and thus the coding content, vary through the use of different 3′ or 5′ splice sites. Two other low-frequency events are the inclusion of one of two mutually exclusive exons or the retention of intronic sequence in the mRNA.

1.4.1 Functional Consequences of Alternative Splicing

A single primary gene transcript can be differentially spliced and yield distinct splice variants. One impressive example is the human Titin gene, encoding a structural sarcomere component. The gene is composed of 363 coding exons and at least 498 splicing variants were detected in human muscle [112]. Although such a complexity of transcripts derived from a single gene is not the rule, over 95% of human genes can generate at least two transcript variants through alternative splicing [15, 62, 103].

Splice variants can have two major outcomes on gene expression regulation. First, the expression levels of the affected gene can be rapidly reduced. So-called poison exons can be included or skipped, leading to a frame-shift and a premature stop codon in the resulting mRNA. This will be recognized by the cellular quality control machinery during the initial round of translation and lead to mRNA degradation through the nonsense-mediated RNA decay (NMD) pathway [113] (as reviewed in Chap. 3). In such cases, alternative splicing is an effective means to rapidly downregulate transcript levels expressed from a given gene without the requirement to alter the transcription rate. Prominent examples for this regulatory mechanism are the transcripts for some splicing factors of the SR (serine and arginine-rich) or hnRNPs (heterogeneous nuclear ribonucleoproteins) protein families [114, 115]. Moreover, a variety of alternative splicing events became detectable when NMD was impaired by depletion of UPF proteins in HeLa cells [116]. Alternative splicing coupled to the use of alternative polyadenylation sites (see below) can also result in the down-regulation of gene expression through the generation of transcripts with alternative 3′UTRs that differ in the presence of miR binding sites and, consequently, in the degree of translation inhibition [117, 118]. In addition, the presence of binding sites for RNA-binding proteins in specific exons can determine the efficiency with which a transcript variant is exported from the nucleus to the cytoplasm, is translated at the ribosome, or the efficiency of its turnover, thus affecting the resulting gene expression levels [119]. Second, alternative splicing can generate mRNA transcripts that encode protein isoforms with altered functional properties. For example, an exon can encode a functionally relevant protein domain or a subcellular localization signal. In this case, its inclusion or skipping will generate distinct in-frame transcripts encoding two related but functionally distinct proteins. Recent genome-wide data suggest that about 75% of alternative variants detected in human cells, especially the cassette-exon type, were also captured at ribosomes, suggesting their translation into protein-coding isoforms [120]. Many examples of functionally distinct variant proteins were described in relation to their development of cancer [121, 122]. For example, (i) the skipping of exon 3b to generate a hyperactive survival-promoting variant of the small GTPase RAC1 [123], (ii) the retention of exon 2 yields the anti-apoptotic variant BCL-X(L) [124], (iii) the skipping of exon 11 generates a receptor tyrosine kinase RON that stimulates invasion and cell motility [125], or (iv) the replacement of exon 9 with exon 10 in pyruvate kinase variant PKM2 promotes metabolic reprogramming of tumour cell glycolysis [126]. Through these possibilities, genes can encode different functional protein isoforms and this greatly expands the protein-coding capacity of the genome. Since nearly all human genes are alternatively spliced, cell-type-specific protein isoforms can modify cell fate or viability during normal and pathologic cell physiology.

1.4.2 Recent Findings on Recursive- and Back-Splicing

Two novel regulatory layers in transcripts processing have been more recently discovered. First, the consensus sequence motifs that define exon-intron borders were also found deep within introns. Here, they serve to remove long introns in a stepwise manner during the splicing process before, in a final step, the two flanking exons are joined together. This process is called recursive splicing and was identified as a conserved and

widespread mechanism in human [127, 128] and *Drosophila* tissues [129]. Although there is still much to be discovered about the molecular basis of recursive splicing, it can be expected to represent another regulatory step during alternative splicing, besides being a potential mutation target leading to human genetic diseases [130]. Second, a novel type of long non-coding RNAs (lncRNAs) can be generated by alternative splicing. Downstream splice donor sites were found to become reversely joined with an upstream splice acceptor site by the spliceosome (back-splicing). This results in a covalently closed circular RNA (circRNAs), which may contain intronic sequence or display alternative splicing patterns [128, 131]. CircRNAs are highly nuclease resistant, ubiquitously expressed in eukaryotic cells and display cell-type, tissue-type and developmental-stage specific expression patterns. It is yet unclear how the splicing machinery discriminates between either canonical, alternative or back-splicing [132–134] or even trans-splicing between transcripts from different genes [135]. Notably, circRNAs were found to have regulatory functions in gene expression, including the regulation of transcription, alternative splicing or translation. In particular, they can operate as competing molecular sponges that absorb and sequester miRs or form protein complexes with RNA-binding proteins or splicing factors [133, 136]. Like miRs, circRNAs were reported to exhibit altered expression under pathological conditions and may have great potential as novel diagnostic biomarkers as well as therapeutic agents for various diseases [137, 138].

1.4.3 Dynamic Regulation of Alternative Splicing Profiles

The identification of which genes are alternatively spliced at a given moment in a cell or tissue may convey the impression of a static and persistent pattern of the expressed transcriptome. However, the past few years have provided substantial evidence for a much more dynamic transcriptome plasticity, in particular due to the connection established between signal transduction pathways and corresponding responses at the splicing level [139–141]. From the mechanistic point of view, this dynamic nature relies on the combinatorial control of the splicing process. As mentioned above, splice site choice depends not only on consensus RNA sequence motifs at the exon-intron borders but is also enhanced or silenced by means of sequence-specific splicing factors that strengthen or weaken the recruitment of the spliceosome, respectively. The corresponding splice regulatory enhancer or silencer sequences can be located within exons or introns and will influence a given alternative splicing event in two possible manners: first, a regulatory RNA sequence can contain two overlapping binding motifs for different competing splicing factors, sometimes with antagonistic effects [142]. In this case, the absolute or locally available concentration of splicing factors in the cell will determine whether an enhancer or silencer effect prevails. For example, the sequence motif recognized by the enhancer SRSF2 overlaps that of the silencer hnRNP A1 in exon 2 of the HIV tat gene [143]. Second, a variety of regulatory RNA sequence elements are present to regulate a given alternative splicing event and the effective splicing outcome is determined by the net result of enhancing versus silencing factors bound in a given cell at any given moment. This was elegantly documented in experimentally designed exons containing different enhancer and silencer combinations [144]. Thus, the effective combination of enhancing and silencing splicing factors is what controls the cellular decision on alternative splicing [111]. This mechanism also provides an explanation for how simple single-nucleotide alterations can lead to disease-causing loss-of-function splicing mutations in patients [145, 146]. To completely understand how cells regulate alternative splicing, we need to determine both how individual SFs recognize and regulate a given splice event and how the combination of multiple SFs becomes integrated.

How can cells modulate the balance between enhancing and silencing splicing factors in order to modulate alternative splicing decisions?

First and on a longer term scale, alternative splicing profiles of a tissue or cell differentiation state can be stabilized by regulating the relative expression levels of antagonistic splicing factors. This implies transcriptional regulation of the respective genes but also includes the above described autoregulation by alternative splicing of SR protein or hnRNP family members [114, 115]. The same may apply to diseased tissues, and indeed altered SF levels have been widely described, for example, during tumorigenesis [121, 122, 147, 148]. An impressive observation was that overexpression of the SRSF1 gene exerts a clear oncogenic effect by disturbing the balance between cooperating or antagonizing splicing factors in genome-wide splicing decisions [149].

Second, epigenetic changes induced, for example, by environmental cues, were shown to influence splicing events. They operate through changes in DNA methylation and post-translational histone modifications to modulate the binding affinity among histones and between histones and DNA. This alters chromatin density and nucleosome occupancy, allowing transition between a more open or more condensed chromatin structure, which then affects the elongation rate of transcription by RNA polymerase II. A decreased elongation rate allows more time for the co-transcriptional binding of lower-affinity splicing regulators and results in alternative splicing [150]. In addition, certain histone modifications such as H3K36me3 can directly participate in the recruitment of splicing factors through intermediate chromatin-binding proteins [150, 151].

Third and of more dynamic nature, alternative splicing can be regulated by post-translational modification of SFs, especially their phosphorylation. These modifications generally represent a response to a cell signalling event. Growth factors, inflammatory cytokines or cellular stress conditions activate diverse signal transduction pathways including the RAS/ERK, PI3K or Wnt pathways and lead to changes in the phosphorylation state of a variety of proteins, including SFs [140, 141, 152]. For example, growth factor stimulation of the mitogen-activated protein kinase (MAPK) ERK phosphorylates the SF SAM68 and enhance its RNA-binding activity to the 3′UTR of the SRSF1 transcript [153]. This promotes intron retention and assures generation of a full-length SRSF1 transcript so that ERK activation results in increased levels of SRSF1 protein. Another outcome of SF phosphorylation can be changes in nuclear localization. Stress signals, for example, were shown to activate the p38-MAPK pathway and induce hnRNP A1 phosphorylation in the nucleus followed by its export into the cytoplasm [154, 155]. Such a decrease in nuclear SF abundance is sufficient to change alternative splicing patterns.

Drug treatment can also trigger a dynamic splicing response via signal transduction pathways. When colon cancer cells were treated with the non-steroidal anti-inflammatory drug ibuprofen, the tumour-related alternative splicing of RAC1b was rapidly repressed by a mechanism involving reduced phosphorylation of SRSF1 by its protein kinase SRPK1 [156].

1.5 Functional Transcript Networks via Alternative mRNA Processing

The ability to determine genome-wide changes in the transcriptome has led to the hypothesis that specific subsets of alternative splicing variants are co-regulated and that their functions in the cell contribute to the same biological processes or pathways. This has been designated as splicing networks. Indeed, alternative exons from such networks have been shown to contribute to critical biological functions [157, 158].

A common molecular mechanism underlying splicing networks may reside in the observation that alternative exons frequently encode disordered protein regions that are enriched in post-translational modification sites or contain conserved protein binding motifs. This way, tissue-specific splicing may allow tissue-specific

protein-protein interaction networks that are important for maintaining tissue identity. Accordingly, genes containing tissue-specific exons were found to occupy central positions in protein interaction networks or select distinct interaction partners [159, 160]. On one hand, splicing networks can define tissue-specific properties. In the brain, for example, tissue-specific expression of some RNA-binding proteins can regulate a wide spectrum of neuron-specific alternative splicing events [161, 162]. When analyzed in different tissues, the majority of alternative splicing or polyadenylation events presented with tissue-specific expression levels of the various isoforms, involved in defining tissue identity [62, 163, 164]. Other tissue-specific or developmental alternative splicing profiles were described in heart or skeletal muscle, and also in spermatogenesis, adipogenesis or stem cell pluripotency [164–166]. Tissue-specific splicing might not only contribute to the functional versatility of proteins in different tissues of multicellular organisms but also promote novel phenotypes or regulatory complexity during organismal evolution. On the other hand, the pathways targeted by splicing networks characterize disease states, including cancer development. Altered splicing networks were found in some tumour types to result from mutations in genes encoding splicing factors or core spliceosomal subunits [167, 168]. More frequently, however, tumours show a disrupted balance of the expression of antagonistic RNA-binding proteins and this affects a concerted set of transcripts acting in specific cellular responses. For example, overexpression of the SRSF1 gene was reported in tumours of the colon, breast and lung [149] and led to specific changes in alternative splicing. Also, the identification of epithelial splicing regulatory proteins 1 and 2 (ESRP1 and ESRP2) revealed a central role as regulators of a transcript network activated when cells undergo the epithelial–mesenchymal transition [169, 170]. Similarly, the transcription factor Myc is frequently overexpressed in tumours and was found to promote expression of the genes encoding the alternative splicing factors PTBP1 and HNRNPF. As a consequence, several splicing isoforms with a known role in colon tumorigenesis were upregulated [171]. These cancer-specific findings hold promise for the identification of biomarkers for tumour-stage, disease progression or therapy-response, but may also identify crucial targets for innovative RNA-based therapies able to correct splicing alterations. Similar principles apply to other diseases [172].

Altogether, genome-wide splicing networks have now been recognized as an important layer of post-transcriptional control of gene expression, besides the layers of transcriptional and epigenetic regulation.

1.6 Differential 3′-End Processing

A significant number of post-transcriptional regulatory events involve the 3′UTR region of mRNA molecules [173]. Although 3′UTRs are generally less conserved than protein coding regions, their degree of conservation supersedes that of other non-coding regions, such as promoters, introns and 5′UTRs, being notably conserved among vertebrates [174, 175]. Unlike DNA-based regulatory elements, which operate through their sequence, RNA-based regulatory motifs can exert their regulatory activity either through primary sequence motifs, secondary structure, or through a combination of both. These sequence and structural elements in the 3′UTRs have been shown to influence mRNA stability, transport, localization and translation efficiency primarily through interactions with regulatory RNAs and RBPs [173].

The most widely recognized group of regulatory RNAs to function in this context is a class of noncoding RNAs termed microRNAs (miRs). As the name suggests, miRs are small RNA molecules that target short sequence motifs (6–8 nucleotides) with perfect base-pairing to the so-called miR ‘seed’ region [176]. miR-3′UTR interactions typically lead to an inhibition of protein production, with miRs functioning predominantly by inducing mRNA decay, concomitant with translational inhibition [176–178]. Alongside with transcription and splicing,

miR-dependent regulation has been clearly shown to contribute to the establishment of regulatory network interactions that define critical developmental and cell differentiation programs, often tightly intertwined with transcription factor regulation. The structure and functional significance of these networks has been widely reviewed in the literature and will not be discussed in detail in the present chapter [179–181].

Interactions between 3′UTRs and RBPs are mediated by regulatory motifs whose sequence and structural features vary considerably. RBP-3′UTR interactions can bring about either a stimulation or an inhibition of protein production, with different RBPs exerting different effects over mRNA stability and translation efficiency, in a context-dependent manner [173, 182]. Furthermore, RBP-3′UTR interactions can facilitate mRNA transport and thus regulate transcript localization to the correct subcellular compartment [183]. More recently, 3′UTRs have been implicated in protein localization through the formation of scaffolds, mediating the establishment of protein-protein interactions [173, 184].

1.6.1 Alternative Polyadenylation

An additional level of complexity associated with post-transcriptional regulation ties in with the fact that a considerable number of genes give rise to mRNAs with alternative 3′UTRs through a process termed alternative polyadenylation (ApA). mRNA polyadenylation occurs co-transcriptionally and is triggered by a sequence motif termed the polyadenylation signal (PAS), regulated by upstream and downstream sequence elements [185]. The recognition of this signal by the polyadenylation machinery leads to the endonucleolytic cleavage of the precursor mRNA at the polyadenylation site, located 10–30 nucleotides downstream of the PAS, and subsequent addition of the poly-A tail [185]. Alternative polyadenylation occurs when more than one PAS is present. In most cases, these alternative PAS are located in tandem in the 3′UTR region, with ApA leading to the production of alternative transcripts that differ exclusively in the length of their 3′UTRs. These 3′UTRs are referred to as alternative 3′UTRs or ‘tandem 3′UTRs’ [185, 186]. Some transcript isoforms display alternative 3′-end formation due to the presence of PAS upstream of the last terminal exon, often in an alternatively spliced intron. When this intron is retained, the transcript will be cleaved and terminated at the first PAS; if an alternative splicing event takes place and the intron is removed, the transcript will be processed at a downstream PAS. These alternative UTRs resulting from ‘upstream-exon ApA’ are usually referred to as ‘alternative terminal exons’ [185, 186]. The selection of alternative polyadenylation signals determines the sequence content of the 3′UTR and thus the landscape of regulatory motifs in the mRNA that are available to interact with miRs and RBPs. Since longer 3′UTRs tend to have additional regulatory motifs, their regulatory potential is greater than that of shorter 3′UTRs. Thus, ApA can have a substantial functional impact on gene expression.

With the advent of transcriptomics and the subsequent development of RNA-Seq 3′-end focused methods, the high conservation and widespread prevalence of ApA has become apparent [186, 187]. Current estimates suggest that approximately half of the protein coding genes in the human, mouse, zebrafish, fly and worm present alternative 3′UTRs through ApA. Moreover, these studies have revealed that alternative PAS usage is regulated in both a developmental and tissue-specific manner [187]. The characterization of ApA during Drosophila [188], zebrafish [189] and mouse [190] development suggested the existence of a regulatory program that promotes a switch from proximal do distal polyadenylation sites, with longer 3′UTRs correlating with more differentiated states. Additionally, specific tissues appeared to follow a global pattern favoring mRNAs with shorter or longer 3′UTR isoforms [188–190]. This was particularly significant in the brain, which has longer 3′UTRs than any other adult tissue [34, 188]. Conversely, analysis of ApA events in cancer and proliferating cells revealed an apparent shortening of 3′UTR lengths [117, 191, 192]. These observations led to the proposal that global

programs for distal to proximal ApA selection would be in place to promote proliferative phenotypes, acting to liberate mRNA molecules from the regulatory control of miRs through the loss of miR targeting motifs that results from 3′UTR shortening. In contrast, transcripts expressed in differentiated cells would be subjected to a tighter control. This model was further substantiated by observations connecting shorter 3′UTR lengths to higher protein outputs, with striking examples regarding the expression of oncogenes, which on average show an increase of tenfold in protein levels estimated to result from the use of proximal PAS [117]. However, a more careful analysis of the prevalence and consequences of ApA based on the development of a highly quantitative 3′UTR RNA-seq method argues strongly against this [34].

The results obtained from the profiling of 14 human tissues and cell lines confirm the prevalence of ApA in human genes, with 7020 out of 13,718 genes (51%) classified as multi-UTR genes. 3′UTR isoforms were found to be present across all tissues. However, within a specific tissue or cell type, the relative proportion of these isoforms was found to change dramatically for some genes, which were termed "poly-adeno-modulated" (pAM) multi-UTR genes [34]. These changes occur both towards longer and shorter 3′UTRs, with the exception of the brain, where no tissue-specific usage of proximal PAS was observed. Most strikingly, this study revealed that tissue specific genes tend to have single 3′UTRs and be transcriptionally regulated, whereas pAM genes were found to encode ubiquitous regulatory proteins with tissue-specific functional signatures [34]. These observations support the current vision that ApA usage is regulated for specific subsets of genes at a tissue specific level, correlating with target mRNA escape from the regulation of ubiquitously expressed miRs. Of note, available evidence does not show a significant connection between UTR length, mRNA stability and protein expression levels [193, 194]. This is in agreement with the fact that, in spite of the well-established relevance of miR-3′UTR interactions, systematic surveys of 3′UTR sequence elements identify both positive and negative regulatory motifs regulating mRNA stability and protein translation [195, 196]. Furthermore, it is possible that the function of the prevalent longer alternative 3′UTRs found in some tissues, in particular in brain, is primarily related to RNA localization and localized translation. Taken together, these studies point to a central role for 3′-end processing in the establishment of coordinated gene expression programs at the tissue-specific level, involving a combination of distinct events depending on target and context, dismissing the concept of a global regulatory program controlling ApA [187].

1.6.2 Post-Transcriptional Processing at the 3′-End

Recently, the occurrence of post-transcriptional cleavage at proximal 3′UTRs coupled to polyadenylation was described [197]. This event leads to the formation of a functional mRNA with a shortened 3′UTR, and an autonomous, uncapped, 3′UTR RNA fragment in the cytoplasm. To determine the scale of this type of events, the authors performed a parallel sequencing study using 5′end and 3′end enrichment methods coupled to RNA-sequencing and compared the relative abundance of the coding and 3′UTR regions of the same gene. The results obtained suggest that thousands of autonomous stable, uncapped "RNA tails" generated by post-transcriptional cleavage of mRNA at alternative poly-A sites exist in the cytoplasm of human cells. These results are supported by previous observations reporting a non-random imbalance between 3′UTR and coding sequence and 3′ abundances in *in situ* hybridization studies in mouse embryos [198]. In several cases, the authors observed that 3′UTR abundance reached a threefold excess compared to the coding sequence, in a robust pattern that varied according to the gene and tissue, thus suggesting a regulated event. In line with this is the recent identification of a widespread accumulation of ribosome-associated 3′UTRs in neuronal cells, evidenced by a relative enrichment of UTR versus coding transcript regions [199]. However, in this case the authors associate

this event to brain aging and oxidative stress, in connection with ribosome stalling and mRNA cleavage by No-Go decay under conditions of deficient ribosome recycling.

To determine the potential functional impact of their observations, Malka and co-workers explored the interaction of two miRs, well known to be expressed in HEK cells and having multiple validated targets, with cellular mRNAs assuming their intact or cleaved nature. This was assessed by analyzing the effect of over-expressing the two miRs on the transcriptome using RNA-seq and correlating the observed impact on the expression levels of mRNA targets predicted considering the presence of miR binding sites on the mRNA molecule as a whole, or on the separate "body" and UTR "tail" parts. The authors found that mRNAs predicted to be miR targets based only on the presence of binding sites on the "body" part showed a significantly higher decrease in expression than the subset that was predicted to be targeted based on the presence of sites in the whole (uncleaved) mRNA molecule. These intriguing observations suggest that estimation of mRNA expression levels, in particular when based on microarray approaches that predominantly target the 3′UTR, may be profoundly biased, as well as studies focusing on the characterization mRNA-miR regulatory networks. Whether the 3′UTR tails have the ability to exert any regulatory roles remains to be established, but one possibility is that they may act as competing endogenous RNAs for post-transcriptional regulation [200]. Together with the complementary observations of cytoplasmic re-capping, these recent results shed new light on the transcriptome and call for a new perspective on the post-transcriptional regulation of genes involving a highly dynamic processing of mRNA molecules to generate new layers of regulatory complexity.

1.7 Future Tendencies

The ingenious application of RNA-seq methodologies to the characterization of transcriptome composition and the dynamics of transcriptional and post-transcriptional processes has generated a wealth of unexpected insights into regulatory networks underlying the robust but highly responsive behaviour of molecular processes in eukaryotic systems. Recent results have consolidated previous observations pinpointing the relevance of network-based gene expression modules and multi-layered interactions in the definition of cellular programs for normal cellular function, as well as for the establishment of disease phenotypes, in particular during cellular transformation. Additionally, they are revealing a whole wealth of unexpected diversity in both the players and the processes that sustain them. As the technologies for characterizing system components at a genome wide scale improve, new trends related to a more precise mapping of players, interactions and quantitative measurements of dynamic processes are likely to gain momentum. These include the development of methods for direct sequencing of intact mRNA molecules with identification of their nucleotide modifications, for example through the use of nanopore-based sequencing [22]; the expansion of methods for real-time and single-cell measurement; and the ability to integrate the multiple layers of information that have been generated into comprehensive models that explain and predict complex network behaviour. Given the current state of the art, it is highly likely that we will also witness the expansion of the current universe of post-transcriptional regulatory events, including the identification of new types of non-coding RNA elements – like the still mysterious circular RNAs [201] or the intronic Alu elements that regulate the activation of the anti-tumoral dsRNA-induced interferon response [202]; the identification of non-genetic roles of RNA – like the ability to modulate phase transitions in the cell [203]; and of new types of regulatory events – like the identification of novel information flows such as the recently reported transmission of 3′UTR-encoded genetic information to proteins [204]. The surprising nature of these recent observations foreshadows all the exciting discoveries that are waiting ahead in the quest for understanding the function and roles of post-transcriptional regulatory networks.

References

1. Weake VM, Workman JL (2010) Inducible gene expression: diverse regulatory mechanisms. Nat Publ Group 11:426–437. https://doi.org/10.1038/nrg2781
2. Pope SD, Medzhitov R (2018) Emerging principles of gene expression programs and their regulation. Mol Cell 71:389–397. https://doi.org/10.1016/j.molcel.2018.07.017
3. Heintzman ND, Ren B (2007) The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome. Cell Mol Life Sci 64:386–400. https://doi.org/10.1007/s00018-006-6295-0
4. Soshnev AA, Josefowicz SZ, Allis CD (2016) Greater than the sum of parts: complexity of the dynamic epigenome. Mol Cell 62:681–694. https://doi.org/10.1016/j.molcel.2016.05.004
5. Corbett AH (2018) Post-transcriptional regulation of gene expression and human disease. Curr Opin Cell Biol 52:96–104. https://doi.org/10.1016/j.ceb.2018.02.011
6. Schmid M, Jensen TH (2018) Controlling nuclear RNA levels. Nat Publ Group 19:518–529. https://doi.org/10.1038/s41576-018-0013-2
7. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484
8. Blencowe BJ, Ahmad S, Lee LJ (2009) Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. Genes Dev 23:1379–1386. https://doi.org/10.1101/gad.1788009
9. Furey TS (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Publ Group 13:840–852. https://doi.org/10.1038/nrg3306
10. Daniel B, Nagy G, Nagy L (2014) The intriguing complexities of mammalian gene regulation: how to link enhancers to regulated genes. Are we there yet? FEBS Lett 588:2379–2391. https://doi.org/10.1016/j.febslet.2014.05.041
11. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270:484–487
12. Velculescu VE, Zhang L, Zhou W et al (1997) Characterization of the yeast transcriptome. Cell 88:243–251. https://doi.org/10.1016/S0092-8674(00)81845-0
13. Clark TA, Sugnet CW, Ares M (2002) Genome wide analysis of mRNA processing in yeast using splicing-specific microarrays. Science 296:907–910. https://doi.org/10.1126/science.1069415
14. Johnson JM, Castle J, Garrett-Engele P et al (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science 302:2141–2144. https://doi.org/10.1126/science.1090100
15. Pan Q, Shai O, Lee LJ et al (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40:1413–1415. https://doi.org/10.1038/ng.259
16. Consortium TEP (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74. https://doi.org/10.1038/nature11247
17. Consortium TF, PMI TR, DGT C (2014) A promoter-level mammalian expression atlas. Nature 507:462–470. https://doi.org/10.1038/nature13182
18. GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) project. Nat Genet 45:580–585. https://doi.org/10.1038/ng.2653
19. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA et al (2013) The cancer genome atlas pan-cancer analysis project. Nat Genet 45:1113–1120. https://doi.org/10.1038/ng.2764
20. Gerstein MB, Bruce C, Rozowsky JS et al (2007) What is a gene, post-ENCODE? History and updated definition. Genome Res 17:669–681. https://doi.org/10.1101/gr.6339607
21. Sharon D, Tilgner H, Grubert F, Snyder M (2013) A single-molecule long-read survey of the human transcriptome. Nat Biotechnol 31:1009–1014. https://doi.org/10.1038/nbt.2705
22. Garalde DR, Snell EA, Jachimowicz D et al (2018) Highly parallel direct RNA sequencing on an array of nanopores. Nat Methods 15:201–206. https://doi.org/10.1038/nmeth.4577
23. Kolodziejczyk AA, Kim JK, Svensson V et al (2015) The technology and biology of single-cell RNA sequencing. Mol Cell 58:610–620. https://doi.org/10.1016/j.molcel.2015.04.005
24. Amaral AJ, Brito FF, Chobanyan T et al (2014) Quality assessment and control of tissue specific RNA-seq libraries of Drosophila transgenic RNAi models. Front Genet 5:43. https://doi.org/10.3389/fgene.2014.00043
25. Hrdlickova R, Toloue M, Tian B (2017) RNA-Seq methods for transcriptome analysis. Wiley Interdiscip Rev RNA 8:e1364. https://doi.org/10.1002/wrna.1364
26. SEQC/MAQC-III Consortium (2014) A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat Biotechnol 32:903–914. https://doi.org/10.1038/nbt.2957
27. Sims D, Sudbery I, Ilott NE et al (2014) Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet 15:121–132. https://doi.org/10.1038/nrg3642
28. Adey A, Morrison HG, Asan et al (2010) Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol 11:R119. https://doi.org/10.1186/gb-2010-11-12-r119
29. Poptsova MS, Il'icheva IA, Nechipurenko DY et al (2014) Non-random DNA fragmentation in next-

generation sequencing. Sci Rep 4:4532. https://doi.org/10.1038/srep04532
30. Aird D, Ross MG, Chen W-S et al (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol 12:R18. https://doi.org/10.1186/gb-2011-12-2-r18
31. Parekh S, Ziegenhain C, Vieth B et al (2016) The impact of amplification on differential expression analyses by RNA-seq. Sci Rep 6:25533. https://doi.org/10.1038/srep25533
32. Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet 12:671–682. https://doi.org/10.1038/nrg3068
33. Conesa A, Madrigal P, Tarazona S et al (2016) A survey of best practices for RNA-seq data analysis. Genome Biol 17:13. https://doi.org/10.1186/s13059-016-0881-8
34. Lianoglou S, Garg V, Yang JL et al (2013) Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev 27:2380–2396. https://doi.org/10.1101/gad.229328.113
35. Carninci P, Kasukawa T, Katayama S et al (2005) The transcriptional landscape of the mammalian genome. Science 309:1559–1563. https://doi.org/10.1126/science.1112014
36. FANTOM Consortium, Suzuki H, Forrest ARR et al (2009) The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. Nat Genet 41:553–562. https://doi.org/10.1038/ng.375
37. Hoskins RA, Landolin JM, Brown JB et al (2011) Genome-wide analysis of promoter architecture in Drosophila melanogaster. Genome Res 21:182–192. https://doi.org/10.1101/gr.112466.110
38. Takahashi H, Lassmann T, Murata M, Carninci P (2012) 5′ end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. Nat Protoc 7:542–561. https://doi.org/10.1038/nprot.2012.005
39. Arribere JA, Gilbert WV (2013) Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. Genome Res 23:977–987. https://doi.org/10.1101/gr.150342.112
40. Kruesi WS, Core LJ, Waters CT et al (2013) Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. elife 2:e00808. https://doi.org/10.7554/eLife.00808
41. Core LJ, Martins AL, Danko CG et al (2014) Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet 46:1311–1320. https://doi.org/10.1038/ng.3142
42. Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322:1845–1848. https://doi.org/10.1126/science.1162228
43. Duttke SHC, Lacadie SA, Ibrahim MM et al (2015) Human promoters are intrinsically directional. Mol Cell 57:674–684. https://doi.org/10.1016/j.molcel.2014.12.029
44. Schwanhäusser B, Busse D, Li N et al (2011) Global quantification of mammalian gene expression control. Nature 473:337–342. https://doi.org/10.1038/nature10098
45. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223. https://doi.org/10.1126/science.1168978
46. Arava Y, Wang Y, Storey JD et al (2003) Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc Natl Acad Sci 100:3889–3894. https://doi.org/10.1073/pnas.0635171100
47. Mata J, Marguerat S, Bähler J (2005) Post-transcriptional control of gene expression: a genome-wide perspective. Trends Biochem Sci 30:506–514. https://doi.org/10.1016/j.tibs.2005.07.005
48. Gama-Carvalho M, Barbosa-Morais NL, Brodsky AS et al (2006) Genome-wide identification of functionally distinct subsets of cellular mRNAs associated with two nucleocytoplasmic-shuttling mammalian splicing factors. Genome Biol 7:R113. https://doi.org/10.1186/gb-2006-7-11-r113
49. Zhao J, Ohsumi TK, Kung JT et al (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell 40:939–953. https://doi.org/10.1016/j.molcel.2010.12.011
50. Ule J, Jensen K, Mele A, Darnell RB (2005) CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods 37:376–386. https://doi.org/10.1016/j.ymeth.2005.07.018
51. König J, Zarnack K, Luscombe NM, Ule J (2011) Protein-RNA interactions: new genomic technologies and perspectives. Nat Rev Genet 13:77–83. https://doi.org/10.1038/nrg3141
52. Nussbacher JK, Batra R, Lagier-Tourenne C, Yeo GW (2015) RNA-binding proteins in neurodegeneration: Seq and you shall receive. Trends Neurosci 38:226–236. https://doi.org/10.1016/j.tins.2015.02.003
53. Keene JD (2007) RNA regulons: coordination of post-transcriptional events. Nat Rev Genet 8:533–543. https://doi.org/10.1038/nrg2111
54. Ule J, Stefani G, Mele A et al (2006) An RNA map predicting Nova-dependent splicing regulation. Nature 444:580–586. https://doi.org/10.1038/nature05304
55. Barash Y, Calarco JA, Gao W et al (2010) Deciphering the splicing code. Nature 465:53–59. https://doi.org/10.1038/nature09000
56. Zhao BS, Roundtree IA, He C (2017) Post-transcriptional gene regulation by mRNA modifications. Nat Rev Mol Cell Biol 18:31–42. https://doi.org/10.1038/nrm.2016.132

57. Meyer KD, Saletore Y, Zumbo P et al (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell 149:1635–1646. https://doi.org/10.1016/j.cell.2012.05.003
58. Lovejoy AF, Riordan DP, Brown PO (2014) Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. PLoS One 9:e110799. https://doi.org/10.1371/journal.pone.0110799
59. Carlile TM, Rojas-Duran MF, Zinshteyn B et al (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature 515:143–146. https://doi.org/10.1038/nature13802
60. Roundtree IA, Evans ME, Pan T, He C (2017) Dynamic RNA modifications in gene expression regulation. Cell 169:1187–1200. https://doi.org/10.1016/j.cell.2017.05.045
61. Sultan M, Schulz M, Richard H et al (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321:956. https://doi.org/10.1126/science.1160342
62. Wang ET, Sandberg R, Luo S et al (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456:470–476. https://doi.org/10.1038/nature07509
63. Shiraki T, Kondo S, Katayama S et al (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci 100:15776–15781. https://doi.org/10.1073/pnas.2136655100
64. Katayama S, Tomaru Y, Kasukawa T et al (2005) Antisense transcription in the mammalian transcriptome. Science 309:1564–1566. https://doi.org/10.1126/science.1112009
65. Lizio M, Harshbarger J, Shimoji H et al (2015) Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol 16:22. https://doi.org/10.1186/s13059-014-0560-6
66. Almada AE, Wu X, Kriz AJ et al (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature 499:360–363. https://doi.org/10.1038/nature12349
67. Pal S, Gupta R, Kim H et al (2011) Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. Genome Res 21:1260–1272. https://doi.org/10.1101/gr.120535.111
68. Reyes A, Huber W (2018) Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. Nucleic Acids Res 46:582–592. https://doi.org/10.1093/nar/gkx1165
69. Ramanathan A, Robb GB, Chan S-H (2016) mRNA capping: biological functions and applications. Nucleic Acids Res 44:7511–7526. https://doi.org/10.1093/nar/gkw551
70. Galloway A, Cowling VH (2018) mRNA cap regulation in mammalian cell function and fate. Biochim Biophys Acta Gene Regul Mech. https://doi.org/10.1016/j.bbagrm.2018.09.011
71. Daffis S, Szretter KJ, Schriewer J et al (2010) 2′-O methylation of the viral mRNA cap evades host restriction by IFIT family members. Nature 468:452–456. https://doi.org/10.1038/nature09489
72. Hyde JL, Diamond MS (2015) Innate immune restriction and antagonism of viral RNA lacking 2′-O methylation. Virology 479–480:66–74. https://doi.org/10.1016/j.virol.2015.01.019
73. Cowling VH, Cole MD (2007) The Myc transactivation domain promotes global phosphorylation of the RNA polymerase II carboxy-terminal domain independently of direct DNA binding. Mol Cell Biol 27:2059–2073. https://doi.org/10.1128/MCB.01828-06
74. Mezquita P, Parghi SS, Brandvold KA, Ruddell A (2005) Myc regulates VEGF production in B cells by stimulating initiation of VEGF mRNA translation. Oncogene 24:889–901. https://doi.org/10.1038/sj.onc.1208251
75. Aregger M, Cowling VH (2014) E2F1-dependent methyl cap formation requires RNA pol II phosphorylation. Cell Cycle 11:2146–2148. https://doi.org/10.4161/cc.20620
76. Lombardi O, Varshney D, Phillips NM, Cowling VH (2016) c-Myc deregulation induces mRNA capping enzyme dependency. Oncotarget 7:82273–82288. https://doi.org/10.18632/oncotarget.12701
77. Aregger M, Kaskar A, Varshney D et al (2016) CDK1-cyclin B1 activates RNMT, coordinating mRNA cap methylation with G1 phase transcription. Mol Cell 61:734–746. https://doi.org/10.1016/j.molcel.2016.02.008
78. Grasso L, Suska O, Davidson L et al (2016) mRNA cap methylation in pluripotency and differentiation. CellReports 16:1352–1365. https://doi.org/10.1016/j.celrep.2016.06.089
79. Inesta-Vaquera F, Chaugule VK, Galloway A et al (2018) DHX15 regulates CMTR1-dependent gene expression and cell proliferation. Life Sci Alliance 1:e201800092. https://doi.org/10.26508/lsa.201800092
80. Furuichi Y, Morgan M, Shatkin AJ et al (1975) Methylated, blocked 5 termini in HeLa cell mRNA. Proc Natl Acad Sci 72:1904–1908. https://doi.org/10.1073/pnas.72.5.1904
81. Perry RP, Kelley DE (1976) Kinetics of formation of 5′ terminal caps in mRNA. Cell 8:433–442. https://doi.org/10.1016/0092-8674(76)90156-2
82. Werner M, Purta E, Kaminska KH et al (2011) 2′-O-ribose methylation of cap2 in human: function and evolution in a horizontally mobile family. Nucleic Acids Res 39:4756–4768. https://doi.org/10.1093/nar/gkr038
83. Wei C, Gershowitz A, Moss B (1975) N6, O2"-dimethyladenosine a novel methylated ribonucleoside next to the 5" terminal of animal cell and virus mRNAs. Nature 257:251–253. https://doi.org/10.1038/257251a0

84. Kruse S, Zhong S, Bodi Z et al (2011) A novel synthesis and detection method for cap-associated adenosine modifications in mouse mRNA. Sci Rep 1:54. https://doi.org/10.1038/srep00126
85. Tamarkin-Ben-Harush A, Vasseur J-J, Debart F et al (2017) Cap-proximal nucleotides via differential eIF4E binding and alternative promoter usage mediate translational response to energy stress. elife 6:977. https://doi.org/10.7554/eLife.21907
86. Abdelhamid RF, Plessy C, Yamauchi Y et al (2014) Multiplicity of 5′ cap structures present on short RNAs. PLoS One 9:e102895. https://doi.org/10.1371/journal.pone.0102895
87. Wetzel C, Limbach PA (2016) Mass spectrometry of modified RNAs: recent developments. Analyst 141:16–23. https://doi.org/10.1039/c5an01797a
88. Bird JG, Zhang Y, Tian Y et al (2016) The mechanism of RNA 5′ capping with NAD+, NADH and desphospho-CoA. Nature 535:444–447. https://doi.org/10.1038/nature18622
89. Jiao X, Doamekpor SK, Bird JG et al (2017) 5′ end nicotinamide adenine dinucleotide cap in human cells promotes RNA decay through DXO-mediated deNADding. Cell 168:1015–1027.e10. https://doi.org/10.1016/j.cell.2017.02.019
90. Kiledjian M (2018) Eukaryotic RNA 5′-end NAD + capping and DeNADding. Trends Cell Biol 28:454–464. https://doi.org/10.1016/j.tcb.2018.02.005
91. Schoenberg DR, Maquat LE (2012) Regulation of cytoplasmic mRNA decay. Nat Publ Group 13:246–259. https://doi.org/10.1038/nrg3160
92. Jiao Y, Riechmann JL, Meyerowitz EM (2008) Transcriptome-wide analysis of uncapped mRNAs in Arabidopsis reveals regulation of mRNA degradation. Plant Cell 20:2571–2585. https://doi.org/10.1105/tpc.108.062786
93. Karginov FV, Cheloufi S, Chong MMW et al (2010) Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases. Mol Cell 38:781–788. https://doi.org/10.1016/j.molcel.2010.06.001
94. Mercer TR, Dinger ME, Bracken CP et al (2010) Regulated post-transcriptional RNA cleavage diversifies the eukaryotic transcriptome. Genome Res 20:1639–1650. https://doi.org/10.1101/gr.112128.110
95. Otsuka Y, Kedersha NL, Schoenberg DR (2009) Identification of a cytoplasmic complex that adds a cap onto 5′-monophosphate RNA. Mol Cell Biol 29:2155–2167. https://doi.org/10.1128/MCB.01325-08
96. Mukherjee C, Bakthavachalu B, Schoenberg DR (2014) The cytoplasmic capping complex assembles on adapter protein nck1 bound to the proline-rich C-terminus of Mammalian capping enzyme. PLoS Biol 12:e1001933. https://doi.org/10.1371/journal.pbio.1001933
97. Trotman JB, Schoenberg DR (2018) A recap of RNA recapping. Wiley Interdiscip Rev RNA 10:e1504. https://doi.org/10.1002/wrna.1504
98. Mukherjee C, Patil DP, Kennedy BA et al (2012) Identification of cytoplasmic capping targets reveals a role for cap homeostasis in translation and mRNA stability. CellReports 2:674–684. https://doi.org/10.1016/j.celrep.2012.07.011
99. Gregory BD, O'Malley RC, Lister R et al (2008) A link between RNA metabolism and silencing affecting Arabidopsis development. Dev Cell 14:854–866. https://doi.org/10.1016/j.devcel.2008.04.005
100. Kiss DL, Oman KM, Dougherty JA et al (2016) Cap homeostasis is independent of poly(A) tail length. Nucleic Acids Res 44:304–314. https://doi.org/10.1093/nar/gkv1460
101. Chen P, Zhou Z, Yao X et al (2017) Capping enzyme mRNA-cap/RNGTT regulates hedgehog pathway activity by antagonizing protein kinase A. Sci Rep 7:2891. https://doi.org/10.1038/s41598-017-03165-2
102. Sharp PA (1994) Split genes and RNA splicing. Cell 77:805–815
103. Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. Nature 463:457–463. https://doi.org/10.1038/nature08909
104. Wahl MC, Will CL, Lührmann R (2009) The spliceosome: design principles of a dynamic RNP machine. Cell 136:701–718. https://doi.org/10.1016/j.cell.2009.02.009
105. Matera AG, Wang Z (2014) A day in the life of the spliceosome. Nat Rev Mol Cell Biol 15:108–121. https://doi.org/10.1038/nrm3742
106. Will CL, Lührmann R (2011) Spliceosome structure and function. Cold Spring Harb Perspect Biol 3:a003707–a003707. https://doi.org/10.1101/cshperspect.a003707
107. Baralle M, Baralle FE (2018) The splicing code. Biosystems 164:39–48. https://doi.org/10.1016/j.biosystems.2017.11.002
108. Smith CW, Valcárcel J (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem Sci 25:381–388
109. Hertel KJ (2008) Combinatorial control of exon recognition. J Biol Chem 283:1211–1215. https://doi.org/10.1074/jbc.R700035200
110. Wang Z, Burge CB (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA 14:802–813. https://doi.org/10.1261/rna.876308
111. Lee Y, Rio DC (2015) Mechanisms and regulation of alternative pre-mRNA splicing. Annu Rev Biochem 84:291–323. https://doi.org/10.1146/annurev-biochem-060614-034316
112. Savarese M, Jonson PH, Huovinen S et al (2018) The complexity of titin splicing pattern in human adult skeletal muscles. Skelet Muscle 8:11. https://doi.org/10.1186/s13395-018-0156-z
113. da Costa PJ, Menezes J, Romão L (2017) The role of alternative splicing coupled to nonsense-mediated mRNA decay in human disease. Int J Biochem Cell Biol 91:168–175. https://doi.org/10.1016/j.biocel.2017.07.013

114. Ni JZ, Grate L, Donohue JP et al (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev 21:708–718. https://doi.org/10.1101/gad.1525507
115. Lareau LF, Inada M, Green RE et al (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature 446:926–929. https://doi.org/10.1038/nature05676
116. Saltzman AL, Kim YK, Pan Q et al (2008) Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. Mol Cell Biol 28:4320–4330. https://doi.org/10.1128/MCB.00361-08
117. Mayr C, Bartel DP (2009) Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell 138:673–684. https://doi.org/10.1016/j.cell.2009.06.016
118. Ogorodnikov A, Kargapolova Y, Danckwardt S (2016) Processing and transcriptome expansion at the mRNA 3′ end in health and disease: finding the right end. Pflugers Arch 468:993–1012. https://doi.org/10.1007/s00424-016-1828-3
119. Long JC, Cáceres JF (2009) The SR protein family of splicing factors: master regulators of gene expression. Biochem J 417:15–27. https://doi.org/10.1042/BJ20081501
120. Weatheritt RJ, Sterne-Weiler T, Blencowe BJ (2016) The ribosome-engaged landscape of alternative splicing. Nat Struct Mol Biol 23:1117–1123. https://doi.org/10.1038/nsmb.3317
121. Venables JP (2006) Unbalanced alternative splicing and its significance in cancer. BioEssays 28:378–386. https://doi.org/10.1002/bies.20390
122. Srebrow A, Kornblihtt AR (2006) The connection between splicing and cancer. J Cell Sci 119:2635–2641. https://doi.org/10.1242/jcs.03053
123. Matos P, Collard JG, Jordan P (2003) Tumor-related alternatively spliced Rac1b is not regulated by Rho-GDP dissociation inhibitors and exhibits selective downstream signaling. J Biol Chem 278:50442–50448. https://doi.org/10.1074/jbc.M308215200
124. Scherr A-L, Gdynia G, Salou M et al (2016) Bcl-xL is an oncogenic driver in colorectal cancer. Cell Death Dis 7:e2342–e2342. https://doi.org/10.1038/cddis.2016.233
125. Ghigna C, Giordano S, Shen H et al (2005) Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. Mol Cell 20:881–890. https://doi.org/10.1016/j.molcel.2005.10.026
126. Chen M, Zhang J, Manley JL (2010) Turning on a fuel switch of cancer: hnRNP proteins regulate alternative splicing of pyruvate kinase mRNA. Cancer Res 70:8977–8980. https://doi.org/10.1158/0008-5472.CAN-10-2513
127. Sibley CR, Emmett W, Blazquez L et al (2015) Recursive splicing in long vertebrate genes. Nature 521:371–375. https://doi.org/10.1038/nature14466
128. Kelly S, Greenman C, Cook PR, Papantonis A (2015) Exon skipping is correlated with exon circularization. J Mol Biol 427:2414–2417. https://doi.org/10.1016/j.jmb.2015.02.018
129. Duff MO, Olson S, Wei X, Garrett SC, Osman A, Bolisetty M et al (2015) Genome-wide identification of zero nucleotide recursive splicing in Drosophila. Nature 521:376–379. https://doi.org/10.1038/nature14475
130. Georgomanolis T, Sofiadis K, Papantonis A (2016) Cutting a long intron short: recursive splicing and its implications. Front Physiol 7:598. https://doi.org/10.3389/fphys.2016.00598
131. Starke S, Jost I, Rossbach O et al (2015) Exon circularization requires canonical splice signals. CellReports 10:103–111. https://doi.org/10.1016/j.celrep.2014.12.002
132. Jeck WR, Sharpless NE (2014) Detecting and characterizing circular RNAs. Nat Biotechnol 32:453–461. https://doi.org/10.1038/nbt.2890
133. Barrett SP, Salzman J (2016) Circular RNAs: analysis, expression and potential functions. Development 143:1838–1847. https://doi.org/10.1242/dev.128074
134. Wang M, Yu F, Wu W et al (2017) Circular RNAs: a novel type of non-coding RNA and their potential implications in antiviral immunity. Int J Biol Sci 13:1497–1506. https://doi.org/10.7150/ijbs.22531
135. Chuang T-J, Chen Y-J, Chen C-Y et al (2018) Integrative transcriptome sequencing reveals extensive alternative trans-splicing and cis-backsplicing in human cells. Nucleic Acids Res 46:3671–3691. https://doi.org/10.1093/nar/gky032
136. Chen L-L (2016) The biogenesis and emerging roles of circular RNAs. Nat Rev Mol Cell Biol 17:205–211. https://doi.org/10.1038/nrm.2015.32
137. Zhang Z-C, Guo X-L, Li X (2018) The novel roles of circular RNAs in metabolic organs. Genes Dis 5:16–23. https://doi.org/10.1016/j.gendis.2017.12.002
138. Han B, Chao J, Yao H (2018) Circular RNA and its mechanisms in disease: from the bench to the clinic. Pharmacol Ther 187:31–44. https://doi.org/10.1016/j.pharmthera.2018.01.010
139. Xiao SH, Manley JL (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. Genes Dev 11:334–344
140. Stamm S (2008) Regulation of alternative splicing by reversible protein phosphorylation. J Biol Chem 283:1223–1227. https://doi.org/10.1074/jbc.R700034200
141. Gonçalves V, Pereira JFS, Jordan P (2017) Signaling pathways driving aberrant splicing in cancer cells. Gene 9:9. https://doi.org/10.3390/genes9010009
142. Goren A, Kim E, Amit M et al (2010) Overlapping splicing regulatory motifs--combinatorial effects on splicing. Nucleic Acids Res 38:3318–3327. https://doi.org/10.1093/nar/gkq005
143. Zahler AM, Damgaard CK, Kjems J, Caputi M (2004) SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic

splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. J Biol Chem 279:10077–10084. https://doi.org/10.1074/jbc.M312743200
144. Zhang XH-F, Arias MA, Ke S, Chasin LA (2009) Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. RNA 15:367–376. https://doi.org/10.1261/rna.1498509
145. Pagani F, Baralle FE (2004) Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet 5:389–396. https://doi.org/10.1038/nrg1327
146. Baralle D, Baralle M (2005) Splicing in action: assessing disease causing sequence changes. J Med Genet 42:737–748. https://doi.org/10.1136/jmg.2004.029538
147. Ghigna C, Moroni M, Porta C et al (1998) Altered expression of heterogenous nuclear ribonucleoproteins and SR factors in human colon adenocarcinomas. Cancer Res 58:5818–5824
148. David CJ, Manley JL (2010) Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes Dev 24:2343–2364. https://doi.org/10.1101/gad.1973010
149. Karni R, de Stanchina E, Lowe SW et al (2007) The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat Struct Mol Biol 14:185–193. https://doi.org/10.1038/nsmb1209
150. Naftelberg S, Schor IE, Ast G, Kornblihtt AR (2015) Regulation of alternative splicing through coupling with transcription and chromatin structure. Annu Rev Biochem 84:165–198. https://doi.org/10.1146/annurev-biochem-060614-034242
151. Luco RF, Pan Q, Tominaga K et al (2010) Regulation of alternative splicing by histone modifications. Science 327:996–1000. https://doi.org/10.1126/science.1184208
152. Blaustein M, Pelisch F, Srebrow A (2007) Signals, pathways and splicing regulation. Int J Biochem Cell Biol 39:2031–2048. https://doi.org/10.1016/j.biocel.2007.04.004
153. Valacca C, Bonomi S, Buratti E et al (2010) Sam68 regulates EMT through alternative splicing-activated nonsense-mediated mRNA decay of the SF2/ASF proto-oncogene. J Cell Biol 191:87–99. https://doi.org/10.1083/jcb.201001073
154. van der Houven, van Oordt W, Diaz-Meco MT, Lozano J et al (2000) The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. J Cell Biol 149:307–316
155. Allemand E, Guil S, Myers M et al (2005) Regulation of heterogenous nuclear ribonucleoprotein A1 transport by phosphorylation in cells stressed by osmotic shock. Proc Natl Acad Sci 102:3605–3610. https://doi.org/10.1073/pnas.0409889102
156. Gonçalves V, Henriques A, Pereira J et al (2014) Phosphorylation of SRSF1 by SRPK1 regulates alternative splicing of tumor-related Rac1b in colorectal cells. RNA. https://doi.org/10.1261/rna.041376.113
157. Licatalosi DD, Darnell RB (2010) RNA processing and its regulation: global insights into biological networks. Nat Rev Genet 11:75–87. https://doi.org/10.1038/nrg2673
158. Kalsotra A, Cooper TA (2011) Functional consequences of developmentally regulated alternative splicing. Nat Rev Genet 12:715–729. https://doi.org/10.1038/nrg3052
159. Buljan M, Chalancon G, Eustermann S et al (2012) Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. Mol Cell 46:871–883. https://doi.org/10.1016/j.molcel.2012.05.039
160. Ellis JD, Barrios-Rodiles M, Çolak R et al (2012) Tissue-specific alternative splicing remodels protein-protein interaction networks. Mol Cell 46:884–892. https://doi.org/10.1016/j.molcel.2012.05.037
161. Barbosa-Morais NL, Irimia M, Pan Q et al (2012) The evolutionary landscape of alternative splicing in vertebrate species. Science 338:1587–1593. https://doi.org/10.1126/science.1230612
162. Traunmüller L, Gomez AM, Nguyen T-M, Scheiffele P (2016) Control of neuronal synapse specification by a highly dedicated alternative splicing program. Science 352:982–986. https://doi.org/10.1126/science.aaf2397
163. Calarco JA, Zhen M, Blencowe BJ (2011) Networking in a global world: establishing functional connections between neural splicing regulators and their target transcripts. RNA 17:775–791. https://doi.org/10.1261/rna.2603911
164. Baralle FE, Giudice J (2017) Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol 18:437–451. https://doi.org/10.1038/nrm.2017.27
165. Gallego-Paez LM, Bordone MC, Leote AC et al (2017) Alternative splicing: the pledge, the turn, and the prestige: the key role of alternative splicing in human biological systems. Hum Genet 136:1015–1042. https://doi.org/10.1007/s00439-017-1790-y
166. Li M, Izpisua Belmonte JC (2018) Deconstructing the pluripotency gene regulatory network. Nat Cell Biol 20:382–392. https://doi.org/10.1038/s41556-018-0067-6
167. Abdel-Wahab O, Levine R (2011) The spliceosome as an indicted conspirator in myeloid malignancies. Cancer Cell 20:420–423. https://doi.org/10.1016/j.ccr.2011.10.004
168. Scott LM, Rebel VI (2013) Acquired mutations that affect pre-mRNA splicing in hematologic malignancies and solid tumors. J Natl Cancer Inst 105:1540–1549. https://doi.org/10.1093/jnci/djt257
169. Dittmar KA, Jiang P, Park JW et al (2012) Genome-wide determination of a broad ESRP-regulated posttranscriptional network by high-throughput sequencing. Mol Cell Biol 32:1468–1482. https://doi.org/10.1128/MCB.06536-11
170. Warzecha CC, Carstens RP (2012) Complex changes in alternative pre-mRNA splicing play a central role in the epithelial-to-mesenchymal

transition (EMT). Semin Cancer Biol 22:417–427. https://doi.org/10.1016/j.semcancer.2012.04.003
171. Hollander D, Donyo M, Atias N et al (2016) A network-based analysis of colon cancer splicing changes reveals a tumorigenesis-favoring regulatory pathway emanating from ELK1. Genome Res 26:541–553. https://doi.org/10.1101/gr.193169.115
172. Kim HK, Pham MHC, Ko KS et al (2018) Alternative splicing isoforms in health and disease. Pflugers Arch 470:995–1016. https://doi.org/10.1007/s00424-018-2136-x
173. Mayr C (2017) Regulation by 3′-untranslated regions. Annu Rev Genet 51:171–194. https://doi.org/10.1146/annurev-genet-120116-024704
174. Xie X, Lu J, Kulbokas EJ et al (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature 434:338–345. https://doi.org/10.1038/nature03441
175. Siepel A, Bejerano G, Pedersen JS et al (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 15:1034–1050. https://doi.org/10.1101/gr.3715005
176. Bartel DP (2018) Metazoan MicroRNAs. Cell 173:20–51. https://doi.org/10.1016/j.cell.2018.03.006
177. Guo H, Ingolia NT, Weissman JS, Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466:835–840. https://doi.org/10.1038/nature09267
178. Eichhorn SW, Guo H, McGeary SE et al (2014) mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. Mol Cell 56:104–115. https://doi.org/10.1016/j.molcel.2014.08.028
179. Marson A, Levine SS, Cole MF et al (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell 134:521–533. https://doi.org/10.1016/j.cell.2008.07.020
180. Liu N, Olson EN (2010) MicroRNA regulatory networks in cardiovascular development. Dev Cell 18:510–525. https://doi.org/10.1016/j.devcel.2010.03.010
181. Gama-Carvalho M, Andrade J, Bras-Rosario L (2014) Regulation of cardiac cell fate by microRNAs: implications for heart regeneration. Cell 3:996–1026. https://doi.org/10.3390/cells3040996
182. Chen C-YA, Shyu A-B (2017) Emerging themes in regulation of global mRNA turnover in cis. Trends Biochem Sci 42:16–27. https://doi.org/10.1016/j.tibs.2016.08.014
183. Andreassi C, Riccio A (2009) To localize or not to localize: mRNA fate is in 3′UTR ends. Trends Cell Biol 19:465–474. https://doi.org/10.1016/j.tcb.2009.06.001
184. Berkovits BD, Mayr C (2015) Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. Nature 522:363–367. https://doi.org/10.1038/nature14321
185. Tian B, Manley JL (2017) Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol 18:18–30. https://doi.org/10.1038/nrm.2016.116
186. Tian B, Hu J, Zhang H, Lutz CS (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res 33:201–212. https://doi.org/10.1093/nar/gki158
187. Mayr C (2016) Evolution and biological roles of alternative 3′ UTRs. Trends Cell Biol 26:227–237. https://doi.org/10.1016/j.tcb.2015.10.012
188. Smibert P, Miura P, Westholm JO et al (2012) Global patterns of tissue-specific alternative polyadenylation in Drosophila. CellReports 1:277–289. https://doi.org/10.1016/j.celrep.2012.01.001
189. Ulitsky I, Shkumatava A, Jan CH et al (2012) Extensive alternative polyadenylation during zebrafish development. Genome Res 22:2054–2066. https://doi.org/10.1101/gr.139733.112
190. Ji Z, Lee JY, Pan Z et al (2009) Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc Natl Acad Sci U S A 106:7028–7033. https://doi.org/10.1073/pnas.0900028106
191. Sandberg R, Neilson JR, Sarma A et al (2008) Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. Science 320:1643–1647. https://doi.org/10.1126/science.1155390
192. Elkon R, Drost J, van Haaften G et al (2012) E2F mediates enhanced alternative polyadenylation in proliferation. Genome Biol 13:R59. https://doi.org/10.1186/gb-2012-13-7-r59
193. Spies N, Burge CB, Bartel DP (2013) 3′ UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. Genome Res 23:2078–2090. https://doi.org/10.1101/gr.156919.113
194. Gruber AR, Martin G, Müller P et al (2014) Global 3′ UTR shortening has a limited effect on protein abundance in proliferating T cells. Nat Commun 5:5465. https://doi.org/10.1038/ncomms6465
195. Geisberg JV, Moqtaderi Z, Fan X et al (2014) Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. Cell 156:812–824. https://doi.org/10.1016/j.cell.2013.12.026
196. Oikonomou P, Goodarzi H, Tavazoie S (2014) Systematic identification of regulatory elements in conserved 3′ UTRs of human transcripts. Cell Rep 7:281–292. https://doi.org/10.1016/j.celrep.2014.03.001
197. Malka Y, Steiman-Shimony A, Rosenthal E et al (2017) Post-transcriptional 3′-UTR cleavage of mRNA transcripts generates thousands of stable uncapped autonomous RNA fragments. Nat Commun 8:2029. https://doi.org/10.1038/s41467-017-02099-7
198. Kocabas A, Duarte T, Kumar S, Hynes MA (2015) Widespread differential expression of coding region and 3′ UTR sequences in neurons and other tissues.

Neuron 88:1149–1156. https://doi.org/10.1016/j.neuron.2015.10.048
199. Sudmant PH, Lee H, Dominguez D et al (2018) Widespread accumulation of ribosome-associated isolated 3′ UTRs in neuronal cell populations of the aging brain. CellReports 25:2447–2456.e4. https://doi.org/10.1016/j.celrep.2018.10.094
200. Salmena L, Poliseno L, Tay Y et al (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell 146:353–358. https://doi.org/10.1016/j.cell.2011.07.014
201. Li X, Yang L, Chen L-L (2018) The biogenesis, functions, and challenges of circular RNAs. Mol Cell 71:428–442. https://doi.org/10.1016/j.molcel.2018.06.034
202. Wu Y, Zhao W, Liu Y et al (2018) Function of HNRNPC in breast cancer cells by controlling the dsRNA-induced interferon response. EMBO J 37:e99017. https://doi.org/10.15252/embj.201899017
203. Maharana S, Wang J, Papadopoulos DK et al (2018) RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science 360:918–921. https://doi.org/10.1126/science.aar7366
204. Ma W, Mayr C (2018) A membraneless organelle associated with the endoplasmic reticulum enables 3′UTR-mediated protein-protein interactions. Cell 175:1492–1506.e19. https://doi.org/10.1016/j.cell.2018.10.007

The Diverse Roles of RNA-Binding Proteins in Glioma Development

2

Mitzli X. Velasco, Adam Kosti, Luiz O. F. Penalva, and Greco Hernández

Abstract

Post-transcriptional regulation of gene expression is fundamental for all forms of life, as it critically contributes to the composition and quantity of a cell's proteome. These processes encompass splicing, polyadenylation, mRNA decay, mRNA editing and modification and translation and are modulated by a variety of RNA-binding proteins (RBPs). Alterations affecting RBP expression and activity contribute to the development of different types of cancer. In this chapter, we discuss current research shedding light on the role of different RBPs in gliomas. These studies place RBPs as modulators of critical signaling pathways, establish their relevance as prognostic markers and open doors for new therapeutic strategies.

Keywords

ADARs · Cancer · Glioblastoma · hnRNP · HuR · IGF2BPs · Musashi1 · PTBP · RNA-binding proteins · Translation initiation

M. X. Velasco
Laboratory of Translation and Cancer, Unit of Biomedical Research on Cancer, Instituto Nacional de Cancerología (INCan), Mexico City, Mexico

Biomedical Sciences Ph.D. Program, National Autonomous University of Mexico, UNAM, Mexico City, Mexico

Greehey Children's Cancer Research Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

A. Kosti · L. O. F. Penalva (✉)
Greehey Children's Cancer Research Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA
e-mail: penalva@uthscsa.edu

G. Hernández (✉)
Laboratory of Translation and Cancer, Unit of Biomedical Research on Cancer, Instituto Nacional de Cancerología (INCan), Mexico City, Mexico
e-mail: ghernandezr@incan.edu.mx

Abbreviations

RBPs	RNA-binding proteins
MS	Mass spectrometry
UTRs	Untranslated regions
GBM	Glioblastoma multiforme or Glioblastoma
hnRNP	Heterogeneous nuclear ribonucleoproteins
PTBP	Polypyrimidine-tract-binding protein
ADARs	Adenosine deaminases that act on RNA
HuR	Hu antigen R
MSI1	Musashi1
IGF2BPs/IMPs	Insulin-like growth factor II mRNA binding proteins
PK	Pyruvate-kinase

L. Romão (ed.), *The mRNA Metabolism in Human Disease*, Advances in Experimental Medicine and Biology 1157, https://doi.org/10.1007/978-3-030-19966-1_2

EGFR	Epidermal growth factor receptor
HK2	Hexokinase2
GLUT3	Glucose Transporter 3
IG20	Insuloma-glucagonoma protein 20
RON	Recepteur d'Origine Nantais
EMT	Epithelial-mesenchymal transition
PDCD4	Programmed cell death protein 4
MMP	Matrix metalloproteinases
STAT3	Signal transducer and activator of transcription 3
NPC	Neuronal precursor cells
RRMs	RNA recognition motifs
FGFR	fibroblast growth factor receptor-1, -2
FBN	fibrilin
CASP2	caspase 2
ABCC1	ATP binding cassette subfamily C member 1
RTN4	Reticulon 4
MARK4	Microtubule affinity-regulating kinase 4
GluR	Glutamate receptors
ELAV	*Embryonic lethal abnormal visual*
AREs	*AU-rich elements*
VEGF	Vascular endothelial growth factor
SIRT1	Silent mating type information regulation 2 homolog 1
BCL-2	B-cell lymphoma 2
ProTα	Prothymosin α
PCM	pericentriolar matrix
SCs	Stem cells
GSCs	Glioblastoma stem cells
MVD	microvessel density
DNA-PKcs	DNA-Protein Kinase Catalytic Subunit
PMAs	Pilomyxoid astrocytomas
QKI	Quaking
INF	Interferon
PI3K/MAPK	Phosphatidylinositol 3-Kinase/Mitogen-activated Protein Kinase

2.1 Introduction

The large majority of cancer studies dedicated to the identification of "driving events" and genes contributing to tumor development have focused on processes occurring at the DNA level, such as mutations and chromosomal defects, changes in methylation status, and transcriptional regulation. Recent studies performed in different species and scenarios have shown a poor correlation between mRNA and protein levels for the majority of the transcriptome. These results support the role of "RNA-based" mechanisms as key modulators of gene expression in both, normal and tumorigenic cells [1, 2]. Such regulatory mechanisms are primarily driven by RNA-binding proteins (RBPs). Large-scale quantitative methods, next-generation sequencing, and modern protein mass spectrometry (MS) have been employed recently to expand the RBP catalogue, to identify their protein co-factors and target transcripts. The number of known RBPs in the human genome is over 1500 [3, 4], which represent ~7.5% of human coding genes. RBPs form complexes with pre-mRNA, mRNA, and a variety of ncRNAs (lncRNAs, miRNAs, rRNAs, etc.), to modulate an array of processes that include splicing, polyadenylation, maturation, modification, transport, stability, localization, and translation. RBPs exert their effect by recognizing specific sequences and/or secondary structures present in untranslated regions (UTRs), coding sequences and/or introns [3, 5].

Mutations and alterations in the expression levels of numerous RBPs are observed across tumor tissues and known to impact the expression of large set of genes, contributing to tumor initiation and growth [6–8]. This scenario is observed in gliomas, an heterogenous group of brain tumors encompassing astrocytomas, oligodendrogliomas, oligoastrocytomas and glioblastoma multiforme [9]. Gliomas are responsible for the majority of deaths caused by primary brain tumors. The absence or presence of anaplastic features is used by the World Health Organization (WHO) for assigning grades of malignancy [9, 10]. Among them, glioblastoma multiforme

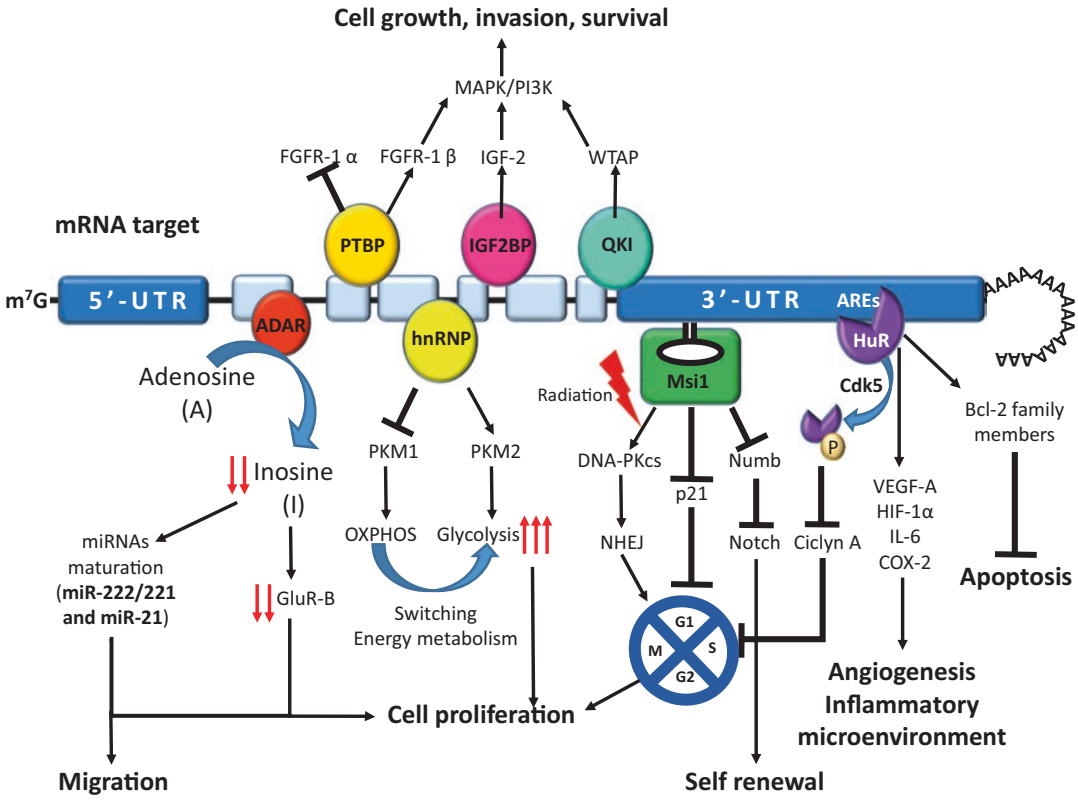

Fig. 2.1 Schematic representation of the involvement of RNA-binding proteins (RBPs) in glioblastoma development. We show the RBP families discussed in this article, the region of the mRNA where they make target, their most relevant target genes, and both the processes and pathways affected by the described regulation

(GBM, grade IV) is the most aggressive and accounts for 45.2% of all malignant primary brain and CNS tumors, and 54% of all gliomas [11]. The number of well-characterized oncogenic RBPs in gliomas is still relatively small. As indicated by a recent genomic study from our lab, aberrant RBP expression is a common feature in GBMs and dozens of these proteins potentially contribute to the acquisition of cancer phenotypes [12].

In the next sections, we discuss the role of some of the best characterized RBPs in the context of gliomas, namely Heterogeneous nuclear ribonucleoproteins (hnRNP), Polypyrimidine-tract-binding protein (PTBP), Adenosine deaminases that act on RNA (ADARs), Hu antigen R (HuR), Musashi1 (MSI1), Insulin-like growth factor II mRNA binding proteins (IGF2BPs/IMPs), and Quaking (QKI) – In Fig. 2.1 we depict the global biology of these RBPs in glioma development.

2.2 Heterogeneous Nuclear Ribonucleoproteins

Heterogeneous nuclear ribonucleoproteins are a family of RBPs that includes approximately 20 genes termed hnRNPs A1-U, which range in size from 34 to 120 kDa [13]. Several hnRNPs have been implicated in the development of various tumor types and their expression levels have been linked to patient survival [14]. HnRNPs regulate different aspects of pre-mRNA processing in gliomas and their expression levels are altered in both low and high grade astrocytomas [15–17].

In gliomas, hnRNPA1/A2 is associated with the regulation of glucose metabolism, where it influences Pyruvate-kinase (PK) splicing and function. PK is a rate-limiting enzyme in glucose metabolism and is encoded by two paralogous genes, PKLR and PKM. Both genes are alternatively spliced; therefore, resulting in four PK isoforms in mammals [18, 19]. The PKM gene consists of 12 exons, of which exons 9 and 10 are alternatively spliced in a mutually exclusive fashion, producing the PKM1 and PKM2 mRNA isoforms, respectively [19]. Interestingly, the expression of hnRNP A1/A2 and PTBP is regulated by the oncogenic transcription factor c-Myc [20]. In cancer cells, hnRNP A1/A2 and PTBP bind to PKM pre-mRNA and repress the inclusion of exon 9. When expression levels of hnRNPA1/A2 and PTBP are reduced, PKM exon 9 inclusion is promoted and more PKM1 transcripts are produced; the end result is a decrease in glucose consumption and lactate production [21]. This is an example of a critical dialogue between RBP-mediated splicing and modulation of metabolism and cell proliferation.

Astrocytomas display mutations in growth-factor-receptor genes, such as epidermal growth factor receptor (EGFR). Alterations in the splicing profile of EGFR are also frequently observed. The most relevant one gives rise to a transcript encoding EGFR variant III (EGFRvIII) [22]; the resulting protein is constitutively active and contributes to a major shift in GBM cell metabolism [23]. Babic et al. (2015) demonstrated that in GBM cells expressing EGFRvIII, the Myc-binding partner Max is alternatively spliced. hnRNPA1 binds upstream of exon 5 of the Max pre-mRNA and facilitates its inclusion, leading to the production of a truncated Max protein referred to as Delta Max. Increased Delta Max production contributes to the expression of the glucose transporter GLUT3 and HK2 and promotes GBM cell proliferation in glucose-containing media [24].

Evasion of programmed cell death, tissue invasion, and metastasis are three important hallmarks of cancer [25]. A number of studies have described hnRNPs as regulators of apoptosis and cell invasion via its impact on RNA processing. For instance, hnRNPH regulates the splicing of Insuloma-glucagonoma protein 20 (IG20) and Recepteur d'Origine Nantais (RON), a death-domain adaptor protein and a tyrosine kinase receptor that participate in apoptosis and cell invasion, respectively. Alternative 5′ splice site usage in exon 13, plus inclusion/skipping of exon 16 generates four main splicing isoforms of IG20 (MADD), which have been described to be aberrantly expressed in tumors [26, 27]. In the case of RON, exon 11 exclusion generates RONΔ11, a transcript that gives rise to a protein missing part of the extracellular domain. This active isoform promotes cell motility and mediates epithelial-mesenchymal transition (EMT) [28] . hnRNPH binds to UGGG elements in the 5′ region of exons 16 and 11 of IG20 and RON pre-mRNAs, respectively, and inhibits its inclusion [29]. Knockdown of hnRNPH promotes inclusion of the exons that are normaly spliced out, and is associated with a decrease in cell viability and migration in gliomas [29]. Binding of hnRNPC to pre-miR-21, which promotes maturation, represents an example of interactions between hnRNPs and miRNAs. Silencing of hnRNPC decreases miR-21 levels, upregulating expression of Programmed cell death protein 4 (PDCD4), thus affecting the proliferative and metastatic potential of GBM cells [30].

2.3 Polypyrimidine-Tract-Binding Protein

Polypyrimidine tract-binding proteins (PTBPs) are a family of RNA binding proteins whose most of its members are preferentially expressed in the brain [31, 32]. PTBP2 (also known as neural PTB) for instance is specifically expressed in post-mitotic neurons [33]. PTBP3 is the least studied of the three PTB paralogs. In rat and human it is encoded by the gene *ROD1*. In rats, it is predominantly expressed in hematopoietic cells or organs of embryonic and adult individuals. In humans, *ROD1* overexpression inhibits pharmacologically-induced differentiation of both megakaryocytic and erythroid K562 leukemia cells, supporting the notion that *ROD1* plays

a role in differentiation control in mammalian cells [34]. These RBPs shuttle between the nucleus and cytoplasm, functioning in large and diverse number of cellular processes, including mRNA splicing, polyadenylation, stability and translation. PTBPs preferentially bind to polypyrimidine-rich stretches through its four RNA recognition motifs (RRMs) [35, 36]. As a splicing factor, PTBP1 induces exon skipping in pre-mRNAs encoding proteins involved in proliferation (FGFR1, FGFR2), invasion (CSRC), motility (ACTN, FBN), apoptosis (FAS, CASP2), and multi-drug resistance (ABCC1) [37].

Several studies have reported significant differences in the expression of PTBP1 in gliomas [37, 38]. PTBP1 regulates numerous splicing events relevant to gliomagenesis. Global exon array analysis identified Nogo (also known as RTN4) as one of the main targets of PTBP1. Nogo triggers a rearrangement of actin filament extensions in neighboring cells to inhibit neurite outgrowth [39]. Nogo has multiple mRNA isoforms, but only three characterized protein variants: Nogo-A, Nogo-B and Nogo-C [40]. Cheung *et al*, (2009) showed that Nogo-B is the predominant mRNA isoform expressed in glioma cells, where high levels of PTBP1 ensure that exon 3 is skipped [37]. Functional assays suggest that regulation of Nogo splicing by PTBP1 plays a role in proliferation and migration [37]. Another study identified the FGFR-1 pre-mRNA as PTBP1 target [41]. FGFR1 is implicated in growth and differentiation pathways and precise regulation of its splicing is critical [42]. PTBP1 interacts in a sequence-specific manner with the intronic RNA sequence, termed ISS-1 element, located upstream of the α exon of FGFR-1 pre-mRNA. This interaction induces exon α exclusion and leads to the production of a receptor with enhanced affinity for fibroblast growth factor (FGFR-1β) [41]. Beta form of FGFR-1 is the predominant isoform and has been described to drive tumor progression [42]. Another PTBP1 target is the kinase MARK4, which belongs to the family of AMP protein kinases. The two MARK4 splicing isoforms (MARK4L and MARK4S) differ in relation to their C-terminal end [43]. While MARK4L is upregulated in glioma cells and is expressed at high levels in neural progenitor cells, MARK4S is found predominantly in normal brain tissue and terminally differentiated neurons [44]. Bioinformatic and biochemical approaches identified PTBP1 binding sites in intron 15 of the MARK4 pre-mRNA, indicating that PTBP1 regulates the inclusion of exon 16. This process influences splicing of exon 18 which contains a stop codon; the result are two alternative protein products containing or not the C-terminal kinase 1 domain [45].

2.4 Adenosine Deaminases that Act on RNA

RNA editing is an important mechanism in which RNA modifications are used to modulate gene expression and function [46]. Adenosine deaminases that act on RNA (ADARs) are key regulators of RNA editing and changes in ADAR expression levels may lead to the development of different neoplasms, including gliomas. They are responsible for adenosine (A) to inosine (I) conversion in RNAs. These RNA modifications often lead to changes in amino acid incorporation that could ultimately affect protein properties [47]. In vertebrates, three members of the ADAR family have been identified. ADAR1 and ADAR2 are ubiquitously expressed and exhibit catalytic activity, whereas ADAR3 is specifically expressed in the brain and has no catalytic activity. ADAR3 has been revealed to competitively inhibit the deaminase activity of other ADARs by binding to dsRNA [48]. In particular, ADAR2 editing activity is crucial for the function of many proteins expressed in central nervous system and is essential for normal brain development [49].

In mammals, the catalogue of genes affected by RNA editing has been expanding. Among them are the glutamate receptors (GluR) sensitive to AMP (B, C and D subunits) and those sensitive to kainate (5 and 6 subunits). A to I editing causes the substitution of the neutral amino acid glutamine (Q) for an arginine (R), located in the pore domain of GluR. The positively charged Arginine prevents the passage of Ca^{2+}. This editing occurs in a majority of AMPA receptors in the central

nervous system and acts as a protective mechanism against excitotoxicity mediated by massive Ca^{2+} influxes [50]. Due to the critical importance of accurate RNA editing in normal brain function, deregulation of editing may influence the progression of pathophysiological processes, such as neuro-degeneration and tumorigenesis [51].

Analyses of GluR-B transcripts in normal brain tissue showed that 100% of Q codons are edited to R codons in both gray and white matter. In contrast, GBM tissue showed a reduction in Q/R-site editing (12–31% decrease). This comes as a result of a decrease in the ADAR2 enzymatic activity [51]. Another study linked a reduction in editing of GluR-B (sites R/G and Y/C) and GluR-6 (sites I/V, Y/C and Q/R) transcripts to grade of malignancy in pediatric astrocytomas [47]. This editing has been attributed to alterations in ADAR2 catalytic activity. It has also been observed that elevated levels of ADAR2 inhibits proliferation and migration of astrocytoma cell lines [47]. Galeano et al. (2012) reported that ADAR2 editing affects proliferation of astrocytoma cells *in vitro* and *in vivo* via cell cycle modulation. Increased ADAR2 expression considerably prolonged survival and significantly inhibited astrocytoma growth *in vivo* [49].

Alternative splicing is also a critical mechanism that regulates ADAR2 activity [49, 50]. A splicing variant of ADAR2, which contains a 47-nucleotide insertion, is increased in gliomas in comparison to normal brain. Furthermore, the presence of this splicing variant correlates with malignancy [52]. Alternative splicing of exon 5a is also an important event in gliomas [53]. Transcripts containing exon 5a encode a protein with ~50% reduced activity that is predominantly expressed in gliomas [53].

ADAR2 is also essential for the editing and modulation of 90 miRNAs in glioblastomas. Rescue of ADAR2 activity in glioblastoma cells recovered the expression of onco-miRNAs and tumor suppressor miRNAs to levels observed in normal human brain. Wild-type ADAR2 activity was shown to inhibit miR-221, miR-222, and miR-21 maturation, causing accumulation of their precursors in different cell lines, impacting cell proliferation and migration [54].

2.5 Hu Antigen R

Hu antigen R (HuR) is a member of the *embryonic lethal abnormal visual* (ELAV) family, initially identified in *Drosophila* as a factor involved in neuronal development, plasticity and memory [55, 56]. HuR is ubiquitously expressed, whereas other members of the family (HuB, HuC and HuD) are tissue specific [56]. HuR regulates a variety of target mRNAs associated with processes like inflamation [57], cell cycle [58], angiogenesis [59], cell survival, and apotosis [60]. HuR preferently binds uracil-rich sequences and has been shown to regulate mRNA stability, splicing and translation [61, 62]. Recent studies have also linked HuR activity with microRNA metabolism and function [63, 64].

Angiogenesis is required for glioma development. Vascular endothelial growth factor (VEGF)-A is considered the major mediator of angiogenesis in malignant tumors, including high-grade astrocytomas [59]. HuR regulates VEGF-A expression and, under hypoxic conditions, inhibition of HuR cytoplasmic translocation by leptomycin B reduces VEGF-A upregulation in astrocytic tumor cells [59, 65]. In gliomas, it was shown that HuR also regulates the expression of angiogenic factors via mRNA stability [66, 67].

Several studies support a role of HuR in cell survival. In GBM, HuR binds the 3′UTR of bcl-2 family members and promotes both mRNA stability and translation. Using a mouse model, Filippova et al. (2011) showed that silencing of HuR promotes apoptosis by decreasing bcl-2 family members levels [68].

The development of an inflammatory microenvironment has long been considered important for the initiation and progression of glioblastoma. GBMs display high levels of the *pro-inflammatory cytokines like* IL-1β, IL-6 and IL-8 [69]. HuR increases the stability of IL-6 mRNA, leading to increased *pro-inflammatory cytokine* IL-6 protein production and secretion [57, 67].

Several studies have shown that HuR regulates the cell cycle in a variety of ways. One of them involves HuR phosphorylation at serine 202 by CDK5. Prolonged disruption of the balance between phosphorylated and unphosphorylated

HuR provokes an arrest of cell cycle progression in glioma cells by altered cyclin A levels [70]. Oscillation of cyclin A levels occurs during mitosis and is critical for DNA replication and centrosome duplication [71]. Moreover, SRC and c-Abl kinases regulate HuR sub-cellular trafficking and influence its accumulation in the pericentriolar matrix (PCM) via a growth factor dependent signaling mechanism [72]. Finally, HuR phosphorylation in the nucleus by Pyruvate kinase M2 (PKM2), results in increased glioma cell growth. The loss of the nuclear interaction between PKM2 and HuR leads to cytoplasmic redistribution of HuR and subsequently an increase in cap-independent mRNA translation of cyclin-dependent kinase inhibitor p27 mRNA, resulting in cell cycle arrest [73]. Some of the molecular mechanisms involved in the expression or activity of HuR in gliomas are mediated by growth factors. AKT/HSF1/HuR axis impacts Rictor expression, a component of the mTORC2 complex. EGF and IGF stimulation increases HuR transcription mediated by HSF1. HuR in turn, enhances Rictor mRNA translation which leads to elevated mTORC2 activity, tumor growth and invasion in GBM [74]. HuR is also subject to microRNA regulation. Yang et al. showed that miR-146b-5p overexpression could reverse HuR overexpression in glioma stem cells (GSCs). This regulation affects GSCs viability and cell cycle progression [75].

2.6 Musashi1

Musashi1 (MSI1) is a highly conserved RBP that controls the balance between self-renewal and differentiation [76]. MSI1 has been described as a pro-oncogenic factor in multiple tumor types [77]. High levels of MSI1 have been observed in several cancers, including medulloblastoma, hepatocarcinoma, cervical carcinoma, breast cancer and gliomas, and is linked to poor survival [78–82]. MSI1 regulates both translation and mRNA decay by binding to UAG motifs present in stem loops [83].

Uren et al. (2015) identified more than 1000 MSI1 target mRNAs in glioblastoma cells, using individual-nucleotide resolution cross-linking and immunoprecipitation (iCLIP) [83]. These targets are preferentially located in cancer relevant pathways such as focal adhesion, adherens junction, Wnt, JAK/STAT, p53, MAPK, VEGF, and ErbB. Functional assays showed that MSI1 knockdown impairs cell adhesion, migration, invasion, apoptosis, proliferation, and cell cycle regulation [83].

MSI1 has been linked to radio- and chemoresistance. MSI1 expression increases in response to DNA damage in glioblastoma cells [84]. MSI1 knockdown increases radiosensitivity by affecting DNA damage repair through regulation of DNA-activated catalytic polypeptide (DNA-PKcs). DNA-PKcs is a key enzyme involved in the classic nonhomologous end-joining (NHEJ) pathway of DNA double-strand break repair in mammals [84]. In addition, it has been demonstrated that overexpression of MSI1 effectively protects GBM cells from drug-induced apoptosis, like cisplatin, via down-regulation of pro-apoptotic genes [85].

2.7 Insulin-Like Growth Factor II mRNA Binding Proteins

The Insulin-like growth factor 2 (IGF2BP/IMPs) family of proteins consists of three members, IMP1 (IGF2BP1), IMP2 (IGF2BP2) and IMP3 (IGF2BP3), which are mainly expressed in early stages of embryogenesis [86]. High expression of IMP proteins has been observed in a broad range of cancer types, including pancreatic, lung, renal cell, ovarian, endometrial, cervical, and glioblastoma [87].

IMP3 is involved in the activation of MAPK and PI3K pathways through the activation of IGF-2, affecting cell proliferation and invasion of GBM cells [88]. Multivariate analysis identified high IMP3 as an unfavorable prognostic factor for pediatric [89] and adult astrocytoma patients [88, 90, 91].

Recent studies have shown that IMPs promote mRNA stability by preventing miRNA-mediated silencing [87]. In gliomas for instance, IMP2 protects mRNAs from let-7-dependent silencing by binding to the corresponding miRNA-binding sites [92].

2.8 Quaking

Quaking (QKI) belongs to the signaling transduction and activation of RNA (STAR) family of proteins. QKI pre-mRNA undergoes extensive alternative splicing to generate at least four transcripts producing isoforms termed QKI-5, QKI-6, QKI-7, and QKI-7b. These QKI isoforms share an RNA-binding KH domain, but differ by several amino acids at the C-terminus [93]. Analysis of glioma tumors found a high incidence of expression alterations in the human quaking gene [94]. It has also been reported that lost of the QKI-7 isoform decreases expression of genes involved in interferon (INF) induction, suggesting a role for QKI-7 as a regulator of the inflammatory pathway in glioblastoma [95]. QKI-6 expression correlates positively with glioma grade and promotes migration and invasion of glioblastoma cells via activation of PI3K/AKT and ERK pathways [96]. Recently, an in-frame MYB-QKI gene fusion was identified as hallmark genetic alteration in the majority of angiocentric gliomas [97–99]. The MYB-QKI rearrangement disrupts both MYB and QKI, resulting in hemizygous deletion of 3′ portion of MYB and the 5′ portion of QKI [99]. *In vitro* and *in vivo* studies show that the MYB-QKI fusion promotes tumorigenesis [97].

Acknowledgments Mitzli X. Velasco belongs to the "Programa de Doctorado en Ciencias Biomédicas, Facultad de Medicina, Universidad Nacional Autónoma de México, CONACYT. No. de becario 270268".

References

1. Maier T, Güell M, Serrano L (2009) Correlation of mRNA and protein in complex biological samples. FEBS Lett 583:3966–3973
2. de Sousa Abreu R, Penalva LO, Marcotte EM, Vogel C (2014) Global signatures of protein and mRNA expression levels. Mol BioSyst 5:1512–1526
3. Castello A, Fischer B, Eichelbaum K et al (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149:1393–1406
4. Gerstberger S, Hafner M, Tuschl T (2014) A census of human RNA-binding proteins. Nat Rev Genet 15:829–845
5. Mitchell SF, Parker R (2014) Principles and properties of eukaryotic mRNPs. Mol Cell 54:547–588
6. Galante PAF, Sandhu D, Abreu RDS, Gradassi M, Vogel C, De Souza SJ, Penalva LOF (2009) A comprehensive in silico expression analysis of RNA binding proteins in normal and tumor tissue: identification of potential players in tumor formation. RNA Biol 6:426–433
7. Kechavarzi B, Janga SC (2014) Dissecting the expression landscape of RNA-binding proteins in human cancers. Genome Biol 15:R14
8. Wang J, Liu Q, Shyr Y (2015) Dysregulated transcription across diverse cancer types reveals the importance of RNA-binding protein in carcinogenesis. BMC Genomics 16:S5
9. Weller M, Wick W, Aldape K, Brada M, Berger M, Nishikawa R, Rosenthal M, Wen PY, Stupp R (2015) Glioma. Natl Rev 1:1–18
10. Thakkar JP, Dolecek TA, Horbinski C, Ostrom QT, Lightner DD, Barnholtz-Sloan JS, Villano JL (2014) Epidemiologic and molecular prognostic review of glioblastoma. Cancer Epidemiol Biomark Prev 23:1985–1996
11. Ostrom QT, Gittleman H, Farah P, Ondracek A, Chen Y, Wolinsky Y, Stroup NE, Kruchko C, Barnholtz-sloan JS (2013) CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2006–2010. J Neuro-Oncol 15:788–796
12. Correa BR, De Araujo PR, Qiao M, Burns SC, Chen C, Schlegel R, Agarwal S, Galante PAF, Penalva LOF (2016) Functional genomics analyses of RNA-binding proteins reveal the splicing regulator SNRPB as an oncogenic candidate in glioblastoma. Genome Biol 17:1–16
13. Han N, Li W, Zhang M (2013) The function of the RNA-binding protein hnRNP in cancer metastasis. J Cancer Res Ther 9:129–134
14. Deng J, Chen S, Wang F, Zhao H (2016) Effects of hnRNP A2/B1 knockdown on inhibition of glioblastoma cell invasion, growth and survival. Mol Neurobiol 53:1132–1144
15. Sueoka E, Sueoka N, Iwanaga K, Sato A, Suga K, Hayashi S, Nagasawa K, Nakachi K (2005) Detection of plasma hnRNP B1 mRNA, a new cancer biomarker, in lung cancer patients by quantitative real-time polymerase chain reaction. Lung Cancer 48:77–83
16. Chen S, Zhang J, Duan L, Zhang Y, Li C, Liu D, Ouyang C, Lu F, Liu X (2014) Identification of HnRNP M as a novel biomarker for colorectal carcinoma by quantitative proteomics. Am J Physiol Gastrointest Liver Physiol 306:394–403
17. Loh TJ, Moon H, Cho S et al (2015) CD44 alternative splicing and hnRNP A1 expression are associated with the metastasis of breast cancer. Oncol Rep 34:1231–1238
18. Noguchi T, Yamada K, Inoue H, Matsuda T, Tanaka T (1987) The L- and R-type isozymes of rat pyruvate kinase are produced from a single gene by use of different promoters. J Biol Chem 262:14366–14371

19. Noguchi T, Inoue H, Tanaka T (1986) The M1- and M2-type isozymes of rat pyruvate kinase are produced from the same gene by alternative RNA splicing. J Biol Chem 261:13807–13812
20. David CJ, Chen M, Assanah M, Canoll P, Manley JL (2010) HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. Nature 463:364–368
21. Clower CV, Chatterjee D, Wang Z, Cantley LC (2010) The alternative splicing repressors hnRNP A1/A2 and PTB influence pyruvate kinase isoform expression and cell metabolism. Proc Natl Acad Sci U S A 107:10–15
22. Parsons DW, Jones S, Zhang X et al (2008) An integrated genomic analysis of human glioblastoma multifome. Science 321:1807–1813
23. Guo D, Hildebrandt IJ, Prins RM et al (2009) The AMPK agonist AICAR inhibits the growth of EGFRvIII-expressing glioblastomas by inhibiting lipogenesis. Proc Natl Acad Sci 106:12932–12937
24. Babic I, Anderson ES, Tanaka K et al (2013) EGFR mutation-induced alternative splicing of max contributes to growth of glycolytic tumors in brain cancer. Cell Metab 17:1000–1008
25. Hanahan D, Weinberg RA, Francisco S (2000) The hallmarks of cancer. Cell 100:57–70
26. Efimova EV, Al-zoubi AM, Martinez O, Kaithamana S, Lu S, Arima T, Prabhakar BS (2004) IG20, in contrast to DENN-SV, (MADD splice variants) suppresses tumor cell survival, and enhances their susceptibility to apoptosis and cancer drugs. Oncogene 23:1076–1087
27. Prabhakar BS, Mulherkar N, Prasad KV (2008) Role of IG20 splice variants inTRAIL resistance. Mol Pathways 14:347–352
28. Ghigna C, Giordano S, Shen H, Benvenuto F, Castiglioni F, Comoglio PM, Green MR, Riva S, Biamonti G, Ferrata V (2005) Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. Mol Cell 20:881–890
29. Lefave CV, Squatrito M, Vorlova S, Rocco GL, Brennan CW, Holland EC, Pan Y, Cartegni L (2011) Splicing factor hnRNPH drives an oncogenic splicing switch in gliomas. EMBO J 30:4084–4097
30. Park YM, Hwang J, Masuda K, Choi K, Jeong M, Nam D, Gorospe M (2012) Heterogeneous nuclear ribonucleoprotein C1/C2 controls the metastatic potential of glioblastoma by regulating PDCD4. Mol Cell Biol 32:4237–4244
31. Ling JP, Chhabra R, Merran JD, Paul M, Wheelan SJ, Jeffry L, Wong PC (2016) PTBP1 and PTBP2 repress nonconserved cryptic exons. Cell Rep 17:104–113
32. Romanelli MG, Diani E, Lievens PM-J (2013) New insights into functional roles of the polypyrimidine tract-binding protein. Int J Mol Sci 14:22906–22932
33. Boutz PL, Stoilov P, Li Q, Lin C, Chawla G, Ostrow K, Shiue L, Jr MA, Black DL (2007) A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. Genes Dev 21:1636–1652
34. Yamamoto H, Tsukahara K, Kanaoka Y, Jinno S (1999) Isolation of a mammalian homologue of a fission yeast differentiation regulator. Mol Cell Biol 19:3829–3841
35. Garcia-Blanco MA, Jamison SF, Sharp PA (1989) Identification and purification of a 62,000-Dalton protein that binds specifically to the polypyrimidine tract of introns. Genes Dev 3:1874–1886
36. Sawicka K, Bushell M, Spriggs KA, Willis AE (2008) Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. Biochem Soc Trans 36:641–647
37. Cheung HC, Hai T, Zhu W, Baggerly KA, Tsavachidis S, Krahe R, Cote GJ (2009) Splicing factors PTBP1 and PTBP2 promote proliferation and migration of glioma cell lines. Brain J Neurol 132:2277–2288
38. Cheung HC, Corley LJ, Fuller GN, Mccutcheon IE, Cote GJ (2006) Polypyrimidine tract binding protein and Notch1 are independently re-expressed in glioma. Mod Pathol 19:1034–1041
39. Schwab JM, Tuli SK, Failli V (2006) The Nogo receptor complex: confining molecules to molecular mechanisms. Trends Mol Med 12:293–297
40. Oertle T, Huber C, van der Putten H, Schwab M (2003) Genomic structure and functional characterisation of the promoters of human and mouse Nogo/Rtn-4. J Mol Biol 325:299–323
41. Jin W, Mccutcheon IE, Fuller GN, Huang ES, Cote GJ (2000) Fibroblast growth factor receptor-1 a-exon exclusion and polypyrimidine tract- binding protein in glioblastoma multiforme tumors. Cancer Res 60:1221–1224
42. Yamaguchi F, Saya H, Bruner JM, Morrison R (1994) Differential expression of two fibroblast growth factor-receptor genes is associated with malignant progression in human astrocytomas. Proc Natl Acad Sci 91:484–488
43. Marx A, Nugoor C, Panneerselvam S, Mandelkow E (2010) Structure and function of polarity-inducing kinase family MARK/Par-1 within the branch of AMPK /Snf1-related kinases. FASEB J 24:1637–1648
44. Moroni RF, Biasi SDE, Colapietro P, Larizza L, Beghini A (2006) Distinct expression pattern of microtubule-associated protein/microtubule affinity-regulating kinase 4 in differentiated neurons. Neuroscience 143:83–94
45. Fontana L, Rovina D, Novielli C, Maffioli E, Tedeschi G, Magnani I (2015) Suggestive evidence on the involvement of polypyrimidine-tract binding protein in regulating alternative splicing of MAP/microtubule affinity-regulating kinase 4 in glioma. Cancer Lett 359:87–96
46. Farajollahi S, Maas S (2010) Molecular diversity through RNA editing: a balancing act. Trends Genet 26:221–230
47. Cenci C, Barzotti R, Galeano F, Corbelli S, Rota R, Massimi L, Di Rocco C, Connell MAO, Gallo A (2008) Down-regulation of RNA editing in pediatric astrocytomas. ADAR2 editing activity inhibits cell migration and proliferaron. J Biol Chem 283:7251–7260

48. Fu YAO, Zhao X, Li Z, Wei JUN, Tian YU (2016) Splicing variants of ADAR2 and ADAR2-mediated RNA editing in glioma (review). Oncol Lett 12:788–792
49. Galeano F, Rossetti C, Tomaselli S, Cifaldi L, Lezzerini M, Pezzullo M, Boldrini R, Massimi L, Di Rocco CM (2012) ADAR2-editing activity inhibits glioblastoma growth through the modulation of the CDC14B/Skp2/p21/p27 axis. Oncogene 32:998–1009
50. Brusa R, Zimmermann F, Koh D, Feldmeyer D, Gass P, Seeburg PH, Sprengel R (1995) Early-onset epilepsy and postnatal lethality associated with an editing-deficient GluR-8 allele in mice. Science 270:1677–1680
51. Maas S, Patt S, Schrey M, Rich A (2001) Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. Neurobiology 98:14687–14692
52. Wei J, Li Z, Du C et al (2014) Abnormal expression of an ADAR2 alternative splicing variant in gliomas downregulates adenosine-to-inosine RNA editing. Acta Neurochir 156:1135–1142
53. Li Z, Tian Y, Tian N, Zhao X, Du C, Han L, Zhang H (2015) Aberrant alternative splicing pattern of ADAR2 downregulates adenosine-to-inosine editing in glioma. Oncol Rep 33:2845–2852
54. Tomaselli S, Galeano F, Alon S et al (2015) Modulation of microRNA editing, expression and processing by ADAR2 deaminase in glioblastoma. Genome Biol 16:5
55. Srikantan S, Gorospe M (2015) HuR function in disease. Front Biosci 17:189–205
56. Hinman MN, Lou H (2008) Diverse molecular functions of Hu proteins. Cell Mol Life Sci 65:3168–3181
57. Gurgis FMS, Yeung YT, Tang MXM, Heng B, Buckland M, Ammit AJ, Haapasalo J, Haapasalo H, Guillemin GJ, Grewal T (2015) The p38-MK2-HuR pathway potentiates EGFRvIII-IL-1β-driven IL-6 secretion in glioblastoma cells. Oncogene 34:2934–2942
58. Wang W, Caldwell MC, Lin S, Furneaux H, Gorospe M (2000) HuR regulates cyclin A and cyclin B1 mRNA stability during cell proliferation. EMBO J 19:2340–2350
59. Ido K, Nakagawa T, Sakuma T, Takeuchi H, Sato K (2008) Expression of vascular endothelial growth factor-A and mRNA stability factor HuR in human astrocytic tumors. Neuropathology 28:604–611
60. Abdelmohsen K, Gorospe M (2007) Posttranscriptional orchestration of an anti-apoptotic program by HuR. Cell Cycle 6:1288–1292
61. López de Silanes I, Zhan M, Lal A, Yang X, Gorospe M (2004) Identification of a target RNA motif for RNA-binding protein HuR. Proc Natl Acad Sci U S A 101:2987–2992
62. Uren PJ, Burns SC, Ruan J, Singh KK, Smith AD, Penalva LOF (2011) Genomic analyses of the RNA-binding protein hu antigen R (HuR) identify a complex network of target genes and novel characteristics of its binding sites. J Biol Chem 286:37063–37066
63. Mukherjee K, Ghoshal B, Ghosh S, Chakrabarty Y, Das S, Bhattacharyya SN (2016) Reversible HuR-microRNA binding controls extracellular export of miR-122 and augments stress response. EMBO Rep 17:1184–1203
64. Young LE, Moore AE, Sokol L, Meisner-kober N, Dixon DA (2012) The mRNA stability factor HuR inhibits MicroRNA-16 targeting of cyclooxygenase-2. Mol Cancer Res 10:167–180
65. Goldberg-cohen I, Furneauxb H, Levy AP (2002) A 40-bp RNA element that mediates stabilization of vascular endothelial growth factor mRNA by HuR. J Biol Chem 277:13635–13640
66. Bolognani F, Gallani A-I, Sokol L, Baskin DS, Meisner-Kober N (2012) mRNA stability alterations mediated by HuR are necessary to sustain the fast growth of glioma cells. J Neuro-Oncol 106:531–542
67. Nabors LB, Gillespie GY, Harkins L, King PH (2001) HuR, a RNA stability factor, is expressed in malignant brain tumors and binds to adenine- and uridine-rich elements within the 3′ untranslated regions of cytokine and angiogenic factor mRNAs. Cancer Res 61:2154–2161
68. Filippova N, Yang X, Wang Y, Gillespie GY, Langford C (2011) The RNA-binding protein HuR promotes glioma growth and treatment resistance. Mol Cancer Res 9:648–660
69. Yeung YT, Mcdonald KL, Grewal T, Munoz L (2013) Interleukins in glioblastoma pathophysiology: implications for therapy. Br J Pharmacol 168:591–606
70. Filippova N, Yang X, King P, Nabors LB (2012) Phosphoregulation of the RNA-binding protein Hu antigen R (HuR) by Cdk5 affects centrosome function. J Biol Chem 287:32277–32287
71. De Boer L, Oakes V, Beamish H, Giles N, Stevens F, Desouza C (2008) Cyclin A/cdk2 coordinates centrosomal and nuclear mitotic events. Oncogene 27:4261–4268
72. Filippova N, Yang X, Nabors LB (2015) Growth factor dependent regulation of centrosome function and genomic instability by HuR. Biomol Ther 5:263–281
73. Mukherjee J, Ohba S, See WL, Phillips JJ, Molinaro AM, Pieper RO (2016) PKM2 uses control of HuR localization to regulate p27 and cell cycle progression in human glioblastoma cells. Int J Cancer 139:99–111
74. Holmes B, Benavides-Serrato A, Freeman RS, Landon KA, Bashir T, Nishimura RN, Gera J (2017) mTORC2/AKT/HSF1/HuR constitute a feed-forward loop regulating Rictor expression and tumor growth in glioblastoma. Oncogene 37:732–743
75. Yang W, Yu H, Shen Y, Liu Y, Yang Z (2016) MiR-146b-5p overexpression attenuates stemness and radioresistance of glioma stem cells by targeting HuR/lincRNA-p21/β-catenin pathway. Oncotarget 7:41505–41526
76. Okano H, Kawahara H, Toriya M, Nakao K, Shibata S, Imai T (2005) Function of RNA-binding protein Musashi-1 in stem cells. Exp Cell Res 306:349–356
77. Kudinov AE, Karanicolas J, Golemis EA, Boumber Y (2017) Musashi RNA-binding proteins as cancer

drivers and novel therapeutic targets. Clin Cancer Res 23:2143–2154
78. Shu H-J, Saito T, Watanabe H, Ito J-I, Takeda H, Okano H, Kawata S (2002) Expression of the Musashi1 gene encoding the RNA-binding protein in human hepatoma cell lines. Biochem Biophys Res Commun 293:150–154
79. Ye F, Zhou C, Cheng Q, Shen J, Chen H (2008) Stem-cell-abundant proteins nanog, nucleostemin and Musashi1 are highly expressed in malignant cervical epithelial cells. BMC Cancer 8:108
80. Wang X-Y, Penalva LO, Yuan H, Linnoila RI, Lu J, Okano H, Glazer RI (2010) Musashi1 regulates breast tumor cell proliferation and is a prognostic indicator of poor survival. Mol Cancer 9:221
81. Dahlrot RH, Hansen S, Hjelmborg J, Kristensen BW (2013) Prognostic value of Musashi-1 in gliomas. J Neuro-Oncol 115:453–461
82. Vo DT, Subramaniam D, Remke M et al (2012) The RNA-binding protein Musashi1 affects medulloblastoma growth via a network of cancer- related genes and is an indicator of poor prognosis. Am J Pathol 181:1762–1772
83. Uren PJ, Vo DT, De Araujo R et al (2015) RNA-binding protein Musashi1 is a central regulator of adhesion pathways in glioblastoma. Mol Cell Biol 35:2965–2978
84. de Araujo PR, Grothi A, da Silva AE, Tonapi SS, Burns SC, Qiao M, Uren PJ, Yuan Z-M, Bishop AJR, Penalva LOF (2016) Musashi1 impacts radioresistance in glioblastoma by controlling DNA-protein kinase catalytic subunit. Am J Pathol 186:2271–2278
85. Chen H, Lin L, Wang M et al (2016) Musashi-1 regulates AKT-derived IL-6 autocrinal/paracrinal malignancy and chemoresistance in glioblastoma. Oncotarget 7:42485–42501
86. Nielsen J, Christiansen JAN, Lykke-andersen J, Johnsen AH, Wewer UM, Nielsen FC (1999) A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development. Mol Cell Biol 19:1262–1270
87. Degrauwe N, Suvà M, Janiszewska M, Riggi N, Stamenkovic I (2016) IMPs: an RNA-binding protein family that provides a link between stem cell maintenance in normal development and cancer. Genes Dev 30:2459–2474
88. Suvasini R, Shruti B, Thota B et al (2011) Insulin growth factor-2 binding protein 3 (IGF2BP3) is a glioblastoma-specific marker that activates phosphatidylinositol 3-kinase/mitogen-activated protein kinase (PI3K/MAPK) pathways by modulating IGF-2. J Biol Chem 286:25882–25890
89. Barton VN, Donson AM, Birks DK, Kleinschmidt-demasters BK, Handler MH, Foreman NK, Rush SZ (2013) Insulin-like growth factor 2 mRNA binding protein 3 expression is an independent prognostic factor in pediatric pilocytic and pilomyxoid astrocytoma. J Neuropathol Exp Neurol 72:442–449
90. Serão NVL, Delfino KR, Southey BR, Beever JE, Rodriguez-zas SL (2011) Cell cycle and aging, morphogenesis, and response to stimuli genes are individualized biomarkers of glioblastoma progression and survival. BMC Med Genet 4:49
91. Shu C, Wang Q, Yan X, Wang J (2018) Whole-genome expression microarray combined with machine learning to identify prognostic biomarkers for high-grade glioma. J Mol Neurosci 64:491–500
92. Degrauwe N, Schlumpf TB, Janiszewska M, Martin P, Cauderay A, Provero P, Riggi N, Suvà M, Paro R, Stamenkovic I (2016) The RNA binding protein IMP2 preserves glioblastoma stem cells by preventing let-7 target gene silencing. Cell Rep 15:1634–1647
93. Darbelli L, Richard S (2016) Emerging functions of the quaking RNA-binding proteins and link to human diseases. Wiley Interdiscip Rev RNA 7:399–412
94. Li ZZ, Kondo T, Murata T, Ebersole TA, Nishi T, Ushio Y, Yamamura K, Abe K (2002) Expression of Hqk encoding a KH RNA binding protein is altered in human glioma. Jpn J Cancer Res 93:167–177
95. Jiang L, Saetre P, Radomska KJ, Jazin E, Carlstrom EL (2010) QKI-7 regulates expression of interferon-related genes in human astrocyte glioma cells. PLoS One 5:e13079
96. Xi Z, Wang P, Xue Y, Shang C, Liu X, Ma J, Li Z, Li Z, Bao M, Liu Y (2017) Overexpression of miR-29a reduces the oncogenic properties of glioblastoma stem cells by downregulating quaking gene isoform 6. Oncotarget 8:24949–24963
97. Bandopadhayay P, Ramkissoon LA, Jain P et al (2016) MYB-QKI rearrangements in angiocentric glioma drive tumorigenicity through a tripartite mechanism. Nat Genet 48:273–282
98. Aronco LD, Rouleau C, Gayden T, Crevier L, Claude J, Sébastien D, Nada P, Bandopadhayay P, Ligon KL, Ellezam B (2017) Brainstem angiocentric gliomas with MYB – QKI rearrangements. Acta Neuropathol 134:667–669
99. Chan E, Bollen AW, Sirohi D, Van Ziffle J, Grenert JP, Kline CN, Tihan T, Perry A, Gupta N, Solomon DA (2017) Angiocentric glioma with MYB-QKI fusion located in the brainstem, rather than cerebral cortex. Acta Neuropathol 134:671–673

Nonsense-Mediated mRNA Decay in Development, Stress and Cancer

3

Rafael Fernandes, Gonçalo Nogueira, Paulo J. da Costa, Francisco Pinto, and Luísa Romão

Abstract

Nonsense-mediated mRNA decay (NMD) is a well characterized eukaryotic mRNA degradation pathway, responsible for the identification and degradation of transcripts harboring translation termination codons in premature contexts. Transcriptome-wide studies revealed that NMD is not only an mRNA surveillance pathway as initially thought, but is also a post-transcriptional regulatory mechanism of gene expression, as it fine-tunes the transcript levels of many wild-type genes. Hence, NMD contributes to the regulation of many essential biological processes, including pathophysiological mechanisms. In this chapter we discuss the importance of NMD and of its regulation to organism development and its link to the cellular stress responses, like the unfolded protein response (UPR) and the integrated stress response (ISR). Additionally, we describe how tumor cells have explored both NMD functions to promote tumorigenesis. Using published data and databases, we have also performed a network-based approach that further supports the link between NMD and these (patho) physiological processes.

Keywords

Nonsense-mediated mRNA decay · mRNA surveillance · Gene expression regulation · Unfolded protein response · Integrated stress response · Tumorigenesis · Human disease

3.1 Introduction

As already discussed in the previous chapter, eukaryotic mRNAs go through a complex process as they pass the genetic information from DNA to protein. This includes many stages, such as transcription, capping, pre-mRNA splicing, polyadenylation, transport, translation and degradation. Some of these steps may determine the production of different mRNAs with the consequent generation of distinct proteins encoded by the same gene. Cells have developed many surveillance mechanisms to control the quality of the mRNAs throughout this process to ensure correct protein synthesis [1, 2]. Among them is

R. Fernandes · G. Nogueira · P. J. da Costa · L. Romão (✉)
Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal
e-mail: luisa.romao@insa.min-saude.pt

F. Pinto
Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

L. Romão (ed.), *The mRNA Metabolism in Human Disease*, Advances in Experimental Medicine and Biology 1157, https://doi.org/10.1007/978-3-030-19966-1_3

nonsense-mediated mRNA decay (NMD), a translation-dependent surveillance mechanism that detects and rapidly degrades transcripts with premature translation termination codons (PTCs) produced by mutated genes or by errors in the mRNA processing. By doing so, NMD protects the cell from the production of potentially harmful truncated proteins [3, 4].

The development of high throughput technologies revealed an additional function of NMD that was unknown for many years. Transcriptome profiling of cells depleted of NMD factors showed that NMD also modulates the levels of many transcripts that encode normal, full-length proteins [5–15]. Thus, NMD arises as a mechanism of gene expression regulation. This new feature suggests that NMD might play an important role in the regulation of many essential biological processes. In this chapter, we briefly discuss the role of NMD during mammalian development and cellular stress responses, and how its time-dependent regulation is important for these mechanisms. We also describe how tumor cells have explored both the quality and regulatory NMD functions to leverage tumorigenesis in their own microenvironment. In the end, to further explore and support the links between NMD and development, stress and cancer, we show a network-based approach using publicly available data.

3.2 Nonsense-Mediated mRNA Decay: The Mechanism and Its Dual Function in the Cell

Since its discovery, NMD has been described as a surveillance mechanism responsible for targeting and rapidly degrading eukaryotic mRNAs harboring PTCs. This protects the cell from the potential deleterious effects that could arise from the generation of C-terminally truncated proteins [16, 17]. NMD was also implicated in human diseases, such as β-thalassemia, cystic fibrosis, Duchenne's muscular dystrophy and cancer [17, 18]. This is not surprising, since it is estimated that around 30% of the genetic diseases arise as a consequence of PTC-introducing mutations that, therefore, can be affected by NMD [18].

There are two main models explaining the NMD mechanism: the classical EJC-dependent and the EJC-independent model (Fig. 3.1) [19]. The EJC-dependent model states that NMD depends on the interaction of the translation termination complex with a multiprotein complex, called exon junction complex (EJC) (Fig. 3.1a) [20–23]. The EJCs are deposited 20–24 nucleotides (nts) upstream of most exon-exon junctions during splicing, remaining associated with the mRNA during its transport to the cytoplasm [20, 22, 23]. It is well established that translating ribosomes displace the EJCs from the open reading frame (ORF) during the pioneer round of translation [24, 25]. However, the ribosome is not able to displace the EJCs whenever an mRNA contains a PTC located more than 50–54 nts upstream to the last exon-exon junction (Fig. 3.1a.1) [20, 26]. Consequently, when the ribosome stops at a PTC, the translation eukaryotic release factors (eRF) 1 and 3 interact with the EJC(s) [27]. The ribosome-EJC interaction is mediated by a multiprotein complex called "SURF", composed by the eRF1 and eRF3, the phosphatidylinositol 3-kinase (PI3K) related kinase, SMG1, associated to SMG8 and SMG9, and UPF1 (Fig. 3.1a.2) [27, 28]. SMG8 and SMG9 bind to the SMG1 in its inactive conformation, mediating its transition to an active state. At the same time, the EJC containing UPF3B interacts with UPF2, which then links the SURF complex to the EJC through its interaction with UPF1, forming the decay-inducing complex (DECID) (Fig. 3.1a.3) [29, 30]. This "SURF-DECID" model was recently challenged by data from Neu-Yilik et al., showing that UPF3B rather than UPF1 interacts with eRF3 during premature translation termination to promote peptide release and dissociation of the termination complexes [31]. UPF1 may also be present at the terminating ribosomes but in an inactive state. They have also shown that UPF3B can directly interact with UPF1, contradicting the previous idea that UPF2 bridges this interaction [31]. Despite this model is consistent with some mechanistically aspects of NMD, it requires further testing and

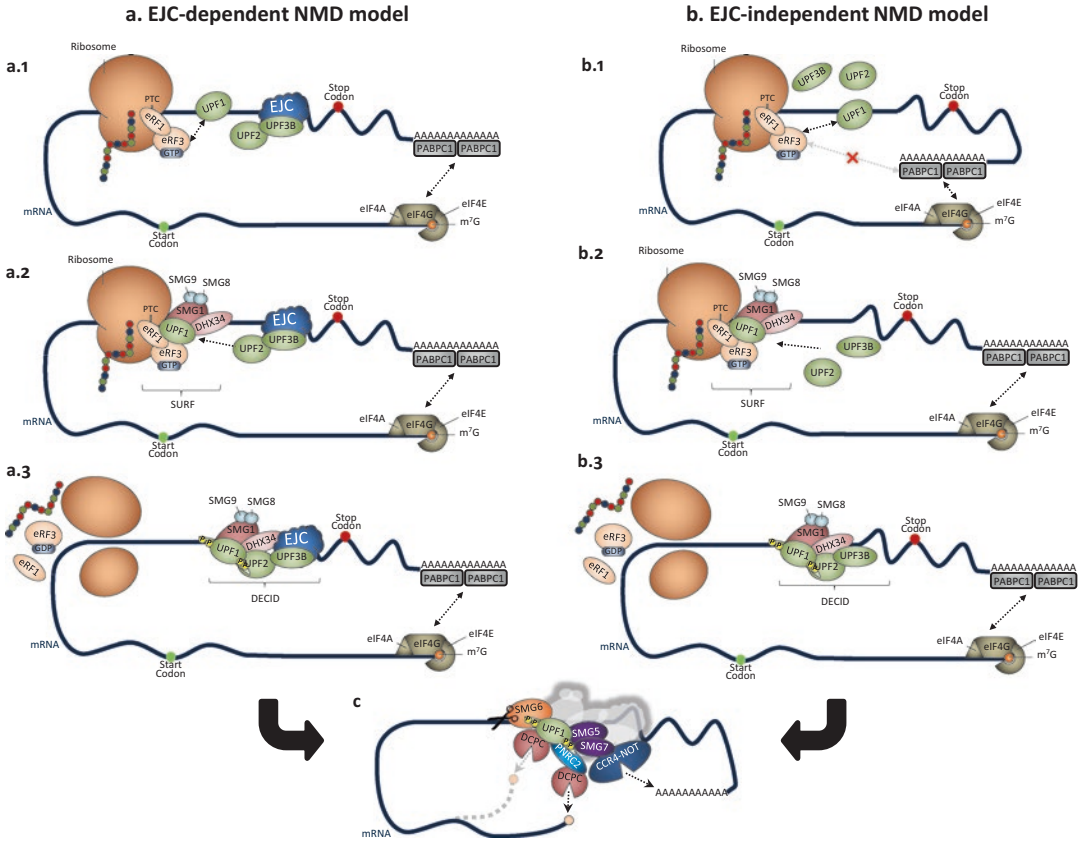

Fig. 3.1 Simplified representation of the NMD models. (**a**) The EJC-dependent model. When the ribosome stops at a PTC with a downstream EJC, the interaction of UPF1 and eRF3 induces premature translation termination and triggers NMD (**a.1**). After this interaction, the SURF complex is formed by eRF1, eRF3, SMG1 associated with SMG8 and SMG9, DHX34 and UPF1 (**a.2**). UPF1 interacts with UPF2-UPF3B bound to the EJC downstream of the PTC to form the DECID complex, which allows the SMG1-mediated phosphorylation of UPF1. At this point, translation has terminated with the dissociation of the ribosomal subunits, the release factors and the nascent peptide (**a.3**). (**b**) The EJC-independent model. In this model, when the ribosome reaches a PTC, the physical distance between eRF3 and the PABPC1 favors the interaction of eRF3 with UPF1, determining a premature translation termination process that triggers NMD (**b.1**). After this interaction, the SURF complex is formed (**b.2**). UPF2 and UPF3B diffused in the cytoplasm interact with UPF1, favoring its phosphorylation by SMG1 and forming the DECID complex (**b.3**). (**c**) The NMD decay phase. Phosphorylated UPF1 recruits the factors that lead to mRNA degradation: SMG6, which produces an endonucleolytic cleavage, SMG5-SMG7 dimer, which recruits the CCR4-NOT deadenylase complex, and/or PNRC2, which recruits the decapping complex (DCPC)

establishes contradictions with previously published data that need to be clarified [32].

According to the prevailing model for the decay phase of NMD, the interaction between UPF1, UPF2 and UPF3B leads to a conformational change in UPF1, allowing its phosphorylation at serine and threonine residues of the C-terminal domain, accomplished by the SMG1 protein (Fig. 3.1a.3) [27, 29, 33–37]. More recently, a new relevant player for the SURF-DECID transition was unveiled, the DHX34 protein. The DHX34 was shown not only to operate as a scaffold protein for UPF1-SMG1 interaction, but also to promote the interaction between UPF1 and UPF2 [38, 39]. Thus, DHX34 appears to be relevant for UPF1 phosphorylation that ultimately leads to NMD triggering by the recruitment of the NMD factors SMG6, SMG5–7 heterodimer, or SMG5 and the proline-rich nuclear receptor co-regulatory protein 2 (PNRC2)

(Fig. 3.1c) [37–40]. The SMG6 recruitment to PTC-containing mRNAs induces an endonucleolytic cleavage in the vicinity of the PTC [41, 42]. The SMG6 endonucleolytic cleavage, is mediated by its C-terminal PilT N terminus (PIN) [41, 43], which generates unprotected mRNA ends that condemn it to degradation [44–46]. SMG5-SMG7 or SMG5-PNRC2 complexes further recruit decapping enzymes (DCP1 and DCP2) and the CCR4-NOT deadenylation complex that remove the 5′ and 3′ modifications (Fig. 3.1c), allowing 5′-to-3′ and 3′-to-5’ RNA degradation by XRN1 and the RNA exosome, respectively [41, 42, 47–50].

In contrast with the EJC-dependent NMD model, the EJC-independent NMD pathway (Fig. 3.1b) postulates that the recognition of a stop codon as a PTC depends on the physical distance between the termination complex at the PTC, and the cytoplasmic poly(A)-binding protein 1 (PABPC1) bound to the poly(A) tail. This seems to rely on the PABPC1 and UPF1 competition for the interaction with the eRF3 at the terminating ribosome (Fig. 3.1b.1) [51–54]. If PABPC1 is in close proximity to the termination complex at the PTC, it represses NMD by interacting with eRF3 and preventing the UPF1-eRF3 interaction [51–54]. The eRF3 interaction with PABPC1, is proposed to be crucial for normal termination [27, 55, 56]. However, other studies have found that NMD suppression by PABPC1 may occur even without the interaction with eRF3 [57, 58]. Instead, it can suppress NMD through interaction with the eIF4G [57, 59]. Nevertheless, according to this model, since there is no EJC to potentiate NMD, the interaction between UPF1, UPF2 and UPF3 may occur by diffusion of the two last factors present in the cytoplasm to the mRNA-bound UPF1 (Fig. 3.1b.2) [19]. The following steps of NMD are the same from the EJC-dependent model (Fig. 3.1a.3, 3.1b.3, 3.1c).

Interestingly, while UPF1 is required for both the EJC-dependent and EJC-independent NMD pathways, UPF2 and UPF3B abundance appear to be more relevant for the EJC-independent NMD pathway [19]. Indeed, Metze et al. showed that silencing UPF2 and UPF3B alone or together strongly inhibits the EJC-independent NMD, while NMD of EJC-containing reporters is only slightly affected. The authors suggest that lower cellular concentrations of UPF2 and UPF3B affect mainly the diffusion of these factors to the target mRNA in the cytoplasm rather than the EJC-bounded factors [19]. This dependence on distinct NMD factors suggested the existence of NMD branches. In higher eukaryotes, it was observed that NMD could be either UPF2-independent, UPF3-independent or, as mentioned above, EJC-independent [23, 60, 61]. Supporting this, Huang et al. found that NMD owns an autoregulatory feedback loop mechanism that is dependent on the different branches of NMD. Apparently, the UPF3B-dependent branch regulates UPF1 and SMG7 transcripts, the EJC-dependent branch regulates UPF1 and SMG5 transcripts, and the UPF2-dependent branch regulates SMG1 mRNA levels [19, 62]. Recently, it was found that the SRSF1 splicing factor promotes NMD by binding directly to UPF1, enhancing the binding of the latter to the mRNA. Moreover, the SRSF1 action is UPF3B- and UPF2-independent, which could constitute a new branch of NMD [63].

Although initial studies identified NMD as a quality control mechanism that degrades aberrant transcripts derived from mutated genes or from errors in the mRNA processing, in the recent years transcriptome-wide studies revealed that NMD also degrades many mRNAs from normal, functional protein-coding genes. NMD, thus, arises also as a mechanism of post-transcriptional control of gene expression, being directly or indirectly responsible for the regulation of ~3–20% of transcripts in eukaryotes from yeast to mammals [5–14]. These endogenous, natural NMD targets may present specific features – NMD-inducing features – that are responsible for eliciting their decay. As seen above, the presence of an exon-exon junction downstream of a stop codon allows the interaction between the SURF complex recruited to a PTC and the EJC, triggering NMD (reviewed in [64]). Upstream open reading frames (uORFs) in the 5′ untranslated region

(UTR) of the transcripts can also trigger NMD, possibly because the stop codon of the uORF is at the 5′-end of the mRNA with downstream EJCs, placing it in a premature context [65]. Another NMD-inducing feature is a long 3'UTR. How NMD targets these transcripts to decay is not completely understood, but it may involve a mechanism that depends on the physical distance between the stop codon and the PABPC1 at the poly(A) tail, which has been related to translation termination efficacy and to NMD [53, 54, 66–68], as mentioned above.

The ability to regulate a subset of normal transcripts, suggested that NMD could play important roles in the regulation of normal biological pathways beyond its quality control function. Indeed, intensive research during the past years implicated NMD activity in organism development, cell differentiation, cell stress and immune responses [69]. In the following sections we will discuss these physiological and pathophysiological roles of NMD in mammalian cells, focusing on organism development, cell stress and cancer.

3.3 NMD is Required for Embryonic Viability and Influences Mammalian Development

The impairment of many core NMD factors, including UPF1, UPF2, UPF3A, SMG1, SMG6 and SMG9, revealed that these proteins are essential for mammalian embryogenesis (reviewed in [69]). As an example, while mouse embryos heterozygous for Upf1 knockout present no phenotypic abnormalities, embryos with a complete knockout of Upf1 with total loss of NMD are only viable before implantation [embryonic day 3.5 (E3.5)] [69, 70]. In addition, cultured Upf1$^{-/-}$ preimplantation blastocysts show strong apoptosis induction after 5 days in culture [70]. In agreement, Smg6 knockout mouse embryos die soon after implantation, apparently because Smg6 deletion induces sustained expression of pluripotency genes normally repressed by NMD, thus impairing embryonic stem cell differentiation [71]. Also, Upf2 and Smg1 knockout experiments showed that mouse embryos lose viability between E3.5 and E9.5 [8, 69] and before E12.5 [72], respectively. In recent studies, Upf3a$^{-/-}$ mouse embryos died between E4.5 and E8.5, showing poor morphology and defects at E3.5 [30]. Additionally, intercrosses of heterozygous mice for a null allele of Smg9 yielded no homozygous progeny, suggesting an embryonic lethal phenotype. Smg9$^{-/-}$ embryos harvested at E15.5 presented variable and incompletely penetrant phenotypes including edema, hemorrhage and exencephaly [73]. Together, these findings point out the relevance of NMD for normal development. However, as some NMD factors also play a role in other NMD-independent biological processes, it is still important to confirm whether the observed embryonic lethality is related to NMD inhibition *per se*, or if it is a consequence of the impairment of non-NMD roles [69]. For example, there is evidence that UPF1, SMG1 and SMG6 are involved in telomere maintenance [74, 75]. Additionally, UPF1 also plays roles in cell cycle, histone mRNA degradation and Staufen (STAU) 1-mediated mRNA decay (SMD) (reviewed in [75]). To confirm that the developmental defects observed after the knockout of NMD factors were due to NMD impairment, Li et al. used the cultured Smg6$^{-/-}$ mice embryonic stem cells (mESCs), that proliferate normally but present a differentiation blockage, to reintroduce variants of the Smg6 protein either lacking domains important for NMD or for telomere maintenance [3, 71]. Their observations revealed that mESCs differentiation process only requires NMD-proficient mutants of Smg6 [71], suggesting that NMD inactivation is the one responsible for the developmental defects. One possible explanation for the embryonic lethality observed after the knockout of NMD factors comes from the ability of NMD to degrade specific endogenous targets. For instance, in *Drosophila melanogaster* it was found that depletion of the growth arrest and DNA damage inducible 45 (*Gadd45*) gene restores the viability of *Upf1*- and *Upf2*-null mutant flies [76]. Gadd45 is a pro-apoptotic protein encoded by an mRNA that is normally degraded by NMD in

order to dampen its deleterious effects. Interestingly, the knockdown of the NMD target, GADD45β, partially rescues the apoptosis induced by UPF1 depletion in mammalian cells, suggesting that the NMD-mediated regulation of GADD45 expression is important to maintain mammalian cell viability [76]. Additionally, another pro-apoptotic factor targeted by NMD is the growth arrest-specific 5 (*GAS5*) RNA [77]. GAS5 is a long non-coding RNA (despite being translationally active) that induces cell-cycle arrest and apoptosis by binding to the glucocorticoid receptor, antagonizing its transcriptional activity [77]. The siRNA-mediated downregulation of GAS5 in several human lymphocyte cell lines relieves the growth arrest observed after UPF1-knokdown [78], indicating that, similarly to what is observed for GADD45β, the NMD-mediated regulation of GAS5 is important for cell viability and development.

Given the observed importance of NMD to normal development, it can be reasoned that its magnitude is regulated throughout the developmental process. Different NMD efficiencies at specific stages would allow the organism to alter the expression levels of many direct or indirect NMD-targeted transcripts, and thus determine the levels of several proteins with developmental implications [79]. Supporting this, Lou et al. demonstrated that differentiation of human embryonic stem cells (hESCs) into endoderm, mesoderm and ectoderm is accompanied by divergent NMD modulation [80]. The authors showed that NMD regulates the expression of many genes involved in signaling cascades, and suggested a model where NMD downregulation drives endoderm differentiation by stimulating the transforming growth factor-β (TGF-β) signaling, and NMD upregulation drives mesoderm/ectoderm differentiation through bone morphogenetic protein (BMP) signaling activation [80].

The notion that NMD is a developmentally regulated process is further supported by many studies reporting several biological contexts (described below) in which NMD activity and/or its regulation are necessary for tissue-specific differentiation programs, not only during embryonic development, but also after birth.

3.3.1 Liver Development and Regeneration

Liver-specific UPF2 knockout in mice embryos results in perinatal lethality. Histological analyses of embryonic livers at E16.5 and E18.5 demonstrated that UPF2 loss during liver development leads to the presence of cells with abnormal nuclear morphology arrested in mitosis. However, the overall cell proliferation is not affected. Gene expression profiling of RNA from livers at these developmental stages indicated that UPF2 ablation in the developing liver determines the dysregulation of hundreds of genes related to liver function. Among them are genes involved in metabolic pathways, in the complement and coagulation cascades, in mRNA processing, and in the DNA damage response pathway. These findings suggest that UPF2 depletion during fetal organogenesis impairs the development of a metabolically functional liver, which is incompatible with postnatal life [81]. Additionally, it was also found that UPF2 is essential for maintenance of the adult liver homeostasis and for regeneration after partial hepatectomy, and that liver-specific loss of UPF2 leads to liver damage and death shortly after its deletion [81]. The authors of this study inferred that these observations are due to impairment of the NMD activity because of UPF2 depletion. However, the direct link remains to be tested.

3.3.2 Spermatogenesis

There are several studies providing evidence that NMD plays a role in male germ line development and fertility. In one of them, conditional ablation of *Upf2* in mice embryonic Sertoli cells (SC) causes severe testicular atrophy and male sterility in adulthood due to complete depletion of SC and germ cells during prepubertal testicular development [82]. RNA-sequencing analyses of *Upf2* knockout testes revealed impaired transcriptomic homeostasis, with accumulation of PTC-containing transcripts and transcriptome-wide dysregulation of genes essential for SC fate control [82]. On a second study, Bao et al. focused

their attention to the germ cells, generating prospermatogonia-specific and spermatocyte-specific *Upf2* conditional knockout mice. In both cases the adult males were infertile, presenting a drastic reduction in testis size and with seminiferous tubules almost depleted of germ cells [69, 83]. The authors observed that Upf2 knockout spermatocytes and round spermatids express a reservoir of mRNAs with short 3'UTRs that are essential for spermatogenesis and male fertility, and proposed that this reservoir is, at least in part, generated via elimination of longer 3'UTR transcripts derived from ubiquitously expressed genes by UPF2-mediated NMD [83]. Similarly, spermatocytes and round spermatids homozygous for a null allele of Tudor domain-containing protein 6 (TDR6) show an accumulation of transcripts with long 3'UTRs [84]. TDR6 is a protein of the chromatoid bodies that is essential for UPF1 localization to these perinuclear organelles of the spermatids, and for UPF1-UPF2 interaction. It was observed that depletion of TDR6 leads to UPF1 mislocalization, blocking its interaction with UPF2 and inhibiting the long 3'UTR-triggered NMD pathway [84]. These findings indicate that NMD is required for normal spermatogenesis, and that UPF2 plays an important role in the process.

Another NMD factor that seems to be essential for spermatogenesis is UPF3A. For several years, the UPF3A was believed to exert the same function as its paralog, UPF3B, on the DECID complex [85]. However, this protein was recently suggested to be a NMD repressor that competes with UPF3B for UPF2 interaction. Apparently, in conditions where UPF3B is absent, UPF3A inhibits NMD by sequestering UPF2 from the NMD machinery [30]. Interestingly, the testis is the only adult tissue in which UPF3A protein and *UPF3A* mRNA are highly expressed, probably because of the transcriptional silencing of *UPF3B* in spermatocytes (that leads to low *UPF3B* mRNA levels in these cells) [30, 86]. Given these evidences, Shum et al. tested UPF3A function in male germ cells. They have shown that Upf3a knockout mice present reduced sperm counts and defective spermatocyte progression [30]. Additionally, they found that UPF3A levels must be tightly controlled to allow normal spermatogenesis, based on the observation that heterozygous mice for a null allele of *Upf3a* also present impaired spermatocyte progression [30]. Together, these findings support the notion that UPF3-dependent NMD and its tight regulation are necessary for normal spermatogenesis. However, we cannot rule out the hypothesis that other UPF3-independent mechanisms might also be involved in this process.

3.3.3 Myogenesis

As mentioned above, in addition to its participation in the NMD pathway, UPF1 is also required for SMD. This translation-dependent mRNA decay pathway has several common features with NMD and is triggered when STAU1, together with UPF1, bind to an mRNA 3'UTR [79, 87], at least 25 nucleotides downstream of the stop codon of the main ORF [87]. RNA analysis of STAU1-depleted HeLa cells revealed that, like NMD, SMD regulates gene expression, influencing the levels of approximately 2% of the mRNAs [79, 87]. Interestingly, Gong et al. found that SMD and NMD are competitive pathways. The first evidence supporting this idea came from the observation that STAU1 and UPF2 binding to UPF1 appear to be mutually exclusive, since the STAU1-binding site within UPF1 overlaps with the UPF2-binding site. Furthermore, SMD inhibition by RNA interference-mediated STAU1 downregulation, stimulates NMD activity, whereas NMD inhibition due to UPF2 downregulation, increases SMD efficiency [88]. The competition between NMD and SMD was proven to be physiologically relevant, particularly to the myogenic process, during which SMD efficiency was seen to be increased, while NMD activity was decreased [87, 88]. Consistent with this, during myogenesis, the interaction of UPF1 with STAU1 increases while its interaction with UPF2 decreases [88]. Accordingly, Gong et al. found that the promyogenic transcription factor, myogenin, is encoded by an NMD target that is upregulated during C2C12 (mouse-derived myoblasts) cell differentiation from myoblasts to

myotubes, as a result of NMD inhibition. Conversely, they found that the antimyogenic factor, paired box protein Pax-3 (PAX3), is a SMD target that is downregulated during the differentiation process, due to SMD enhanced activity [88]. Together, these findings suggest that the UPF2-dependent NMD branch and SMD are antagonistically regulated to drive the myogenic process.

Curiously, Gong et al. also found that, in differentiating C2C12 cells, UPF3B appears upregulated and not downregulated as UPF2, and its interaction with UPF1 is higher in myotubes than in myoblasts [88]. These observations suggest that the UPF3-dependent branch of NMD must be active during muscle differentiation. Indeed, UPF3B-dependent NMD targets were found to be decreased in differentiated cells [88]. Altogether, these *in vitro* observations allow us to reason that these different mRNA decay mechanisms and their respective mRNA substrates are tightly and conversely regulated to drive a specific developmental process [79].

3.3.4 Neural Development

While studying the possibility of NMD being regulated by microRNAs (miRNAs) during development, Bruno et al. identified two similar miRNAs – miR-128-1 and miR-128-2 – that directly target two NMD factors: UPF1 and metastatic lymph node gene 51 protein (MLN51, a core component of the EJC with a role in the EJC-dependent branch of NMD) [60, 89]. The authors showed that expression of miR-128 in HEK293 and HeLa cells decreases the protein levels of UPF1 and MLN51, acting through this ability to repress NMD in mammalian cells [89]. Interestingly, miR-128 is brain-enriched and its expression is induced between E9.5 and E12.5 in the mouse brain, continually increasing during postnatal development into adulthood [89]. Its upregulation during brain development is accompanied by a decrease in the levels of UPF1 [89, 90]. Ectopic expression of miR-128 in mouse neural stem cells (mNCSs) revealed that it upregulates hundreds of transcripts encoding proteins important for neural development and function, many of which apparently being NMD targets [89]. Based on gain- and loss-of-function experiments in P19 (mouse-derived embryonal carcinoma) cells subjected to neural differentiating conditions with retinoic acid (RA) and in mNSCs, Bruno et al. suggested that miR-128 expression and its ability to repress NMD is critical for the neural differentiation process [89], proposing a model where the induction of miR-128 during neuron differentiation decreases NMD magnitude, allowing the upregulation of specific transcripts that encode pro-neural factors, and thus triggering the necessary neural program to drive neuron differentiation [79, 89]. Supporting this model and the notion that NMD is reduced during neural development, Lou et al. found a decrease in the mRNA levels of many NMD factors, including UPF1, UPF2, UPF3B, SMG1 and SMG6, during *in vitro* neural differentiation of mNSCs and human neural progenitor cells (hNPCs) [90]. In order to address the impact of NMD regulation in the neural differentiation process, Lou et al. overexpressed UPF1 in differentiating P19 cells treated with RA. This UPF1 overexpression inhibits neural differentiation and sustains a proliferative state, as indicated by the upregulation of stem-cell markers and downregulation of differentiation markers [90]. On the other hand, UPF1 knockdown revealed to be sufficient to elicit neural differentiation and maturation of P19 cells in the absence of RA [90]. Furthermore, isolated mNSCs from UPF3B-null mice presented high levels of early neural markers, suggesting that UPF3B suppresses the differentiation ability of these cells [90]. These findings indicate that NMD modulation is important to determine the proliferation/differentiation state of the cells during brain development. This can be explained by the fact that NMD targets subsets of mRNAs encoding proliferation inhibitors and neural differentiation factors [3, 90]. Altogether, these results suggest that NMD downregulation induces normal differentiation and maturation of

neural cells, which is achieved, at least in part, by miR-128-mediated UPF1 downregulation during neural development.

In addition to its role in neuron differentiation, NMD is also important for axon guidance in later stages of neural development. During this process, when the commissural neurons pass the spinal cord midline, floor plate signals rapidly induce translation of a splice variant of the roundabout homologue 3 (*ROBO3*) gene, the *ROBO3.2*, which contains a PTC in a retained intron predicted to induce NMD [91, 92]. The resulting protein is a receptor that contributes to the repulsion of the axon from the midline, guiding its correct trajectory [3, 91]. Depletion of UPF2 in mouse commissural neurons leads to increased levels of ROBO3.2 receptor, which determines an abnormal migration of the axons with enlarged repulsion from the floor plate [92]. This suggests that NMD activity is essential to degrade *ROBO3.2* mRNA after a few rounds of translation, fine-tuning the expression of ROBO3.2 protein in a spatiotemporal manner [92].

The relevance of NMD for neural developmental is further confirmed by several studies reporting mutations in the *UPF3B* gene that are related to X-linked mental retardation, schizophrenia and autism [93–96]. The consequent loss of a functional UPF3B protein is accompanied by the dysregulation of genes involved in neural function [93], which can explain some characteristic phenotypes of the mental disorders. Copy number variations in other NMD factor-coding genes, including deletions in *UPF2* and deletions and/or duplications in *UPF3A*, *SMG6*, *EIF4A3* and *RNPS1*, were also seen in patients with various forms of intellectual disability [96]. Interestingly, most part of the genes found dysregulated in patients with *UPF2* deletions overlap with the ones found in patients with *UPF3B* mutations [96]. These data suggest that the phenotypes of intellectual disability found in patients with *UPF2* and *UPF3B* alterations are similar as a consequence of impaired NMD, which determines the dysregulation of identical genes related to neural function.

3.4 NMD and Stress Responses are Linked by a Dynamic Regulatory Circuit

A consistent finding in many transcriptome-wide studies where depletion of NMD factors was performed is the upregulation of mRNAs that encode stress-related proteins [6, 8, 11, 14, 81, 97]. This upregulation could either indicate that these mRNAs are NMD targets or that NMD impairment induces stress and consequently potentiate their expression [3]. Oren et al. provided evidence supporting the latter hypothesis by showing that NMD inhibition, through siRNA-mediated knockdown of UPF1, increases the phosphorylation level of the alpha subunit of eukaryotic initiation factor 2 (eIF2α) [98]. Phosphorylation of eIF2α is a phenomenon that occurs during stress conditions due to the activation of specific kinases in response to diverse stress stimuli. For instance, amino acid deprivation induces general control nonderepressible 2 (GCN2) kinase, endoplasmic reticulum (ER) stress, hypoxia and reactive oxygen species (ROS) induce protein kinase R (PKR)-like endoplasmic reticulum kinase (PERK), and viral infection induces PKR kinase. Phosphorylation of eIF2α in all these conditions causes a reduction in global protein translation while allowing the selective synthesis of proteins, including the master transcription factor, activating transcription factor 4 (ATF4), and its downstream target, CCAAT-enhancer-binding protein homologous protein (CHOP), responsible for transcribing genes involved in cell recovery and survival, as part of an adaptive pathway termed integrated stress response (ISR) (reviewed in [99]).

The PERK-eIF2α-ATF4 cascade is also part of another stress-response pathway, called unfolded protein response (UPR). The UPR is induced by ER stress when misfolded proteins accumulate in the ER due to protein-coding mutations, abnormal high translation in the ER and/or inefficient protein folding capacity as a consequence of decreased expression of chaperones or perturbations in the cell energy levels,

calcium homeostasis or redox status [5, 98]. The misfolded proteins are sensed by three ER-resident proteins, PERK (the common factor between ISR and UPR), inositol-requiring enzyme 1 (IRE1) and ATF6, which correspond to the three branches of the UPR. When activated, they induce the expression of downstream targets, initiating a cascade of events aiming cell homeostasis restoration or, in the case of a strong/prolonged stress stimulus, cell death induction (reviewed in [100]). Briefly, when IRE1 is activated its α subunit (IRE1α) undergoes trans-autophosphorylation that then mediates the cytoplasmic excision of a 26-nucleotide intron of the pre-mRNA encoding the X-box-binding protein 1 (XBP1) [101]. This produces an active transcription factor termed spliced XBP1 (XBP1s), that then translocates to the nucleus to induce the upregulation of genes encoding chaperones and the ones involved in the ER-associated degradation (ERAD) of misfolded proteins [100, 101]. On the other hand, ATF6 activation during the UPR induces its translocation to the Golgi, where it is cleaved to produce an active transcription factor responsible for the expression of genes encoding ERAD factors, XBP1 and chaperones, like binding immunoglobulin protein (BIP) [100]. Finally, and as mentioned above, PERK activation by ER stress induces eIF2α phosphorylation, decreasing the rate of protein translation initiation and allowing the cell to cope with the high content of unfolded proteins present in the ER. Simultaneously, the specific translation of ATF4 and CHOP induces the expression of genes important for autophagy and, eventually, apoptosis [5, 100]. Curiously, and refuting the idea that NMD impairment induces cell stress, supported by Oren et al., Karam et al. found that depletion of NMD factors do not activate the ISR nor the UPR, since the mRNA levels of *BIP*, *XBP1s* and *CHOP* (markers of ATF6, IRE1 and PERK activation, respectively) do not significantly increase after UPF1 or UPF3A/B depletion [102]. While this controversy is not solved, there are already many evidences showing that subsets of stress-related mRNAs are indeed NMD targets in unstressed cells [97, 98, 102–105], suggesting a link between NMD function and stress.

3.4.1 Stress-Responses Regulate NMD Magnitude

Many of the stress-related-NMD-sensitive transcripts present sequence features known to induce NMD, including uORFs and long 3'UTRs [11, 102]. Interestingly, in stress conditions these mRNAs evade NMD and become upregulated [97, 102–104]. This is possible because stress can modulate NMD activity, repressing it. The first evidence for this was brought by Mendell et al., in a study where amino acid starvation induced NMD inhibition and the upregulation of transcripts required for amino acid homeostasis [11]. Later, it was demonstrated that other stresses, including hypoxia, ROS and ER stress could also inhibit NMD [97, 102–106]. Indeed, a recent study reported a strong positive correlation between ER stress and NMD inhibition, in Na2 cells treated with the potent ER stressor, thapsigargin, in a dose-dependent manner [106].

There are many experiments showing that the stress-mediated NMD inhibition depends on the phosphorylation of eIF2α [97, 102, 103, 105, 106]. Nevertheless, the exact mechanism by which eIF2α phosphorylation does this is not clear yet. Since NMD depends on protein translation to recognize the stop codons in premature contexts to promote rapid mRNA decay, it could be reasoned that the ability of eIF2α phosphorylation to reduce translation would be the one responsible for NMD inhibition during stress. However, there is evidence showing that eIF2α phosphorylation only attenuates around 20 to 45% of the global protein synthesis [97, 107]. Additionally, NMD only requires a small percentage of translation to occur [108] and several NMD targets are translated even in the presence of stress and phosphorylated eIF2α [97, 103]. Hence, it is unlikely that suppression of translation is the only mechanism by which eIF2α phosphorylation mediates NMD repression [109]. In order to further explore this issue, Gardner studied and proposed another possibility for the eIF2α-mediated NMD inhibition, based on the spatial separation of the mRNAs and the RNA degradation machinery during stress [103]. Processing bodies (p-bodies) are foci scattered

through the cytoplasm lacking a membrane or physical delimitation, that contain RNA-degrading enzymes and RNAs, including PTC-containing transcripts [110]. Stress granules, on the other hand, are locals of translational repression formed in response to eIF2α phosphorylation during stress conditions [111, 112], that contain many translation initiation factors, untranslated mRNAs [112] and NMD factors, such as UPF1, SMG1 and UPF2 [113]. Using immunofluorescence assays, Gardner demonstrated that in hypoxic conditions, eIF2α phosphorylation induces the formation of stress granules, that sequester UPF1 and, possibly, NMD targets, while the decapping enzymes required for RNA decay are localized to p-bodies, providing a plausible explanation for the impairment of NMD activity [103]. This theory is supported by a recent report showing that, in unstressed cells, NMD activity is inhibited by cytoskeleton disruptors, with accumulation of UPF1, UPF3B and PTC-containing transcripts in p-bodies [114]. This indicates that, in fact, the p-bodies and the transit of NMD factors/substrates to or from these cytoplasmic foci are important for normal NMD activity [114]. Although well supported, the model proposed by Gardner requires further investigation to answer some still open questions.

The presence of uORFs in some stress-related transcripts may provide an alternative mechanism by which eIF2α phosphorylation inhibits NMD activity. In this case this is accomplished by specific, rather than global regulation of NMD [5]. One possible example of this is the ATF4 mRNA that contains two uORFs, one of them overlapped with the main ORF. In unstressed cells, the low levels of phosphorylated eIF2α allow the translation of the uORFs, thus inhibiting the translation of the downstream main ORF. However, in stress conditions, the increased levels of phosphorylated eIF2α induce translation reinitiation at the main ORF after translation of the first uORF, permitting ATF4 expression [115]. Indeed, it was demonstrated that ATF4 mRNA is a direct NMD target and that the translation of the uORFs is related to its decay in unstressed cells, and to its increased stability in hypoxic cells [103]. Other likely examples of this type of regulation could be the stress-responsive proteins, CHOP and general control protein GCN4 (GCN4), both encoded by uORF-harboring transcripts [116, 117].

3.4.2 NMD Regulates the Magnitude of the Stress-Responses

Among the stress-related transcripts targeted by NMD in normal conditions, we can find some mRNAs that encode important factors of the ISR and UPR, such as PERK, ATF4, ATF3 and CHOP [11, 97, 103]. More recently, it was found that NMD also targets many components specific of the UPR, including IRE1α, fibronectin type III and SPRY domain containing 1 like (FSD1L), homocysteine-induced endoplasmic reticulum protein (HERP), toll-like receptor-specific co-chaperone for HSP90B1 (TNRC5) and TNF receptor associated factor 2 (TRAF2) [102]. Confirming the ability of NMD to target these stress-responsive genes, a recent proteomic study revealed that several UPR proteins upregulated by the ER stressor, dithiothreitol, were also upregulated after UPF1 knockdown [107]. This is particularly relevant, because it shows that NMD-mediated degradation of the mRNAs that encode these UPR factors is reflected at the protein level, suggesting a physiological role of NMD in cell homeostasis. Given this, it could be reasoned that NMD has the ability to regulate the magnitude of the ISR and the UPR.

The first evidence that NMD is able to regulate stress-responses came from a study conducted by Gardner [103]. The author has shown that in U2OS (human-derived bone osteosarcoma) cells rendered hypoxic, the eIF2α-dependent inhibition of NMD results in the upregulation and stabilization of the mRNAs that encode ATF4, ATF3 and CHOP. Additionally, the protein levels of these factors are higher in UPF1-depleted U2OS cells treated with tunicamycin, when compared to normal cells treated with the same ER-stressor. On the contrary, UPF1 overexpression decreases the mRNA levels of the ISR

factors. These findings indicate that NMD activity suppresses ISR magnitude, while its inhibition during stress increases the ISR [103]. On another study, Karam et al. assessed the physiological impact of NMD in the UPR [102]. They found that UPF3B-depleted HeLa cells and mice exhibited higher UPR activation than control cells or littermate mice in response to the same doses of tunicamycin. Furthermore, they also studied the temporal kinetics of the UPR in the same cells treated with a low dose of tunicamycin and concluded that NMD-deficient cells exhibited more rapid and increased expression of mRNAs encoding markers of UPR activation. Together, these data provided evidence that NMD raises the activation threshold of the UPR, and that by targeting and downregulating UPR factors, NMD prevents the triggering of a stress response in conditions of innocuous ER stress [102]. When the stimulus ceases and the stress is resolved, the UPR must be shut down, or otherwise, it can activate pro-apoptotic pathways [100]. To test if NMD has an effect in the attenuation of the UPR, Karam et al. induced a severe UPR with high doses of tunicamycin in the same UPF3B-depleted cells as above. They observed that in the termination phase of the stress response the NMD-deficient cells fail to downregulate the mRNAs encoding UPR activation markers, indicating that NMD normally promotes the termination of the UPR, in a time-dependent manner [102].

The ability of NMD to shape stress responses allied to the previously discussed fact that it regulates pro-apoptotic factors, suggest that NMD can also be involved in the apoptotic pathways derived from prolonged stress. Indeed, Sakaki et al. found that siRNA-mediated SMG6 knockdown in HeLa (human-derived cervix cancer) cells and murine immortalized hepatocytes treated with tunicamycin induces a ~50% reduction in cell survival, when compared to control cells. In contrast, SMG6 overexpression increases survival of HeLa cells treated in the same conditions [118]. Notably, this overexpression decreases the induction of CHOP that activates UPR-related pro-apoptotic pathways in situations of prolonged stress, further supporting that the pro-apoptotic response is reduced by overexpression of SMG6 [118]. In agreement with the observations from Sakaki et al., Karam et al. have shown, both *in vitro* and *in vivo*, that UPF3B depletion increases the apoptotic percentage of cells also treated with tunicamycin [102], indicating that NMD is critical for survival in response to ER stress.

Altogether, these results unveil a bi-directional control between NMD and the UPR with important physiological implications for cell homeostasis and survival. This regulatory link comprises several aspects: (i) by suppressing UPR, NMD imposes a threshold that filters the harmful stimuli; (ii) when *bona fide* stress is encountered, the UPR inhibits NMD activity, so that the cell can upregulate stress-responsive components and properly respond to the stress; (iii) when the stress is solved, NMD resumes and the UPR is terminated. It is worth noting that one mechanism by which NMD accomplishes this, is through its ability to degrade the IRE1α mRNA, as tested by Karam et al. through rescue and gain-of-function experiments [102]. This is further supported by the proteomic study conducted by Sieber et al., where UPF1 depletion upregulated target-proteins of the IRE1α and the PERK pathways, while proteins activated by ATF6 were either downregulated or simply just did not respond to NMD inhibition [107]. This suggests that NMD and the UPR are linked by the PERK and IRE1α branches, but not by the ATF6 branch of the UPR.

3.5 NMD Presents a Pathophysiological Role in Cancer

So far, we have discussed the physiological implications of the gene expression regulation exerted by NMD. However, NMD activity and, sometimes, its regulation, may also contribute to phenotypic aspects characteristic of several human diseases, being implicated in many pathophysiological mechanisms. One such example is cancer, in which the tumor cells have exploited both the quality-control and the gene expression

regulatory facets of NMD to potentiate the development of the disease [2]. On one hand, the selective acquisition of PTCs in tumor-suppressor genes allows the unconstrained growth of tumor cells. On the other hand, by fine-tuning NMD activity, these cells can favor the upregulation of specific genes required for the adaptation to the tumor environment [2, 119].

3.5.1 Cancer Selects Tumor-Favoring Mutations Through NMD

After performing a meta-analysis of thousands of mutations in human genes, Mort et al. revealed that while oncogenes present mostly missense mutations, tumor-suppressor genes exhibit a higher number of nonsense mutations, many of which predicted to induce NMD [120]. Interestingly, several examples of PTC-introducing mutations in tumor suppressor genes have been reported in a diversity of cancers. For instance, microarray analysis after NMD inhibition with the translational-repressor, emetine, in mantle-cell lymphoma cell lines revealed that mutations in the tumor suppressor genes, *p53* and retinoblastoma 1 (*RB1*), originate PTC-harboring transcripts. These were stabilized by emetine, suggesting that they are targeted by NMD [121]. Additionally, a PTC-harboring mutant transcript of p53 found in breast cancer presented increased mRNA stability after NMD inhibition, as did the resulting C-terminal truncated protein, when compared to the correspondent wild-type variants [122]. In patients with hereditary diffuse gastric cancer, around 80% of the mutations in the tumor suppressor E-cadherin gene generate a PTC. The resulting E-cadherin mRNAs are upregulated in response to protein translational inhibitors and to knockdown of NMD factors, suggesting that their normal downregulation is mediated by NMD [123]. Similarly, around 80% of the PTC-bearing alleles found in breast cancer type 1 susceptibility protein (*BRCA1*) gene, in breast and ovarian cancers, are capable of triggering NMD [124]. The same was observed for breast cancer type 2 susceptibility protein (*BRCA2*) gene in breast cancer, where most of the PTC-introducing mutations apparently destabilize the generated transcripts [125]. In these examples, the quality-control function of NMD is important to degrade the mutated transcripts and prevent the production of proteins with potential dominant-negative activity that could induce tumorigenesis [2], as has been shown for the tumor-suppressor genes *WT1*, in kidney cancers [126], and *p53*, in pancreatic adenosquamous carcinoma (ASC) [127]. However, there are scenarios where this protective NMD-function may work in the opposite direction. For instance, if the PTC-bearing allele encodes a protein without dominant-negative properties that, instead, preserves part of the normal function, NMD activity would potentiate the development of cancer by degrading the correspondent mRNA [2].

While trying to understand the rules that govern NMD targeting in human cells by matching exome and transcriptome data from human tumors, Lindeboom et al. found that PTCs are enriched in regions of tumor-suppressor genes predicted to trigger NMD [128]. This study has also exposed how tumors use NMD activity to impair tumor-suppressor gene function and promote cancer. The first possibility is the combination of a PTC-bearing allele that is targeted by NMD with an heterozygous deletion of the wild-type allele, achieving biallelic inactivation of the tumor-suppressor. Another possibility is the combination of the PTC-bearing allele with a haploinsufficient version of the wild-type allele. Alternatively, PTCs positioned in regions that fail to trigger NMD may lead to truncated proteins that can be functionally inactive, degraded, or, in the worst case, present dominant-negative activity [2, 128].

3.5.2 NMD Normally Works as a Tumor Suppressor Pathway

There is both experimental and clinical evidence suggesting that NMD, as an expression regulator of some wild-type transcripts, can suppress tumorigenesis. Firstly, Wang et al. found that

overexpression of UPF1 in many cancer cell lines, including prostate cancer (PC3), colon cancer (HCT116) and melanoma (A375), reduced the number and size of the tumor cell colonies formed in soft agar, when compared to the control cells [97]. Similarly, no significant tumor growth was verified after injection of UPF1-overexpressing PC3 cells as tumor explants in nude mice [97]. After performing an expression array analysis in U2OS cells with repressed NMD, the authors found that NMD targets a wide variety of transcripts that encode important factors for tumorigenesis, including proteins involved in cell growth, cell cycle, growth factor signaling, apoptosis and cell migration [97]. In addition to these findings, clinical studies have reported *UPF1* somatic mutations in patients with ASC [127] and inflammatory myofibroblastic tumors (IMT) [129], with relevant consequences for the patients. These mutations were found to cause alternative splicing of the *UPF1* pre-mRNA, leading to low or undetectable levels of UPF1. Consequently, NMD efficiency is decreased in these tumors, as suggested by an increase in the mRNA levels of specific NMD targets [127, 129]. Interestingly, one of the increased NMD substrates found in IMTs was the mRNA encoding the mitogen activated protein kinase kinase kinase 14 (MAP3K14 or NIK), a potent activator of the proinflammatory NF-κB signaling pathway. IMTs with *UPF1* mutations presented elevated levels of chemokines and immune cells, supporting a model in which *UPF1* mutations impair NMD activity, leading to NIK upregulation and NF-κB induction, which contributes to the immune infiltration characteristic of IMTs. [129]. Reduced UPF1 expression, due to promoter hypermethylation, was also found in hepatocellular carcinoma (HCC) cells and tissues, and was correlated with poor prognosis [130]. Similarly, in lung adenocarcinoma (ADC) UPF1 was found to be downregulated with the consequent upregulation of various TGF-β signaling components, which are essential for the epithelial-to-mesenchymal transition that drives the disease [131]. Altogether, these studies suggest that, by targeting transcripts encoding important factors for cell growth, differentiation, proliferation and survival, NMD acts as a tumor suppressor pathway. On the contrary, in tumors with impaired NMD, these transcripts can be stabilized, creating favorable conditions for tumor proliferation.

3.5.3 The Tumor Microenvironment Inhibits NMD and Promotes Tumorigenesis

The presence of *UPF1* mutations in many cancers with favorable implications to the tumorigenic process may indicate that NMD impairment is part of the adaptive mechanisms that tumor cells have adopted to leverage survival and proliferation. In fact, NMD inhibition appears to be a consequence of the tumor microenvironment [97], where the increased cell mass and the insufficient vasculature induce cell stresses that include hypoxia, nutrient deprivation, ROS production and ER stress [132]. As discussed above, these type of stimuli determine the phosphorylation of eIF2α and the consequent decrease in the NMD magnitude [97, 102, 103, 105]. Supporting this notion, Wang et al. found that PC3 cells grown as three-dimensional tumor explants in mice present decreased NMD efficiency, when compared to PC3 cells cultured in monolayers [97]. Additionally, the tumor explants presented areas of significant cytoplasmic eIF2α phosphorylation [97]. On another line of investigation, Wang et al. showed that overexpression of the oncogene, *MYC*, in a B-cell tumor line also induces eIF2α phosphorylation and NMD inhibition, apparently through induction of ROS [105]. A major implication of this stress-mediated NMD inhibition in tumors is the stabilization and upregulation of transcripts encoding stress-responsive factors, including the ones from the ISR and the UPR. This triggers a set of adaptive mechanisms that allow the tumor cells to proliferate in the harmful microenvironment they generate [97, 133, 134].

Tumor formation is frequently accompanied by the production of ROS as a result of homeostatic imbalances or from mutations in mitochondrial genes or oncogenes, like *MYC* [104, 105, 132]. The ROS-mediated oxidative damage can be avoided with the production of Glutathione (GSH), a tripeptide comprised of cystine (oxidized form of cysteine), glutamic acid, and glycine, that neutralizes free radicals and reactive oxygen compounds [5, 104]. The cystine/glutamate exchanger SLC7A11 is a subunit of the xCT amino acid transporter system, a rate-limiting channel responsible for the cellular uptake of cystine for GSH production [104]. Interestingly, *SLC7A11* mRNA is an NMD target that is upregulated in stress conditions [104]. It was shown that the NMD inhibition mediated by the stress-induced eIF2α phosphorylation stabilizes the *SLC7A11* transcript. This is reflected into an increase in the SLC7A11 protein expression and the intracellular GSH levels, which confers cell resistance to oxidative stress [104].

Another stress experienced by tumors is the insufficient supply of amino acids. This induces NMD inhibition through phosphorylation of eIF2α by the GCN2 and PERK kinases as part of the ISR and the UPR, allowing the upregulation of ATF4 and its downstream targets that include genes involved in amino acid synthesis and transport [11, 134]. In these conditions, NMD inhibition also induces autophagy, an adaptive recycling mechanism that allows the cell to replenish its amino acid and energy stocks in conditions of metabolic stress [135]. During autophagy, the cell involves the targeted organelle or protein aggregate in auto phagosomes that then fuse with lysosomes, responsible for the recycling process [134, 135]. The stabilization of *ATF4* mRNA due to NMD inhibition is, at least in part, responsible for autophagy induction during stress. Indeed, Wengrod et al. have shown that ATF4 depletion rescues autophagy inhibition in the face of impaired NMD [135]. The authors have also found that inhibition of autophagy with chloroquine in HCT116 cells depleted of UPF1 or UPF2 present decreased cell viability and increased apoptosis. In contrast, UPF1 overexpression reduces the sensitivity of the cells to the effects of chloroquine, suggesting an intolerance to NMD inhibition in the absence of autophagy [135]. Together, these results support the idea that autophagy is an adaptive response to NMD inhibition, which can be used by tumor cells to survive in conditions of metabolic stress.

Although NMD inhibition appears to be an important mechanism for tumor progression, it seems to be contradicting with the fact that UPF1 depletion and reduced NMD activity are also related to cell death. Further densifying this paradox, two recent studies have shown that different pro-apoptotic agents, including chemotherapeutics, have the ability to inhibit NMD through caspase-mediated cleavage of UPF1 and UPF2 [136, 137]. These truncated forms present dominant-negative activity that lead to the upregulation of apoptosis-related NMD targets, creating a positive feedback loop responsible for driving cell death [136, 137]. One possible explanation for the opposite outcomes of NMD inhibition during tumorigenesis may rely on the tumor background and on the combination of mutations that it has acquired overtime, which may help tumor cells to overcome deleterious aspects derived from the normal biological processes [2].

3.6 Integrative Network Approach to Explore Links Between NMD, Development, Stress and Cancer

There are significant and important links between biological processes that can be unveiled through protein physical interactions, and by regulatory modules that control the condition specific activation of the related protein complexes [138]. These physical interactions may be the basis of information flow in signaling events (phosphorylation, dephosphorylation or other protein modifications), or be relevant for correct protein function by controlling protein subcellular local-

ization, conformation or stability. The search in molecular interaction databases for an enrichment in physical interactions between proteins involved in NMD and other processes may confirm already known mechanisms, but can also highlight new plausible links between processes that are not yet mechanistically understood. By reviewing published data about protein-protein interactions, we present an independent network-based approach to uncover complementary evidences of interactive and cooperative relationships between NMD, development, stress and cancer. The data integrated in this approach was extracted from different publicly available databases.

The first step of our analysis was to search for proteins associated with each molecular category: 703 cancer-related proteins extracted from the *Catalogue Of Somatic Mutations In Cancer* (*COSMIC*) database [139, 140]; 293 stress- and 963 development-related proteins extracted from *Reactome* database [141, 142]; 24 NMD-factors extracted from published work [143]. Each group of proteins was used to query a human protein-protein interactome to obtain protein-protein interactions (PPI) between NMD-factors and the proteins associated with the other categories. The interactome is composed by an overlay of different human PPI databases: The *Human Reference Protein Interactome* (*HuRI)* [144–148], *Agile Protein Interactome DataAnalyzer* (*APID)* [149], *OmniPath* [150] and *inBio Map* [151]. These interactions were submitted to a statistical analysis (hypergeometric test) to avoid non-specific interactions. In this way, only proteins that were statistically enriched in interactions with NMD-factors were included in the NMD-cancer-stress-development network. From this query, 526 interactions between NMD-factors and 193 cancer-related proteins, 471 between NMD-factors and 133 stress-related proteins and 718 between NMD-factors and 225 development-related proteins were obtained. The fact that these proteins have in average 3 interactions with NMD-factors support the relation between these processes. Additionally, it is important to note that some of these PPI are shared between categories, since some proteins are associated simultaneously with more than one of the three studied processes (development, cancer and stress).

Since PPIs depend on the presence of both proteins in the same place at the same time, coregulation events and coexpression among such proteins strengthens the functional relevance of the interactions. Transcription factors (TFs) and miRNAs are important regulators of gene expression and their action can translate into highly coordinated coregulation events. This coregulation can also be analyzed via a network approach [152, 153], looking for regulators in common between NMD-factors and other proteins in the NMD-cancer-stress-development network. For a stricter analysis, only regulators targeting more than two NMD-factors were considered (miRNA [154]; TFs [155–157]). In average, these proteins have 1 miRNA and 8 TFs in common with three or more NMD-factors. A complementary view of coregulation is coexpression. Therefore, coexpression data [158] was used to detect which proteins are coexpressed with NMD-factors. Interestingly, we detect a coexpression interaction with 4 NMD-factors, in average. Altogether, the coregulation and coexpression interactions increase the strength of molecular links already revealed by the PPIs.

The analysis, above described, produced a large network of interactions displaying how NMD-factors are linked to proteins associated with cancer, stress and/or development (a table with the network proteins and attributes can be found in Annex 1). Considering the size of the network, a set of cutoffs was applied in each type of data to present a subnetwork with the most relevant proteins (Fig. 3.2). Apart from NMD-factors, the proteins presented must interact with at least 3 NMD-factors, are coregulated with NMD via 1 miRNA and 5 transcription factors (or more) and are coexpressed with at least 2 NMD-factors. The subnetwork is composed of 37 proteins related to different cancer types (11 germline tumors and 63 somatic tumors), 45 proteins related to stress (3 associated with the unfolded protein response, 34 with heat stress

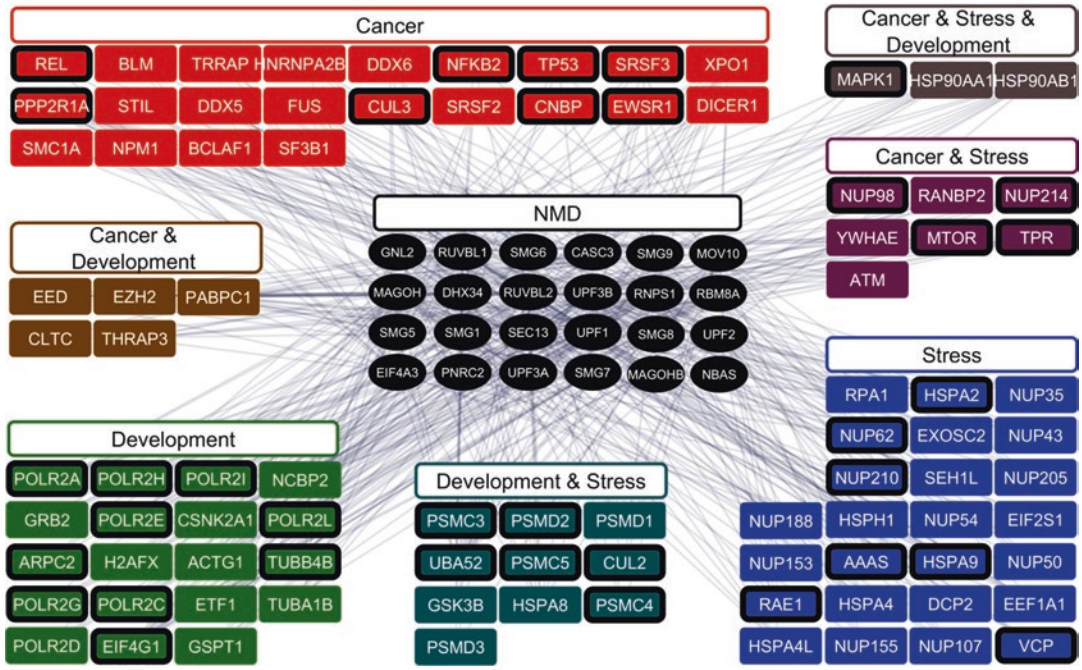

Fig. 3.2 A subnetwork of the analysis output that shows the NMD factors at the center and the proteins related to the other biological processes at the periphery. Protein-protein interactions are represented by straight lines (edges) that link pairs of proteins (nodes). Proteins are labeled with their official gene symbols. The NMD targets are highlighted with black colored borders

and 8 with hypoxia-induced stress) and 37 related to development (11 associated with activation of HOX genes during differentiation, 25 with axon guidance, 8 with transcriptional regulation of pluripotent stem cells and 1 with transcription regulation of white adipocyte differentiation). The statistical significance of these findings can be assessed by a randomization test, repeating the same analysis 1000 times, using random groups of 24 (the number of NMD-factors) neighbors of cancer, development and stress proteins. After applying the above-mentioned thresholds, the results showed a maximum of 13 cancer-, 13 stress- and 13 development-related proteins supporting that the number of cancer-, stress- and development-related proteins with strong interactions with NMD is significantly higher than expected by chance ($p < 0.001$). Additionally, 75% of the random analysis showed 2 or less cancer- and stress-related proteins, and only 1 or none of development-related proteins. This confirms that the number of proteins associated with these processes that are linked to NMD-factors is extremely unlikely to be observed by chance, suggesting its biological significance. Furthermore, since many NMD-factors are also auto-regulated through NMD, additional information regarding NMD targeting was introduced in the analysis. Taking advantage of published expression data in NMD inhibited conditions (accomplished via UFP1 knockdown) it is possible to identify proteins targeted directly or indirectly by NMD [137]. This step revealed the presence of NMD-targets in the subnetwork: 13 cancer-, 18 stress- and 17 development-related proteins.

Looking into the subnetwork proteins (Fig. 3.2), we can observe proteins already known

as modulators of the NMD pathway, such as PABPC1 [52], ETF1 (also known as eRF1) [27, 52], GSPT1 [27, 52] (also known as eRF3A), nuclear cap-binding protein subunit 2 (NCBP2) [159], DCP2 [49], EIF4G1 [160] and serine/threonine-protein phosphatase 2A regulatory subunit A (PPP2R1A) [161], suggesting that Fig. 3.2 depicts some links that represent mechanisms already studied in the NMD context. An example related to cancer is the cellular tumor antigen p53 protein (TP53), where NMD plays a critical role regulating the expression of p53β isoform. Since the aberrant expression of p53β and dysfunctional NMD are implicated in cancers, this regulation contribute to explain the underlying mechanism of tumorigenesis [162]. An example related to stress is the serine-protein kinase ATM, an important kinase in DNA damage response, which belongs to the same protein family of SMG1, and like the latter, ATM is capable of phosphorylating UPF1 [163]. Although, the functional consequences of UPF1 phosphorylation by this protein remains to be elucidated, it opens the possibility that phosphorylation of UPF1 may not only modulate its activity in NMD, but also recruit this protein into the stress response network [164]. A development-related example is the actin-related protein 2 (ACTR2), tubulin alpha-1B chain (TUBA1B) and beta-4B chain (TUBB4B), which encode for a protein related with actin polymerization [165] and tubulin chains (the major constituent of microtubules), respectively. As mentioned above in this chapter, the NMD factors and substrates are transported through cytoplasmic foci, the p-bodies [110, 114]. Such transport is modulated by the cytoskeleton and, interestingly, the disruption of actin filaments or microtubules leads to the inhibition of NMD [114]. In conclusion, this analysis evidences a broader, multi-scale perspective for studying NMD and related biological processes, which can point to new molecular players through which NMD can modulate or be modulated by other physiological processes.

3.7 Conclusions

As discussed in this chapter, NMD is important to the mammalian cell, not only as a vigilant of the quality of the transcriptome, but also as a regulator of many essential biological processes that occur from the early stages of the embryonic development to adulthood and beyond. For the correct biological outcome to be achieved, NMD activity must be tightly regulated to act on the proper targets at the right time and place. Although most of the regulatory mechanisms/pathways that affect NMD magnitude are still unknown, some of them already started to be unveiled. This is the case, for instance, of the miR-128 expression and the phosphorylation of eIF2α that, as we discussed above, are important regulatory events of the NMD activity during neural development and stress conditions, respectively. In these cases, NMD downregulation has a positive outcome. However, there are situations where NMD modulation can have a pathophysiological result. This is the case of cancer, where NMD inhibition mediated by the tumor microenvironment provides the adaptive conditions necessary for tumor cells to survive and proliferate. The number of examples given here are far from the extensive list of biological processes and diseases where NMD can be found to play important roles. Knowing them and understanding the complexity of the molecular mechanisms and interactions they present with NMD is crucial for the development of new and more personalized therapies for related pathologies.

Acknowledgements This work was partially supported by Fundação para a Ciência e a Tecnologia (FCT; Portugal) (PTFC/BIM-MEC/3749/2014 to LR and UID/MULTI/04046/2013 Research Unit grant to BioISI), and by National Institute of Health Dr. Ricardo Jorge. RF is recipient of a fellowship from BioSys PhD programme (SFRH/BD/114392/2016) from FCT (Portugal). GN is recipient of a fellowship from BioSys PhD programme (PD/BD/130959/2017) from FCT (Portugal). PJC is recipient of a fellowship from BioSys PhD programme (PD/BD/52495/2014) from FCT (Portugal).

Annex 1: Table of Network Proteins and Their Attributes

Uniprot ID	Gene Symbol	Biological Process	NMD links	NMD-factors	Coexpression	miRNA Coregulation	TF Coregulation	NMD Target
Q9NQ94	A1CF	Cancer	4	UPF3B / RNPS1 / RBM8A / MAGOH	0	2	1	0
Q9NRG9	AAAS	Stress	4	UPF3B / RNPS1 / RBM8A / MAGOH	9	2	10	1
P49748	ACADVL	Stress	1	EIF4A3	2	1	4	1
O95573	ACSL3	Cancer	3	UPF2 / RUVBL1 / RUVBL2	1	1	13	1
P60709	ACTB	Development	7	UPF1 / MAGOH / EIF4A3 / SEC13 / GNL2 / RUVBL1 / RUVBL2	0	5	12	1
P63261	ACTG1	Development	5	UPF1 / SEC13 / GNL2 / RUVBL1 / RUVBL2	3	7	5	1
P61160	ACTR2	Development	3	GNL2 / RUVBL1 / RUVBL2	1	2	19	0
P61158	ACTR3	Development	2	GNL2 / RUVBL1	4	0	15	0
Q9UHB7	AFF4	Cancer	1	MAGOH	2	2	17	0
O00468	AGRN	Development	1	MOV10	3	1	2	1
P05091	ALDH2	Cancer	3	GNL2 / RUVBL1 / RUVBL2	0	2	4	1
O95782	AP2A1	Development	2	UPF1 / RNPS1	4	1	6	1
O94973	AP2A2	Development	2	UPF1 / MOV10	3	3	11	1
Q96BI3	APH1A	Development	1	MOV10	3	0	9	1
P10275	AR	Cancer	2	RUVBL1 / PNRC2	0	1	4	0
Q8N6T3	ARFGAP1	Stress	1	UPF1	6	0	13	1
O14497	ARID1A	Cancer	2	MOV10 / RUVBL1	7	5	5	1
Q8NFD5	ARID1B	Cancer	1	RUVBL1	5	1	4	1
Q68CP9	ARID2	Cancer	1	RUVBL1	2	0	1	0
Q92747	ARPC1A	Development	2	GNL2 / RUVBL1	2	0	12	1
O15143	ARPC1B	Development	3	GNL2 / MOV10 / RUVBL1	2	0	8	1
O15144	ARPC2	Development	3	GNL2 / RUVBL1 / RUVBL2	2	2	6	1
O15145	ARPC3	Development	2	GNL2 / RUVBL1	1	1	10	1
P59998	ARPC4	Development	3	GNL2 / MOV10 / RUVBL1	1	0	10	1
O15511	ARPC5	Development	2	GNL2 / RUVBL1	1	1	6	0
O43681	ASNA1	Stress	2	RUVBL1 / RUVBL2	6	3	8	1
P18848	ATF4	Stress	1	MOV10	5	0	11	1
Q13315	ATM	Cancer Stress	5	UPF1 / SMG1 / MOV10 / RUVBL1 / RUVBL2	5	1	16	0

(continued)

Uniprot ID	Gene Symbol	Biological Process	NMD links	NMD-factors	Coexpression	miRNA Coregulation	TF Coregulation	NMD Target
P05023	ATP1A1	Cancer	1	RUVBL2	1	5	6	1
Q13535	ATR	Cancer Stress	4	UPF1 / SMG1 / RUVBL1 / RUVBL2	3	0	19	0
P61769	B2M	Cancer	1	SEC13	1	1	11	0
O95817	BAG3	Stress	2	SEC13 / RUVBL2	0	1	1	1
O95429	BAG4	Stress	1	MOV10	6	1	7	0
P20749	BCL3	Cancer	2	RUVBL1 / RUVBL2	2	1	9	1
Q9NYF8	BCLAF1	Cancer	5	RNPS1 / MAGOH / EIF4A3 / MOV10 / PNRC2	5	3	21	0
P11274	BCR	Cancer	1	SEC13	3	0	8	1
Q13489	BIRC3	Cancer	2	RUVBL1 / RUVBL2	1	0	14	1
P54132	BLM	Cancer	3	UPF2 / MAGOH / EIF4A3	7	1	7	0
Q12982	BNIP2	Development	1	MOV10	5	2	8	1
Q15059	BRD3	Cancer	1	RUVBL2	4	1	2	1
O60885	BRD4	Cancer	1	MAGOH	4	3	1	1
O60566	BUB1B	Cancer	1	SEC13	7	1	12	0
P27797	CALR	Cancer Stress	1	UPF2	3	1	15	1
Q8WVQ1	CANT1	Cancer	1	MOV10	2	0	11	1
P04632	CAPNS1	Development	1	RUVBL1	1	2	12	1
Q14790	CASP8	Cancer	1	CASC3	3	1	12	1
P04040	CAT	Stress	2	UPF1 / MOV10	0	0	12	0
Q8N163	CCAR2	Stress	1	RUVBL2	10	0	13	0
P04233	CD74	Cancer	1	SEC13	0	0	4	1
Q9NYV4	CDK12	Cancer	1	RNPS1	3	1	12	1
Q00534	CDK6	Cancer	1	CASC3	0	6	9	0
P49715	CEBPA	Cancer Development	2	RNPS1 / EIF4A3	0	0	14	1
P23528	CFL1	Development	2	RUVBL1 / RUVBL2	4	3	5	1
O14647	CHD2	Cancer	2	RUVBL1 / RUVBL2	4	0	6	1
O96017	CHEK2	Cancer	2	RUVBL1 / RUVBL2	5	1	10	1
Q7Z460	CLASP1	Development	1	SEC13	2	1	12	1
O75122	CLASP2	Development	2	SEC13 / MOV10	0	0	17	0
P30622	CLIP1	Cancer	1	SEC13	0	0	9	1
Q92989	CLP1	Cancer	5	UPF3B / RNPS1 / RBM8A / MAGOH / SEC13	7	0	10	1

P09496	CLTA	Development	1	RBM8A	2	2	10	1
P09497	CLTB	Development	1	SEC13	1	0	8	1
Q00610	CLTC	Cancer Development	4	SEC13 / GNL2 / RUVBL1 / RUVBL2	3	5	24	0
P53675	CLTCL1	Cancer Development	4	GNL2 / MOV10 / RUVBL1 / RUVBL2	0	0	1	1
P62633	CNBP	Cancer	4	UPF2 / MAGOH / GNL2 / RUVBL1	8	1	18	1
O75175	CNOT3	Cancer	1	RUVBL2	9	0	6	1
Q9ULM6	CNOT6	Development	2	RUVBL1 / RUVBL2	7	1	11	0
Q92600	CNOT9	Development	1	RUVBL2	1	0	0	0
Q7Z7A1	CNTRL	Cancer	1	UPF1	4	1	0	1
Q01955	COL4A3	Development	1	MOV10	0	0	1	0
O43889	CREB3	Stress	1	MOV10	2	0	19	1
Q8TEY5	CREB3L4	Stress	1	MOV10	1	0	7	0
Q9BZJ0	CRNKL1	Cancer	5	UPF1 / MAGOH / EIF4A3 / GNL2 / RUVBL2	6	0	11	1
P68400	CSNK2A1	Development	5	RNPS1 / SEC13 / MOV10 / RUVBL1 / RUVBL2	9	2	11	0
P19784	CSNK2A2	Development	3	RNPS1 / SEC13 / RUVBL2	8	0	2	1
P67870	CSNK2B	Development	6	UPF1 / RNPS1 / SEC13 / MOV10 / RUVBL1 / RUVBL2	3	0	6	1
P35222	CTNNB1	Cancer Development	3	MOV10 / RUVBL1 / RUVBL2	1	1	19	0
Q13617	CUL2	Stress Development	3	SMG6 / RNPS1 / MOV10	6	2	11	1
Q13618	CUL3	Cancer	8	UPF1 / SMG6 / RNPS1 / EIF4A3 / SEC13 / MOV10 / RUVBL1 / RUVBL2	5	2	6	1
Q14999	CUL7	Stress	5	RNPS1 / EIF4A3 / GNL2 / RUVBL1 / RUVBL2	4	0	8	1
P13498	CYBA	Stress	1	UPF2	2	0	3	1
Q8IU60	DCP2	Stress	5	UPF1 / UPF2 / UPF3A / UPF3B / PNRC2	7	1	8	0
Q14203	DCTN1	Cancer Stress	2	UPF1 / MOV10	1	1	6	1
Q13206	DDX10	Cancer	1	UPF1	4	1	4	1
Q96FC9	DDX11	Stress	2	SMG9 / DHX34	9	1	10	0
O00571	DDX3X	Cancer	4	MAGOH / EIF4A3 / RUVBL1 / RUVBL2	4	3	4	0
P17844	DDX5	Cancer	5	UPF1 / EIF4A3 / SEC13 / RUVBL1 / RUVBL2	4	2	17	1
P26196	DDX6	Cancer	4	UPF1 / GNL2 / RUVBL2 / PNRC2	4	7	19	0

(continued)

Uniprot ID	Gene Symbol	Biological Process	NMD links	NMD-factors	Coexpression	miRNA Coregulation	TF Coregulation	NMD Target
P35659	DEK	Cancer	1	UPF1	10	0	13	0
Q8WYQ5	DGCR8	Cancer	3	UPF1 / GNL2 / MOV10	9	0	17	1
Q9UPY3	DICER1	Cancer	6	UPF1 / EIF4A3 / GNL2 / MOV10 / RUVBL1 / RUVBL2	4	3	12	0
Q9Y2L1	DIS3	Stress	1	RUVBL2	4	1	9	1
P25685	DNAJB1	Cancer Stress	3	UPF2 / GNL2 / RUVBL2	0	1	13	1
Q99615	DNAJC7	Stress	1	SMG8	10	1	7	1
P50570	DNM2	Cancer Development	1	RUVBL2	5	2	18	1
Q7L190	DPPA4	Development	2	UPF1 / MAGOH	0	1	1	1
Q02413	DSG1	Development	1	RUVBL1	0	0	0	0
P15924	DSP	Development	2	RUVBL1 / RUVBL2	0	2	2	1
O75530	EED	Cancer Development	9	UPF1 / RNPS1 / RBM8A / MAGOH / MAGOHB / EIF4A3 / MOV10 / RUVBL1 / RUVBL2	6	2	25	0
P68104	EEF1A1	Stress	8	UPF1 / MAGOH / EIF4A3 / SEC13 / GNL2 / MOV10 / RUVBL1 / RUVBL2	2	8	20	1
P52798	EFNA4	Development	1	MOV10	2	1	3	1
P47813	EIF1AX	Cancer	2	RBM8A / MAGOH	6	3	0	0
P05198	EIF2S1	Stress	3	UPF1 / EIF4A3 / RUVBL2	7	2	15	0
Q14240	EIF4A2	Cancer	9	UPF3B / SMG1 / RNPS1 / RBM8A / MAGOH / MAGOHB / EIF4A3 / RUVBL1 / RUVBL2	1	3	14	0
Q04637	EIF4G1	Development	15	UPF1 / UPF2 / UPF3A / UPF3B / SMG1 / SMG5 / SMG6 / SMG7 / SMG8 / SMG9 / RNPS1 / RBM8A / MAGOH / CASC3 / EIF4A3	8	2	10	1
Q99814	EPAS1	Cancer Stress Development	1	RUVBL1	0	0	3	1
P21709	EPHA1	Development	3	SMG1 / SMG8 / SMG9	0	0	1	1
P18074	ERCC2	Cancer	1	RUVBL2	7	0	3	1
Q92889	ERCC4	Cancer	1	RUVBL2	3	0	10	0
Q96HE7	ERO1A	Stress	2	EIF4A3 / SEC13	0	0	0	0

P03372	ESR1	Cancer	6	UPF1 / EIF4A3 / GNL2 / RUVBL1 / RUVBL2 / PNRC2	0	2	7	1
P62495	ETF1	Development	15	UPF1 / UPF2 / UPF3A / UPF3B / SMG1 / SMG5 / SMG6 / SMG7 / SMG8 / SMG9 / RNPS1 / RBM8A / MAGOH / CASC3 / EIF4A3	8	2	14	0
P41161	ETV5	Cancer	1	MOV10	1	0	10	1
Q01844	EWSR1	Cancer	5	UPF1 / RBM8A / CASC3 / SEC13 / RUVBL1	11	2	11	1
Q9Y3B2	EXOSC1	Stress	1	UPF2	11	1	11	0
Q13868	EXOSC2	Stress	5	UPF1 / UPF2 / UPF3A / UPF3B / RUVBL2	11	2	11	1
Q9NPD3	EXOSC4	Stress	4	UPF1 / UPF2 / UPF3A / UPF3B	8	0	24	1
Q5RKV6	EXOSC6	Stress	1	UPF2	4	1	8	1
Q96B26	EXOSC8	Stress	2	UPF2 / MOV10	4	0	8	1
Q15910	EZH2	Cancer Development	4	UPF1 / MAGOH / RUVBL1 / RUVBL2	11	4	12	0
P25445	FAS	Cancer	1	RUVBL2	0	0	6	1
Q969H0	FBXW7	Cancer	2	EIF4A3 / MOV10	4	1	6	1
P39748	FEN1	Cancer	1	MOV10	11	2	16	1
O95684	FGFR1OP	Cancer	1	UPF3A	1	3	17	1
P22607	FGFR3	Cancer	1	SMG7	1	1	0	1
P22455	FGFR4	Cancer	1	CASC3	1	1	5	1
P49789	FHIT	Cancer	1	UPF1	1	1	4	1
Q6UN15	FIP1L1	Cancer	1	RNPS1	3	1	7	1
P21333	FLNA	Cancer	3	SEC13 / RUVBL1 / RUVBL2	0	8	1	1
P58012	FOXL2	Cancer	1	MOV10	0	0	0	0
Q96AE4	FUBP1	Cancer	1	PNRC2	9	2	13	0
P35637	FUS	Cancer	8	UPF1 / UPF2 / UPF3A / UPF3B / RNPS1 / RBM8A / MAGOH / SEC13	10	2	20	0
P35557	GCK	Development	3	SEC13 / RUVBL1 / RUVBL2	0	0	2	1
Q06210	GFPT1	Stress	2	RUVBL1 / RUVBL2	3	1	11	1
Q9Y2X7	GIT1	Development	1	RNPS1	3	1	4	1
Q9HD26	GOPC	Cancer	1	MOV10	4	0	10	0
P51654	GPC3	Cancer	2	UPF2 / RNPS1	0	0	0	0
P78333	GPC5	Cancer	2	UPF2 / RNPS1	0	0	2	0
Q96SL4	GPX7	Stress	1	UPF1	0	1	3	1

(continued)

Uniprot ID	Gene Symbol	Biological Process	NMD links	NMD-factors	Coexpression	miRNA Coregulation	TF Coregulation	NMD Target
Q8TED1	GPX8	Stress	1	UPF1	1	0	3	0
P62993	GRB2	Development	4	CASC3 / EIF4A3 / RUVBL1 / RUVBL2	5	2	12	1
P49840	GSK3A	Stress	1	RUVBL2	5	1	13	1
P49841	GSK3B	Stress Development	3	UPF3A / DHX34 / RUVBL2	5	4	15	0
P15170	GSPT1	Development	6	UPF1 / UPF2 / UPF3A / SMG1 / RBM8A / RUVBL1	3	1	17	1
Q8IYD1	GSPT2	Development	16	UPF1 / UPF2 / UPF3A / UPF3B / SMG1 / SMG5 / SMG6 / SMG7 / SMG8 / SMG9 / RNPS1 / RBM8A / MAGOH / CASC3 / EIF4A3 / MOV10	0	1	0	1
P0C5Y9	H2AFB1	Development	2	RUVBL1 / RUVBL2	0	0	0	0
Q9BTM1	H2AFJ	Development	4	RBM8A / SEC13 / RUVBL1 / RUVBL2	1	0	8	1
Q71UI9	H2AFV	Development	2	RUVBL1 / RUVBL2	3	2	24	0
P16104	H2AFX	Development	4	UPF1 / SEC13 / RUVBL1 / RUVBL2	7	2	24	1
P0C0S5	H2AFZ	Development	2	RUVBL1 / RUVBL2	10	0	9	0
P84243	H3F3A	Cancer Development	3	GNL2 / RUVBL1 / RUVBL2	6	0	8	0
P51858	HDGF	Stress	2	MOV10 / RUVBL1	5	6	16	0
Q16665	HIF1A	Cancer Stress	1	RUVBL2	0	2	10	1
P04908	HIST1H2AB	Development	4	SEC13 / MOV10 / RUVBL1 / RUVBL2	0	1	6	1
Q93077	HIST1H2AC	Development	3	SEC13 / RUVBL1 / RUVBL2	0	1	8	1
P20671	HIST1H2AD	Development	5	MAGOH / EIF4A3 / SEC13 / RUVBL1 / RUVBL2	0	1	5	1
Q99878	HIST1H2AJ	Development	4	RBM8A / SEC13 / RUVBL1 / RUVBL2	0	2	3	1
Q96A08	HIST1H2BA	Development	2	RUVBL1 / RUVBL2	0	0	0	1
P33778	HIST1H2BB	Development	2	RUVBL1 / RUVBL2	0	0	1	1
P62807	HIST1H2BC	Development	5	UPF3B / MAGOH / EIF4A3 / RUVBL1 / RUVBL2	0	2	9	1
P58876	HIST1H2BD	Development	2	RUVBL1 / RUVBL2	0	4	10	1
Q93079	HIST1H2BH	Development	2	RUVBL1 / RUVBL2	0	0	1	1
P06899	HIST1H2BJ	Development	2	RUVBL1 / RUVBL2	0	2	13	1
O60814	HIST1H2BK	Development	2	RUVBL1 / RUVBL2	4	2	12	1
Q99880	HIST1H2BL	Development	2	RUVBL1 / RUVBL2	0	1	4	1
Q99879	HIST1H2BM	Development	2	RUVBL1 / RUVBL2	0	0	3	1

Q99877	HIST1H2BN	Development	2	RUVBL1 / RUVBL2	0	0	5	1
P23527	HIST1H2BO	Development	3	MOV10 / RUVBL1 / RUVBL2	0	0	7	1
P68431	HIST1H3A	Cancer Development	3	GNL2 / RUVBL1 / RUVBL2	0	0	3	1
P62805	HIST1H4A	Cancer Development	7	UPF1 / MAGOH / EIF4A3 / SEC13 / MOV10 / RUVBL1 / RUVBL2	3	0	8	0
Q6FI13	HIST2H2AA3	Development	3	SEC13 / RUVBL1 / RUVBL2	0	3	1	0
Q16777	HIST2H2AC	Development	3	SEC13 / RUVBL1 / RUVBL2	0	0	12	1
Q16778	HIST2H2BE	Development	2	RUVBL1 / RUVBL2	0	3	9	1
Q71DI3	HIST2H3A;	Development	3	GNL2 / RUVBL1 / RUVBL2	0	0	0	0
Q8N257	HIST3H2BB	Development	2	RUVBL1 / RUVBL2	1	0	3	1
P17096	HMGA1	Cancer	1	UPF2	3	6	8	1
P52926	HMGA2	Cancer	6	RBM8A / MAGOH / EIF4A3 / SEC13 / RUVBL1 / RUVBL2	1	3	4	0
P41235	HNF4A	Development	8	UPF1 / UPF3B / SMG9 / MAGOH / NBAS / MOV10 / RUVBL2 / PNRC2	0	0	7	1
Q14541	HNF4G	Development	1	PNRC2	0	0	0	1
P22626	HNRNPA2B1	Cancer	4	UPF3B / RNPS1 / RBM8A / MAGOH	6	4	20	1
P09016	HOXD4	Development	1	RNPS1	0	0	0	1
P07900	HSP90AA1	Cancer Stress Development	7	SMG1 / SMG5 / SMG6 / SEC13 / GNL2 / RUVBL1 / RUVBL2	7	4	13	1
P08238	HSP90AB1	Cancer Stress Development	9	UPF1 / UPF2 / EIF4A3 / DHX34 / SEC13 / GNL2 / MOV10 / RUVBL1 / RUVBL2	3	5	16	1
P14625	HSP90B1	Stress	2	SMG1 / RUVBL2	1	3	12	1
P0DMV8	HSPA1A	Stress	8	UPF1 / MAGOH / EIF4A3 / SEC13 / GNL2 / MOV10 / RUVBL1 / RUVBL2	0	1	1	1
P0DMV9	HSPA1B	Stress	9	UPF1 / UPF2 / MAGOH / EIF4A3 / SEC13 / GNL2 / MOV10 / RUVBL1 / RUVBL2	0	8	3	1
P34931	HSPA1L	Stress	7	UPF1 / UPF2 / EIF4A3 / SEC13 / GNL2 / RUVBL1 / RUVBL2	1	2	0	1
P54652	HSPA2	Stress	7	UPF1 / UPF2 / EIF4A3 / SEC13 / GNL2 / RUVBL1 / RUVBL2	2	1	7	1
P34932	HSPA4	Stress	7	UPF1 / UPF2 / SMG1 / GNL2 / MOV10 / RUVBL1 / RUVBL2	11	2	14	0

(continued)

Uniprot ID	Gene Symbol	Biological Process	NMD links	NMD-factors	Coexpression	miRNA Coregulation	TF Coregulation	NMD Target
O95757	HSPA4L	Stress	6	UPF1 / UPF2 / MAGOH / GNL2 / RUVBL1 / RUVBL2	2	2	8	0
P11021	HSPA5	Stress	4	GNL2 / MOV10 / RUVBL1 / RUVBL2	0	1	15	1
P17066	HSPA6	Stress	7	UPF1 / UPF2 / EIF4A3 / SEC13 / GNL2 / RUVBL1 / RUVBL2	0	0	2	1
P11142	HSPA8	Stress Development	9	UPF1 / UPF2 / RNPS1 / MAGOH / EIF4A3 / SEC13 / GNL2 / RUVBL1 / RUVBL2	4	7	16	1
P38646	HSPA9	Stress	7	MAGOH / EIF4A3 / DHX34 / GNL2 / MOV10 / RUVBL1 / RUVBL2	10	2	19	1
Q92598	HSPH1	Stress	5	UPF1 / UPF2 / GNL2 / RUVBL1 / RUVBL2	5	3	11	0
Q9Y4L1	HYOU1	Stress	1	MOV10	2	4	15	1
O75874	IDH1	Cancer	2	UPF1 / MOV10	0	0	10	1
P48735	IDH2	Cancer	1	UPF1	0	1	8	1
Q13422	IKZF1	Cancer	1	MAGOHB	3	1	12	1
P56199	ITGA1	Development	1	MAGOH	0	0	5	0
P23458	JAK1	Cancer	2	MAGOH / EIF4A3	2	1	10	1
P52333	JAK3	Cancer	1	RNPS1	1	0	3	0
Q86VZ6	JAZF1	Cancer	1	RUVBL1	1	0	1	0
P05412	JUN	Cancer Development	8	UPF1 / UPF2 / UPF3B / RNPS1 / EIF4A3 / MOV10 / RUVBL1 / RUVBL2	0	3	11	1
Q92794	KAT6A	Cancer	1	RNPS1	5	2	16	0
Q8WYB5	KAT6B	Cancer	2	RUVBL1 / RUVBL2	2	0	10	0
P29375	KDM5A	Cancer	1	RBM8A	7	0	11	1
O95239	KIF4A	Development	1	MOV10	2	2	4	1
P33176	KIF5B	Cancer	1	EIF4A3	2	2	11	0
Q9BQ90	KLHDC3	Stress	1	MOV10	1	2	8	0
Q8NG31	KNL1	Cancer	2	SEC13 / RUVBL1	2	0	0	0
Q9Y448	KNSTRN	Cancer	1	MOV10	8	1	3	0
P04264	KRT1	Development	2	MAGOH / EIF4A3	0	0	0	1
P13645	KRT10	Development	2	MAGOH / EIF4A3	0	1	1	0
P02533	KRT14	Development	2	MAGOH / EIF4A3	0	0	0	1
P08779	KRT16	Development	1	EIF4A3	0	0	1	1
P05783	KRT18	Development	1	UPF2	0	1	4	1

P08727	KRT19	Development	2	MAGOH / EIF4A3	0	1	3	1
P35908	KRT2	Development	2	MAGOH / EIF4A3	0	0	1	1
O76011	KRT34	Development	1	SMG9	1	0	1	1
P13647	KRT5	Development	2	MAGOH / EIF4A3	0	0	2	1
P04259	KRT6B	Development	1	EIF4A3	0	0	0	1
Q3SY84	KRT71	Development	1	EIF4A3	0	0	0	1
O95678	KRT75	Development	2	MAGOHB / EIF4A3	0	0	0	1
P35527	KRT9	Development	2	MAGOH / EIF4A3	0	0	0	1
Q07627	KRTAP1–1	Development	1	MAGOHB	1	0	0	0
Q52LG2	KRTAP13–2	Development	1	MAGOHB	0	0	0	0
Q3LI66	KRTAP6–2	Development	1	CASC3	0	0	0	0
Q9UJU2	LEF1	Cancer	1	RUVBL1	1	1	0	1
P53671	LIMK2	Development	1	RUVBL2	2	0	8	1
Q9H9Z2	LIN28A	Development	2	UPF1 / MOV10	0	2	0	1
P02545	LMNA	Cancer Stress	4	SMG1 / GNL2 / MOV10 / RUVBL1	0	1	13	1
Q8N653	LZTR1	Cancer	2	EIF4A3 / RUVBL1	5	0	19	1
P36507	MAP2K2	Cancer Development	2	RUVBL1 / RUVBL2	5	1	7	1
P28482	MAPK1	Cancer Stress Development	3	UPF1 / RUVBL1 / RUVBL2	6	2	20	1
P27361	MAPK3	Stress Development	3	UPF1 / RUVBL1 / RUVBL2	2	0	2	1
P49137	MAPKAPK2	Stress	1	MOV10	1	3	1	1
Q14703	MBTPS1	Stress	1	MOV10	5	1	13	1
Q93074	MED12	Cancer Development	1	CASC3	7	1	0	0
Q9UHV7	MED13	Development	1	PNRC2	8	2	1	0
Q71F56	MED13L	Development	1	PNRC2	2	2	3	1
A0JLT2	MED19	Development	2	RNPS1 / RUVBL1	5	1	8	1
Q9H204	MED28	Development	1	MOV10	9	2	2	0
Q9NX70	MED29	Development	1	MOV10	5	1	11	1
Q9Y3C7	MED31	Development	1	MOV10	6	1	4	0
Q9NPJ6	MED4	Development	2	RNPS1 / SEC13	6	0	8	0

(continued)

Uniprot ID	Gene Symbol	Biological Process	NMD links	NMD-factors	Coexpression	miRNA Coregulation	TF Coregulation	NMD Target
Q96G25	MED8	Development	1	MOV10	5	2	8	1
Q14814	MEF2D	Development	1	MOV10	2	4	14	1
P40692	MLH1	Cancer	2	RUVBL1 / RUVBL2	3	0	18	1
Q9BVC4	MLST8	Stress	4	SEC13 / GNL2 / MOV10 / RUVBL2	8	0	14	1
P43246	MSH2	Cancer	2	RUVBL1 / RUVBL2	10	1	15	0
P52701	MSH6	Cancer	1	RUVBL2	10	4	12	0
P42345	MTOR	Cancer Stress	7	UPF1 / UPF3B / SMG1 / EIF4A3 / MOV10 / RUVBL1 / RUVBL2	10	1	6	1
P10242	MYB	Cancer	1	UPF2	1	2	7	1
P12524	MYCL	Cancer	2	RUVBL1 / RUVBL2	0	0	1	0
Q969H8	MYDGF	Stress	1	MOV10	3	0	0	0
P35580	MYH10	Development	3	MAGOH / EIF4A3 / RUVBL2	1	0	3	1
Q7Z406	MYH14	Development	1	RUVBL2	0	3	1	1
P35579	MYH9	Cancer Development	3	MAGOH / EIF4A3 / RUVBL2	1	7	16	1
Q9H9S0	NANOG	Development	4	SEC13 / MOV10 / RUVBL1 / RUVBL2	3	0	1	1
O60934	NBN	Cancer	1	CASC3	4	0	17	0
Q09161	NCBP1	Development	16	UPF1 / UPF2 / UPF3A / UPF3B / SMG1 / SMG5 / SMG6 / SMG7 / SMG8 / SMG9 / RNPS1 / RBM8A / MAGOH / CASC3 / EIF4A3 / RUVBL1	1	1	15	0
P52298	NCBP2	Development	15	UPF1 / UPF2 / UPF3A / UPF3B / SMG1 / SMG5 / SMG6 / SMG7 / SMG8 / SMG9 / RNPS1 / RBM8A / MAGOH / CASC3 / EIF4A3	9	1	24	0
Q92597	NDRG1	Cancer	2	UPF1 / RUVBL2	0	1	19	1
Q00653	NFKB2	Cancer	3	MOV10 / RUVBL1 / RUVBL2	3	1	13	1
O00221	NFKBIE	Cancer	1	RUVBL2	0	1	5	1
P23511	NFYA	Stress	1	MOV10	3	0	16	0
Q8N4C6	NIN	Cancer	1	UPF1	3	2	6	0
Q15155	NOMO1	Development	1	UPF2	2	0	1	1
Q5JPE7	NOMO2	Development	1	MOV10	1	0	1	1
Q15233	NONO	Cancer	1	RNPS1	11	4	9	0
P06748	NPM1	Cancer	6	UPF1 / EIF4A3 / GNL2 / MOV10 / RUVBL1 / RUVBL2	8	3	19	0

Q13285	NR5A1	Development	2	EIF4A3 / PNRC2	1	0	2	0
O00482	NR5A2	Development	1	PNRC2	0	0	1	1
O96028	NSD2	Cancer	1	RUVBL2	7	0	0	0
P78549	NTHL1	Cancer	2	MOV10 / RUVBL2	10	3	18	1
O95631	NTN1	Development	6	UPF1 / SMG5 / EIF4A3 / DHX34 / GNL2 / RUVBL2	0	0	0	0
O00634	NTN3	Development	6	UPF1 / SMG5 / EIF4A3 / DHX34 / GNL2 / RUVBL2	0	0	1	0
Q9HB63	NTN4	Development	1	MOV10	0	0	1	1
P04629	NTRK1	Cancer	9	UPF1 / UPF2 / SMG8 / EIF4A3 / GNL2 / NBAS / MOV10 / RUVBL1 / RUVBL2	0	0	4	1
Q14980	NUMA1	Cancer	2	MOV10 / RUVBL1	5	0	13	1
P57740	NUP107	Stress	5	UPF3B / RNPS1 / RBM8A / MAGOH / SEC13	8	1	16	0
Q8WUM0	NUP133	Stress	5	UPF3B / RNPS1 / RBM8A / MAGOH / SEC13	2	0	13	0
P49790	NUP153	Stress	4	UPF3B / RNPS1 / RBM8A / MAGOH	8	3	19	0
O75694	NUP155	Stress	4	UPF3B / RNPS1 / RBM8A / MAGOH	5	4	12	0
Q12769	NUP160	Stress	5	UPF3B / RNPS1 / RBM8A / MAGOH / SEC13	10	1	4	0
Q5SRE5	NUP188	Stress	4	UPF3B / RNPS1 / RBM8A / MAGOH	10	1	8	1
Q92621	NUP205	Stress	5	UPF3B / RNPS1 / RBM8A / MAGOH / MOV10	10	2	24	0
Q8TEM1	NUP210	Stress	4	UPF3B / RNPS1 / RBM8A / MAGOH	8	1	8	1
P35658	NUP214	Cancer Stress	4	UPF3B / RNPS1 / RBM8A / MAGOH	5	3	18	1
Q8NFH5	NUP35	Stress	4	UPF3B / RNPS1 / RBM8A / MAGOH	2	2	10	0
Q8NFH4	NUP37	Stress	5	UPF3B / RNPS1 / RBM8A / MAGOH / SEC13	2	1	4	0
Q8NFH3	NUP43	Stress	5	UPF3B / RNPS1 / RBM8A / MAGOH / SEC13	2	2	5	0
Q9UKX7	NUP50	Stress	4	UPF3B / RNPS1 / RBM8A / MAGOH	6	3	13	1
Q7Z3B4	NUP54	Stress	4	UPF3B / RNPS1 / RBM8A / MAGOH	10	1	9	0
Q9BVL2	NUP58	Stress	4	UPF3B / RNPS1 / RBM8A / MAGOH	8	0	0	0
P37198	NUP62	Stress	5	UPF3B / RNPS1 / RBM8A / MAGOH / EIF4A3	2	1	19	1
Q9BW27	NUP85	Stress	5	UPF3B / RNPS1 / RBM8A / MAGOH / SEC13	12	0	12	1
Q99567	NUP88	Stress	4	UPF3B / RNPS1 / RBM8A / MAGOH	8	0	13	1
Q8N1F7	NUP93	Stress	5	UPF3B / RNPS1 / RBM8A / MAGOH / GNL2	9	0	6	1
P52948	NUP98	Cancer Stress	5	UPF3B / RNPS1 / RBM8A / MAGOH / SEC13	3	2	20	1
O15504	NUPL2	Stress	4	UPF3B / RNPS1 / RBM8A / MAGOH	8	0	11	0

(continued)

Uniprot ID	Gene Symbol	Biological Process	NMD links	NMD-factors	Coexpression	miRNA Coregulation	TF Coregulation	NMD Target
P07237	P4HB	Stress	2	RBM8A / RUVBL2	1	3	1	1
P11940	PABPC1	Cancer Development	18	UPF1 / UPF2 / UPF3A / UPF3B / SMG1 / SMG5 / SMG6 / SMG7 / SMG8 / SMG9 / RNPS1 / RBM8A / MAGOH / CASC3 / EIF4A3 / MOV10 / RUVBL1 / RUVBL2	4	2	12	1
O75914	PAK3	Development	1	SMG6	0	0	0	1
Q86YC2	PALB2	Cancer	1	MOV10	4	0	16	0
O95453	PARN	Stress	4	UPF1 / UPF2 / UPF3A / UPF3B	2	0	5	1
P23759	PAX7	Cancer	1	MAGOHB	0	0	1	0
Q86U86	PBRM1	Cancer	1	RUVBL1	6	2	8	0
Q15365	PCBP1	Cancer	6	UPF3B / RNPS1 / RBM8A / MAGOH / RUVBL1 / RUVBL2	5	0	7	0
P35227	PCGF2	Development	1	MOV10	0	0	1	1
Q15154	PCM1	Cancer	2	SMG5 / SMG7	4	0	17	1
P01127	PDGFB	Cancer	1	GNL2	0	2	0	1
P16234	PDGFRA	Cancer	1	RNPS1	0	3	1	1
Q8IWS0	PHF6	Cancer	1	MAGOH	7	1	1	0
Q13492	PICALM	Cancer	1	SEC13	2	0	17	0
O60331	PIP5K1C	Development	1	MOV10	1	3	3	1
P30613	PKLR	Development	2	EIF4A3 / GNL2	1	0	3	1
P55347	PKNOX1	Development	1	MOV10	0	1	12	1
Q9Y446	PKP3	Development	1	RBM8A	0	0	0	1
Q99569	PKP4	Development	1	MOV10	4	0	6	1
P29590	PML	Cancer	1	RBM8A	2	0	4	1
P54277	PMS1	Cancer	1	MOV10	7	1	10	0
P54278	PMS2	Cancer	2	RUVBL1 / RUVBL2	7	0	8	0
P28340	POLD1	Cancer	2	UPF1 / UPF2	9	1	9	1
Q07864	POLE	Cancer	1	UPF1	8	4	16	1
P24928	POLR2A	Development	8	UPF1 / UPF3B / RNPS1 / RBM8A / MAGOH / MOV10 / RUVBL1 / RUVBL2	7	5	22	1
P30876	POLR2B	Development	7	UPF3B / RNPS1 / RBM8A / MAGOH / GNL2 / RUVBL1 / RUVBL2	9	1	2	1
P19387	POLR2C	Development	5	UPF3B / RNPS1 / RBM8A / MAGOH / RUVBL2	6	2	5	1
O15514	POLR2D	Development	4	UPF3B / RNPS1 / RBM8A / MAGOH	8	1	11	1

P19388	POLR2E	Development	8	UPF3B / SMG1 / RNPS1 / RBM8A / MAGOH / SEC13 / RUVBL1 / RUVBL2	7	1	13	1
P61218	POLR2F	Development	4	UPF3B / RNPS1 / RBM8A / MAGOH	7	1	4	1
P62487	POLR2G	Development	4	UPF3B / RNPS1 / RBM8A / MAGOH	9	1	8	1
P52434	POLR2H	Development	5	UPF3B / RNPS1 / RBM8A / MAGOH / RUVBL2	7	2	10	1
P36954	POLR2I	Development	4	UPF3B / RNPS1 / RBM8A / MAGOH	6	1	8	1
P52435	POLR2J	Development	5	UPF3B / RNPS1 / RBM8A / MAGOH / RUVBL1	2	0	1	1
P53803	POLR2K	Development	5	UPF3B / RNPS1 / RBM8A / MAGOH / MOV10	8	0	8	0
P62875	POLR2L	Development	5	UPF3B / RNPS1 / RBM8A / MAGOH / RUVBL2	2	3	12	1
Q96HA1	POM121	Stress	4	UPF3B / RNPS1 / RBM8A / MAGOH	7	4	4	1
Q86W92	PPFIBP1	Cancer	1	MOV10	0	0	6	0
P30153	PPP2R1A	Cancer	17	UPF1 / UPF2 / UPF3A / UPF3B / SMG1 / SMG5 / SMG6 / SMG7 / SMG8 / SMG9 / RNPS1 / RBM8A / MAGOH / CASC3 / EIF4A3 / GNL2 / RUVBL1	6	3	10	1
Q15173	PPP2R5B	Stress	1	RUVBL1	0	0	6	1
P16298	PPP3CB	Development	1	RUVBL2	0	0	7	1
Q06830	PRDX1	Stress	2	MAGOH / EIF4A3	6	1	16	1
P32119	PRDX2	Stress	2	RUVBL1 / RUVBL2	2	1	8	1
Q6NWY9	PRPF40B	Cancer	1	RUVBL2	4	1	7	1
O75475	PSIP1	Cancer	2	EIF4A3 / PNRC2	6	1	13	0
P25788	PSMA3	Stress Development	2	RUVBL1 / RUVBL2	8	0	12	1
P20618	PSMB1	Stress Development	2	RBM8A / SEC13	8	1	7	1
P49721	PSMB2	Stress Development	2	GNL2 / RUVBL2	11	0	6	1
P49720	PSMB3	Stress Development	2	GNL2 / RUVBL2	0	0	17	1
P62191	PSMC1	Stress Development	2	RUVBL1 / RUVBL2	2	2	2	1
P35998	PSMC2	Stress Development	2	RUVBL1 / RUVBL2	5	2	1	1
P17980	PSMC3	Stress Development	3	MOV10 / RUVBL1 / RUVBL2	2	1	12	1

(continued)

Uniprot ID	Gene Symbol	Biological Process	NMD links	NMD-factors	Coexpression	miRNA Coregulation	TF Coregulation	NMD Target
P43686	PSMC4	Stress Development	3	MOV10 / RUVBL1 / RUVBL2	10	4	17	1
P62195	PSMC5	Stress Development	3	GNL2 / RUVBL1 / RUVBL2	9	2	17	1
P62333	PSMC6	Stress Development	2	RUVBL1 / RUVBL2	6	1	9	0
Q99460	PSMD1	Stress Development	4	GNL2 / MOV10 / RUVBL1 / RUVBL2	9	1	6	1
O75832	PSMD10	Stress Development	2	RUVBL1 / RUVBL2	8	1	3	0
O00231	PSMD11	Stress Development	2	RUVBL1 / RUVBL2	8	6	13	0
O00232	PSMD12	Stress Development	3	UPF1 / RUVBL1 / RUVBL2	1	1	11	1
Q9UNM6	PSMD13	Stress Development	2	RUVBL1 / RUVBL2	8	1	3	1
O00487	PSMD14	Stress Development	5	UPF1 / SEC13 / GNL2 / RUVBL1 / RUVBL2	10	0	9	0
Q13200	PSMD2	Stress Development	4	GNL2 / MOV10 / RUVBL1 / RUVBL2	6	3	15	1
O43242	PSMD3	Stress Development	3	MOV10 / RUVBL1 / RUVBL2	5	1	21	1
Q15008	PSMD6	Stress Development	2	RUVBL1 / RUVBL2	3	1	9	1
P51665	PSMD7	Stress Development	2	RUVBL1 / RUVBL2	10	2	12	1
P48556	PSMD8	Stress Development	2	GNL2 / RUVBL2	5	2	11	1
P61289	PSME3	Stress Development	2	RNPS1 / MOV10	8	5	10	1
Q15185	PTGES3	Stress	2	SMG5 / SMG6	9	1	22	0
P78406	RAE1	Stress	5	UPF3B / RNPS1 / RBM8A / MAGOH / SEC13	9	1	12	1
P49792	RANBP2	Cancer Stress	5	UPF3B / RNPS1 / RBM8A / MAGOH / SEC13	2	1	15	0
Q96S59	RANBP9	Development	1	RNPS1	2	0	7	0

P52306	RAP1GDS1	Cancer	1	MOV10	3	0	4	0
P10276	RARA	Cancer Development	2	RUVBL1 / PNRC2	2	2	5	1
P13631	RARG	Development	1	PNRC2	1	0	0	1
Q09028	RBBP4	Development	2	RUVBL1 / RUVBL2	10	1	19	0
Q16576	RBBP7	Development	2	RUVBL1 / RUVBL2	3	2	3	1
P98175	RBM10	Cancer	5	MAGOH / EIF4A3 / MOV10 / RUVBL1 / RUVBL2	8	3	3	1
Q96T37	RBM15	Cancer	1	UPF1	8	0	16	1
Q04864	REL	Cancer	3	SMG9 / MAGOHB / EIF4A3	3	2	7	1
Q04206	RELA	Development	2	RUVBL1 / RUVBL2	3	1	14	0
Q6PCD5	RFWD3	Cancer	1	MOV10	12	2	7	0
A6NKT7	RGPD3	Cancer	1	SEC13	2	0	0	0
P62745	RHOB	Development	1	SEC13	0	1	9	0
P08134	RHOC	Development	1	SEC13	1	1	4	1
P27694	RPA1	Stress	3	MOV10 / RUVBL1 / RUVBL2	5	1	8	0
P15927	RPA2	Stress	2	MOV10 / RUVBL2	4	2	16	0
P35244	RPA3	Stress	2	MOV10 / RUVBL2	9	2	6	0
P04843	RPN1	Cancer	2	SEC13 / RUVBL1	7	1	4	1
Q9UK32	RPS6KA6	Development	1	DHX34	0	1	1	1
Q8N122	RPTOR	Stress	3	MOV10 / RUVBL1 / RUVBL2	9	0	9	1
P19793	RXRA	Development	1	PNRC2	2	0	0	1
Q9NSC2	SALL1	Development	1	RUVBL2	0	2	1	1
Q9UJQ4	SALL4	Cancer Development	1	RUVBL2	3	0	3	0
P21912	SDHB	Cancer	1	RUVBL2	5	1	2	1
Q99643	SDHC	Cancer	1	UPF1	2	2	5	0
O94979	SEC31A	Stress	2	SEC13 / MOV10	5	2	14	1
P61619	SEC61A1	Stress	2	UPF3B / MOV10	4	3	13	1
Q96EE3	SEH1L	Stress	5	UPF3B / RNPS1 / RBM8A / MAGOH / SEC13	12	1	18	1
Q99719	SEPT5	Cancer	1	RUVBL2	0	0	12	0
Q14141	SEPT6	Cancer	1	DHX34	1	0	2	0
O75533	SF3B1	Cancer	8	UPF1 / UPF3B / RNPS1 / RBM8A / MAGOH / EIF4A3 / MOV10 / RUVBL2	8	2	12	0

(continued)

Uniprot ID	Gene Symbol	Biological Process	NMD links	NMD-factors	Coexpression	miRNA Coregulation	TF Coregulation	NMD Target
P23246	SFPQ	Cancer	2	RNPS1 / MOV10	9	4	10	1
Q96B97	SH3KBP1	Development	1	MOV10	0	0	1	1
P29353	SHC1	Stress Development	3	MOV10 / RUVBL1 / RUVBL2	2	0	13	0
Q12824	SMARCB1	Cancer	2	EIF4A3 / RUVBL1	10	1	7	1
Q96GM5	SMARCD1	Cancer	2	UPF2 / RUVBL1	8	1	5	1
Q6STE5	SMARCD3	Development	2	UPF2 / RUVBL1	0	0	2	1
Q969G3	SMARCE1	Cancer	1	MOV10	6	1	10	0
Q14683	SMC1A	Cancer	5	UPF3B / RNPS1 / RBM8A / MAGOH / RUVBL2	10	3	11	0
Q8TEQ0	SNX29	Cancer	1	MAGOH	1	0	5	1
P04179	SOD2	Stress	2	GNL2 / RUVBL2	0	3	14	0
P48431	SOX2	Cancer Development	3	SMG6 / RUVBL1 / RUVBL2	2	0	2	0
O60271	SPAG9	Development	1	PNRC2	5	3	9	1
Q13813	SPTAN1	Development	2	MAGOH / EIF4A3	1	1	5	1
Q01082	SPTBN1	Development	3	RBM8A / MAGOH / EIF4A3	0	4	9	0
P36956	SREBF1	Development	2	SEC13 / RUVBL2	2	1	1	1
Q12772	SREBF2	Development	1	SEC13	5	3	19	1
Q01130	SRSF2	Cancer	6	UPF3B / RNPS1 / RBM8A / MAGOH / EIF4A3 / RUVBL1	11	2	16	0
P84103	SRSF3	Cancer	5	UPF3B / RNPS1 / RBM8A / MAGOH / EIF4A3	12	1	19	1
P43307	SSR1	Stress	1	RUVBL2	3	3	11	0
P40763	STAT3	Cancer Development	2	RBM8A / MOV10	2	3	16	1
Q15468	STIL	Cancer	3	UPF1 / SMG7 / SMG9	9	1	9	0
Q16623	STX1A	Development	1	SEC13	0	1	0	1
P61266	STX1B	Development	1	SEC13	0	0	0	1
Q15022	SUZ12	Cancer Development	6	UPF1 / RNPS1 / RBM8A / MAGOH / EIF4A3 / RUVBL2	9	0	7	0
Q93075	TATDN2	Stress	1	MOV10	5	2	11	1
Q9BZK7	TBL1XR1	Cancer Development	1	RUVBL2	5	2	8	0
Q99081	TCF12	Cancer Development	2	RBM8A / MOV10	1	1	19	0

P15923	TCF3	Cancer Development	2	RUVBL1 / RUVBL2	6	2	6	0
P15884	TCF4	Development	1	MAGOHB	0	1	8	1
Q9NQB0	TCF7L2	Cancer	2	MAGOHB / RUVBL1	4	2	4	1
O14746	TERT	Cancer	5	UPF1 / SMG5 / SMG6 / RUVBL1 / RUVBL2	0	1	7	1
Q92734	TFG	Cancer	1	SEC13	3	3	9	1
P0C1Z6	TFPT	Cancer	2	RUVBL1 / RUVBL2	4	1	15	1
Q96RS0	TGS1	Development	1	MAGOH	7	0	19	0
Q9Y2W1	THRAP3	Cancer Development	6	UPF3B / RNPS1 / RBM8A / MAGOH / CASC3 / EIF4A3	8	4	21	1
Q92956	TNFRSF14	Cancer	1	RUVBL2	2	0	7	1
P04637	TP53	Cancer	4	SMG1 / SMG5 / SMG7 / MOV10	4	2	20	1
Q9H3D4	TP63	Cancer	1	UPF2	0	1	3	0
P06753	TPM3	Cancer	1	MOV10	3	4	12	1
P12270	TPR	Cancer Stress	8	UPF1 / UPF3B / RNPS1 / RBM8A / MAGOH / SEC13 / GNL2 / RUVBL2	8	1	19	1
P14373	TRIM27	Cancer	3	SMG9 / MAGOHB / EIF4A3	13	0	1	1
Q9UPN9	TRIM33	Cancer	1	RUVBL1	3	0	15	0
Q9Y4A5	TRRAP	Cancer	3	SMG1 / RUVBL1 / RUVBL2	11	1	10	0
Q92574	TSC1	Cancer	1	MOV10	2	2	7	1
Q71U36	TUBA1A	Development	5	SMG1 / SEC13 / GNL2 / RUVBL1 / RUVBL2	0	3	2	1
P68363	TUBA1B	Development	5	MAGOH / SEC13 / GNL2 / RUVBL1 / RUVBL2	2	6	9	1
Q13748	TUBA3C	Development	4	SEC13 / GNL2 / RUVBL1 / RUVBL2	1	0	0	1
Q6PEY2	TUBA3E	Development	4	SEC13 / GNL2 / RUVBL1 / RUVBL2	0	0	0	1
P68366	TUBA4A	Development	6	RNPS1 / SEC13 / GNL2 / MOV10 / RUVBL1 / RUVBL2	0	0	2	1
Q9NY65	TUBA8	Development	4	SEC13 / GNL2 / RUVBL1 / RUVBL2	1	0	1	1
A6NHL2	TUBAL3	Development	3	GNL2 / RUVBL1 / RUVBL2	1	0	0	1
Q9H4B7	TUBB1	Development	4	EIF4A3 / GNL2 / RUVBL1 / RUVBL2	3	0	1	1
Q13885	TUBB2A	Development	5	UPF2 / EIF4A3 / GNL2 / RUVBL1 / RUVBL2	0	3	2	1
Q9BVA1	TUBB2B	Development	4	EIF4A3 / GNL2 / RUVBL1 / RUVBL2	1	3	1	1
Q13509	TUBB3	Development	4	EIF4A3 / GNL2 / RUVBL1 / RUVBL2	0	3	0	1
P04350	TUBB4A	Development	4	EIF4A3 / GNL2 / RUVBL1 / RUVBL2	0	0	0	1
P68371	TUBB4B	Development	4	EIF4A3 / GNL2 / RUVBL1 / RUVBL2	8	3	5	1

(continued)

Uniprot ID	Gene Symbol	Biological Process	NMD links	NMD-factors	Coexpression	miRNA Coregulation	TF Coregulation	NMD Target
Q9BUF5	TUBB6	Development	5	EIF4A3 / GNL2 / MOV10 / RUVBL1 / RUVBL2	1	0	6	1
Q3ZCM7	TUBB8	Development	4	EIF4A3 / GNL2 / RUVBL1 / RUVBL2	0	0	0	1
Q99757	TXN2	Stress	1	MOV10	4	2	7	1
Q01081	U2AF1	Cancer	4	UPF3B / RNPS1 / RBM8A / MAGOH	9	0	14	1
P62987	UBA52	Stress Development	18	UPF1 / UPF2 / UPF3A / UPF3B / SMG1 / SMG5 / SMG6 / SMG7 / SMG8 / SMG9 / RNPS1 / RBM8A / MAGOH / CASC3 / EIF4A3 / MOV10 / RUVBL1 / RUVBL2	2	2	20	1
P0CG48	UBC	Stress Development	23	UPF1 / UPF2 / UPF3B / SMG1 / SMG5 / SMG6 / SMG7 / SMG8 / SMG9 / RNPS1 / RBM8A / MAGOH / MAGOHB / CASC3 / EIF4A3 / DHX34 / SEC13 / GNL2 / NBAS / MOV10 / RUVBL1 / RUVBL2 / PNRC2	0	3	13	1
P35125	USP6	Cancer	1	RUVBL1	2	1	0	0
P50552	VASP	Development	1	MOV10	2	1	8	1
P55072	VCP	Stress	4	SEC13 / GNL2 / RUVBL1 / RUVBL2	6	2	13	1
Q96AJ9	VTI1A	Cancer	1	SEC13	8	1	9	1
O00401	WASL	Development	1	RUVBL2	0	4	23	0
Q9H6R7	WDCP	Cancer	2	RUVBL1 / RUVBL2	8	0	0	0
P61964	WDR5	Development	2	RUVBL1 / RUVBL2	14	2	5	1
O14980	XPO1	Cancer	12	UPF1 / UPF2 / UPF3A / UPF3B / SMG7 / RNPS1 / CASC3 / EIF4A3 / SEC13 / GNL2 / RUVBL1 / RUVBL2	14	3	12	0
P62258	YWHAE	Cancer Stress	5	UPF1 / SEC13 / GNL2 / RUVBL1 / RUVBL2	6	3	10	1
P25490	YY1	Development	2	RUVBL1 / RUVBL2	7	5	7	0
Q05516	ZBTB16	Cancer	1	PNRC2	0	1	10	0
Q14202	ZMYM3	Cancer	1	MOV10	8	0	3	1

References

1. da Costa PJ, Menezes J, Romão L (2017) The role of alternative splicing coupled to nonsense-mediated mRNA decay in human disease. Int J Biochem Cell Biol 91:168–175
2. Popp MW, Maquat LE (2018) Nonsense-mediated mRNA Decay and Cancer. Curr Opin Genet Dev 48:44–50
3. Lykke-Andersen S, Jensen TH (2015) Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nat Rev Mol Cell Biol 16:665–677
4. Popp MW-L, Maquat LE (2013) Organizing principles of mammalian nonsense-mediated mRNA decay. Annu Rev Genet 47:139–165
5. Goetz AE, Wilkinson M (2017) Stress and the nonsense-mediated RNA decay pathway. Cell Mol Life Sci 74:3509–3531
6. Tani H, Imamachi N, Salam KA, Mizutani R, Ijiri K, Irie T, Yada T, Suzuki Y, Akimitsu N (2012) Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. RNA Biol 9:1370–1379
7. Chan W-K, Bhalla AD, Le Hir H, Nguyen LS, Huang L, Gécz J, Wilkinson MF (2009) A UPF3-mediated regulatory switch that maintains RNA surveillance. Nat Struct Mol Biol 16:747–753
8. Weischenfeldt J, Damgaard I, Bryder D, Theilgaard-Monch K, Thoren LA, Nielsen FC, Jacobsen SEW, Nerlov C, Porse BT (2008) NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. Genes Dev 22:1381–1396
9. Chan W-K, Huang L, Gudikote JP, Chang Y-F, Imam JS, MacLean JA, Wilkinson MF (2007) An alternative branch of the nonsense-mediated decay pathway. EMBO J 26:1820–1830
10. Wittmann J, Hol EM, Jäck H-M (2006) hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. Mol Cell Biol 26:1272–1287
11. Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet 36:1073–1078
12. He F, Li X, Spatrick P, Casillo R, Dong S, Jacobson A (2003) Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5′ to 3' mRNA decay pathways in yeast. Mol Cell 12:1439–1452
13. Lelivelt MJ, Culbertson MR (1999) Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. Mol Cell Biol 19:6710–6719
14. Yepiskoposyan H, Aeschimann F, Nilsson D, Okoniewski M, Muhlemann O (2011) Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. RNA 17:2108–2118
15. Colombo M, Karousis ED, Bourquin J, Bruggmann R, Mühlemann O (2017) Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways. RNA 23:189–201
16. Losson R, Lacroute F (1979) Interference of nonsense mutations with eukaryotic messenger RNA stability. Proc Natl Acad Sci 76:5134–5137
17. Maquat LE, Kinniburgh AJ, Rachmilewitz EA, Ross J (1981) Unstable β-globin mRNA in mRNA-deficient β0 thalassemia. Cell 27:543–553
18. Miller JN, Pearce DA (2014) Nonsense-mediated decay in genetic disease: Friend or foe? Mutat Res Rev Mutat Res 762:52–64
19. Metze S, Herzog VA, Ruepp M-D, Mühlemann O (2013) Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways. RNA 19:1432–1448
20. Le Hir H, Izaurralde E, Maquat LE, Moore MJ (2000) The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J 19:6860–6869
21. Alexandrov A, Colognori D, Shu M-D, Steitz JA (2012) Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay. Proc Natl Acad Sci 109:21313–21318
22. Le Hir H, Gatfield D, Izaurralde E, Moore MJ (2001) The exon–exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J 20:4987–4997
23. Chan W, Huang L, Gudikote JP, Chang Y, Imam JS, MacLean JA, Wilkinson MF (2007) An alternative branch of the nonsense-mediated decay pathway. EMBO J 26:1820–1830
24. Ishigaki Y, Li X, Serin G, Maquat LE (2001) Evidence for a Pioneer Round of mRNA Translation. Cell 106:607–617
25. Lejeune F, Ishigaki Y, Li X, Maquat LE (2002) The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. EMBO J 21:3536–3545
26. Gehring NH, Lamprinaki S, Kulozik AE, Hentze MW (2009) Disassembly of exon junction complexes by PYM. Cell 137:536–548
27. Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, Ohno M, Dreyfuss G, Ohno S (2006) Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev 20:355–367
28. Buchwald G, Ebert J, Basquin C, Sauliere J, Jayachandran U, Bono F, Le Hir H, Conti E (2010) Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. Proc Natl Acad Sci 107:10050–10055

29. Yamashita A, Izumi N, Kashima I, Ohnishi T, Saari B, Katsuhata Y, Muramatsu R, Morita T, Iwamatsu A, Hachiya T, Kurata R, Hirano H, Anderson P, Ohno S (2009) SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. Genes Dev 23:1091–1105
30. Shum EY, Jones SH, Shao A, Dumdie J, Krause MD, Chan W-K, Lou C-H, Espinoza JL, Song H-W, Phan MH, Ramaiah M, Huang L, McCarrey JR, Peterson KJ, De Rooij DG, Cook-andersen H, Wilkinson MF (2016) The antagonistic gene paralogs Upf3a and Upf3b govern nonsense-mediated RNA decay. Cell 165:382–395
31. Neu-Yilik G, Raimondeau E, Eliseev B, Yeramala L, Amthor B, Deniaud A, Huard K, Kerschgens K, Hentze MW, Schaffitzel C, Kulozik AE (2017) Dual function of UPF3B in early and late translation termination. EMBO J 36:2968–2986
32. Mühlemann O, Karousis ED (2017) New functions in translation termination uncovered for NMD factor UPF3B. EMBO J 36:2928–2930
33. Arias-Palomo E, Yamashita A, Fernández IS, Núñez-Ramírez R, Bamba Y, Izumi N, Ohno S, Llorca O (2011) The nonsense-mediated mRNA decay SMG-1 kinase is regulated by large-scale conformational changes controlled by SMG-8. Genes Dev 25:153–164
34. Chamieh H, Ballut L, Bonneau F, Le Hir H (2008) NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. Nat Struct Mol Biol 15:85–93
35. Deniaud A, Karuppasamy M, Bock T, Masiulis S, Huard K, Garzoni F, Kerschgens K, Hentze MW, Kulozik AE, Beck M, Neu-Yilik G, Schaffitzel C (2015) A network of SMG-8, SMG-9 and SMG-1 C-terminal insertion domain regulates UPF1 substrate recruitment and phosphorylation. Nucleic Acids Res 43:7600–7611
36. Melero R, Uchiyama A, Castaño R, Kataoka N, Kurosawa H, Ohno S, Yamashita A, Llorca O (2014) Structures of SMG1-UPFs complexes: SMG1 contributes to regulate UPF2-dependent activation of UPF1 in NMD. Structure 22:1105–1119
37. Okada-Katsuhata Y, Yamashita A, Kutsuzawa K, Izumi N, Hirahara F, Ohno S (2012) N-and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. Nucleic Acids Res 40:1251–1266
38. Hug N, Cáceres JF (2014) The RNA Helicase DHX34 Activates NMD by promoting a transition from the surveillance to the decay-inducing complex. Cell Rep 8:1845–1856
39. Melero R, Hug N, Lopez-Perrote A, Yamashita A, Caceres JF, Llorca O (2016) The RNA helicase DHX34 functions as a scaffold for SMG1-mediated UPF1 phosphorylation. Nat Commun 7:10585
40. Cho H, Kim KM, Kim YK (2009) Human proline-rich nuclear receptor coregulatory protein 2 mediates an interaction between mRNA surveillance machinery and decapping complex. Mol Cell 33:75–86
41. Eberle AB, Lykke-Andersen S, Mühlemann O, Jensen TH (2009) SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. Nat Struct Mol Biol 16:49–55
42. Huntzinger E, Kashima I, Fauser M, Sauliere J, Izaurralde E (2008) SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. RNA 14:2609–2617
43. Glavan F, Behm-Ansmant I, Izaurralde E, Conti E (2006) Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. EMBO J 25:5117–5125
44. Boehm V, Haberman N, Ottens F, Ule J, Gehring NH (2014) 3′ UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. Cell Rep 9:555–568
45. Schmidt SA, Foley PL, Jeong DH, Rymarquis LA, Doyle F, Tenenbaum SA, Belasco JG, Green PJ (2015) Identification of SMG6 cleavage sites and a preferred RNA cleavage motif by global analysis of endogenous NMD targets in human cells. Nucleic Acids Res 43:309–323
46. Lykke-Andersen S, Chen Y, Ardal BR, Lilje B, Waage J, Sandelin A, Jensen TH (2014) Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. Genes Dev 28:2498
47. Fukuhara N, Ebert J, Unterholzner L, Lindner D, Izaurralde E, Conti E (2005) SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. Mol Cell 17:537–547
48. Gatfield D, Izaurralde E (2004) Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila. Nature 429:575–578
49. Lejeune F, Li X, Maquat LE (2003) Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. Mol Cell 12:675–687
50. Unterholzner L, Izaurralde E (2004) SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. Mol Cell 16:587–596
51. Eberle AB, Stalder L, Mathys H, Orozco RZ, Mühlemann O (2008) Posttranscriptional gene regulation by spatial rearrangement of the 3′ untranslated region. PLoS Biol 6:849–859
52. Ivanov PV, Gehring NH, Kunz JB, Hentze MW, Kulozik AE (2008) Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. EMBO J 27:736–747
53. Singh G, Rebbapragada I, Lykke-Andersen J (2008) A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. PLoS Biol 6:e111
54. Silva AL, Ribeiro P, Inácio A, Liebhaber SA, Romão L (2008) Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. RNA 14:563–576

55. Wang W, Czaplinski K, Rao Y, Peltz SW (2001) The role of Upf proteins in modulating the translation read-through of nonsense-containing transcripts. EMBO J 20:880–890
56. Kononenko AV, Mitkevich VA, Atkinson GC, Tenson T, Dubovaya VI, Frolova LY, Makarov AA, Hauryliuk V (2010) GTP-dependent structural rearrangement of the eRF1:eRF3 complex and eRF3 sequence motifs essential for PABP binding. Nucleic Acids Res 38:548–558
57. Fatscher T, Boehm V, Weiche B, Gehring NH (2014) The interaction of cytoplasmic poly(A)-binding protein with eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay. RNA 20:1579–1592
58. Roque S, Cerciat M, Gaugué S, Mora L, Floch AG, De Zamaroczy M, Heurgué-Hamard V, Kervestin S (2015) Interaction between the poly(A)-binding protein Pab1 and the eukaryotic release factor eRF3 regulates translation termination but not mRNA decay in Saccharomyces cerevisiae. RNA 21:124–134
59. Joncourt R, Eberle AB, Rufener SC, Mühlemann O (2014) Eukaryotic initiation factor 4G suppresses nonsense-mediated mRNA decay by two genetically separable mechanisms. PLoS One 9:e104391
60. Gehring NH, Kunz JB, Neu-Yilik G, Breit S, Viegas MH, Hentze MW, Kulozik AE (2005) Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. Mol Cell 20:65–75
61. Bühler M, Steiner S, Mohn F, Paillusson A, Mühlemann O (2006) EJC-independent degradation of nonsense immunoglobulin-μ mRNA depends on 3′ UTR length. Nat Struct Mol Biol 13:462–464
62. Huang L, Lou C-H, Chan W, Shum EY, Shao A, Stone E, Karam R, Song H-W, Wilkinson MF (2011) RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. Mol Cell 43:950–961
63. Aznarez I, Nomakuchi TT, Tetenbaum-Novatt, Jaclyn Rahman MA, Fregoso O, Rees H, Krainer AR (2018) Mechanism of Nonsense-Mediated mRNA Decay Stimulation by Splicing Factor SRSF1. Cell Rep 23:2186–2198
64. Maquat LE (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol 5:89–99
65. Oliveira CC, McCarthy JE (1995) The relationship between eukaryotic translation and mRNA stability. A short upstream open reading frame strongly inhibits translational initiation and greatly accelerates mRNA degradation in the yeast Saccharomyces cerevisiae. J Biol Chem 270:8936–8943
66. Hogg JR, Goff SP (2010) Upf1 senses 3'UTR length to potentiate mRNA decay. Cell 143:379–389
67. Amrani N, Ghosh S, Mangus DA, Jacobson A (2008) Translation factors promote the formation of two states of the closed-loop mRNP. Nature 453:1276–1280
68. Peixeiro I, Inácio Â, Barbosa C, Silva AL, Liebhaber SA, Romão L (2012) Interaction of PABPC1 with the translation initiation complex is critical to the NMD resistance of AUG-proximal nonsense mutations. Nucleic Acids Res 40:1160–1173
69. Nasif S, Contu L, Mühlemann O (2018) Beyond quality control: the role of nonsense-mediated mRNA decay (NMD) in regulating gene expression. Semin Cell Dev Biol 75:78–87
70. Medghalchi SM, Frischmeyer PA, Mendell JT, Kelly AG, Lawler AM, Dietz HC (2001) Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. Hum Mol Genet 10:99–105
71. Li T, Shi Y, Wang P, Guachalla LM, Sun B, Joerss T, Chen Y-S, Groth M, Krueger A, Platzer M, Yang Y-G, Rudolph KL, Wang Z-Q (2015) Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. EMBO J 34:1630–1647
72. McIlwain DR, Pan Q, Reilly PT, Elia AJ, McCracken S, Wakeham AC, Itie-Youten A, Blencowe BJ, Mak TW (2010) Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay. Proc Natl Acad Sci 107:12186–12191
73. Shaheen R, Anazi S, Ben-Omran T, Seidahmed MZ, Caddle LB, Palmer K, Ali R, Alshidi T, Hagos S, Goodwin L, Hashem M, Wakil SM, Abouelhoda M, Colak D, Murray SA, Alkuraya FS (2016) Mutations in SMG9, encoding an essential component of nonsense-mediated decay machinery, cause a multiple congenital anomaly syndrome in humans and mice. Am J Hum Genet 98:643–652
74. Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. Science 318:798–801
75. Imamachi N, Tani H, Akimitsu N (2012) Up-frameshift protein 1 (UPF1): multitalented entertainer in RNA decay. Drug Discov Ther 6:55–61
76. Nelson JO, Moore KA, Chapin A, Hollien J, Metzstein MM (2016) Degradation of Gadd45 mRNA by nonsense-mediated decay is essential for viability. elife 5:e12876
77. Tani H, Torimura M, Akimitsu N (2013) The RNA degradation pathway regulates the function of GAS5 a non-coding RNA in mammalian cells. PLoS One 8:e55684
78. Mourtada-Maarabouni M, Williams GT (2013) Growth arrest on inhibition of nonsense-mediated decay is mediated by noncoding RNA GAS5. Biomed Res Int 2013:358015
79. Huang L, Wilkinson MF (2012) Regulation of nonsense-mediated mRNA decay. Wiley Interdiscip Rev RNA 3:807–828
80. Lou CH, Dumdie J, Goetz A, Shum EY, Brafman D, Liao X, Mora-Castilla S, Ramaiah M, Cook-Andersen H, Laurent L, Wilkinson MF (2016) Nonsense-mediated RNA decay influences human

embryonic stem cell fate. Stem Cell Reports 6:844–857
81. Thoren LA, Nørgaard GA, Weischenfeldt J, Waage J, Jakobsen JS, Damgaard I, Bergström FC, Blom AM, Borup R, Bisgaard HC, Porse BT (2010) UPF2 is a critical regulator of liver development, function and regeneration. PLoS One 5:e11650
82. Bao J, Tang C, Yuan S, Porse BT, Yan W (2015) UPF2, a nonsense-mediated mRNA decay factor, is required for prepubertal Sertoli cell development and male fertility by ensuring fidelity of the transcriptome. Development 142:352–362
83. Bao J, Vitting-Seerup K, Waage J, Tang C, Ge Y, Porse BT, Yan W (2016) UPF2-dependent nonsense-mediated mRNA decay pathway is essential for spermatogenesis by selectively eliminating longer 3'UTR transcripts. PLoS Genet 12:e1005863
84. Fanourgakis G, Lesche M, Akpinar M, Dahl A, Jessberger R (2016) chromatoid body protein TDRD6 supports long 3' UTR triggered nonsense mediated mRNA decay. PLoS Genet 12:e1005857
85. Kunz JB, Neu-Yilik G, Hentze MW, Kulozik AE, Gehring NH (2006) Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation. RNA 12:1015–1022
86. Turner JMA (2007) Meiotic sex chromosome inactivation. Development 134:1823–1831
87. Kim YK, Furic L, Parisien M, Major F, DesGroseillers L, Maquat LE (2007) Staufen1 regulates diverse classes of mammalian transcripts. EMBO J 26:2670–2681
88. Gong C, Kim YK, Woeller CF, Tang Y, Maquat LE (2009) SMD and NMD are competitive pathways that contribute to myogenesis: effects on PAX3 and myogenin mRNAs. Genes Dev 23:54–66
89. Bruno IG, Karam R, Huang L, Bhardwaj A, Lou CH, Shum EY, Song H-W, Corbett MA, Gifford WD, Gecz J, Pfaff SL, Wilkinson MF (2011) Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay. Mol Cell 42:500–510
90. Lou CH, Shao A, Shum EY, Espinoza JL, Huang L, Karam R, Wilkinson MF (2014) Posttranscriptional control of the stem cell and neurogenic programs by the nonsense-mediated RNA decay pathway. Cell Rep 6:748–764
91. Preitner N, Quan J, Flanagan JG (2013) This message will self-destruct: NMD regulates axon guidance. Cell 153:1185–1187
92. Colak D, Ji S-J, Porse BT, Jaffrey SR (2013) Regulation of axon guidance by compartmentalized nonsense-mediated mRNA decay. Cell 153:1252–1265
93. Nguyen LS, Jolly L, Shoubridge C, Chan WK, Huang L, Laumonnier F, Raynaud M, Hackett A, Field M, Rodriguez J, Srivastava AK, Lee Y, Long R, Addington AM, Rapoport JL, Suren S, Hahn CN, Gamble J, Wilkinson MF, Corbett MA, Gecz J (2012) Transcriptome profiling of UPF3B/NMD-deficient lymphoblastoid cells from patients with various forms of intellectual disability. Mol Psychiatry 17:1103–1115
94. Laumonnier F, Shoubridge C, Antar C, Nguyen LS, Van Esch H, Kleefstra T, Briault S, Fryns JP, Hamel B, Chelly J, Ropers HH, Ronce N, Blesson S, Moraine C, Gécz J, Raynaud M (2010) Mutations of the UPF3B gene, which encodes a protein widely expressed in neurons, are associated with nonspecific mental retardation with or without autism. Mol Psychiatry 15:767–776
95. Tarpey PS, Lucy Raymond F, Nguyen LS, Rodriguez J, Hackett A, Vandeleur L, Smith R, Shoubridge C, Edkins S, Stevens C, O'Meara S, Tofts C, Barthorpe S, Buck G, Cole J, Halliday K, Hills K, Jones D, Mironenko T, Perry J, Varian J, West S, Widaa S, Teague J, Dicks E, Butler A, Menzies A, Richardson D, Jenkinson A, Shepherd R, Raine K, Moon J, Luo Y, Parnau J, Bhat SS, Gardner A, Corbett M, Brooks D, Thomas P, Parkinson-Lawrence E, Porteous ME, Warner JP, Sanderson T, Pearson P, Simensen RJ, Skinner C, Hoganson G, Superneau D, Wooster R, Bobrow M, Turner G, Stevenson RE, Schwartz CE, Andrew Futreal P, Srivastava AK, Stratton MR, Gécz J (2007) Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. Nat Genet 39:1127–1133
96. Nguyen LS, Kim H-G, Rosenfeld JA, Shen Y, Gusella JF, Lacassie Y, Layman LC, Shaffer LG, Gécz J (2013) Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neuro-developmental disorders. Hum Mol Genet 22:1816–1825
97. Wang D, Zavadil J, Martin L, Parisi F, Friedman E, Levy D, Harding H, Ron D, Gardner LB (2011) Inhibition of nonsense-mediated RNA decay by the tumor microenvironment promotes tumorigenesis. Mol Cell Biol 31:3670–3680
98. Oren YS, McClure ML, Rowe SM, Sorscher EJ, Bester AC, Manor M, Kerem E, Rivlin J, Zahdeh F, Mann M, Geiger T, Kerem B (2014) The unfolded protein response affects readthrough of premature termination codons. EMBO Mol Med 6:685–701
99. Pakos-Zebrucka K, Koryga I, Mnich K, Ljujic M, Samali A, Gorman AM (2016) The integrated stress response. EMBO Rep 17:1374–1395
100. Hetz C (2012) The unfolded protein response: controlling cell fate decisions under ER stress and beyond. Nat Rev Mol Cell Biol 13:89–102
101. Yoshida H, Matsui T, Yamamoto A, Okada T, Mori K (2001) XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. Cell 107:881–891
102. Karam R, Lou C-H, Kroeger H, Huang L, Lin JH, Wilkinson MF (2015) The unfolded protein response is shaped by the NMD pathway. EMBO Rep 16:599–609
103. Gardner LB (2008) Hypoxic inhibition of nonsense-mediated RNA decay regulates gene expression

and the integrated stress response. Mol Cell Biol 28:3729–3741
104. Martin L, Gardner LB (2015) Stress-induced inhibition of nonsense-mediated RNA decay regulates intracellular cystine transport and intracellular glutathione through regulation of the cystine/glutamate exchanger SLC7A11. Oncogene 34:4211–4218
105. Wang D, Wengrod J, Gardner LB (2011) Overexpression of the c-myc oncogene inhibits nonsense-mediated RNA decay in B lymphocytes. J Biol Chem 286:40038–40043
106. Li Z, Vuong JK, Zhang M, Stork C, Zheng S (2017) Inhibition of nonsense-mediated RNA decay by ER stress. RNA 23:378–394
107. Sieber J, Hauer C, Bhuvanagiri M, Leicht S, Krijgsveld J, Neu-Yilik G, Hentze MW, Kulozik AE (2016) Proteomic analysis reveals branch-specific regulation of the unfolded protein response by nonsense-mediated mRNA decay. Mol Cell Proteomics 15:1584–1597
108. Buzina A, Shulman MJ (1999) Infrequent translation of a nonsense codon is sufficient to decrease mRNA level. Mol Biol Cell 10:515–524
109. Karam R, Wengrod J, Gardner LB, Wilkinson MF (2013) Regulation of nonsense-mediated mRNA decay: implications for physiology and disease. Biochim Biophys Acta 1829:624–633
110. Zheng D, Chen C-YA, Shyu A-B (2011) Unraveling regulation and new components of human P-bodies through a protein interaction framework and experimental validation. RNA 17:1619–1634
111. McEwen E, Kedersha N, Song B, Scheuner D, Gilks N, Han A, Chen JJ, Anderson P, Kaufman RJ (2005) Heme-regulated inhibitor kinase-mediated phosphorylation of eukaryotic translation initiation factor 2 inhibits translation, induces stress granule formation, and mediates survival upon arsenite exposure. J Biol Chem 280:16925–16933
112. Kimball SR, Horetsky RL, Ron D, Jefferson LS, Harding HP (2003) Mammalian stress granules represent sites of accumulation of stalled translation initiation complexes. Am J Physiol Physiol 284:C273–C284
113. Brown JAL, Roberts TL, Richards R, Woods R, Birrell G, Lim YC, Ohno S, Yamashita A, Abraham RT, Gueven N, Lavin MF (2011) A novel role for hSMG-1 in stress granule formation. Mol Cell Biol 31:4417–4429
114. Jia J, Werkmeister E, Gonzalez-Hilarion S, Leroy C, Gruenert DC, Lafont F, Tulasne D, Lejeune F (2017) Premature termination codon readthrough in human cells occurs in novel cytoplasmic foci and requires UPF proteins. J Cell Sci 130:3009–3022
115. Vattem KM, Wek RC (2004) Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. Proc Natl Acad Sci U S A 101:11269–11274
116. Palam LR, Baird TD, Wek RC (2011) Phosphorylation of eIF2 facilitates ribosomal bypass of an inhibitory upstream ORF to enhance CHOP translation. J Biol Chem 286:10939–10949
117. Abastado JP, Miller PF, Jackson BM, Hinnebusch AG (1991) Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for GCN4 translational control. Mol Cell Biol 11:486–496
118. Sakaki K, Yoshina S, Shen X, Han J, DeSantis MR, Xiong M, Mitani S, Kaufman RJ (2012) RNA surveillance is required for endoplasmic reticulum homeostasis. Proc Natl Acad Sci U S A 109:8079–8084
119. Gardner LB (2010) Nonsense-mediated RNA decay regulation by cellular stress: implications for tumorigenesis. Mol Cancer Res 8:295–308
120. Mort M, Ivanov D, Cooper DN, Chuzhanova NA (2008) A meta-analysis of nonsense mutations causing human genetic disease. Hum Mutat 29:1037–1047
121. Pinyol M, Bea S, Pla L, Ribrag V, Bosq J, Rosenwald A, Campo E, Jares P (2007) Inactivation of RB1 in mantle-cell lymphoma detected by nonsense-mediated mRNA decay pathway inhibition and microarray analysis. Blood 109:5422–5429
122. Anczuków O, Ware MD, Buisson M, Zetoune AB, Stoppa-Lyonnet D, Sinilnikova OM, Mazoyer S (2008) Does the nonsense-mediated mRNA decay mechanism prevent the synthesis of truncated BRCA1, CHK2, and p53 proteins? Hum Mutat 29:65–73
123. Karam R, Carvalho J, Bruno I, Graziadio C, Senz J, Huntsman D, Carneiro F, Seruca R, Wilkinson MF, Oliveira C (2008) The NMD mRNA surveillance pathway downregulates aberrant E-cadherin transcripts in gastric cancer cells and in CDH1 mutation carriers. Oncogene 27:4255–4260
124. Perrin-Vidoz L, Sinilnikova OM, Stoppa-Lyonnet D, Lenoir GM, Mazoyer S (2002) The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature termination codons. Hum Mol Genet 11:2805–2814
125. Ware MD, DeSilva D, Sinilnikova OM, Stoppa-Lyonnet D, Tavtigian SV, Mazoyer S (2006) Does nonsense-mediated mRNA decay explain the ovarian cancer cluster region of the BRCA2 gene? Oncogene 25:323–328
126. Reddy JC, Morris JC, Wang J, English MA, Haber DA, Shi Y, Licht JD (1995) WT1-mediated transcriptional activation is inhibited by dominant negative mutant proteins. J Biol Chem 270:10878–10884
127. Liu C, Karam R, Zhou Y, Su F, Ji Y, Li G, Xu G, Lu L, Wang C, Song M, Zhu J, Wang Y, Zhao Y, Foo WC, Zuo M, Valasek MA, Javle M, Wilkinson MF, Lu Y (2014) The UPF1 RNA surveillance gene is commonly mutated in pancreatic adenosquamous carcinoma. Nat Med 20:596–598
128. Lindeboom RGH, Supek F, Lehner B (2016) The rules and impact of nonsense-mediated mRNA decay in human cancers. Nat Genet 48:1112–1118
129. Lu J, Plank T-D, Su F, Shi X, Liu C, Ji Y, Li S, Huynh A, Shi C, Zhu B, Yang G, Wu Y, Wilkinson MF, Lu Y

(2016) The nonsense-mediated RNA decay pathway is disrupted in inflammatory myofibroblastic tumors. J Clin Invest 126:3058–3062
130. Chang L, Li C, Guo T, Wang H, Ma W, Yuan Y, Liu Q, Ye Q, Liu Z (2016) The human RNA surveillance factor UPF1 regulates tumorigenesis by targeting Smad7 in hepatocellular carcinoma. J Exp Clin Cancer Res 35:8
131. Cao L, Qi L, Zhang L, Song W, Yu Y, Xu C, Li L, Guo Y, Yang L, Liu C, Huang Q, Wang Y, Sun B, Meng B, Zhang B, Cao W (2017) Human nonsense-mediated RNA decay regulates EMT by targeting the TGF-ß signaling pathway in lung adenocarcinoma. Cancer Lett 403:246–259
132. Giampietri C, Petrungaro S, Conti S, Facchiano A, Filippini A, Ziparo E (2015) Cancer microenvironment and endoplasmic reticulum stress response. Mediat Inflamm 2015:417281
133. Degenhardt K, Mathew R, Beaudoin B, Bray K, Anderson D, Chen G, Mukherjee C, Shi Y, Gélinas C, Fan Y, Nelson DA, Jin S, White E (2006) Autophagy promotes tumor cell survival and restricts necrosis, inflammation, and tumorigenesis. Cancer Cell 10:51–64
134. Ye J, Kumanova M, Hart LS, Sloane K, Zhang H, De Panis DN, Bobrovnikova-Marjon E, Diehl JA, Ron D, Koumenis C (2010) The GCN2-ATF4 pathway is critical for tumour cell survival and proliferation in response to nutrient deprivation. EMBO J 29:2082–2096
135. Wengrod J, Martin L, Wang D, Frischmeyer-Guerrerio P, Dietz HC, Gardner LB (2013) Inhibition of nonsense-mediated RNA decay activates autophagy. Mol Cell Biol 33:2128–2135
136. Jia J, Furlan A, Gonzalez-Hilarion S, Leroy C, Gruenert DC, Tulasne D, Lejeune F (2015) Caspases shutdown nonsense-mediated mRNA decay during apoptosis. Cell Death Differ 22:1754–1763
137. Popp MW, Maquat LE (2015) Attenuation of nonsense-mediated mRNA decay facilitates the response to chemotherapeutics. Nat Commun 6:6632
138. Dotan-Cohen D, Letovsky S, Melkman AA, Kasif S (2009) Biological process linkage networks. PLoS One 4:e5313
139. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, Stefancsik R, Harsha B, Kok CY, Jia M, Jubb H, Sondka Z, Thompson S, De T, Campbell PJ (2017) COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res 45:D777–D783
140. COSMIC [Internet]. [cited 2018 Jul 18]. Available from: https://cancer.sanger.ac.uk / cosmic
141. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, Milacic M, Roca CD, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Viteri G, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P (2018) The reactome pathway knowledgebase. Nucleic Acids Res 46:D649–D655
142. Reactome project [Internet]. [cited 2018 Jul 18]. Available from: https://reactome.org / download-data
143. Hug N, Longman D, Cáceres JF (2016) Mechanism and regulation of the nonsense-mediated decay pathway. Nucleic Acids Res 44:1483–1495
144. Yang X, Coulombe-Huntington J, Kang S, Sheynkman GM, Hao T, Richardson A, Sun S, Yang F, Shen YA, Murray RR, Spirohn K, Begg BE, Duran-Frigola M, MacWilliams A, Pevzner SJ, Zhong Q, Trigg SA, Tam S, Ghamsari L, Sahni N, Yi S, Rodriguez MD, Balcha D, Tan G, Costanzo M, Andrews B, Boone C, Zhou XJ, Salehi-Ashtiani K, Charloteaux B, Chen AA, Calderwood MA, Aloy P, Roth FP, Hill DE, Iakoucheva LM, Xia Y, Vidal M (2016) Widespread expansion of protein interaction capabilities by alternative splicing. Cell 164:805–817
145. Rolland T, Taşan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, Kamburov A, Ghiassian SD, Yang X, Ghamsari L, Balcha D, Begg BE, Braun P, Brehme M, Broly MP, Carvunis A-R, Convery-Zupan D, Corominas R, Coulombe-Huntington J, Dann E, Dreze M, Dricot A, Fan C, Franzosa E, Gebreab F, Gutierrez BJ, Hardy MF, Jin M, Kang S, Kiros R, Lin GN, Luck K, MacWilliams A, Menche J, Murray RR, Palagi A, Poulin MM, Rambout X, Rasla J, Reichert P, Romero V, Ruyssinck E, Sahalie JM, Scholz A, Shah AA, Sharma A, Shen Y, Spirohn K, Tam S, Tejeda AO, Trigg SA, Twizere J-C, Vega K, Walsh J, Cusick ME, Xia Y, Barabási A-L, Iakoucheva LM, Aloy P, De Las Rivas J, Tavernier J, Calderwood MA, Hill DE, Hao T, Roth FP, Vidal M (2014) A proteome-scale map of the human interactome network. Cell 159:1212–1226
146. Venkatesan K, Rual J-F, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh K-I, Yildirim MA, Simonis N, Heinzmann K, Gebreab F, Sahalie JM, Cevik S, Simon C, de Smet A-S, Dann E, Smolyar A, Vinayagam A, Yu H, Szeto D, Borick H, Dricot A, Klitgord N, Murray RR, Lin C, Lalowski M, Timm J, Rau K, Boone C, Braun P, Cusick ME, Roth FP, Hill DE, Tavernier J, Wanker EE, Barabási A-L, Vidal M (2009) An empirical framework for binary interactome mapping. Nat Methods 6:83–90
147. Yu H, Tardivo L, Tam S, Weiner E, Gebreab F, Fan C, Svrzikapa N, Hirozane-Kishikawa T, Rietman E, Yang X, Sahalie J, Salehi-Ashtiani K, Hao T, Cusick ME, Hill DE, Roth FP, Braun P, Vidal M (2011) Next-generation sequencing to generate interactome datasets. Nat Methods 8:478–480
148. Rual J-F, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-

Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M (2005) Towards a proteome-scale map of the human protein–protein interaction network. Nature 437:1173–1178
149. Alonso-López D, Gutiérrez MA, Lopes KP, Prieto C, Santamaría R, De Las Rivas J (2016) APID interactomes: providing proteome-based interactomes with controlled quality for multiple species and derived networks. Nucleic Acids Res 44:W529–W535
150. Türei D, Korcsmáros T, Saez-Rodriguez J (2016) OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat Methods 13:966–967
151. Li T, Wernersson R, Hansen RB, Horn H, Mercer J, Slodkowicz G, Workman CT, Rigina O, Rapacki K, Stærfeldt HH, Brunak S, Jensen TS, Lage K (2017) A scored human protein-protein interaction network to catalyze genomic interpretation. Nat Methods 14:61–64
152. Martinez NJ, Walhout AJM (2009) The interplay between transcription factors and microRNAs in genome-scale regulatory networks. BioEssays 31:435–445
153. Xu Y, Yue W, Yao Shugart Y, Li S, Cai L, Li Q, Cheng Z, Wang G, Zhou Z, Jin C, Yuan J, Tian L, Wang J, Zhang K, Zhang K, Liu S, Song Y, Zhang F (2016) Exploring transcription factors-microRNAs co-regulation networks in schizophrenia. Schizophr Bull 42:1037–1045
154. Chou C-H, Shrestha S, Yang C-D, Chang N-W, Lin Y-L, Liao K-W, Huang W-C, Sun T-H, Tu S-J, Lee W-H, Chiew M-Y, Tai C-S, Wei T-Y, Tsai T-R, Huang H-T, Wang C-Y, Wu H-Y, Ho S-Y, Chen P-R, Chuang C-H, Hsieh P-J, Wu Y-S, Chen W-L, Li M-J, Wu Y-C, Huang X-Y, Ng FL, Buddhakosai W, Huang P-C, Lan K-C, Huang C-Y, Weng S-L, Cheng Y-N, Liang C, Hsu W-L, Huang H-D (2018) miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res 46:D296–D302
155. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database 2016:baw100
156. Li YF, Altman RB (2018) Systematic target function annotation of human transcription factors. BMC Biol 16:4
157. Han H, Cho J-W, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E, Lee S, Kang B, Jeong D, Kim Y, Jeon H-N, Jung H, Nam S, Chung M, Kim J-H, Lee I (2018) TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res 46:D380–D386
158. van Dam S, Craig T, de Magalhães JP (2015) GeneFriends: a human RNA-seq-based gene and transcript co-expression database. Nucleic Acids Res 43:D1124–D1132
159. Ishigaki Y, Li X, Serin G, Maquat LE (2001) Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell 106:607–617
160. Lejeune F, Ranganathan AC, Maquat LE (2004) eIF4G is required for the pioneer round of translation in mammalian cells. Nat Struct Mol Biol 11:992–1000
161. Ohnishi T, Yamashita A, Kashima I, Schell T, Anders KR, Grimson A, Hachiya T, Hentze MW, Anderson P, Ohno S (2003) Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. Mol Cell 12:1187–1200
162. Cowen LE, Tang Y (2017) Identification of nonsense-mediated mRNA decay pathway as a critical regulator of p53 isoform β. Sci Rep 7:17535
163. Brumbaugh KM, Otterness DM, Geisen C, Oliveira V, Brognard J, Li X, Lejeune F, Tibbetts RS, Maquat LE, Abraham RT (2004) The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. Mol Cell 14:585–598
164. Abraham RT (2004) The ATM-related kinase, hSMG-1, bridges genome and RNA surveillance pathways. DNA Repair (Amst) 3:919–925
165. Welch MD, Iwamatsu A, Mitchison TJ (1997) Actin polymerization is induced by Arp 2/3 protein complex at the surface of Listeria monocytogenes. Nature 385:265–269

4 The Implication of mRNA Degradation Disorders on Human DISease: Focus on DIS3 and DIS3-Like Enzymes

Margarida Saramago, Paulo J. da Costa, Sandra C. Viegas, and Cecília M. Arraiano

Abstract

RNA degradation is considered a critical post-transcriptional regulatory checkpoint, maintaining the correct functioning of organisms. When a specific RNA transcript is no longer required in the cell, it is signaled for degradation through a number of highly regulated steps. Ribonucleases (or simply RNases) are key enzymes involved in the control of RNA stability. These enzymes can perform the RNA degradation alone or cooperate with other proteins in RNA degradation complexes. Important findings over the last years have shed light into eukaryotic RNA degradation by members of the RNase II/RNB family of enzymes. DIS3 enzyme belongs to this family and represents one of the catalytic subunits of the multiprotein complex exosome. This RNase has a diverse range of functions, mainly within nuclear RNA metabolism. Humans encode two other DIS3-like enzymes: DIS3L (DIS3L1) and DIS3L2. DIS3L1 also acts in association with the exosome but is strictly cytoplasmic. In contrast, DIS3L2 acts independently of the exosome and shows a distinctive preference for uridylated RNAs. These enzymes have been shown to be involved in important cellular processes, such as mitotic control, and associated with human disorders like cancer. This review shows how the impairment of function of each of these enzymes is implicated in human disease.

Keywords

Cancer · DIS3 · DIS3L1 · DIS3L2 · Exoribonuclease · Exosome · Polyadenylation · RNase · RNA degradation · Uridylation

4.1 Introduction

RNA is a labile molecule, by its chemical nature, and a plethora of events and factors ensure its protection or decay. The termini of cytoplasmic mRNAs are usually protected by a 7-methyl guanosine (m^7GpppG) at the 5′ end cap, and by a long terminal poly(A) tail at the 3′ end, which promotes the association of Poly(A) Binding Proteins (PABPs) (Fig. 4.1a). These RNA modifications in the ends of the mRNA molecules warrant their stability. Specific protein factors

M. Saramago · S. C. Viegas (✉)
C. M. Arraiano (✉)
Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal
e-mail: sviegas@itqb.unl.pt; cecilia@itqb.unl.pt

P. J. da Costa
Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

L. Romão (ed.), *The mRNA Metabolism in Human Disease*, Advances in Experimental Medicine and Biology 1157, https://doi.org/10.1007/978-3-030-19966-1_4

associate with each of the structures in the mRNA extremities, and its physical interaction holds the mRNA in a circular conformation [1, 2]. This closed-loop mRNA structure is recognized by specific translation initiation complexes [2–5]. When a specific RNA is no longer required for the cellular metabolism, the molecule is signaled for degradation through a number of highly regulated steps (Fig. 4.1a).

4.2 Mechanisms of Cytoplasmic mRNA Degradation in Humans

The stability of mRNAs depends on intrinsic features of its sequence and on the cellular demands for the protein it encodes. For instance, specific features such as AU rich elements (AREs) that consist on stretches of adenine and uracil nucleo-

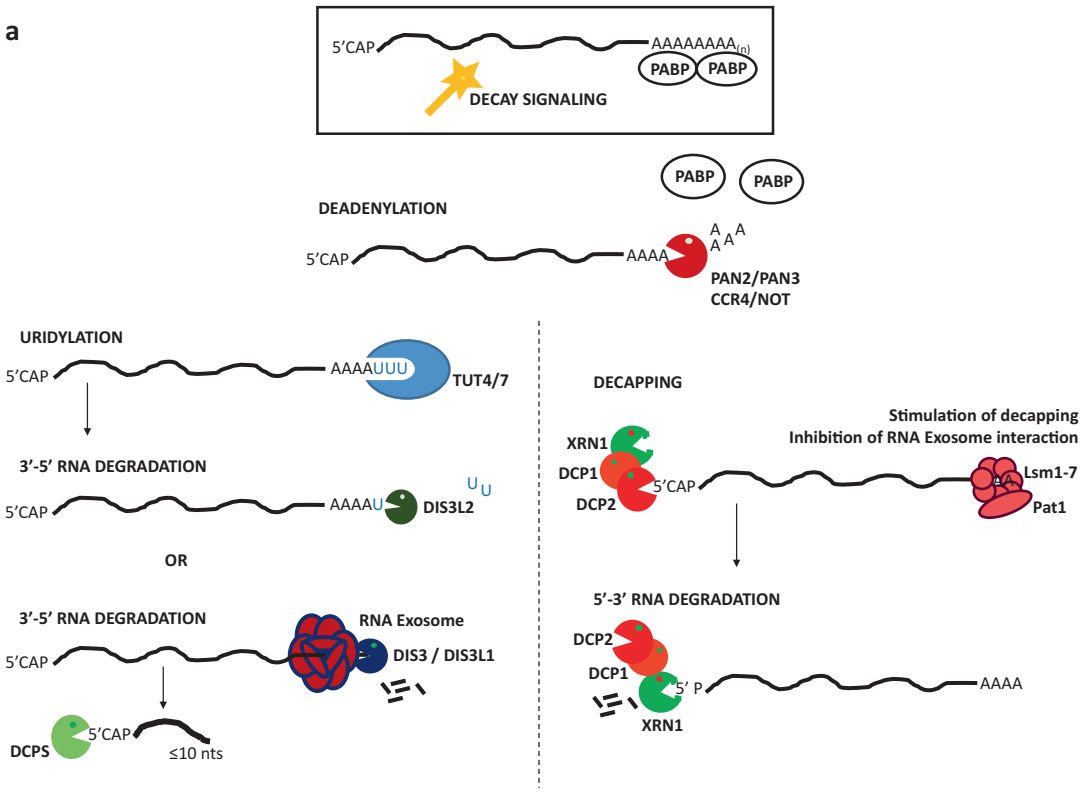

Fig. 4.1 (**a**) Overview of RNA degradation pathways in humans. RNA transcripts first undergo removal of the 3′ poly(A) tail through a deadenylation step (performed by PAN2/PAN3 complex followed by the CCR4/NOT). Following this step, the RNA becomes vulnerable and can be degraded 3′-5′ by the Exosome-DIS3/DIS3L1 or by DIS3L2 (that prefers uridylated RNA substrates), independently of the exosome. Following deadenylation, RNAs can also be decapped by DCP1/DCP2 (removal of the 5′-cap), a step stimulated by Lsm1–7–Pat1 complex, exposing the transcripts to the 5′-3′ exoribonuclease XRN1. (**b**) Correlation between DIS3 enzymes and human disease. Scheme of human diseases that have been related with overexpression or dysfunction of DIS3, DIS3L1 and DIS3L2. Dashed arrows correspond to correlations that need further confirmation. Each enzyme-disease association represented in the picture is developed in the main text, with the respective references

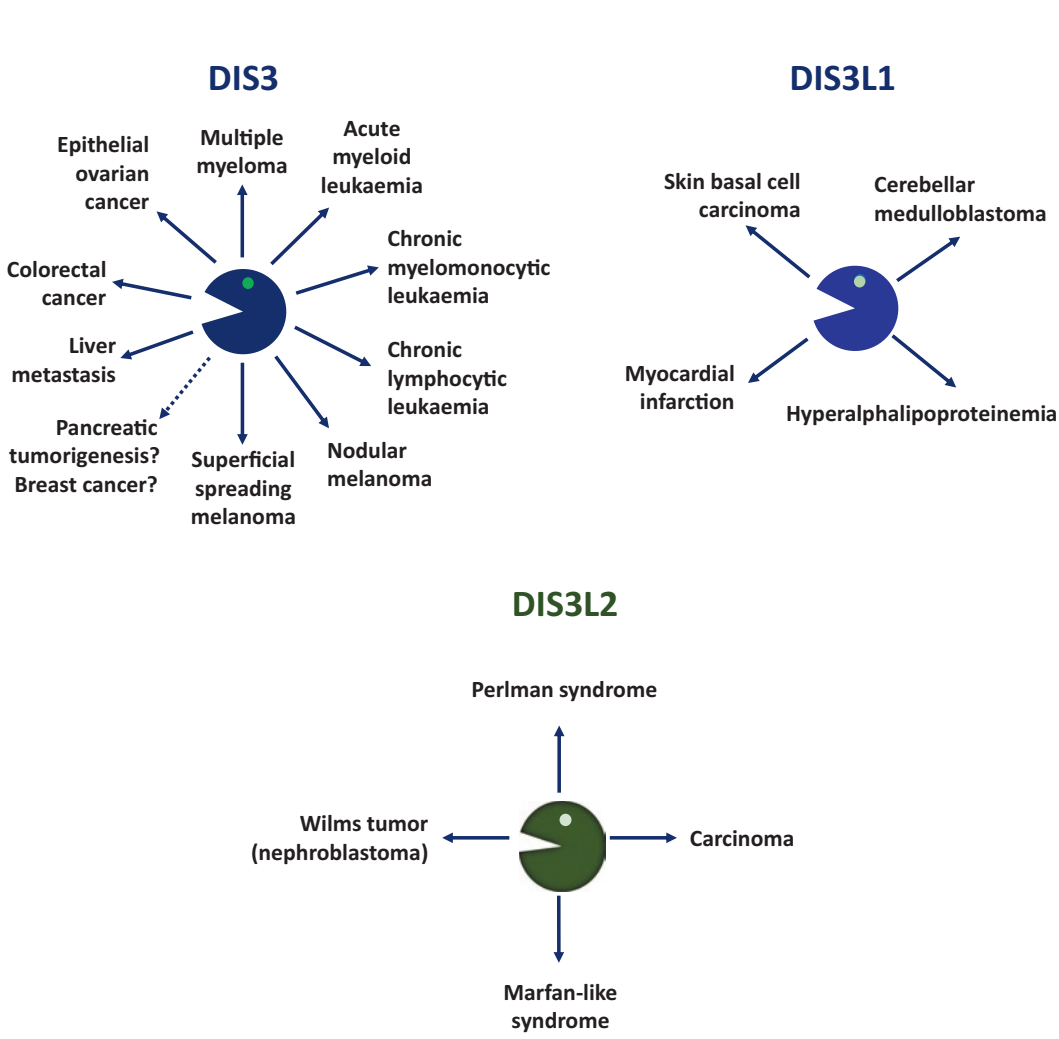

Fig. 4.1 (continued)

tides within the 3′UTR of some mRNAs [6–9] are able to promote or protect RNA from degradation [10–12]. The 3′UTR of the mRNAs can also contain other specific sequences that target its decay trough the binding of regulatory microRNAs (miRNAs) [13–15].

Under particular circumstances, like in cellular quality control mechanisms, endonucleolytic cleavage can directly disrupt the closed-circle RNA conformation and trigger subsequent decay, from either of the RNA extremities ([16, 17]*;* see other chapters in this volume for details on quality control mechanisms). Though, in general, mRNAs must be first deadenylated or decapped to enable the access of exoribonucleases and trigger degradation.

For many cytoplasmic mRNAs, decay starts with the shortening or removal of the poly(A) tail. Deadenylation releases the PABPs that protect the (A) tail and leaves the 3′ end exposed to exoribonucleases. This is considered the main event that signals mRNAs for degradation from the 3′ end. The length of the poly(A) tail depends on the organism, but the regulation of its extension is always a dynamic process that involves the concerted action of poly(A) polymerases (PAPs) and poly(A) specific 3′ exonucleases (deadenylases). This allows the fine-tuning con-

trol of mRNA stability. Eukaryotic genomes encode a wide variety of deadenylases [18]. In humans, initial deadenylation is performed by the PAN2/PAN3 complex (reviewed in [19–21]) followed by the action of the CCR4/NOT complex (Fig. 4.1a) [22–26].

Following deadenylation, degradation can alternatively proceed from the 5′ end requiring the prior removal of the 5′-cap of the mRNA. The decapping process is regulated by a plethora of activators and inhibitors [27–29]. One of these, the Lsm1–7-Pat1 complex, preferentially binds the shortened 3′-terminal adenosine extensions of the deadenylated mRNAs to stimulate decapping and inhibit exosome attachment [30, 31]. The reaction products of decapping enzymes are the 5′ m^7GDP cap and an unprotected 5′ monophosphate RNA that is then accessible for XRN1 processive and complete degradation [32, 33]. Decapping process commits RNAs to 5′-3′ degradation since XRN1 directly interacts with the decapping enzymes (DCP1/DCP2) [32] (Fig. 4.1a).

In contrast to the 5′-3′ pathway, where XRN1 is the only known cytoplasmic 5′-3′ exoribonuclease, there are several options in the degradation from the 3′ end. After initial deadenylation, the multisubunit RNA exosome complex may further degrade the shortened oligo(A) tail and proceed with the 3′-5′ degradation into the mRNA body. Exosome activity depends on the presence of specific cofactors, called "superkillers" or Ski proteins that regulate its activity. Human homologs of the Ski family – SKIV2L (Ski2), TTC37 (Ski3) and WDR61 (Ski8), associate in the Ski complex [26, 34]. The remaining mRNA fragment with its 5′-cap (m^7GpppG) is hydrolyzed by the scavenger decapping enzyme (DCPS) [35]. DCPS is a m^7G-specific pyrophosphatase that shows specificity towards RNA fragments not longer than 10 nts (Fig. 4.1a) [36–38].

DIS3 enzyme is the essential catalytic subunit of the exosome. The human genome encodes three members of DIS3 family: DIS3, DIS3L1 and DIS3L2 enzymes and their characteristics and specifications will be described therein in this chapter. While both DIS3 and DIS3L1 interact with the exosome ring, in different cellular locations [39–43], DIS3L2 does not. The enzyme represents a 3′-5′ RNA decay pathway alternative to degradation by XRN1 and the exosome [44, 45]. DIS3L2 shows a distinctive preference towards uridylated substrates, which prompted the discovery of new roles for 3′-uridylation in cytoplasmic mRNA decay. It was proposed that the requirement of deadenylation as an mRNA decay signal can be overcome trough 3′ oligouridylation of transcripts. It can either stimulate decapping and consequent degradation in the 5′-3′ direction, through Lsm1–7/Pat1 binding, or directly activate 3′-5′ DIS3L2 dependent degradation [46] (Fig. 4.1a). The importance of this uridylation-dependent pathway in bulk mRNA degradation was highlighted by the substantial technological progresses in RNA analysis in the recent years. Novel approaches as TAIL-seq revealed that 3′ end mRNA modifications such as urydilation, cytidylation or guanylation are also frequent [47]. The use of oligo(dT)-based priming methods, due to the long poly(A) tails in the 3' end of eukaryotic mRNAs, had previously underestimated its presence.

In general, regulation of gene expression in eukaryotic cells occurs through multiple parallel, partially redundant, mRNA decay pathways. This is further illustrated by the multiplicity of enzymes, which are able to catalyze the same reaction, and their functional redundancy. RNases' activity is also important for RNA surveillance and processing. Their high degree of conservation in different organisms and the specific phenotypes following their individual loss suggest defined roles within the cell.

Beyond the advances on the mechanisms whereby these enzymes affect cellular processes, structural information is crucial to explain the mechanism of action and exact function of the protein alone or in the context of a multiprotein complex. Structural changes across species provide insight into the evolution and conservation of the protein architecture. In fact, important findings over the last years have shed new light onto the mechanistic details of RNA degradation by members of the RNase II/RNB family of exoribonucleases, including DIS3 enzymes [45, 48–56]. A phylogenetic comparison of different Dis3 homologues in eukaryotes indicates a clear division of

three Dis3-like protein families, with the Dis3 group being the most conserved and the Dis3L1 and Dis3L2 groups being more divergent [45].

A growing number of publications associate DIS3-enzymes with several human diseases [57–60]. In this review we will try to sum up the mechanistic and structural details of these RNase II-like enzymes, and how human disorders can result from the associated defects.

4.2.1 DIS3

DIS3 (defective in sister chromatid joining) gene is located on human chromosome 13 (q22·1), and it encodes for a highly conserved ribonuclease (also known as Rrp44 in yeast) that contains both 3′-5′ exoribonuclease and endoribonuclease activities [40, 42, 61]. This RNase constitutes an essential catalytic subunit of the exosome [62]. This multiprotein complex is composed of a catalytically inert ring-shaped 9-subunit core with a prominent central channel and associated catalytic subunits [61, 63–65]. The composition of its catalytic subunits varies accordingly to cellular localization. Exosome-associated DIS3 enzyme exists mainly in the nucleus, where it acts on a vast array of different RNAs, and has a diverse range of functions in RNA metabolism including processing, maturation and quality control [66–71].

The specific domains of DIS3 dictate its degradation preferences. DIS3 is composed by two cold shock domains (CSD1 and CDS2), followed by an RNB and an S1 domain. Both CSDs and the S1 domain contribute to RNA binding. At the N-terminal region, it contains a PilT N-terminal (PIN) domain [40, 42, 51, 72] and also a CR3 motif involved in the binding of DIS3 to the exosome [73]. Both RNB and PIN domains are responsible for RNA degradation activity. The RNB catalytic domain is a hallmark of the RNase II protein family, and confers to DIS3 the ability to cleave RNA in a highly processive manner [51, 61, 74]. Arraiano's lab contributed to the resolution of the crystal structure of the family prototype, *E. coli* RNase II, and to its extensive functional characterization [49–51]. This was an important breakthrough in the understanding of the mechanism of action of this ubiquitous family of proteins. Moreover, the determination of the electron microscopy structure of yeast Rrp44 (Dis3) suggested that the RNA recruitment mechanism is conserved [75]. The knowledge acquired by these model organisms was crucial for the construction of the 3D model of human DIS3. Its exoribonuclease activity is dependent on four conserved aspartic acid residues (D488, D487, D485, D489) that coordinate two magnesium ions in the catalytic center [61, 63]. The RNB active site in DIS3 is responsible to hydrolyze single-stranded RNA (ssRNA) in a 3′–5′ direction, releasing one nucleotide at a time and leaving an end product of 4 nts [51, 70, 76–78]. Only ssRNAs with a minimum length of 7 nts can be cleaved [76]. Dis3/Rrp44 is also able to unwind and digest structured RNAs as long as there is an unstructured region of ~4 nts at the 3′ terminus [79].

The PIN domain in DIS3 confers the ability to cleave RNA endoribonucleolytically [40, 42, 43]. PIN-like domains constitute a widespread superfamily of nucleases with representatives in all kingdoms of life [80, 81]. Combination of endoribonuclease and exoribonuclease activities is a widespread feature of RNA-degrading machines from bacteria to humans [48]. Both DIS3 activities cooperate with each other in the degradation of RNA molecules [40, 42, 43, 82]. The PIN domain is able to cleave circular and linear ssRNAs, preferentially with a 5′ monophosphate [40, 42]. Its active site is composed of four acidic amino acids essential for endoribonuclease activity (E97, D69, D177 and D146) that coordinate two divalent metal ions, and Mn^{2+} is the preferred ion for its activity [40, 42]. It was proposed that the role of the PIN domain is to assist in the release of RNA substrates that are stalled at sites with strong secondary structures [83]. Besides its endoribonuclease activity, the PIN domain has also a structural role, being necessary for DIS3 association with the core exosome [39, 42, 43, 84].

Like in other organisms, such as *Drosophila* and yeast, human DIS3 is essential for survival [85, 86]. This RNase together with the exosome complex play a crucial role in maintaining the fidelity of gene expression. In the nucleus, the

exosome-associated DIS3 is involved in the degradation of a vast range of RNAs, including protein-coding RNAs, stable RNA species such as ribosomal RNA (rRNA), transfer RNA (tRNAs) and small nucleolar RNAs (snoRNAs), introns, long non-coding RNAs (ncRNAs), miRNAs and also unstable RNAs products, like Promoter Upstream Transcripts (PROMPTs) [87].

An impaired RNA surveillance system can compromise RNA homeostasis, having detrimental consequences for multiple biological processes, which may result in malignancy [88]. Indeed, an increasing number of publications have associated dysregulation of DIS3 with human disease, namely cancer (reviewed by [60, 89]). Sequencing data have identified *DIS3* gene as one of the most frequently mutated genes in multiple myeloma (on average in 11–18.5% of patients) [58, 90–93]. This constitutes the second most frequent hematologic tumor after lymphomas [94]. In multiple myeloma patients, DIS3 mutations were detected in highly conserved regions along PIN, CSD2, RNB and S1 domains [58, 90, 91, 93]. However, the mutations in the RNB domain seem to be more prevalent for the development of the disease [58, 90, 91], and contain mutational hotspots (D488, E665 and R780) [90, 93].

Tomecki and co-workers have studied Multiple myeloma mutations that abolish or cause dysfunction, without total inactivation, of DIS3 RNB activity *in vitro* [85]. The results indicated that the point mutations D487N, S477R, G766R and R780K cause significant aberrations on exoribonucleolytic activity. DIS3 amino acid changes with significantly decreased activity *in vitro* gave rise to a slower cellular proliferation rate in HEK293-derived cells [85]. DIS3-mutations such as R780K, an amino acid involved in RNA binding, also revealed an abnormal RNA metabolism, with accumulation of 5.8S processing intermediates, tRNAs, RNA polymerase III transcripts and PROMPTs [85].

Szczepińska and co-workers also observed a major role of DIS3 in maintaining RNA polimerase II transcriptome homeostasis and in the regulation of PROMPTs [87]. PROMPTs are transcribed in reverse orientation to most active protein coding genes, and cover ~1% of the human genome [72, 95]. Although the biological role of PROMPTs has yet to be elucidated, there are evidences that these RNAs could serve important functions in human cells [96]. For instance, the PROMPT HIF2PUT was suggested to be a novel regulator of osteosarcoma, the most common primary bone malignancy [96]. The observation that PROMPTs were the most prominent targets of DIS3 (>50-fold increase in DIS3 mutant cells), indicates that there are no alternative pathways for their decay, making their connection with DIS3 and disease necessary to be explored.

Multiple myeloma associated mutations were also mapped on PIN domain, showing only small effects on cell growth [85]. However, when mutations in PIN and RNB domains are combined, a synergistic effect in the proliferation and metabolic activity is observed in human cells [85, 87].

Human cells bear two DIS3 isoforms that differ in the size of the PIN domain. Isoform 1 encodes a full-length PIN domain, whereas the PIN domain of isoform 2 is shorter and misses a segment with conserved amino acids [52]. A study by Robinson and co-workers [52] anticipated that different ratios of the two PIN isoforms could be characteristic of several haematological cancers, namely Multiple Myeloma [52]. Isoform 1 was found in higher levels than isoform 2 in Multiple Myeloma patient samples and all cancer cell lines tested [52]. Contrastingly, healthy donors and Acute Myeloid Leukemia and Chronic Myelomonocytic Leukaemia patients have similar levels of both isoforms. Regarding leukaemias, in Acute Myeloid Leukemia (a cancer of the myeloid line of blood cells) missense mutations in DIS3 account for 4% of patients and were all found in the RNB domain [59]. In patients with Chronic Lymphocytic Leukemia (a monoclonal disorder characterized by a progressive accumulation of abnormal lymphocytes) DIS3 *locus*, 13q22, is often deleted [97]. This evidence together with DIS3 mutations in several cancers, suggest that DIS3 may function as a tumor suppressor gene.

Increased levels of DIS3 mRNA and protein have also been proposed as one of the causes of other types of cancer. This is the case of epithelial ovarian cancer in which DIS3 was observed to be

significantly up-regulated in plasma from patients in the late stage of the disease (FIGO III/IV) [98]. The majority of cancer deaths are due to metastasis of neoplastic cells from the primary tumor to distant organs. Metastasis is thus the most important factor that determines bad prognoses for cancer patients. In 1997, DIS3 was reported to have a 38-fold higher expression in primary tumors and metastatic cells from patients with colorectal cancers and liver metastases (compared to adenomas) [99]. The same study has classified DIS3 as an oncogene, being positively correlated with the incidence of metastasis and consistent with its involvement in the regulation of mitosis in *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae* and *Drosophila* [66, 100–102]. Other studies also reported a significant overexpression of DIS3 in colorectal carcinomas (compared to adenomas) [103, 104]. This overexpression could be explained by an amplification of the DIS3 *locus*, 13q22, frequently observed in colorectal cancer.

DIS3 was also found to be differentially expressed in melanoma cells [105]. Specifically, in superficial spreading melanoma cells, DIS3 has a reduced expression, contrarily to nodular melanoma cells, where DIS3 is overexpressed (compared to normal melanocytes) [105] However, in 2013, a wide-genome analysis resultant from five melanoma microarrays datasets did not recognize DIS3 as a melanoma biomarker [106]. DIS3 has also a role in pancreatic tumorigenesis and breast cancer, however their linkage has to be further explored [107, 108].

All examples presented here strongly suggest that both DIS3 overexpression and lack of function can lead to the manifestation of different cancers. This seems contradictory, but it is in agreement with the fact that DIS3 may function as either oncogene or tumor suppressor. Some genes are known to have both functions, and recently it was reported that most of these genes are transcription factors or kinases that can regulate transcription positively and negatively [109]. Since RNases are responsible to control post-transcriptionally gene expression, it is not surprising that DIS3 would also have a dual function. For instance, DIS3 is known to facilitate the maturation of the tumor suppressor *let-7* miRNA. When the levels of the mature *let-7* miRNA are reduced, translation of oncogenes (MYC and RAS) increases, enhancing tumorigenesis [110]. Human DIS3 also appear to function in the Ran signaling pathway required for nuclear import of proteins [111]. Ran was associated to cancer progression and has been investigated as a target for cancer therapy. In sum, the precise and dual role of DIS3 in cancer is not fully understood lacking further investigation.

4.2.2 DIS3L1

Human DIS3 and DIS3L1 have a similar domain composition, however only the first has an active endoribonuclease domain. Two important residues (E97 and D146N) are absent in DIS3L1 PIN domain rendering it inactive. The aspartic acid D146 is the most conserved in the PIN domains and its single mutation is reported to abolish its activity *in vivo* and *in vitro* [43, 112, 113]. The E97 is not strictly conserved across the PIN-domain family [112].

Both human DIS3 and DIS3L1 associate with the exosome ring. In contrast to the mainly nuclear localization of DIS3, DIS3L1 is strictly cytoplasmic [57, 72, 114]. The stable association of DIS3L1 with the cytoplasmic exosome suggests that it acts in concert with the core of the exosome in the degradation of cytoplasmic RNAs. One of the RNA substrates degraded by the exosome-associated DIS3L1 is the 28S rRNA. The degradation proceeds through polyadenylated intermediates, which accumulate upon DIS3L1 knockdown [115, 116]. DIS3L1 was also implicated in the degradation of intermediary products generated by DNA-based antisense oligonucleotides (ASOs), as part of the RNA surveillance machinery. These agents recruit RNase H1 that after endonucleolytic cleavage of the ASO-targeted mRNAs generates both a 5′ and a 3′ fragments. DIS3L1 appears to be involved, together with the exosome, on the 3′-5′ exoribonucleolytic degradation of the cytoplasmic upstream cleavage products [117].

Less is known about DIS3L1 association with human disease, but there are a few reports showing its implication on diverse pathologies, directly or indirectly related with its function over specific substrates. It was suggested that DIS3L1 could be implicated in the regulation of steady state levels of Y RNAs, an abundant class of small non-coding RNAs with a role in a range of cellular processes, such as RNA quality control, DNA replication, cellular stress responses and histone mRNA processing [118, 119]. Several enzymes are involved in Y RNAs maturation, both on its 3′ end adenylation and subsequent trimming. Poly(A)-specific ribonuclease (PARN) is one of the enzymes involved on its 3′ end processing and, in its absence, oligo(A) tails are degraded by the exoribonuclease DIS3L1. It was seen that PARN mutations cause a severe form of dyskeratosis congenita (DC), a telomere biology disorder characterized by dysplastic nails, lacy reticular pigmentation of the upper chest and/or neck, and oral leukoplakia [120]. The loss of PARN reduces the levels of human Y RNAs. At the same time, low levels of Y RNAs intensify the effect of PARN depletion on telomere maintenance, leading to the same severe DC phenotype. PARN seems to be responsible to stabilize Y RNAs by removing the oligoadenylated tails that recruit DIS3L1 for degradation [121]. Moreover, it was recently demonstrated that PARN is also involved in miRNAs stabilization by removing their oligo(A) tails – the signal for the recruitment of the cytoplasmic exonucleases DIS3L1 or DIS3L2. Therefore, upon PARN knockdown there is decrease in miRNAs' levels, namely several that target p53 mRNA (a gene that plays a central role in cancer). This was the missing link to explain the p53 accumulation that is observed in PARN-defective patients [122].

DIS3L1 also seems to play a role in cancer, similarly to DIS3 (see above) and DIS3L2 (see bellow) homologues. The Hedgehog (Hh) pathway controls cell proliferation and differentiation in response to a gradient of secreted Hh ligands, and its aberrant activation can promote tumorigenesis. The transcriptional factor Zfx is a common cell-intrinsic regulator of diverse Hh-induced tumors. *hDIS3L1*, the human gene encoding DIS3L1 was identified as direct transcriptional target of Zfx, in the context of skin basal cell carcinoma (BCC) and cerebellar medulloblastoma (MB) models *in vivo* and *in vitro* [123].

Two independent exome sequencing studies (technique that sequences all protein-coding genes in a genome) have reported an association of *hDIS3L1* with cardiac risk. First, in a study to identify genetic variants that confer susceptibility to myocardial infarction (MI) in the Asian population (Korean individuals), several single nucleotide polymorphisms (SNPs) on *hDIS3L1* gene were associated with MI risk. However, how the gene influences MI pathogenesis would have to be determined and confirmed in other ethnic populations [124]. In another study, novel SNPs were also identified in the *hDIS3L1* gene of individuals with Hyperalphalipoproteinemia (HALP). This condition of high-density lipoprotein cholesterol (HDL) levels is inversely correlated with coronary heart disease (CHD), and *hDIS3L1* gene was identified as a candidate gene associated with HALP [125].

4.2.3 DIS3L2

Human DIS3L2 is the third member of the RNase II/RNB family of enzymes. This protein is a processive 3′-5′ exoribonuclease (mainly cytoplasmic) able to degrade structured RNA molecules, as long as they possess a 2 nt 3′ overhang as a "landing platform" [44, 45]. Unlike its family counterparts (DIS3 and DIS3L1), DIS3L2 lacks the PIN and the CR3 domains on its structure, both necessary for the interaction with the exosome complex [39, 43, 45, 72, 73].

In mammalian cells, DIS3L2 is involved in miRNA maturation and in the decay of numerous RNA-species, namely bulk mRNA, ARES and ncRNAs [126–129]. Several studies associated DIS3L2 with a degradation pathway that relies on the addition of untemplated uridines to several classes of RNAs, in a process called uridylation [44, 45, 127, 129–132]. This process was reported for the first time in *S. pombe* [133], in which Dis3L2 was shown to degrade uridylated poly(A)-containing mRNAs [45]. Later

on, uridylation was found to be widespread and to have a decisive impact on RNA′s fate. There is a negative correlation between the addition of short (1–4) uridine residues with mRNA stability [47, 130]. Oligo(U) tailed mRNAs are recognized by Lsm1–7 complex stimulating 3′-5′ degradation by DIS3L2, however it can also trigger decapping by DCP2 allowing 5′-3′ degradation by XRN1 [44, 45, 129, 130, 133–135]. DIS3L2 crystal structure unveiled the DIS3L2 RNA pathway, revealing three uracil-specific zones that explain how DIS3L2 recognizes, binds and processes preferentially oligo(U)-tailed RNAs [132].

The uridylation process is achieved by proteins termed uridyltransferases (TUTases). Humans have seven TUTases that are strictly cytoplasmic, except TUTase-1 (TUT1) that can also be found in mitochondria [133, 136]. Two TUTases were implicated in mRNA uridylation at the 3′ end, TUT4 and TUT7 [130]. In the same study, Lim and colleagues showed that these TUTases were able to sense the length of the poly(A) tail, and preferentially uridylate mRNAs with a tail ranging between 0–25 As. On the contrary, PABPs preferentially bind longer poly(A) tails protecting them from the action of TUT4/7 [130].

This so called TUT-DIS3L2 mRNA decay mechanism was found to prevail in cells under apoptosis. Apoptosis is the most common physiological program of cell death, which plays a vital role in pathogen immune defense, removal of damaged cells, cancer surveillance and cancer therapy effectiveness [137]. Thomas and colleagues [138] observed in human apoptotic cells, that apoptosis triggers global mRNA decay, and the RNA products generated are 3′-uridylated by TUT4/7 and subsequently degraded by DIS3L2. Knockdown of the exoribonuclease inhibits mRNA decay and suppresses cell death; conversely, DIS3L2 overexpression enhances apoptosis, supporting that mRNA decay is a hallmark of cell death [138]. Recently, Liu and colleagues [139], brought another player to this process, a mitochondrial exoribonuclease called PNPT1 that evolved from bacterial PNPase [48, 139, 140]. The work performed in human colon cancer cells, demonstrated that, upon apoptosis triggering, PNPT1 and DIS3L2 act in the same pathway. PNPT1 is released from mitochondria, and starts to degrade RNA from the 3′ end. PNPT1 stops whenever it encounters an obstacle (e.g. ribosome, RNA-binding protein or highly structured sequence), being the RNAs further degraded by the TUT-DIS3L2 pathway [139].

DIS3L2 is also involved in the regulation of let-7 miRNA expression in pluripotent cells, establishing a role of this enzyme in cell differentiation [128, 131, 132, 141]. Indeed, let-7 pre-miRNA biogenesis is one of the best characterized DIS3L2-mediated pathways. miRNAs from the let-7 family function as tumor suppressors and are involved in stem cell renewal [128, 131]. In undifferentiated cells, the expression of let-7 miRNAs is blocked by Lin28, a pluripotency factor that also functions as an oncogene in several cancers [142]. This RNA-binding protein binds to let-7 precursors and promotes their uridylation by TUT4/7. These RNA precursors are thus marked for DIS3L2 degradation, leading to inhibition of let-7 biogenesis.

A recent study has also found a role of DIS3L2 in nonsense-mediated decay (NMD), a quality control pathway that degrades aberrant and physiological mRNAs to maintain cellular homeostasis (as discussed in Chap. 3). In this context, DIS3L2 acts over 3′ ends of NMD decay intermediates that were previously subject to uridylation (L. Romão, personal communication and [143]).

The involvement of DIS3L2 in such cellular important processes, like apoptosis, cell differentiation and RNA quality control (NMD) anticipates its role in human disease. In fact, this RNase has been related with several human disorders. DIS3L2 is associated with Perlman syndrome, which is a rare congenital overgrowth disease [57, 144]. Children affected with Perlman syndrome display macrocephaly, facial abnormalities, neurodevelopmental delay, fetal gigantism, kidney abnormal enlargement and high neonatal mortality. These children also present nephroblastomatosis, an important precursor for Wilms' tumor, a kidney cancer also known as nephroblastoma. Astuti et al. [57] demonstrated that the affected children have germline mutations con-

sistent with DIS3L2 loss of function. DIS3L2 mutations were also associated with Wilms tumor susceptibility [57, 144]. It has been recently suggested that regulation of the growth-promoting gene, insulin growth factor 2 (IGF2), by DIS3L2, could be the link between this RNase and Wilms tumorigenesis [145]. Interestingly, Gregory RI and colleagues have found that DIS3L2 has no effect on the steady state mRNAs levels in DIS3L2-deficient cell lines and knockout mouse kidneys. Instead, it rather specifically perturbs endoplasmic reticulum (ER)-mediated translation (R.I. Gregory, personal communication).

Besides its well documented role in Perlman syndrome and Wilms' tumor, DIS3L2 was also found to be mutated in 3–6% of carcinomas [57, 146]. Also, DIS3L2 has been associated with a Marfan-like syndrome with skeletal overgrowth [147]. The Marfan syndrome is a disorder of the connective tissue that causes high mortality for untreated patients, mainly due to aortic complications [148]. Patients in which DIS3L2 gene was affected showed skeletal overgrowth and malformations, including severe scoliosis (abnormal curvature of the spine), arachnodactyly (long, slender fingers, curvature of the hands and feet) and mild syndactyly (interdigital webbing) [147, 149, 150].

4.3 Concluding Remarks

RNA degradation is a set of highly regulated steps that maintain cellular integrity and homeostasis. DIS3-enzymes act over a panoply of RNA substrates in eukaryotic cells and it is clear their role in human disease, namely in cancer development and progression. In this chapter, we explored the consequences of DIS3-enzymes impairment on the physiology of human cells. From this group of proteins, the most well-characterized is DIS3, however its role in cancer is not completely understood. Less is known about the mechanism of action and specific RNA targets of its homologs DIS3L1 and DIS3L2. The molecular mechanisms that link both proteins with disease are still unexplored. Despite the progress that has been made, there is still much work to perform in order to completely understand how DIS3-enzymes regulate cellular pathways, and how they are related with disease progression. Clinical medicine will certainly benefit from this kind of fundamental research.

Acknowledgements This work was supported by project LISBOA-01-0145-FEDER-007660 (Microbiologia Molecular, Estrutural e Celular) funded by FEDER funds through COMPETE2020 – Programa Operacional Competitividade e Internacionalização (POCI) and by national funds from FCT (Fundação para a Ciência e a Tecnologia); project PTDC/BIA-MIC/1399/2014 to CMA and project PTDC/BIM-MEC/3749/2014 to SCV. In addition, FCT provides postdoctoral grant ref. SFRH/BPD/109464/2015 to MS. SCV was financed by program IF of FCT (ref IF/00217/2015).

References

1. Archer SK et al (2015) Probing the closed-loop model of mRNA translation in living cells. RNA Biol 12(3):248–254
2. Wells SE et al (1998) Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell 2(1):135–140
3. Darnell JE Jr (2013) Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture. RNA 19(4):443–460
4. Kahvejian A et al (2005) Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. Genes Dev 19(1):104–113
5. Mangus DA, Evans MC, Jacobson A (2003) Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. Genome Biol 4(7):223
6. Barreau C, Paillard L, Osborne HB (2005) AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res 33(22):7138–7150
7. Chen CY, Shyu AB (1995) AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem Sci 20(11):465–470
8. Eulalio A et al (2009) Deadenylation is a widespread effect of miRNA regulation. RNA 15(1):21–32
9. Wu L, Fan J, Belasco JG (2006) MicroRNAs direct rapid deadenylation of mRNA. Proc Natl Acad Sci U S A 103(11):4034–4039
10. Abdelmohsen K, Gorospe M (2010) Posttranscriptional regulation of cancer traits by HuR. Wiley Interdiscip Rev RNA 1(2):214–229
11. Murray EL, Schoenberg DR (2007) A+U-rich instability elements differentially activate 5′-3′ and 3′-5′ mRNA decay. Mol Cell Biol 27(8):2791–2799

12. Peng SS et al (1998) RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. EMBO J 17(12):3461–3470
13. Huntzinger E, Izaurralde E (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet 12(2):99–110
14. Oliveto S et al (2017) Role of microRNAs in translation regulation and cancer. World J Biol Chem 8(1):45–56
15. Valinezhad Orang A, Safaralizadeh R, Kazemzadeh-Bavili M (2014) Mechanisms of miRNA-mediated gene regulation from common downregulation to mRNA-specific upregulation. Int J Genomics 2014:970607
16. Eberle AB et al (2009) SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. Nat Struct Mol Biol 16(1):49–55
17. Huntzinger E et al (2008) SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. RNA 14(12):2609–2617
18. Pavlopoulou A et al (2013) A comprehensive phylogenetic analysis of deadenylases. Evol Bioinformatics Online 9:491–497
19. Uchida N, Hoshino S, Katada T (2004) Identification of a human cytoplasmic poly(A) nuclease complex stimulated by poly(A)-binding protein. J Biol Chem 279(2):1383–1391
20. Wahle E, Winkler GS (2013) RNA decay machines: deadenylation by the Ccr4-not and Pan2-Pan3 complexes. Biochim Biophys Acta 1829(6–7):561–570
21. Wolf J, Passmore LA (2014) mRNA deadenylation by Pan2-Pan3. Biochem Soc Trans 42(1):184–187
22. Chen CY, Shyu AB (2011) Mechanisms of deadenylation-dependent decay. Wiley Interdiscip Rev RNA 2(2):167–183
23. Doidge R et al (2012) Deadenylation of cytoplasmic mRNA by the mammalian Ccr4-Not complex. Biochem Soc Trans 40(4):896–901
24. Lau NC et al (2009) Human Ccr4-Not complexes contain variable deadenylase subunits. Biochem J 422(3):443–453
25. Temme C, Simonelig M, Wahle E (2014) Deadenylation of mRNA by the CCR4-NOT complex in Drosophila: molecular and developmental aspects. Front Genet 5:143
26. Yamashita A et al (2005) Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. Nat Struct Mol Biol 12(12):1054–1063
27. Coller J, Parker R (2004) Eukaryotic mRNA decapping. Annu Rev Biochem 73:861–890
28. Franks TM, Lykke-Andersen J (2008) The control of mRNA decapping and P-body formation. Mol Cell 32(5):605–615
29. Li Y, Kiledjian M (2010) Regulation of mRNA decapping. Wiley Interdiscip Rev RNA 1(2):253–265
30. Sharif H, Conti E (2013) Architecture of the Lsm1-7-Pat1 complex: a conserved assembly in eukaryotic mRNA turnover. Cell Rep 5(2):283–291
31. Tharun S (2009) Lsm1-7-Pat1 complex: a link between 3′ and 5′-ends in mRNA decay? RNA Biol 6(3):228–232
32. Braun JE et al (2012) A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5′ exonucleolytic degradation. Nat Struct Mol Biol 19(12):1324–1331
33. Nissan T et al (2010) Decapping activators in Saccharomyces cerevisiae act by multiple mechanisms. Mol Cell 39(5):773–783
34. Halbach F et al (2013) The yeast ski complex: crystal structure and RNA channeling to the exosome complex. Cell 154(4):814–826
35. Chen N et al (2005) Crystal structures of human DcpS in ligand-free and m7GDP-bound forms suggest a dynamic mechanism for scavenger mRNA decapping. J Mol Biol 347(4):707–718
36. Liu H et al (2002) The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. EMBO J 21(17):4699–4708
37. Milac AL, Bojarska E, Wypijewska del Nogal A (2014) Decapping Scavenger (DcpS) enzyme: advances in its structure, activity and roles in the cap-dependent mRNA metabolism. Biochim Biophys Acta 1839(6):452–462
38. Shen V et al (2008) DcpS scavenger decapping enzyme can modulate pre-mRNA splicing. RNA 14(6):1132–1142
39. Bonneau F et al (2009) The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. Cell 139(3):547–559
40. Lebreton A et al (2008) Endonucleolytic RNA cleavage by a eukaryotic exosome. Nature 456(7224):993–996
41. Mamolen M, Smith A, Andrulis ED (2010) Drosophila melanogaster Dis3 N-terminal domains are required for ribonuclease activities, nuclear localization and exosome interactions. Nucleic Acids Res 38(16):5507–5517
42. Schaeffer D et al (2009) The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. Nat Struct Mol Biol 16(1):56–62
43. Schneider C et al (2009) The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. Nucleic Acids Res 37(4):1127–1140
44. Lubas M et al (2013) Exonuclease hDIS3L2 specifies an exosome-independent 3′-5′ degradation pathway of human cytoplasmic mRNA. EMBO J 32(13):1855–1868
45. Malecki M et al (2013) The exoribonuclease Dis3L2 defines a novel eukaryotic RNA degradation pathway. EMBO J 32(13):1842–1854
46. Arribas-Layton M et al (2013) Structural and functional control of the eukaryotic mRNA decapping machinery. Biochim Biophys Acta 1829(6–7):580–589
47. Chang H et al (2014) TAIL-seq: genome-wide determination of poly(A) tail length and 3′ end modifications. Mol Cell 53(6):1044–1052

48. Arraiano CM et al (2010) The critical role of RNA processing and degradation in the control of gene expression. FEMS Microbiol Rev 34(5):883–923
49. Barbas A et al (2008) New insights into the mechanism of RNA degradation by ribonuclease II: identification of the residue responsible for setting the RNase II end product. J Biol Chem 283(19):13070–13076
50. Barbas A et al (2009) Determination of key residues for catalysis and RNA cleavage specificity: one mutation turns RNase II into a "SUPER-ENZYME". J Biol Chem 284(31):20486–20498
51. Frazao C et al (2006) Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. Nature 443(7107):110–114
52. Robinson SR et al (2018) DIS3 isoforms vary in their endoribonuclease activity and are differentially expressed within haematological cancers. Biochem J 475(12):2091–2105
53. Viegas SC et al (2015) Surprises in the 3′-end: 'U' can decide too! FEBS J 282(18):3489–3499
54. Matos RG et al (2014) The importance of proteins of the RNase II/RNB-family in pathogenic bacteria. Front Cell Infect Microbiol 4:68
55. Matos RG et al (2012) The rnb gene of Synechocystis PCC6803 encodes a RNA hydrolase displaying RNase II and not RNase R enzymatic properties. PLoS One 7(3):e32690
56. Matos RG et al (2011) Swapping the domains of exoribonucleases RNase II and RNase R: conferring upon RNase II the ability to degrade ds RNA. Proteins 79(6):1853–1867
57. Astuti D et al (2012) Germline mutations in DIS3L2 cause the Perlman syndrome of overgrowth and Wilms tumor susceptibility. Nat Genet 44(3):277–284
58. Chapman MA et al (2011) Initial genome sequencing and analysis of multiple myeloma. Nature 471(7339):467–472
59. Ding L et al (2012) Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature 481(7382):506–510
60. Reis FP et al (2013) The RNase II/RNB family of exoribonucleases: putting the 'Dis' in disease. Wiley Interdiscip Rev RNA 4(5):607–615
61. Dziembowski A et al (2007) A single subunit, Dis3, is essentially responsible for yeast exosome core activity. Nat Struct Mol Biol 14(1):15–22
62. Wasmuth EV, Lima CD (2012) Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel. Mol Cell 48(1):133–144
63. Liu Q, Greimann JC, Lima CD (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell 127(6):1223–1237
64. Makino DL, Baumgartner M, Conti E (2013) Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. Nature 495(7439):70–75
65. Wasmuth EV, Januszyk K, Lima CD (2014) Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. Nature 511(7510):435–439
66. Allmang C et al (1999) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J 18(19):5399–5410
67. Bousquet-Antonelli C, Presutti C, Tollervey D (2000) Identification of a regulated pathway for nuclear pre-mRNA turnover. Cell 102(6):765–775
68. Chen CY et al (2001) AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. Cell 107(4):451–464
69. Milligan L et al (2005) A nuclear surveillance pathway for mRNAs with defective polyadenylation. Mol Cell Biol 25(22):9996–10004
70. Mitchell P et al (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3′-->5′ exoribonucleases. Cell 91(4):457–466
71. Mukherjee D et al (2002) The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. EMBO J 21(1–2):165–174
72. Tomecki R et al (2010) The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. EMBO J 29(14):2342–2357
73. Schaeffer D et al (2012) The CR3 motif of Rrp44p is important for interaction with the core exosome and exosome function. Nucleic Acids Res 40(18):9298–9307
74. Schneider C, Anderson JT, Tollervey D (2007) The exosome subunit Rrp44 plays a direct role in RNA substrate recognition. Mol Cell 27(2):324–331
75. Wang HW et al (2007) Architecture of the yeast Rrp44 exosome complex suggests routes of RNA recruitment for 3′ end processing. Proc Natl Acad Sci U S A 104(43):16844–16849
76. Lorentzen E et al (2008) Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. Mol Cell 29(6):717–728
77. Amblar M et al (2006) Characterization of the functional domains of Escherichia coli RNase II. J Mol Biol 360(5):921–933
78. Amblar M et al (2007) The role of the S1 domain in exoribonucleolytic activity: substrate specificity and multimerization. RNA 13(3):317–327
79. Lee G et al (2012) Elastic coupling between RNA degradation and unwinding by an exoribonuclease. Science 336(6089):1726–1729
80. Matelska D, Steczkiewicz K, Ginalski K (2017) Comprehensive classification of the PIN domain-like superfamily. Nucleic Acids Res 45(12):6995–7020
81. Senissar M, Manav MC, Brodersen DE (2017) Structural conservation of the PIN domain active site across all domains of life. Protein Sci 26(8):1474–1492
82. Drazkowska K et al (2013) The RNA exosome complex central channel controls both exonuclease and endonuclease Dis3 activities in vivo and in vitro. Nucleic Acids Res 41(6):3845–3858
83. Schneider C et al (2012) Transcriptome-wide analysis of exosome targets. Mol Cell 48(3):422–433

84. Malet H et al (2010) RNA channelling by the eukaryotic exosome. EMBO Rep 11(12):936–942
85. Tomecki R et al (2014) Multiple myeloma-associated hDIS3 mutations cause perturbations in cellular RNA metabolism and suggest hDIS3 PIN domain as a potential drug target. Nucleic Acids Res 42(2):1270–1290
86. Hou D, Ruiz M, Andrulis ED (2012) The ribonuclease Dis3 is an essential regulator of the developmental transcriptome. BMC Genomics 13:359
87. Szczepinska T et al (2015) DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts. Genome Res 25(11):1622–1633
88. Morton DJ et al (2018) The RNA exosome and RNA exosome-linked disease. RNA 24(2):127–142
89. Robinson SR et al (2015) The 3′ to 5′ exoribonuclease DIS3: from structure and mechanisms to biological functions and role in human disease. Biomol Ther 5(3):1515–1539
90. Lionetti M et al (2015) A compendium of DIS3 mutations and associated transcriptional signatures in plasma cell dyscrasias. Oncotarget 6(28):26129–26141
91. Lohr JG et al (2014) Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. Cancer Cell 25(1):91–101
92. Walker BA et al (2012) Intraclonal heterogeneity and distinct molecular mechanisms characterize the development of t(4;14) and t(11;14) myeloma. Blood 120(5):1077–1086
93. Weissbach S et al (2015) The molecular spectrum and clinical impact of DIS3 mutations in multiple myeloma. Br J Haematol 169(1):57–70
94. Laubach J, Richardson P, Anderson K (2011) Multiple myeloma. Annu Rev Med 62:249–264
95. Preker P et al (2008) RNA exosome depletion reveals transcription upstream of active human promoters. Science 322(5909):1851–1854
96. Wang Y et al (2015) A novel long non-coding RNA, hypoxia-inducible factor-2alpha promoter upstream transcript, functions as an inhibitor of osteosarcoma stem cells in vitro. Mol Med Rep 11(4):2534–2540
97. Ng D et al (2007) Identification of a novel chromosome region, 13q21.33-q22.2, for susceptibility genes in familial chronic lymphocytic leukemia. Blood 109(3):916–925
98. Pils D et al (2013) A combined blood based gene expression and plasma protein abundance signature for diagnosis of epithelial ovarian cancer--a study of the OVCAD consortium. BMC Cancer 13:178
99. Lim J et al (1997) Isolation of murine and human homologues of the fission-yeast dis3+ gene encoding a mitotic-control protein and its overexpression in cancer cells with progressive phenotype. Cancer Res 57(5):921–925
100. Kinoshita N, Goebl M, Yanagida M (1991) The fission yeast dis3+ gene encodes a 110-kDa essential protein implicated in mitotic control. Mol Cell Biol 11(12):5839–5847
101. Ohkura H et al (1988) Cold-sensitive and caffeine-supersensitive mutants of the Schizosaccharomyces pombe dis genes implicated in sister chromatid separation during mitosis. EMBO J 7(5):1465–1473
102. Towler BP et al (2015) The 3′-5′ exoribonuclease Dis3 regulates the expression of specific microRNAs in Drosophila wing imaginal discs. RNA Biol 12(7):728–741
103. Camps J et al (2013) Genetic amplification of the NOTCH modulator LNX2 upregulates the WNT/beta-catenin pathway in colorectal cancer. Cancer Res 73(6):2003–2013
104. de Groen FL et al (2014) Gene-dosage dependent overexpression at the 13q amplicon identifies DIS3 as candidate oncogene in colorectal cancer progression. Genes Chromosom Cancer 53(4):339–348
105. Rose AE et al (2011) Integrative genomics identifies molecular alterations that challenge the linear model of melanoma progression. Cancer Res 71(7):2561–2571
106. Liu W, Peng Y, Tobin DJ (2013) A new 12-gene diagnostic biomarker signature of melanoma revealed by integrated microarray analysis. PeerJ 1:e49
107. Rozenblum E et al (2002) A genomic map of a 6-Mb region at 13q21-q22 implicated in cancer development: identification and characterization of candidate genes. Hum Genet 110(2):111–121
108. Hoskins JW et al (2016) Functional characterization of a chr13q22.1 pancreatic cancer risk locus reveals long-range interaction and allele-specific effects on DIS3 expression. Hum Mol Genet 25(21):4726–4738
109. Shen L, Shi Q, Wang W (2018) Double agents: genes with both oncogenic and tumor-suppressor functions. Oncogene 7(3):25
110. Segalla S et al (2015) The ribonuclease DIS3 promotes let-7 miRNA maturation by degrading the pluripotency factor LIN28B mRNA. Nucleic Acids Res 43(10):5182–5193
111. Noguchi E et al (1996) Dis3, implicated in mitotic control, binds directly to ran and enhances the GEF activity of RCC1. EMBO J 15(20):5595–5605
112. Glavan F et al (2006) Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. EMBO J 25(21):5117–5125
113. Bleichert F et al (2006) The PINc domain protein Utp24, a putative nuclease, is required for the early cleavage steps in 18S rRNA maturation. Proc Natl Acad Sci U S A 103(25):9464–9469
114. Staals RH et al (2010) Dis3-like 1: a novel exoribonuclease associated with the human exosome. EMBO J 29(14):2358–2367
115. Slomovic S et al (2010) Addition of poly(A) and poly(A)-rich tails during RNA degradation in the cytoplasm of human cells. Proc Natl Acad Sci U S A 107(16):7407–7412
116. Slomovic S et al (2006) Polyadenylation of ribosomal RNA in human cells. Nucleic Acids Res 34(10):2966–2975

117. Lima WF et al (2016) RNA cleavage products generated by antisense oligonucleotides and siRNAs are processed by the RNA surveillance machinery. Nucleic Acids Res 44(7):3351–3363
118. Kohn M et al (2015) The Y3∗∗ ncRNA promotes the 3′ end processing of histone mRNAs. Genes Dev 29(19):1998–2003
119. Kowalski MP, Krude T (2015) Functional roles of non-coding Y RNAs. Int J Biochem Cell Biol 66:20–29
120. Savage SA (1993) Dyskeratosis congenita. In: Adam MP et al (eds) GeneReviews((R)). University of Washington, Seattle
121. Shukla S, Parker R (2017) PARN modulates Y RNA stability and its 3′-end formation. Mol Cell Biol 37(20):e00264
122. Shukla S et al (2019) The RNase PARN controls the levels of specific miRNAs that contribute to p53 regulation. Mol Cell 73:1204
123. Palmer CJ et al (2014) Zfx facilitates tumorigenesis caused by activation of the hedgehog pathway. Cancer Res 74(20):5914–5924
124. Lee JY et al (2017) Genome-based exome sequencing analysis identifies GYG1, DIS3L and DDRGK1 are associated with myocardial infarction in Koreans. J Genet 96(6):1041–1046
125. Oates CP et al (2018) Novel polymorphisms associated with hyperalphalipoproteinemia and apparent cardioprotection. J Clin Lipidol 12(1):110–115
126. Labno A et al (2016) Perlman syndrome nuclease DIS3L2 controls cytoplasmic non-coding RNAs and provides surveillance pathway for maturing snRNAs. Nucleic Acids Res 44(21):10437–10453
127. Pirouz M et al (2016) Dis3l2-mediated decay is a quality control pathway for noncoding RNAs. Cell Rep 16(7):1861–1873
128. Ustianenko D et al (2013) Mammalian DIS3L2 exoribonuclease targets the uridylated precursors of let-7 miRNAs. RNA 19(12):1632–1638
129. Ustianenko D et al (2016) TUT-DIS3L2 is a mammalian surveillance pathway for aberrant structured non-coding RNAs. EMBO J 35(20):2179–2191
130. Lim J et al (2014) Uridylation by TUT4 and TUT7 marks mRNA for degradation. Cell 159(6):1365–1376
131. Chang HM et al (2013) A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. Nature 497(7448):244–248
132. Faehnle CR, Walleshauser J, Joshua-Tor L (2014) Mechanism of Dis3l2 substrate recognition in the Lin28-let-7 pathway. Nature 514(7521):252–256
133. Rissland OS, Mikulasova A, Norbury CJ (2007) Efficient RNA polyuridylation by noncanonical poly(A) polymerases. Mol Cell Biol 27(10):3612–3624
134. Mullen TE, Marzluff WF (2008) Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5′ to 3′ and 3′ to 5′. Genes Dev 22(1):50–65
135. Song MG, Kiledjian M (2007) 3′ terminal oligo U-tract-mediated stimulation of decapping. RNA 13(12):2356–2365
136. Tomecki R et al (2004) Identification of a novel human nuclear-encoded mitochondrial poly(A) polymerase. Nucleic Acids Res 32(20):6001–6014
137. Shimizu S et al (2014) Autophagic cell death and cancer. Int J Mol Sci 15(2):3145–3153
138. Thomas MP et al (2015) Apoptosis triggers specific, rapid, and global mRNA decay with 3′ uridylated intermediates degraded by DIS3L2. Cell Rep 11(7):1079–1089
139. Liu X et al (2018) PNPT1 release from mitochondria during apoptosis triggers decay of Poly(A) RNAs. Cell 174(1):187–201.e12
140. Briani F, Carzaniga T, Deho G (2016) Regulation and functions of bacterial PNPase. Wiley Interdiscip Rev RNA 7(2):241–258
141. Bussing I, Slack FJ, Grosshans H (2008) let-7 microRNAs in development, stem cells and cancer. Trends Mol Med 14(9):400–409
142. Thornton JE, Gregory RI (2012) How does Lin28 let-7 control development and disease? Trends Cell Biol 22(9):474–482
143. Kurosaki T et al (2018) NMD-degradome sequencing reveals ribosome-bound intermediates with 3′-end non-templated nucleotides. Nat Struct Mol Biol 25(10):940–950
144. Schilke K et al (2000) A case of Perlman syndrome: fetal gigantism, renal dysplasia, and severe neurological deficits. Am J Med Genet 91(1):29–33
145. Hunter RW et al (2018) Loss of Dis3l2 partially phenocopies Perlman syndrome in mice and results in up-regulation of Igf2 in nephron progenitor cells. Genes Dev 32(13–14):903–908
146. Morris MR, Astuti D, Maher ER (2013) Perlman syndrome: overgrowth, Wilms tumor predisposition and DIS3L2. Am J Med Genet C: Semin Med Genet 163C(2):106–113
147. Tassano E et al (2013) Genotype-phenotype correlation of 2q37 deletions including NPPC gene associated with skeletal malformations. PLoS One 8(6):e66048
148. von Kodolitsch Y et al (2010) Marfan syndrome and the evolving spectrum of heritable thoracic aortic disease: do we need genetics for clinical decisions? Vasa 39(1):17–32
149. Bocciardi R et al (2007) Overexpression of the C-type natriuretic peptide (CNP) is associated with overgrowth and bone anomalies in an individual with balanced t(2;7) translocation. Hum Mutat 28(7):724–731
150. Moncla A et al (2007) A cluster of translocation breakpoints in 2q37 is associated with overexpression of NPPC in patients with a similar overgrowth phenotype. Hum Mutat 28(12):1183–1188

5 Translational Regulation by Upstream Open Reading Frames and Human Diseases

Joana Silva, Rafael Fernandes, and Luísa Romão

Abstract

Short upstream open reading frames (uORFs) are *cis*-acting elements located within the 5′-leader sequence of transcripts and are defined by an initiation codon in-frame with a termination codon located upstream or downstream of its main ORF (mORF) initiation codon. Recent genome-wide ribosome profiling studies have confirmed the widespread presence of uORFs and have shown that many uORFs can initiate with non-AUG codons. uORFs can impact gene expression of the downstream mORF by triggering mRNA decay or by regulating translation. Thus, disruption or creation of uORFs can elicit the development of several genetic diseases. Here, we review the mechanisms by which AUG- and non-AUG uORFs regulate translation. We also show some examples of uORF deregulation in human genetic diseases, focusing mainly on cancer. The knowledge of how uORF deregulation drives the onset of a disease, points out the need to screen the 5′-leader sequences of the transcripts in search for potential disease-related variants. This information will be relevant for the implementation of new diagnostic and/or therapeutic tools.

Keywords

Genetic disease · Non-AUG upstream open reading frame (uORF) · Stress · Translational regulation · Translatome · uORF · uORF-encoded peptide

5.1 Introduction

Gene expression is largely modulated at the level of mRNA translation, which is itself divided into initiation, elongation, termination and ribosome recycling steps [1–4]. Translation initiation is the rate-limiting step of translation and is tightly controlled by mechanisms that involve different regulatory elements in the 5′-leader sequence of the transcripts [2]. Among them are: (i) internal ribosome entry sites (IRES), which are highly structured RNA regions that recruit ribosomes to or near the translation initiation codon and thus induce translation in a cap-independent manner; (ii) RNA structures (hairpins, G-quadruplexes and pseudoknots) that impair start codon scanning by the ribosome and repress translation initiation; (iii) protein binding sites where different

J. Silva · R. Fernandes · L. Romão (✉)
Department of Human Genetics,
National Institute of Health Doutor Ricardo Jorge,
Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon,
Lisboa, Portugal
e-mail: luisa.romao@insa.min-saude.pt

L. Romão (ed.), *The mRNA Metabolism in Human Disease*, Advances in Experimental Medicine and Biology 1157, https://doi.org/10.1007/978-3-030-19966-1_5

molecular ligands can interact or form stable ribonucleoprotein complexes, thus promoting or repressing translation; (iv) RNA modifications that unfold RNA and are usually associated with an efficient translation; and (v) upstream AUG or non-AUG open reading frames (uORFs) that usually inhibit the downstream translation initiation at the main ORF (mORF) [2, 4–6]. A uORF consists of a well-studied class of small ORFs potentially translated, with its initiation codon within the 5′-leader sequence of a mRNA and its in-frame termination codon upstream or overlapped with the mORF, and ranging typically from two to one hundred codons [2, 7–9]. It is bioinformatically estimated that approximately 49–58% of human transcripts carry at least one uORF [8, 10]. Indeed, uORFs are conspicuous in certain classes of genes, such as transcription factors, cellular receptors, oncogenes and genes involved in cell growth and differentiation control [10–14]. These regulatory elements are usually seen to be evolutionarily conserved, which suggest an important biological function [8, 15].

Under physiological conditions, uORFs are typically described as repressors of translation initiation at the downstream mORF [2, 8, 16]. This is explained by the process of start codon recognition during translation initiation: the 43S pre-initiation complex (PIC) composed of eIF3, eIF1, eIF1A, eIF5, the eIF2·GTP·$tRNA_i^{Met}$ ternary complex, and the small 40S ribosomal subunit, binds to the 5′-cap structure of the mRNA and scans from 5′ to 3′ until it reaches an initiation codon [2, 4, 10]. Then, eIF2 ternary complex and other initiation factors dissociate, which reduces the levels of mORF expression in about 30–80% [8, 15, 17]. However, in response to cellular stress, the presence of uORFs can increase the expression of certain mRNAs, a mechanism used by stress-responsive transcripts to alleviate the cell from stress, and also by oncogenes during the tumor initiating process [1, 18].

With the advent of ribosome profiling (RiboSeq), a highly precise technique for monitoring *in vivo* translation based in RNA deep-sequencing of mRNA fragments covered with ribosomes [19], a significant ribosomal occupancy (signature of active translation) was shown in regions thought to be non-coding, as the case of the 5′-leader sequences, which is consistent with the widespread presence of translatable uORFs among the transcriptome [8, 9, 19]. It was recently shown that, among the uORFs identified by RiboSeq, uORFs containing near-cognate start codons are more frequent compared to AUG-containing uORFs, and their recognition involves alternative translation initiation mechanisms. The most prevalent non-AUG start codon is CUG [9, 18]. Non-AUG uORFs are important for the regulation of protein synthesis of specific transcripts, including several with a relevant role in stress responses, and also oncogenes [9, 18, 20–22]. Given the wide presence and the regulatory function of uORFs, it is not surprising that their deregulation, for instance by mutations that create, delete or modify a uORF, may play a role in the onset of several diseases, including the development and/or predisposition to cancer [2, 11–13].

Here, we intend to highlight the mechanisms that drive uORF-mediated translational regulation during both physiological and stress conditions. In light of the new contributions given by RiboSeq analyses in the widespread detection of translatable uORFs, especially demonstrating the prevalent translation of non-AUG uORFs, we dissect the mechanisms that regulate their recognition, as well as the mechanism by which non-AUG uORFs control their mORF expression. Moreover, we describe examples of the pathophysiological impact of uORF deregulation in human genetic diseases, giving emphasis to cancer. The implication of uORF creation, deletion or deregulation in the onset of several human diseases, highlights the need for the application of high-throughput technologies to systematically search disease-associated variations in 5′-leader sequences.

5.2 uORFs as Translational Regulators

The cap-dependent translation initiation process is described as a scanning mechanism [reviewed in 4]. Briefly, as mentioned above, the pre-formed

43S PIC attaches to the 7-methylguanosine (m^7GTP) cap structure at the 5′ end of the mRNA (5′ cap), forming the 48S preinitiation complex [4, 23]. This process is mediated by the eIF4F complex composed by eIF4E, eIF4G and eIF4A, where eIF4E establishes a link with the mRNA 5′ cap, eIF4G recruits PIC by interaction with eIF3, and eIF4A unwinds the mRNA secondary structure to allow ribosomal scanning [4, 23, 24]. PIC scans the 5′-leader sequence from 5′ to 3′ until it reaches an initiation codon (usually an AUG) at the peptidyl (P) decoding site of the ribosome. At this point, eIF5 mediates hydrolysis of eIF2-GTP, which promotes $tRNA_i^{Met}$ anticodon base-pairing with the start codon, the release of eIF2-GDP along with other eIFs, and the binding of the 60S large ribosomal subunit by eIF5B to form the 80S translating ribosome [4, 23, 25, 26].

One of the factors that can influence AUG recognition is its surrounding context, defined as the Kozak consensus sequence [27, 28]. The optimal sequence surrounding the AUG initiation codon was postulated to be (GCC)A/GCCAUGG, where the positions −3 and +4 relative to the A of the AUG, are the most important ones [28]. Additionally, the nucleotide (nt) G at the position −6, as well as C at positions −1 and −2, seem to improve translation efficiency [28, 29]. The G purines at positions −3 and +4 establish a link with the α subunit of eIF2 (eIF2α) and the 18S ribosomal RNA (rRNA), respectively. However, from these two interactions the most critical is the association between eIF2α and the purine at the position −3, which seems to stabilize the 48S preinitiation complex to evade dissociation by eIF1 [30]. The Kozak sequence context is postulated not only for the main initiation codons but also for the uORF initiation codons. The uAUG of a uORF in a strong or adequate Kozak context can be recognized and translation initiation will occur, which might have a negative impact on the main AUG (mAUG) [28]. In fact, the inhibitory activity of a uORF can be associated with a strong upstream initiation codon context. Other features may also determine the repression activity of a uORF, such as a long distance from the 5′ end of the transcript to the beginning of the uORF, the existence of multiple uORFs (additive effect), the consumption of active pre-initiation complexes, a long uORF and/or a short intercistronic distance (distance from the uORF stop codon to the mORF start codon) [8, 10, 16, 31, 32].

Translational repression can be achieved by ribosome dissociation and recycling after uORF translation or by ribosome stalling of the elongating/terminating ribosome during the process of uORF translation (Fig. 5.1a) [1, 2]. While ribosome dissociation can be regulated either by the nucleotide sequence or the resulting uORF-encoded peptide [33, 34], ribosome stalling is associated with unique features in the 5′-leader sequence, such as the presence of secondary structures, the interaction with *trans*-acting factors, the nascent peptide sequence, or codon usage (rare *versus* common codons) [16, 31, 35–37]. Ribosomal stalling at the uORF termination codon can also trigger nonsense-mediated mRNA decay (NMD), since the uORF stop codon can be recognized as a premature termination codon (PTC) (Fig. 5.1a) [38–40]. As an example, an inhibitory overlapping uORF in human and yeast *STN1* mRNAs was recently described [41]. In yeast, this overlapping uORF targets *STN1* to NMD, maintaining the low abundance of STN1 responsible for the normal activity of telomeres [41].

In addition to their role as major translational repressors, uORFs can also work as mediators of mORF expression, for instance in stress conditions. During a stress stimulus (oxidative and endoplasmic reticulum (ER) stress, hypoxia, nutrient deprivation, UV radiation, among others) the cell tends to respond rapidly by reprogramming its gene expression pattern at the level of protein synthesis [14, 42]. Depending on the stress, four different serine-threonine kinases can be activated, phosphorylating eIF2: PKR-like ER kinase, PERK; protein kinase double-stranded RNA-dependent, PKR; general control non-derepressible-2, GCN2; and heme-regulated inhibitor, HRI [42]. Phosphorylation of eIF2 at serine 51 on its α subunit (eIF2α-P) impairs GDP to GTP exchange by guanine nucleotide exchange factor eIF2B, necessary to obtain a new active form of eIF2α. Therefore, the ternary complex is not formed, which impairs translation initiation

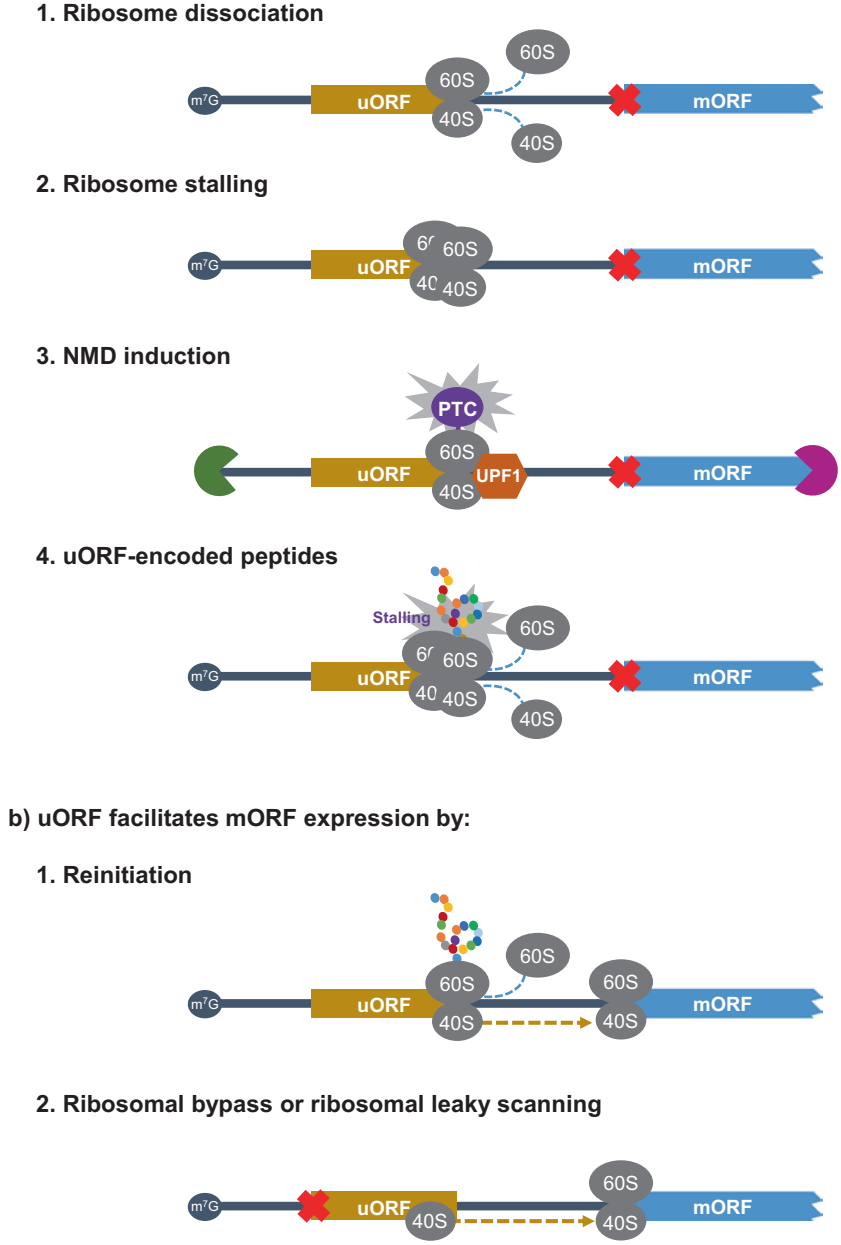

Fig. 5.1 Mechanisms of upstream open reading frame (uORF)-mediated translational regulation. **(a)** uORF translation can mediate main ORF (mORF) repression by different mechanisms: (1) ribosome dissociation with consequent ribosome recycling; (2) ribosome stalling, in which the elongating/terminating ribosomes are blocked usually due to (i) the presence of secondary structures in the uORF, (ii) the uORF nucleotide context, or (iii) the interaction with *trans*-acting factors; (3) nonsense-mediated decay (NMD) induction, due to stalling of the ribosomes at the uORF termination codon that is recognized as a premature termination codon (PTC); and (4) uORF-encoded peptides that depending on their amino acids sequence and their interaction with the translational machinery, can induce ribosome stalling and dissociation. **(b)** uORF facilitates mORF translation by: (1) reinitiation, where the uORF is translated and the 40S ribosomal subunit remains attached, resumes scanning and reinitiates translation in a downstream ORF; and (2) ribossomal bypass or ribosomal leaky scanning, where the scanning ribosomes pass through the uORF initiation codon without recognizing it and initiate translation further downstream, a mechanism mainly associated with the Kozak consensus sequence of the uORF initiation codon

[43, 44]. The abundance of eIF2α-P determines global and gene-specific translation rates [34]. This gene expression reprogramming in conditions of global translation repression, aims to restore cell homeostasis and cell survival by activation of the integrated stress response (ISR). If the stress is too long or severe, cells can be committed to programmed cell death (apoptosis) [42, 45, 46]. Therefore, stressed cells repress global translation to avoid unnecessary protein synthesis, but at the same time, allow translation of specific proteins aiming to solve the stress [1, 14]. Many of these proteins are encoded by uORFs-containing transcripts [4, 14, 34, 39]. However, the existence of a uORF is not a straight indication that the corresponding transcripts will be translated in stress conditions [34]. There are several factors that can influence the balance between translational repression and promotion by a uORF: (i) its length; (ii) the intercistronic distance; (iii) the distance from the 5′ end of the 5′-leader sequence to the uORF start codon; (iv) its start codon context; (v) its secondary structure; (vi) number of uORFs; (vii) the type of initiation codon, and (viii) its stop codon context [2, 8, 33]. These features can operate alone or in combination to define the uORF-mediated mechanism of translational control orchestrated for a specific mRNA under physiological or stress conditions [1, 34].

5.2.1 Translation Reinitiation

As mentioned above, if a scanning ribosome recognizes a uORF initiation codon, it can translate the uORF and dissociate or stall during either the elongation or the termination phase of translation. However, another option exists for the ribosome: it can translate the uORF and remain associated with the mRNA, continue scanning, and reinitiate translation further downstream (Fig. 5.1b). The mechanism of translation reinitiation implies that after uORF translation termination, the ribosomal 60S subunit is released from the mRNA while the 40S subunit remains attached, and resumes scanning until a new competent ribosome is assembled for translation initiation at a downstream initiation codon [1, 3, 23]. One major factor influencing reinitiation is the availability of the ternary complex, more precisely the eIF2-GTP in this complex, which is hydrolyzed and released after uORF start codon recognition and ribosomal assembly. The abundance of the ternary complex can be linked to the distance scanned by the 40S ribosomal subunit until it reaches a new initiation codon [47]. In this case, short uORFs and long intercistronic distances are positively related to an efficient translation reinitiation at a downstream ORF [1, 13, 48]. On the other hand, a decrease in the reinitiation rate, for instance, by longer uORFs and/or the presence of secondary structures that promote ribosome stalling, can be explained by the loss of several eIFs along with the ternary complex after uORF translation is completed [3, 31, 37]. In fact, some of the eIFs are transiently maintained associated with elongating and terminating ribosomes, which in turn are the ones that will resume scanning and reinitiate translation downstream [49]. It seems that if the 40S-eIF3-eIF4F complex, which is responsible for the engagement of the 40S ribosome to the uORF initiation codon, is kept together until termination of uORF translation, it will allow the 40S subunit to resume scanning and initiate translation at a downstream initiation codon [37]. Disruption of this complex before uORF translation is completed, destabilizes the interaction between the ribosome and the mRNA, leading to its dissociation after uORF translation termination, which in the end prevents reinitiation at the downstream mORF [37].

The mammalian activating transcription factor 4 (*ATF4*) and its yeast homolog, the general control protein (*GCN4*), are well described examples of transcripts translated via a mechanism of 'delayed' reinitiation. Both are transcription factors that, when induced, activate several pathways of the ISR aiming to release the cell from stress [47, 50]. *ATF4* 5′-leader sequence bears two uORFs: a 3 aminoacid (aa)-long first uORF (uORF1) that is 5′ proximal, and a lengthened, out-of-frame and coding sequence-overlapping uORF2 of 59 aa, both with an AUG in a strong Kozak consensus context. During basal conditions, the short uORF1 is translated, allowing the

40S ribosomal subunit to maintain attached to the mRNA, which resumes scanning until it acquires a new ternary complex (eIF2·GTP·$tRNA_i^{Met}$) in time to reinitiate translation at the uORF2 start codon. This last event leads to a translational repression of the mORF, since the ribosome terminates uORF2 translation at 3′ of the mORF initiation codon [50]. During stress, phosphorylation of eIF2α reduces the cellular levels of eIF2-GTP, which is a limiting step for the formation of more ternary complexes [1, 47]. In these conditions, after *ATF4* uORF1 translation, there is a delay in the reload of a new ternary complex to the 40S ribosomal subunit, important to initiate translation at the uORF2 initiation codon. Therefore, the ribosome bypasses the inhibitory uORF2 and reinitiates translation at *ATF4* main start codon [50]. As exemplified by the inhibitory role of *ATF4* uORF2, overlapped uORFs are obvious translational repressors since their termination codons are located 3′ of the mORF initiation codon [1, 3]. However, it is speculated that the post-terminating ribosomes can scan backwards in a 3′ to 5′ fashion at least for a critical number of nucleotides (less than ten) and reinitiate translation at an upstream initiation codon, although less efficiently [31].

In the case of *GCN4*, its 5′-leader sequence contains four short uORFs and from them only uORF1 is translated in both unstressed (good nutrition) and stressed (starvation) conditions. In fact, abolishing translation of the uORF1 by mutating its initiation codon, significantly impairs the expression of *GCN4* mORF [47]. In good nutritional conditions, reinitiation occurs in the following uORF-initiation codons after uORF1 translation, since the terminating ribosomes are rapidly reloaded with the necessary factors to initiate translation, which therefore repress *GCN4* expression [11, 47]. From the three downstream uORFs, uORF3 and uORF4 are the most repressive ones [47]. During stress, as for *ATF4* mRNA, the terminating ribosomes can no longer reload the required initiation factors in time to translate the inhibitory uORFs, resulting in the translation of the mORF [11, 47]. Interestingly, it is postulated that reinitiation in yeast diverges from mammalian reinitiation by the existence of specific *cis*-acting elements located upstream or downstream of the uORF [49]. The last codon and the 10 nt AU-rich sequence located at the 3′ side of the uORF1 termination codon promote the continued binding of the ribosome to the mRNA to allow permissive reinitiation. This is explained by the lack of a strong base-pairing between the rRNA and the AU-rich sequence that, in contrast to the GC-rich uORF4 termination codon context, will not allow the terminating ribosome to dissociate, and therefore the ribosome will resume scanning [33]. Additionally, permissive reinitiation is also associated with an upstream sequence that interacts with the initiation factor eIF3, mainly the eIF3a subunit via its N-terminal domain, which stabilizes the post-terminating 40S ribosome subunit in the mRNA, promoting reinitiation at the *GCN4* main initiation codon [51–53].

Another relevant aspect of uORF-mediated reinitiation is the ability to select translation initiation codons responsible for the synthesis of different protein isoforms encoded from the same mRNA molecule [1]. This is illustrated by the transcription factors CCAAT/enhancer-binding protein-α (*C/EBPα*) and -β (*C/EBPβ*), which regulate the proliferation and differentiation of multiple cell types, that have four alternative initiation sites: one for an out-of-frame uORF and three others that encode functionally different isoforms (extended, p42 and p30 for *C/EBPα*, and LAP*, LAP and LIP for *C/EBPβ*) [11, 54]. The presence of a translatable uORF between the first and third initiation codons of both mRNAs regulate the expression of each isoform, being responsible for the translational reinitiation of the truncated isoforms of *C/EBPα* and *C/EBPβ*, respectively, p30 and LIP [11, 54, 55].

5.2.2 Leaky Scanning or Ribosomal Bypass

Ribosomal leaky scanning or ribosomal bypass is a mechanism in which the scanning ribosomes bypass the uORF start codon and initiate translation at the subsequent uORF or at the mORF start codon (Fig. 5.1b) [1–3]. Ribosomal bypass is

usually associated with the proximity of the uAUG to the 5′ end of the mRNA and, to a greater extent, to the surrounding context of the uORF initiation codon [1, 3, 56, 57]. In fact, during ER stress, a weak uORF initiation codon context, together with a stronger initiation codon at the mORF, are seen in mRNAs that are preferentially translated [34, 58, 59]. However, it is unclear if this mechanism is triggered by a reduced availability of the ternary complex or an alteration of other critical factors for the translational process [3]. For instance, in the case of *C/EBPα* and *C/EBPβ* transcripts, the uORF initiation codon is bypassed in conditions of reduced levels of eIF4E and eIF2α (high abundance of eIF2α-P) resulting in the expression of the full-length isoforms in detriment to the truncated ones [54]. Another example of a transcript that is regulated by a bypass mechanism is the growth arrest and DNA damage-inducible protein 34 (*GADD34*). GADD34 is an important protein that regulates the ISR by a negative feedback mechanism. GADD34 controls the magnitude and duration of gene expression reprogramming by controlling the levels of eIF2α-P via interaction with the catalytic subunit of protein phosphatase 1 (PP1c), thus leading to dephosphorylation of eIF2α and ISR attenuation [34, 43]. Both *GADD34* uORFs are bypassed during stress conditions due to their weak Kozak context, thus increasing GADD34 expression [34, 45]. The α isoform of inhibitor of Bruton's tyrosine kinase (*IBTKα*) mRNA is also translationally regulated by a mechanism of bypass during conditions of eIF2α phosphorylation. In the case of the transcript *IBTKα* that encodes a protein important for the adaption and resolution of stress by promoting cell survival, it has a highly conserved 588 nts-long 5′-leader sequence bearing four uORFs. The *IBTKα* 5′-leader sequence cloned upstream of the firefly luciferase reporter construct when transfected into mouse embryonic fibroblast (MEF) cells treated with thapsigargin showed an increase in the luciferase activity, which is associated with a translational deregulation of the inhibitory uORFs. By mutational analyses of the uORFs initiation codons, it was shown that uORF1 is the main inhibitory uORF of IBTKα expression. The weak *IBTKα* uORF1 initiation codon context, together with low levels of eIF2α during stress, allow its bypass and thus, mORF expression [58].

Interestingly, some of the transcripts containing NMD-inducing uORFs in physiological settings, develop NMD-resistance under stress conditions [39, 60]. This is achieved by ribosomal bypass, where the ribosome scans through the uORF initiation codon, not recognizing the uORF termination codon as a PTC, and therefore saving the transcript from NMD [2]. This seems to be the case of the transcript 1 of interferon related developmental regulator 1 (*IFRD1*) that has a 52 aa-long uORF that, when translated in physiological conditions, represses *IFRD1* expression by triggering mRNA decay [60]. In conditions of ER stress induced by tunicamycin, *IFRD1* mRNA stabilizes, as there is ribosomal bypass of its uORF, leading to increased protein levels [60]. Other stress-related mRNAs, like *ATF4* [38, 39], as well as CCAAT-enhancer-binding protein homologous protein (*CHOP*) [39, 48], are also well described examples of uORF-containing transcripts that commit them to NMD. These mRNAs are upregulated in stress conditions, where eIF2α-P causes uORF ribosomal bypass and consequent NMD impairment, thus enhancing the ISR [39]. For instance, *CHOP* 5′-leader sequence contains a uORF with two in-frame uAUGs with inhibitory activity when in physiological conditions [48, 59]. In thapsigargin-induced ER stress, a translational increase of CHOP is reported due to leaky scanning of the uORF initiation codons that are seen to be in a weak Kozak context [48, 59, 61]. Persistent elevated levels of CHOP during a prolonged ISR, where the cells can no longer resolve the stress and survive, will induce the apoptotic signaling cascade by transcriptional activation of pro-apoptotic genes [62].

5.2.3 Recognition of Non-canonical Initiation Codons

A large group of the uORFs identified by RiboSeq are seen to initiate translation at near-cognate AUG codons [9, 18]. However, the real number

of non-canonical codons acting as initiator sites of translation is possibly underestimated, since initiation at these codons shows some resistance to the recognition by translation initiation-inhibitor drugs and also because there are no reliable bioinformatic tools for their discrimination [9, 27, 63]. CUG codon, encoding for a leucine (Leu), is the most prevalent non-AUG codon used in uORFs, followed by GUG (valine; Val) that is present in uORFs at roughly the same extent as AUG, and then by UUG (Leu) [9, 18, 20, 27]. Consistent with this data are the results obtained by a peptidomics study, where small peptides synthesized by non-AUG uORFs were detected [64].

Although translation initiation is poorly efficient when started at non-canonical initiation codons compared to that at the AUG, uORFs containing near-cognate start codons seem to be equally translated and important for the regulation of protein synthesis of specific transcripts during stress, where global translation is repressed [9, 18, 20–22]. uORFs bearing non-AUG codons appear to regulate the expression of their downstream ORFs by a mechanism of ribosomal bypass [1, 3]. For instance, the glutamyl-propyl-tRNA synthase (*EPRS*) mRNA is highly expressed in conditions of eIF2α-P abundance induced by the ER stressor thapsigargin, due to the regulation of two inhibitory uORFs with non-canonical start codons. The first uORF initiates with a CUG and is overlapped and out-of-frame with the mORF, and the second one is a UUG-containing uORF. Mutation of the CUG codon at uORF1 to a AUG in an optimal Kozak context represses the downstream ORF in physiological conditions and during thapsigargin-induced stress. Thus, bypass of the uORF1, due to its non-canonical start codon, is necessary for the translation of the *EPRS* mORF. The same holds for uORF2 that when its initiation codon is mutated to AUG one obtains low downstream expression during normal and stress conditions. Thus, as occurs for uORF1, the presence of the UUG initiation codon in uORF2 facilitates its bypass by some scanning ribosomes that will initiate translation at the mORF, a mechanism exacerbated under stress conditions due to high levels of eIF2α-P [65]. The growth arrest and DNA-damage inducible gamma (*GADD45G*) transcript, with a relevant function in cell growth and apoptosis regulation, also bears a CUG-containing uORF that, due to its overlapping and out-of-frame context, represses GADD45G expression in unstressed conditions. During starvation, the inhibitory uORF is bypassed essentially due to the non-canonical nature of its initiation codon, which increases GADD45G protein levels [66]. One additional example of this mechanism is observed for the WNT signaling pathway regulator *FRAT2* transcript. A construct carrying the bicistronic *FRAT2* mRNA sequence, with the uORF and the mORF tagged differentially, shows co-expression of the uORF-encoded peptide and the main protein in HEK293T cells. The uORF translation is initiated at an ACG codon that when mutated to AUG abolishes the expression of its mORF with concomitant expression of the uORF. This agrees with a bypass of the *FRAT2* uORF to allow downstream translation [64].

The conspicuous translation initiation at non-AUG codons raises the question of how these non-canonical codons are recognized as initiation sites by the translation machinery. It was seen that a CUG initiation codon is translated via a cap-dependent mechanism, although translation initiation seems to be independent on the $tRNA_{i}^{Met}$ [63, 67, 68]. For instance, using methionine sulfamide, an inhibitor of the methionyl-tRNA synthase (responsible for the aminoacylation of the tRNA with methionine), in toeprinting assays, does not produce a significant effect on the CUG codon recognition. This result suggests the 'replacement' of the $tRNA_{i}^{Met}$ by another aminoacyl-tRNA in the ribosomal P site of the pre-initiation complexes for the translation of CUG codons [63]. In fact, previous studies revealed that CUG is translated as a Leu and not as a Met residue [67]. Furthermore, it is suggested that some of the 40S ribosomal subunits are not pre-loaded with any initiator tRNA, contrasting with the current model of translation, and that they are able to load the $tRNA^{Leu}$ when CUG is encountered. Delivery of $tRNA^{Leu}$ to the P site of the ribosome is also independent of

eIF2·GTP·$tRNA_i^{Met}$ and is mediated by the eukaryotic initiation factor 2A (eIF2A) [68]. In conditions of low abundance of eIF2·GTP·$tRNA_i^{Met}$, the levels of eIF2A are found to be elevated [69]. Conversely to eIF2α, depletion of eIF2A does not promote measurable global translational repression, but has a negative impact on uORFs containing CUG or UUG as initiation codons, which supports the function of eIF2A in translation initiation of non-canonical start codons [68, 69]. This is shown for the binding immunoglobulin protein (*BiP*) transcript, where depletion of eIF2A impairs translation of its UUG-containing uORF. During ER stress induced by thapsigargin, this depletion leads to a significant decrease in BiP expression, due to the downregulation of the UUG-uORF [69]. In parallel to eIF2A, it was shown that other factors are involved in non-canonical translational events. Controlling the rate of non-AUG initiation is the equilibrium between the translation initiation factor eIF5 and the eIF5-mimic protein, 5MP, acting as an enhancer and a repressor, respectively, of translation initiation at near-cognate codons [20, 70]. The repression exerted by 5MP, is suggested to be achieved by preventing the interaction of eIF5 with eIF2 in PIC in favor of its own interaction [20]. Moreover, start codon selection is also the result of a cross-regulation between eIF5 and eIF1, where overexpression of eIF1 represses non-AUG initiation [70]. Yet, the interplay between eIF5 and eIF2A in the promotion of translation initiation at non-canonical start codons was not determined [20].

Similarly to what is postulated for the AUG, recognition of the non-AUG codons is associated with a strong Kozak consensus sequence. The optimal sequence context for translation initiation at a CUG is almost the same already described for AUG, being TCCACCCUGG, and as expected, modifications of this sequence will weaken the recognition of the CUG [67]. In the case of *FRAT2* mRNA, when the ACG start codon of the uORF is placed in a less optimal Kozak sequence, the uORF-encoded peptide is expressed at lower levels [64]. As for CUG, efficient translation at the GUG codon requires the presence of a CCACC sequence immediately upstream [71]. Additionally, it was also demonstrated that non-AUG start codons enhance their own translation by blocking the scanning ribosomes in their vicinity. This is explained by a high GC-rich content immediately after the non-AUG codon allowing the formation of hairpins that will impair scanning [21].

As pointed out before, it seems that there are a pool of ribosomes that scan specifically for the non-AUG codons, and that those ribosomes may be different from the ones initiating at the AUG [67]. In fact, ribosomal heterogeneity can be used as an explanation of the differential recognition of AUG and non-AUG codons [63]. This seems to be dependent on the rRNA composition, on ribosomal proteins that can be differentially expressed and post-translationally modified, and on the modification of the translation factors interacting with the ribosome, as well as the tRNAs [63, 72, 73]. These so-called specialized ribosomes can modulate translation of mRNAs depending on the presence of specific features such as uORFs, thus leading to the preferential translation of classes of mRNAs, adding another layer of translational regulation centered on the ribosome [72, 74]. Additionally, even the constitutive components of the ribosome that have little or no variation, can have specialized activities when interacting with regulatory elements in the 5′-leader sequence [72]. Moreover, it is postulated that changes in the ribosomal components is crucial for gene expression reprograming depending on the cell environment, differentiation and development [73].

5.2.4 Functional uORF-Encoded Peptides

The uORF-encoded peptides are usually poorly detected by the conventional proteomic methods based on mass spectrometry (MS) [75]. Moreover, despite the wide detection of uORF translation by RiboSeq approaches, this technique does not give information on the peptide encoded by these elements [9, 64, 76]. Thus, complementation between mRNA deep-sequencing and RiboSeq with proteomics approaches, such as MS, expand

overall protein identification, including the detection of uORF-encoded peptides [64, 75–77]. These small peptides were encountered in cells at concentrations equivalent to the cellular proteins and also with specific subcellular localizations, providing evidence that they have biological functions, others than mORF repression [64].

The uORF-encoded peptides can act as *cis*- or *trans*-regulators of the mORF expression (Fig. 5.1a) [3]. The ability of the peptides to stall the translation machinery in a *cis*-fashion can be explained by their specific sequence and/or their direct or indirect interaction with small molecules ('peptoswitch'), as occurs for S-adenosylmethionine decarboxylase (*AdoMetDC*) mRNA [7, 78]. The AdoMetDC uORF-encoded peptide has 6 aa with the sequence MAGDIS that seems to interact directly, via the fourth and fifth residues, with the last tRNA ($tRNA^{Ser}$), stabilizing the ribosome near the uORF termination codon in the presence of high levels of polyamines [36, 78]. This process stalls the ribosomes and allows their dissociation, which results in the translational repression of the mORF [78]. In the case of the already mentioned *GADD34* mRNA, its 5′-leader sequence contains two inhibitory uORFs, in which uORF2 translation significantly represses mORF expression during basal conditions in a nucleotide-dependent manner [34, 45]. The uORF2-encoded small peptide contains a conserved Pro-Pro-Gly sequence at the C-terminal region that allows the release of the translating ribosomes, thus inhibiting mORF translation [34]. An example of the *trans*-acting regulation activity of the uORF-encoded peptides is represented for the argininosuccinate synthase (*AS*) transcript [79]. Pendleton and co-workers showed that the peptide encoded by the overlapped out-of-frame uORF of the *AS* transcript inhibits the overall expression of endogenous AS protein encoded from different *AS* isoforms, through a process dependent on the uORF length and sequence [79].

The uORF-encoded peptides can also interact with different cellular proteins, and not only with the products of their own coding sequence, functioning as *trans*-acting factors, and thus have several biological functions in the cell [7, 69]. For instance, the small peptides encoded by both non-AUG uORFs of the *BiP* transcript function as human leukocyte antigen (HLA)-presented epitopes recognized by human T cells [69]. The uORF-encoded peptides can also be developmentally and/or spatially regulated, and for that are endowed of distinct functions compared to the protein encoded by the mORF [64, 80, 81]. An example of this developmental regulation was already described for the spliced transcript variant of MYCN (*MYCNΔ1b*) where the presence of a uORF does not influence the translational level of MYCNΔ1b, but instead produces a small peptide (MYCNOT) that is expressed in fetal but not in adult brains [81]. Regarding spatial regulation, there is the McKusick-Kaufman syndrome (*MKKS*) transcript, where two of the three uORFs in its 5′-leader sequence seem to be translated into highly conserved peptides with mitochondrial localization, distinct from the cytoplasmic localization of the MKKS protein. This suggests that the uORF-encoded peptides and the main protein have distinct functions [80].

5.3 uORFs Deregulation in Genetic Disorders

The importance of uORFs in the regulation of different patterns of gene expression under normal and stress conditions highlights their relevance, if altered, for the development of several human diseases, such as metabolic, hematologic and neurologic disorders, inherited syndromes, cancer and its susceptibility [2, 11–14, 22]. In different disease-associated variant databases, more than 3700 variants were identified in the 5′-leader sequence of human transcripts that can alter a uORF [10]. Among these variations, the most harmful are the ones that originate or eliminate an initiation or termination codon, and thus regulate the presence or absence of a uORF [10, 11]. Additionally, other alterations could deregulate uORF features like the Kozak consensus context, the uORF length, the uORF number, and its distance to the 5′ end of the mRNA or to the CDS [11]. Several bioinformatic analyses were performed to map possible variations (polymor-

phisms and mutations) within the 5′-leader sequence of transcripts that can interfere with the regulatory function of uORFs [8, 10]. Calvo and co-workers identified 509 transcripts bearing common polymorphisms associated with the creation or elimination of a uORF [8]. More recently, Wethmar and co-workers identified in 2610 genes, 1375 single nucleotide polymorphisms (SNPs) disrupting uAUGs and 2724 SNPs affecting the Kozak context of a uORF. Additionally, in a small percentage of genes 697 SNPs are present at uORF stop codons. Eight uORF-disruptive SNPs already have clinical association [82]. A well-known example of a disruptive polymorphism is the one identified in the factor XII (*FXII*) gene, that plays a role in the coagulation process [83]. The alteration of a C for a T at the position −4 (c.-4C>T) of the *FXII* 5′-leader sequence forms a novel uORF that negatively regulates the FXII plasma levels, associating this polymorphism to the occurrence of stroke episodes [8, 83, 84]. Calvo and co-workers have also identified 11 novel patient mutations in disease-associated genes that create or eliminate a uORF [8]. One example is the interferon regulatory factor 6 (*IRF6*) gene associated with Van der Woude syndrome (VWS) and the popliteal pterygium syndrome (PPS), two disorders characterized by facial alterations, such as cleft lip and palate. From the several mutations identified in the *IRF6* gene, a frameshift mutation in the position −48 of the 5′-leader sequence that alters a T to an A (c.-48T>A), creates a new initiation codon [85]. This mutation originates a uORF that significantly represses IRF6 protein levels (70–100% of inhibition), which is typically associated with the disease phenotype [8]. Additionally, there are mutations that create another uORF in a uORF-containing mRNA, such as the case of the sex determining region Y (*SRY*) gene in gonadal dysgenesis (c.-75G>A) and serine protease inhibitor, kazal-type 1 (*SPINK1*) gene in hereditary pancreatitis (c.-53C>T). This leads to the almost complete inhibition of mORF expression, which is explained by a cumulative effect of multiple inhibitory uORFs [8, 86, 87].

In the reviews of Barbosa and co-workers (2013) and Silva and co-workers (2017) are listed several examples of uORF creation, elimination or modification in the development of several types of diseases, including rare disorders [2, 88], such as hereditary thrombocythemia (uORF elimination) [89, 90], melanoma predisposition (uORF creation) [91, 92] and Marie Unna hereditary hair loss (uORF modification) [93]. Several new examples of genetic diseases have been described to have an association with uORFs deregulation. For instance, haploinsufficiency of twist-related protein 1 (TWIST1) is correlated with the development of Saethre-Chotzen syndrome (SCS), characterized by a malformation of the skull (craniosynostosis). From a screening of 14 genetically undiagnosed SCS patients, two novel single nucleotide variants (SNVs) were detected in the *TWIST1* 5′-leader sequence (c.-263C>A and c.-255G>A) that contribute to the formation of novel start sites (AUG) in a good Kozak context. The c.-263C>A SNV generates an out-of-frame uORF of 68 codons and the c.-255G>A SNV generates an in-frame uAUG with the mORF that possibly forms an N-extended isoform of TWIST1 protein. Both alterations repress TWIST1 mORF expression, which is associated with the typical disease phenotype, when no common mutations in the main coding sequence are present [94]. Kitano and co-workers identified the presence of three SNPs – rs542483929, rs188349884 and rs759579732 – in the 5′-leader sequence of the histidine receptor H2 (*HRH2*) gene that create transposon-derived upstream ATGs, which originate uORFs. These new formed uORFs downregulate mORF expression and can be potentially associated with gastric cancer susceptibility, a disease already related to mutations in the enhancer region of *HRH2* [95].

Alterations in the pattern and function of uORFs in proto-oncogenes and tumor suppressors genes can explain how cancer is triggered: mutations that lead to a loss-of-function of the uORF in proto-oncogenes resulting in their overexpression or mutations that promote a gain-of-function of the uORF in tumor suppressor genes resulting in a decrease of these protective proteins [11]. RiboSeq analysis studies provide evidence of the widespread presence of exonic

cancer mutations that alter uORF start codons impairing their regulatory role in several oncogenes and tumor-suppressor genes, such as *MYC*, B-cell lymphoma (*BCL-2*), phosphatase and tensin homolog (*PTEN*), tumor protein p53 (*TP53*), MutS homolog 5 (*MSH5*), among others [96].

As mentioned above, familial melanoma predisposition is a well-established uORF-related condition, associated with mutations in the cyclin dependent kinase inhibitor 2A (*CDKN2A*) gene [92]. *CDKN2A* impairs cell cycle progression by encoding two tumor suppressor proteins, $p16^{INK4A}$ and $p14^{ARF}$ [91, 97]. A transversion of the nucleotide G to T at position −34 (c.-34G>T) in the *CDKN2A* 5′-leader sequence creates an out-of-frame uAUG in the mRNA. This uAUG forms a translatable uORF responsible for the low expression levels of the mORF, thus providing predisposition to melanoma [92].

In addition to its role in disease development, uORF creation has also been related to poor prognosis in the response to drug treatment. The creation of a uORF in the mRNA of the excision repair cross-complementation group 5 (*ERCC5*) gene generated by a polymorphic variation (rs751402) confers resistance to platinum-based chemotherapeutics in childhood ependymoma (malignant brain tumor). This is observed by the ERCC5 up-regulation mediated by its uORF following cisplatin-induced bulky adduct DNA damage. The elevated levels of ERCC5 lead to a higher degree of cisplatin-induced DNA damage repair, thus lowering the efficiency of this chemotherapeutic [98].

An example of a genetic alteration that modifies a uORF was reported in cyclin dependent kinase 1B (*CDKN1B*) gene, associated with the induction of inherited multiple endocrine neoplasia syndrome (MEN), which is characterized by several distinct tumors affecting at least two endocrine organs. *CDKN1B* gene encodes for $p27^{KIP1}$, a tumor suppressor that regulates cell cycle and cell proliferation, by promoting cell cycle arrest at G1 phase [97]. A 4-base pair (4 bp) deletion (c.-456-453delCCTT) was identified in the sequence of the *CDKN1B* uORF, which disrupts and shifts its termination codon, resulting in a lengthened uORF sequence and a reduced intercistronic space in relation to the downstream initiation codon. This germline mutation allows the translation of a lengthened uORF-encoded peptide that reduces mORF expression by preventing reinitiation events to occur that in the end can be associated with cancer formation. The patient carrying this germline mutation has pituitary adenomas and tumors in the endocrine pancreas, consistent with the MEN4 phenotype [99]. No other biological functions were associated with this uORF-encoded peptide that can be related to the disease phenotype [100]. A uORF-encoded peptide with an association with the disease phenotype was recently identified in the familial DOPA responsive dystonia (DRD). The c.-22C>T SNP in the 5′-leader sequence of the guanosine triphosphate cyclohydrolase 1 (*GCH1*) gene creates a uAUG that promotes the formation of an overlapped and out-of-frame uORF. This uORF encodes a 73 aa-long peptide responsible for low levels of GCH1 protein, which impairs the dopamine biosynthesis pathway, resulting in reduced levels of dopamine and dopaminergic dysfunction in the brain, characteristic of DRD. This 73 aa peptide also accumulates at considerable levels within the nucleus where it is predicted to be involved in transcription factor activities, promoting cytotoxic effects with reduction of cell viability [101].

A systematic search for cancer-related uORF mutations has been performed, screening for loss-of-function uORF mutations in 404 uORF initiation sites of 132 potential proto-oncogenes in 308 human malignancies. Interestingly, mutations were identified in both the uAUG and the uORF Kozak consensus sequence. Four novel uORF-associated mutations caused the loss of a uAUG in the Src family tyrosine kinase BLK proto-oncogene (*BLK*) in a colon adenocarcinoma; the ephrin receptor B1 (*EPHB1*) in a mammary carcinoma; the Janus kinase 2 (*JAK2*) in chronic lymphocytic leukemia; and the mitogen-activated protein kinase kinase 6 (*MAP2K6*) in a colon adenocarcinoma. There was no detectable function of the mutations found in *BLK* and *JAK2*. Although, in the cases of

EPHB1 and *MAP2K6*, the uAUGs were mutated to a uGUG and a uACG, respectively, which were responsible for the up-regulation of the downstream ORF. It is worth noting that the *MAP2K6* uORF mutation was also present in normal control tissue of the affected patient, which suggests a relationship between this variant and the predisposition to tumor development. In an additional study, a whole exome sequencing computational analysis of datasets of 464 colon adenocarcinomas revealed 53 non-recurrent somatic mutations that delete either the uORF initiation or termination codon [102]. This highlights the importance of uORF mutations in the tumorigenic process, although further functional studies are needed.

In addition to the role of genetic alterations that deregulate uORF-mediated translation, there are other mechanisms that can overcome or take advantage of the repression exerted by naturally occurring uORFs to promote a disease phenotype. For instance, MDM2 is an oncoprotein that antagonizes, in a feedback loop, the function of the tumor suppressor p53 and has been seen overexpressed in many tumors, such as osteosarcomas, gliomas and soft tissue sarcomas. MDM2 can be produced from two spliced isoforms with different 5′-leader sequences as a result of the function of two cryptic promoters: (i) the first promoter (P1) transcribes a long mRNA (L-mdm2) without exon 2, and (ii) the second promoter (P2) transcribes a short mRNA (S-mdm2) with exon 2, but lacking exon 1. L-mdm2 mRNAs contain two inhibitory uORFs in exon 1 that will repress MDM2 expression. However, those uORFs do not exist in the S-mdm2 mRNA. In choriocarcinoma cells, MDM2 expression is translated from the S-mdm2 transcript. Apparently, the transcriptional switch from promoter P1 to P2 is related to the binding of p53 to elements in exon 1 (lacking in L-mdm2) which activates P2 and consequently induces MDM2 expression via transcription of S-mdm2 [103]. Additionally, BRCA1, DNA repair associated gene, is partially regulated by a similar mechanism. *BRCA1* is well known to be the major susceptibility gene for breast and ovarian cancers by showing a reduced expression level. It was shown that two distinct promoters are used to produce two *BRCA1* transcript variants distinct in their 5′-leader sequences (named 5'UTRa and the 5'UTRb). The mRNA containing the 5'UTRb is expressed in breast cancer but not in normal tissues, and it contains three uORFs that negatively modulate mORF expression [104]. Another alternative mechanism of uORF regulation with pathophysiological relevance was described for the oncoprotein Erb-B2 receptor tyrosine kinase 2 (*ERBB2*) transcript, also known as the human epidermal growth factor receptor 2 (*HER-2*). *HER-2* mRNAs have an inhibitory small uORF in which the stop codon is located 5 nt upstream of the mORF initiation codon, impairing translation reinitiation and keeping HER-2 at basal levels under physiological conditions [32, 105]. However, another post-transcriptional mechanism seems to occur to promote overexpression of HER-2 in cancer cells, without having alterations in the mRNA sequence or size. A U-rich translational derepression element (TDE) was identified in the 3'UTR of the *HER-2* mRNA, that associates with several *trans*-acting factors, among them RNA-binding proteins, to repress the inhibitory activity of the translatable uORF, allowing an efficient translation of the mORF in breast cancer cells [105].

Recently, RiboSeq data has shown that, during initiation of the tumorigenic process of epidermal cells, translation of cancer-related mRNAs is dependent on the translation of the uORFs present in their 5′-leader sequences. Moreover, those uORFs initiate to a great extent at non-canonical start codons, with the CUG being the most prevalent. High levels of eIF2A were also detected, and this, as mentioned before, has a role in the translational regulation of those non-AUG uORFs, and thus this alternative translation factor is also associated with tumor progression. Given this, the authors speculate that for tumorigenic initiation the translational apparatus needs to be redirected to the translation of uORFs in a cohort of cancer-related mRNAs such as catenin beta 1 (*CTNNB1*), hypoxia inducible factor 1 subunit alpha (*HIF1α*), Rac family small GTPase 1 (*Rac1*), cyclin-dependent kinase 1 (*Cdk1*), among others [18].

5.4 Conclusions

As broadly evidenced by RiboSeq and transcript-specific studies, AUG- and non-AUG uORFs are ubiquitous *cis*-acting elements in the 5′-leader sequences of human transcripts. These uORFs have a crucial role in the translational control of their downstream mORF, which is relevant for homeostasis maintenance of the pool of proteins in a cell. Furthermore, uORFs-encoded peptides are now recognized as having important and diverse regulatory functions in the cell. Taking into account the widespread distribution of uORFs, as well as their regulatory function, it is expected that potential alterations that create, disrupt or modify the uORF function can be associated with the etiology of diverse pathologies that do not have the coding molecular alterations usually underlying the disease. Thus, it is of outmost importance the systematic screening of 5′-leader sequences in search for novel uORFs or uORF-altering mutations that can be linked to a disease with the final purpose of developing new diagnostic and, potentially, therapeutic approaches.

Acknowledgements Work partially supported by UID/MULTI/04046/2013 centre grant from FCT, Portugal (to BioISI), and by National Institute of Health Dr. Ricardo Jorge. J.F.P.S. is recipient of a fellowship from BioSys PhD programme (SFRH/BD/106081/2015) from FCT (Portugal). R.Q.F. is recipient of a fellowship from BioSys PhD programme (SFRH/BD/114392/2016) from FCT (Portugal).

References

1. Young SK, Wek RC (2016) Upstream open reading frames differentially regulate gene-specific translation in the Integrated Stress Response. J Biol Chem 291:16927–16935
2. Barbosa C, Peixeiro I, Romão L (2013) Gene expression regulation by upstream open reading frames and human disease. PLoS Genet 9:e1003529
3. Somers J, Pöyry T, Willis AE (2013) A perspective on mammalian upstream open reading frame function. Int J Biochem Cell Biol 45:1690–1700
4. Sonenberg N, Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell 136:731–745
5. Leppek K, Das R, Barna M (2018) Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat Rev Mol Cell Biol 19:158–174
6. Lacerda R, Menezes J, Romão L (2017) More than just scanning: the importance of cap-independent mRNA translation initiation for cellular stress response and cancer. Cell Mol Life Sci 74:1659–1680
7. Andrews SJ, Rothnagel JA (2014) Emerging evidence for functional peptides encoded by short open reading frames. Nat Rev Genet 15:193–204
8. Calvo SE, Pagliarini DJ, Mootha VK (2009) Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci USA 106:7507–7512
9. Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147:789–802
10. Ye Y, Liang Y, Yu Q, Hu L, Li H, Zhang Z, Xu X (2015) Analysis of human upstream open reading frames and impact on gene expression. Hum Genet 134:605–612
11. Wethmar K, Smink JJ, Leutz A (2010) Upstream open reading frames: molecular switches in (patho) physiology. Bioessays 32:885–893
12. Morris DR, Geballe AP (2000) Upstream open reading frames as regulators of mRNA translation. Mol Cell Biol 20:8635–8642
13. Onofre C, Tomé F, Barbosa C, Silva AL, Romão L (2015) Expression of human hemojuvelin (HJV) is tightly regulated by two upstream open reading frames in HJV mRNA that respond to iron overload in hepatic cells. Mol Cell Biol 35:1376–1389
14. Sajjanar B, Deb R, Raina SK, Pawar S, Brahmane MP, Nirmale AV, Kurade NP, Manjunathareddy GB, Bal SK, Singh NP (2017) Untranslated regions (UTRs) orchestrate translation reprogramming in cellular stress responses. J Therm Biol 65:69–75
15. Iacono M, Mignone F, Pesole G (2005) uAUG and uORFs in human and rodent 5′untranslated mRNAs. Gene 349:97–105
16. Wethmar K (2014) The regulatory potential of upstream open reading frames in eukaryotic gene expression. Wiley Interdiscip Rev RNA 5:765–768
17. Johnstone TG, Bazzini AA, Giraldez AJ (2016) Upstream ORFs are prevalent translational repressors in vertebrates. EMBO J 35:706–723
18. Sendoel A, Dunn JG, Rodriguez EH, Naik S, Gomez NC, Hurwitz B, Levorse J, Dill BD, Schramek D, Molina H, Weissman JS, Fuchs E (2017) Translation from unconventional 5′ start sites drives tumour initiation. Nature 541:494–499
19. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223
20. Tang L, Morris J, Wan J, Moore C, Fujita Y, Gillaspie S, Aube E, Nanda J, Marques M, Jangal

M, Anderson A, Cox C, Hiraishi H, Dong L, Saito H, Singh CR, Witcher M, Topisirovic I, Qian S-B, Asano K (2017) Competition between translation initiation factor eIF5 and its mimic protein 5MP determines non-AUG initiation rate genome-wide. Nucleic Acids Res 45:11941–11953
21. Kozak M (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc Natl Acad Sci USA 87:8301–8305
22. Wethmar K, Barbosa-Silva A, Andrade-Navarro MA, Leutz A (2014) uORFdb—A comprehensive literature database on eukaryotic uORF biology. Nucleic Acids Res 42:D60–D67
23. Jackson RJ, Hellen CUT, Pestova TV (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol 11:113–127
24. Graff JR, Konicek BW, Carter JH, Marcusson EG (2008) Targeting the eukaryotic translation initiation factor 4E for cancer therapy. Cancer Res 68:631–634
25. Algire MA, Maag D, Lorsch JR (2005) P_i release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. Mol Cell 20:251–262
26. Fringer JM, Acker MG, Fekete CA, Lorsch JR, Dever TE (2007) Coupled release of eukaryotic translation initiation factors 5B and 1A from 80S ribosomes following subunit joining. Mol Cell Biol 27:2384–2397
27. Lee S, Liu B, Lee S, Huang S-X, Shen B, Qian S-B (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc Natl Acad Sci USA 109:E2424–E2432
28. Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283–292
29. Kozak M (1987) At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. J Mol Biol 196:947–950
30. Pisarev AV, Kolupaeva VG, Pisareva VP, Merrick WC, Hellen CUT, Pestova TV (2006) Specific functional interactions of nucleotides at key − 3 and + 4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. Genes Dev 20:624–636
31. Kozak M (2001) Constraints on reinitiation of translation in mammals. Nucleic Acids Res 29:5226–5232
32. Child SJ, Miller MK, Geballe AP (1999) Translational control by an upstream open reading frame in the HER-2/neu transcript. J Biol Chem 274:24335–24341
33. Grant CM, Hinnebusch AG (1994) Effect of sequence context at stop codons on efficiency of reinitiation in GCN4 translational control. Mol Cell Biol 14:606–618
34. Young SK, Willy JA, Wu C, Sachs MS, Wek RC (2015) Ribosome reinitiation directs gene-specific translation and regulates the Integrated Stress Response. J Biol Chem 290:28257–28271
35. Col B, Oltean S, Banerjee R (2007) Translational regulation of human methionine synthase by upstream open reading frames. Biochim Biophys Acta 1769:532–540
36. Law GL, Raney A, Heusner C, Morris DR (2001) Polyamine regulation of ribosome pausing at the upstream open reading frame of S-adenosylmethionine decarboxylase. J Biol Chem 276:38036–38043
37. Poyry TAA, Kaminski A, Jackson RJ (2004) What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? Genes Dev 18:62–75
38. Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet 36:1073–1078
39. Gardner LB (2008) Hypoxic inhibition of nonsense-mediated RNA decay regulates gene expression and the Integrated Stress Response. Mol Cell Biol 28:3729–3741
40. Rebbapragada I, Lykke-Andersen J (2009) Execution of nonsense-mediated mRNA decay: what defines a substrate? Curr Opin Cell Biol 21:394–402
41. Torrance V, Lydall D (2018) Overlapping open reading frames strongly reduce human and yeast STN1 gene expression and affect telomere function. PLoS Genet 14:e1007523
42. Donnelly N, Gorman AM, Gupta S, Samali A (2013) The eIF2α kinases: their structures and functions. Cell Mol Life Sci 70:3493–3511
43. Novoa I, Zeng H, Harding HP, Ron D (2001) Feedback inhibition of the Unfolded Protein Response by GADD34-mediated dephosphorylation of eIF2α. J Cell Biol 153:1011–1021
44. Spilka R, Ernst C, Mehta AK, Haybaeck J (2013) Eukaryotic translation initiation factors in cancer development and progression. Cancer Lett 340:9–21
45. Lee Y-Y, Cevallos RC, Jan E (2009) An upstream open reading frame regulates translation of GADD34 during cellular stresses that induce eIF2α phosphorylation. J Biol Chem 284:6661–6673
46. Harding HP, Zhang Y, Zeng H, Novoa I, Lu PD, Calfon M, Sadri N, Yun C, Popko B, Paules R, Stojdl DF, Bell JC, Hettmann T, Leiden JM, Ron D (2003) An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. Mol Cell 11:619–633
47. Hinnebusch AG (2005) Translational regulation of *GCN4* and the general amino acid control of yeast. Annu Rev Microbiol 59:407–450
48. Young SK, Palam LR, Wu C, Sachs MS, Wek RC (2016) Ribosome elongation stall directs gene-specific translation in the Integrated Stress Response. J Biol Chem 291:6546–6558
49. Skabkin MA, Skabkina OV, Hellen CUT, Pestova TV (2013) Reinitiation and other unconventional post-termination events during eukaryotic translation. Mol Cell 51:249–264

50. Vattem KM, Wek RC (2004) Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. Proc Natl Acad Sci 101:11269–11274
51. Grant CM, Miller PF, Hinnebusch AG (1995) Sequences 5′ of the first upstream open reading frame in *GCN4* mRNA are required for efficient translational reinitiation. Nucleic Acids Res 23:3980–3988
52. Szamecz B, Rutkai E, Cuchalová L, Munzarová V, Herrmannová A, Nielsen KH, Burela L, Hinnebusch AG, Valášek L (2008) eIF3a cooperates with sequences 5′ of uORF1 to promote resumption of scanning by post-termination ribosomes for reinitiation on GCN4 mRNA. Genes Dev 22:2414–2425
53. Munzarová V, Pánek J, Gunišová S, Dányi I, Szamecz B, Valášek LS (2011) Translation reinitiation relies on the interaction between eIF3a/TIF32 and progressively folded *cis*-acting mRNA elements preceding short uORFs. PLoS Genet 7:e1002137
54. Calkhoven CF, Müller C, Leutz A (2000) Translational control of C/EBPα and C/EBPβ isoform expression. Genes Dev 14:1920–1932
55. Wethmar K, Bégay V, Smink JJ, Zaragoza K, Wiesenthal V, Dörken B, Calkhoven CF, Leutz A (2010) C/EBPbetaDeltauORF mice-a genetic model for uORF-mediated translational control in mammals. Genes Dev 24:15–20
56. Kozak M (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. J Biol Chem 266:19867–19870
57. Kozak M (1991) A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. Gene Expr 1:111–115
58. Baird TD, Palam LR, Fusakio ME, Willy JA, Davis CM, McClintick JN, Anthony TG, Wek RC (2014) Selective mRNA translation during eIF2 phosphorylation induces expression of IBTKα. Mol Biol Cell 25:1686–1697
59. Palam LR, Baird TD, Wek RC (2011) Phosphorylation of eIF2 facilitates ribosomal bypass of an inhibitory upstream ORF to enhance *CHOP* translation. J Biol Chem 286:10939–10949
60. Zhao C, Datta S, Mandal P, Xu S, Hamilton T (2010) Stress-sensitive regulation of IFRD1 mRNA decay is mediated by an upstream open reading frame. J Biol Chem 285:8552–8562
61. Chen Y-J, Tan BC-M, Cheng Y-Y, Chen J-S, Lee S-C (2010) Differential regulation of CHOP translation by phosphorylated eIF4E under stress conditions. Nucleic Acids Res 38:764–777
62. Rutkowski DT, Arnold SM, Miller CN, Wu J, Li J, Gunnison KM, Mori K, Akha AAS, Raden D, Kaufman RJ (2006) Adaptation to ER stress is mediated by differential stabilities of pro-survival and pro-apoptotic mRNAs and proteins. PLoS Biol 4:e374
63. Starck SR, Ow Y, Jiang V, Tokuyama M, Rivera M, Qi X, Roberts RW, Shastri N (2008) A distinct translation initiation mechanism generates cryptic peptides for immune surveillance. PLoS ONE 3:e3460
64. Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, Karger AD, Budnik BA, Rinn JL, Saghatelian A (2013) Peptidomic discovery of short open reading frame–encoded peptides in human cells. Nat Chem Biol 9:59–64
65. Young SK, Baird TD, Wek RC (2016) Translation regulation of the glutamyl-prolyl-tRNA synthetase gene EPRS through bypass of upstream open reading frames with noncanonical initiation codons. J Biol Chem 291:10824–10835
66. Gao X, Wan J, Liu B, Ma M, Shen B, Qian S-B (2015) Quantitative profiling of initiating ribosomes *in vivo*. Nat Methods 12:147–153
67. Schwab SR, Shugart JA, Horng T, Malarkannan S, Shastri N (2004) Unanticipated antigens: translation initiation at CUG with leucine. PLoS Biol 2:e366
68. Starck SR, Jiang V, Pavon-Eternod M, Prasad S, McCarthy B, Pan T, Shastri N (2012) Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. Science 336:1719–1723
69. Starck SR, Tsai JC, Chen K, Shodiya M, Wang L, Yahiro K, Martins-Green M, Shastri N, Walter P (2016) Translation from the 5′ untranslated region shapes the integrated stress response. Science 351:aad3867
70. Loughran G, Sachs MS, Atkins JF, Ivanov IP (2012) Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5. Nucleic Acids Res 40:2898–2906
71. Kozak M (1989) Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. Mol Cell Biol 9:5073–5080
72. Xue S, Barna M (2012) Specialized ribosomes: a new frontier in gene regulation and organismal biology. Nat Rev Mol Cell Biol 13:355–369
73. Sauert M, Temmel H, Moll I (2015) Heterogeneity of the translational machinery: variations on a common theme. Biochimie 114:39–47
74. Shi Z, Fujii K, Kovary KM, Genuth NR, Röst HL, Teruel MN, Barna M (2017) Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. Mol Cell 67:71–83
75. Oyama M, Kozuka-Hata H, Suzuki Y, Semba K, Yamamoto T, Sugano S (2007) Diversity of translation start sites may define increased complexity of the human short ORFeome. Mol Cell Proteomics 6:1000–1006
76. Crappé J, Ndah E, Koch A, Steyaert S, Gawron D, De Keulenaer S, De Meester E, De Meyer T, Van Criekinge W, Van Damme P, Menschaert G (2015) PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration. Nucleic Acids Res 43:e29
77. Raj A, Wang SH, Shim H, Harpak A, Li YI, Engelmann B, Stephens M, Gilad Y, Pritchard JK (2016) Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. elife 5:e13328
78. Raney A, Law GL, Mize GJ, Morris DR (2002) Regulated translation termination at the upstream

open reading frame in s-adenosylmethionine decarboxylase mRNA. J Biol Chem 277:5988–5994
79. Pendleton LC, Goodwin BL, Solomonson LP, Eichler DC (2005) Regulation of endothelial argininosuccinate synthase expression and NO production by an upstream open reading frame. J Biol Chem 280:24252–24260
80. Akimoto C, Sakashita E, Kasashima K, Kuroiwa K, Tominaga K, Hamamoto T, Endo H (2013) Translational repression of the McKusick-Kaufman syndrome transcript by unique upstream open reading frames encoding mitochondrial proteins with alternative polyadenylation sites. Biochim Biophys Acta 1830:2728–2738
81. Besançon R, Valsesia-Wittmann S, Locher C, Delloye-Bourgeois C, Furhman L, Tutrone G, Bertrand C, Jallas A-C, Garin E, Puisieux A (2009) Upstream ORF affects MYCN translation depending on exon 1b alternative splicing. BMC Cancer 9:445
82. Wethmar K, Schulz J, Muro EM, Talyan S, Andrade-Navarro MA, Leutz A (2016) Comprehensive translational control of tyrosine kinase expression by upstream open reading frames. Oncogene 35:1736–1742
83. Bersano A, Ballabio E, Bresolin N, Candelise L (2008) Genetic polymorphisms for the study of multifactorial stroke. Hum Mutat 29:776–795
84. Kanaji T, Okamura T, Osaki K, Kuroiwa M, Shimoda K, Hamasaki N, Niho Y (1998) A common genetic polymorphism (46 C to T substitution) in the 5′-untranslated region of the coagulation factor XII gene is associated with low translation efficiency and decrease in plasma factor XII level. Blood 91:2010–2014
85. Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, De Lima RLLF, Daack-Hirsch S, Sander A, McDonald-McGinn DM, Zackai EH, Lammer EJ, Aylsworth AS, Ardinger HH, Lidral AC, Pober BR, Moreno L, Arcos-Burgos M, Valencia C, Houdayer C, Bahuau M, Moretti-Ferreira D, Richieri-Costa A, Dixon MJ, Murray JC (2002) Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat Genet 32:285–289
86. Poulat F, Desclozeaux M, Tuffery S, Jay P, Boizet B, Berta P (1998) Mutation in the 5′ noncoding region of the SRY gene in an XY sex-reversed patient. Hum Mutat Suppl 1:S192–S194
87. Witt H, Luck W, Hennies HC, Claßen M, Kage A, Laß U, Landt O, Becker M (2000) Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. Nat Genet 25:213–216
88. Silva J, Fernandes R, Romão L (2017) Gene expression regulation by upstream open reading frames in rare diseases. J Rare Dis Res Treat 2:33–38
89. Nelson ND, Marcogliese A, Bergstrom K, Scheurer M, Mahoney D, Bertuch AA (2016) Thrombopoietin measurement as a key component in the evaluation of pediatric thrombocytosis. Pediatr Blood Cancer 63:1484–1487
90. Ghilardi N, Wiestner A, Kikuchi M, Ohsaka A, Skoda RC (1999) Hereditary thrombocythaemia in a Japanese family is caused by a novel point mutation in the thrombopoietin gene. Br J Haematol 107:310–316
91. Bisio A, Nasti S, Jordan JJ, Gargiulo S, Pastorino L, Provenzani A, Quattrone A, Queirolo P, Bianchi-Scarrà G, Ghiorzo P, Inga A (2010) Functional analysis of *CDKN2A*/p16^{INK4a} 5′-UTR variants predisposing to melanoma. Hum Mol Genet 19:1479–1491
92. Liu L, Dilworth D, Gao L, Monzon J, Summers A, Lassam N, Hogg D (1999) Mutation of the CDKN2A 5′ UTR creates an aberrant initiation codon and predisposes to melanoma. Nat Genet 21:128–132
93. Wen Y, Liu Y, Xu Y, Zhao Y, Hua R, Wang K, Sun M, Li Y, Yang S, Zhang X-J, Kruse R, Cichon S, Betz RC, Nöthen MM, van Steensel MAM, van Geel M, Steijlen PM, Hohl D, Huber M, Dunnill GS, Kennedy C, Messenger A, Munro CS, Terrinoni A, Hovnanian A, Bodemer C, de Prost Y, Paller AS, Irvine AD, Sinclair R, Green J, Shang D, Liu Q, Luo Y, Jiang L, Chen H-D, Lo WH-Y, McLean WHI, He C-D, Zhang X (2009) Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis. Nat Genet 41:228–233
94. Zhou Y, Koelling N, Fenwick AL, McGowan SJ, Calpena E, Wall SA, Smithson SF, Wilkie AOM, Twigg SRF (2018) Disruption of *TWIST1* translation by 5′ UTR variants in Saethre-Chotzen syndrome. Hum Mutat 39:1360–1365
95. Kitano S, Kurasawa H, Aizawa Y (2018) Transposable elements shape the human proteome landscape via formation of *cis*-acting upstream open reading frames. Genes Cells 23:274–284
96. McGillivray P, Ault R, Pawashe M, Kitchen R, Balasubramanian S, Gerstein M (2018) A comprehensive catalog of predicted functional upstream open reading frames in humans. Nucleic Acids Res 46:3326–3338
97. Sherr CJ, Roberts JM (1999) CDK inhibitors: positive and negative regulators of G_1-phase progression. Genes Dev 13:1501–1512
98. Somers J, Wilson LA, Kilday J-P, Horvilleur E, Cannell IG, Pöyry TAA, Cobbold LC, Kondrashov A, Knight JRP, Puget S, Grill J, Grundy RG, Bushell M, Willis AE (2015) A common polymorphism in the 5′ UTR of ERCC5 creates an upstream ORF that confers resistance to platinum-based chemotherapy. Genes Dev 29:1891–1896
99. Pellegata NS, Quintanilla-Martinez L, Siggelkow H, Samson E, Bink K, Höfler H, Fend F, Graw J, Atkinson MJ (2006) Germ-line mutations in p27^{Kip1} cause a multiple endocrine neoplasia syndrome in rats and humans. Proc Natl Acad Sci USA 103:15558–15563

100. Occhi G, Regazzo D, Trivellin G, Boaretto F, Ciato D, Bobisse S, Ferasin S, Cetani F, Pardi E, Korbonits M, Pellegata NS, Sidarovich V, Quattrone A, Opocher G, Mantero F, Scaroni C (2013) A novel mutation in the upstream open reading frame of the *CDKN1B* gene Causes a MEN4 phenotype. PLoS Genet 9:e1003350
101. Jones L, Goode L, Davila E, Brown A, McCarthy DM, Sharma N, Bhide PG, Armata IA (2017) Translational effects and coding potential of an upstream open reading frame associated with DOPA Responsive Dystonia. Biochim Biophys Acta Mol basis Dis 1863:1171–1182
102. Schulz J, Mah N, Neuenschwander M, Kischka T, Ratei R, Schlag PM, Castaños-Vélez E, Fichtner I, Tunn P-U, Denkert C, Klaas O, Berdel WE, von Kries JP, Makalowski W, Andrade-Navarro MA, Leutz A, Wethmar K (2018) Loss-of-function uORF mutations in human malignancies. Sci Rep 8:2395
103. Brown CY, Mize GJ, Pineda M, George DL, Morris DR (1999) Role of two upstream open reading frames in the translational control of oncogene mdm2. Oncogene 18:5631–5637
104. Sobczak K, Krzyzosiak WJ (2002) Structural determinants of BRCA1 translational regulation. J Biol Chem 277:17349–17358
105. Mehta A, Trotta CR, Peltz SW (2006) Derepression of the Her-2 uORF is mediated by a novel post-transcriptional control mechanism in cancer cells. Genes Dev 20:939–953

6 Alternative Mechanisms of mRNA Translation Initiation in Cellular Stress Response and Cancer

Rafaela Lacerda, Juliane Menezes, and Marco M. Candeias

Abstract

Throughout evolution, eukaryotic cells have devised different mechanisms to cope with stressful environments. When eukaryotic cells are exposed to stress stimuli, they activate adaptive pathways that allow them to restore cellular homeostasis. Most types of stress stimuli have been reported to induce a decrease in overall protein synthesis accompanied by induction of alternative mechanisms of mRNA translation initiation. Here, we present well-studied and recent examples of such stress responses and the alternative translation initiation mechanisms they induce, and discuss the consequences of such regulation for cell homeostasis and oncogenic transformation.

Keywords

mRNA translation · Alternative translation initiation · Cellular stress · Cancer · IRES · Non-AUG

6.1 Introduction

Most eukaryotic mRNAs are translated into proteins through a 5′-end m^7G cap-dependent translation mechanism. Translation initiation is marked by the formation of a ternary complex (TC) composed of eukaryotic initiation factor 2 (eIF2) bound to Met–$tRNA_i^{Met}$ and GTP [1]. eIF2B, a guanine nucleotide exchange factor (GEF), controls TC assembly by converting eIF2–GDP into the active eIF2–GTP complex before each round of translation [2]. Once properly assembled and active, the TC binds the 40S ribosomal subunit, forming the 43S pre-initiation complex (43S PIC). Several initiation factors, such as eIF1, eIF1A, eIF3 and eIF5, help this binding [3–8]. Separately, eIF4E binds the 5′-end m^7G cap of the mRNA and recruits eIF4G and eIF4A forming the eIF4F complex [8] and stimulating eIF4A's helicase activity, which promotes mRNA restructuring [9, 10]. eIF4G is the scaffold for the eIF4F components and binds to poly(A)-binding protein (PABP) and eIF3 at subunits c, d and e, helping recruit the 43S PIC to the transcript [11]. Following recruitment, 43S PIC will often require scanning downstream in order

Authors Rafaela Lacerda and Juliane Menezes have been contributed equally to this chapter.

R. Lacerda · J. Menezes
Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Lisboa, Portugal

M. M. Candeias (✉)
Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Lisboa, Portugal

MaRCU – Molecular and RNA Cancer Unit, Graduate School of Medicine, Kyoto University, Kyoto, Japan

L. Romão (ed.), *The mRNA Metabolism in Human Disease*, Advances in Experimental Medicine and Biology 1157, https://doi.org/10.1007/978-3-030-19966-1_6

to find the initiation codon [12, 13], unless this one is within close reach, in which case scanning is unnecessary and instead a specific Kozak sequence termed TISU can help prevent "leakage" to a downstream AUG [14, 15]. In most cases, however, the AUG is relatively far from the 5′-end and the 5′ untranslated region (UTR) is at least mildly structured, so scanning of – or jumping over, during a phenomenon named ribosomal shunting [16] – the 5′ UTR by the 43S PIC is often a requisite for translation initiation and typically entails the hydrolysis of ATP, eIF1, eIF1A, and DHX29 [3, 17]. ATP hydrolysis can be used by eIF4A to unwind secondary structures in the mRNA, and actively displace the ribosome in a 5′ to 3′ direction. At the same time, the ribosome is prevented from backsliding, because the unwound structures behind it resume their initial winding conformation [18, 19]. Scanning, when necessary, usually stops when the 43S PIC reaches the first AUG codon positioned in a favourable Kozak context (a purine, usually adenine, in position −3 and a guanine in position +4) [12]. At this point the 48S pre-initiation complex (48S PIC) is formed, with pairing of all three nucleotides of the anticodon and release of eIFs from the small subunit. Upon recognition of the initiation site, eIF5 and eIF5B promote the hydrolysis of eIF2–bound GTP [20, 21]. eIF5B–GTP binds to the 40S subunit and stimulates the 60S subunit joining, requiring a second step of GTP hydrolysis in order to make the 80S ribosome, which will be ready to start decoding the message [22, 23]. When this happens, eIF2 is completely released from the ribosome in its GDP-bound form that latter must be reverted to GTP-bound again to allow reassembly of the TC for another round of translation. The formed eIF2–GTP is only stable when Met–$tRNA_i^{Met}$ joins in to form the TC [2].

Although the translation process may be hindered at several stages, the translation initiation phase is the rate-limiting step and involves more factors and sub-steps that are prone to error, and many of them may be inactivated, modified or adjusted under adverse cell conditions [24]. In fact, translation initiation is globally impaired under stress conditions and overall protein synthesis is reduced, a response that has been termed Integrated Stress Response (ISR). Such reduction can happen due to eIF4E's or eIF2's inability to bind eIF4G/5′-end cap or integrate the TC, respectively. The first occurs for example during stress conditions that inhibit the mTOR proliferation and survival pathway; as inactivated mTOR kinase is no longer able to phosphorylate 4E-binding proteins (4E-BP) and the resulting hypophosphorylated 4E-BP binds to eIF4E preventing its association with eIF4G and the formation of eIF4F. The core event of the ISR however, is the second mechanism, involving phosphorylation of serine 51 of the eIF2α subunit by any of several protein kinases activated by a wide range of different stress conditions. This phosphorylation stabilizes the interaction between eIF2B and eIF2–GDP and prevents the formation of the eIF2–GTP–Met–$tRNA_i^{Met}$ complex (TC).

Repression of global protein synthesis is often accompanied by selective translation of mRNA encoding crucial stress-responsive proteins that can lead to either stress recovery and survival or cell death. This selective translation involves alternative initiation elements, often RNA structures or modifications, and may not require some of the more common elements such as the 5′-end m^7G cap. Genes that can maintain their expression under stress conditions may contribute to modify cell fate if they operate on repair, survival or programmed death pathways. Many cancer cells take advantage of this ability in transcripts such as XIAP, HIF1α or VEGF to escape apoptosis, resist to hypoxic conditions, or vascularise the tumour surroundings, respectively. It is well reported that many of these stress-response transcripts can maintain expression because of the alternative mechanisms of initiation that mediate their translation. Out of these mechanisms, internal ribosome entry at internal ribosome entry sites (IRES) has been the most widely studied and is accepted as a backup mechanism for cells to cope with conditions when canonical translation is shut down.

In this chapter, we aim to compare the different mechanisms of translation initiation involved in stress-response. We will also briefly consider to what extent these mechanisms may create an

adaptive advantage to the eukaryotic cell under stress conditions or how sometimes this advantage turns into a burden to the organism when coupled with processes of cellular transformation.

6.2 Translation Initiation in Stress Response

Throughout evolution, eukaryotic cells have devised different mechanisms to cope with stressful environments. When eukaryotic cells are exposed to stress stimuli, they activate a common adaptive pathway that allows them to restore cellular homeostasis: the integrated stress response (ISR) [25]. Most types of stress stimuli have been reported to activate the ISR [26, 27], including hypoxia [28], nutrient deprivation [29], oxidative stress, heat shock [30], viral infection (dsRNA) [31], endoplasmic reticulum (ER) stress / unfolded protein response (UPR) [32], UV irradiation [33], proteasome inhibition [34] and oncogene activation [35]. They all lead to the phosphorylation of eIF2α and consequent inhibition of TC formation, by activating kinase haem-regulated inhibitor (HRI), protein kinase activated by double-stranded RNA (PKR), general control non-derepressible-2 (GCN2) or PKR-like endoplasmic reticulum kinase (PERK) [26, 36]. As a consequence, there is a decrease in overall protein synthesis. However, several selected genes that participate in cellular stress response are still translated under such conditions [26, 37]. Below, we discuss some well-studied examples of stress stimuli, the inhibition they induce and the alternative mechanisms they activate, as well as the consequences of such regulation for cell homeostasis and transformation.

6.2.1 Endoplasmic Reticulum Stress

The endoplasmic reticulum (ER) is a central organelle in which proteins are translated and properly folded, may undergo some post-translational modification and from there are sent to the Golgi complex and sites of action. When perturbations in ER homeostasis occur, causing an accumulation of unfolded proteins in the ER, these perturbations are sensed and transduced to the whole cell and an evolutionarily conserved response is activated—the unfolded protein response (UPR) [38–40]. The UPR consists of three branches, each of which can be distinguished by the action of a different stress sensor protein: inositol-requiring protein-1α (IRE1α), protein kinase RNA (PKR)-like ER kinase (PERK) and activating transcription factor 6 (ATF6) [39, 40]. In normal conditions, the UPR modulators IRE1, PERK and ATF6 interact with molecular chaperone Binding immunoglobulin protein (Bip)/GRP78 and remain inactive [39–42]. However, when unfolded proteins accumulate in the ER, BiP is required to assist with the folding and the sensor proteins are liberated and activated. Other proteins have also been shown to regulate IRE1α and the other sensors [43]. Upon activation, all three stress sensors induce signal transduction in order to deal with the stress and reduce the amount of misfolded proteins. This includes a tight reprogramming of transcription and translation to ensure less and more specific gene expression [44]. The key regulatory pathway for this response is the PERK/eIF2α/ATF4 pathway in which the kinase PERK phosphorylates eIF2α (eIF2α-P). This signal then induces overall translational impairment, but it also enables translation of the transcription factor ATF4 mRNA [44]. ATF4 is a transcription factor whose translation has been shown to be regulated through reinitiation at upstream open reading frames (uORF) [45] and also by an Internal ribosome entry site (IRES) stimulated by eIF2α phosphorylation [46] (see also below section "Translation initiation by internal ribosome entry" and Fig. 6.1).

Translation of HIAP2 (a member of the inhibitor of apoptosis protein family) is also mediated via an inducible IRES element during ER stress [47]. HIAP2 IRES activity is enhanced during ER stress through the caspase-mediated proteolytic processing of eukaryotic initiation factor p97/DAP5/eIF4G2/NAT1 (DAP5), which produces

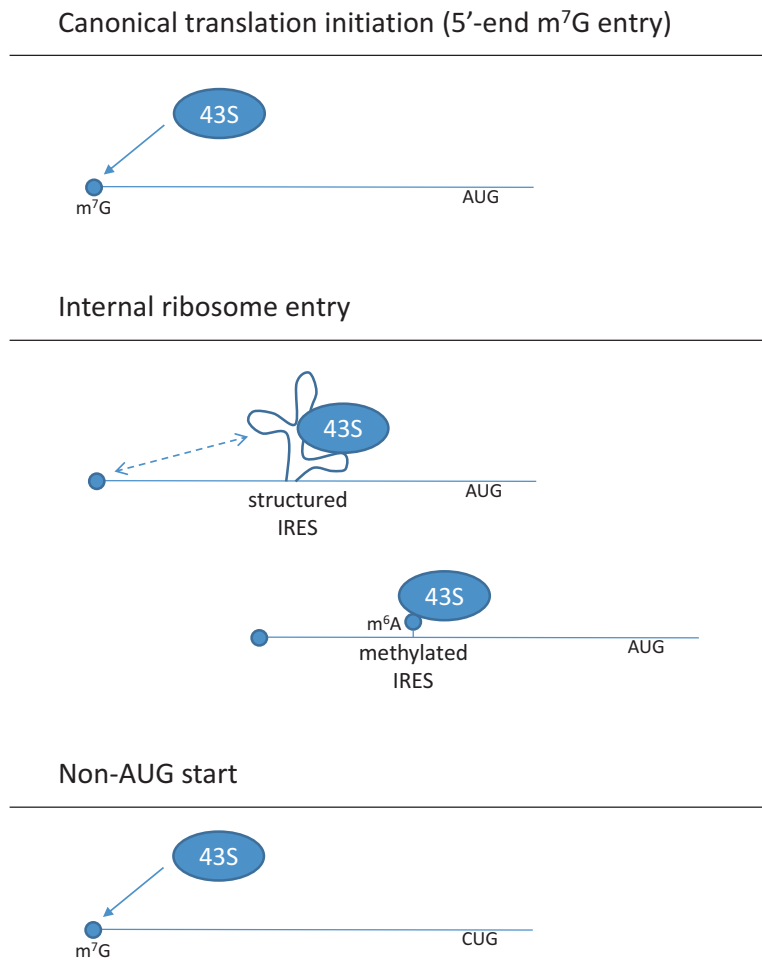

Fig. 6.1 During canonical translation initiation the ribosome is recruited to the 5′-terminal m^7G cap (top). Cell stress and cancer induce alternative methods of translation initiation such as internal ribosome entry at a structured internal ribosome entry site (IRES) or at a methylated site (center). The 5′-end may sometimes still be required for this initiation (indicated by the dashed arrow). Another type of alternative initiation involves the use of non-AUG start codons (bottom)

a fragment that specifically activates IRES [47]. In fact, DAP5 is a translation initiation factor that can regulate the expression of a selected group of mRNAs during ER stress via internal ribosome entry [48, 49]. Other DAP5 targets include IRES in c-Myc, Apaf-1, XIAP and HMGN3 mRNA. In this regard, DAP5-mediated translation seems to be crucial in cell differentiation and death, stressing the role of IRES translation in deciding cell fate [49]. Furthermore, DAP5 translation itself is mediated by an IRES element within its 5′ UTR, thus generating a positive self-regulatory loop that allows its continuous translation under conditions impairing 5′ end ribosome entry–dependent translation [49, 50].

6.2.2 Hypoxia

Hypoxia results from the decrease in the oxygen available to reach the different organs and cells. Under hypoxic stress, a group of transcription factors known as hypoxia inducible factors (HIF) are stabilized and initiate a cascade of cell signals by activating target genes in the nucleus [51]. Inhibition of protein synthesis and consequent energy saving is an advantage for hypoxic cells, so, in order to achieve this, canonical translation initiation is drastically reduced under such conditions. However, hypoxic cells need to translate mRNAs critical for an adaptive response to low oxygen levels [52]. To accomplish this selective

translation, cells use non-canonical mechanisms of translation initiation, such as IRES-mediated translation. IRES elements present within the 5 UTRs of several transcripts have been proved to mediate translation of stress-regulated mRNAs, such as vascular endothelial growth factor (VEGF), HIF-1α and -2α, glucose transporter-like protein 1, p57(Kip2), La, BiP, and triose phosphate isomerase (TPI) transcripts [52]. Many proteins share such characteristics. Phosphofructokinase 1 (PFK1), the major regulatory enzyme of the glycolytic pathway, converts fructose-6-phosphate to fructose-1,6-bisphosphate. This pathway is highly dynamic and may be affected by different stress conditions, such as hypoxia, which is known to significantly influence the glycolytic pathway [53, 54]. Ismail and colleagues showed that PFK1's 5′ UTR includes a hypoxia-responsive IRES element, which was established to be the possible mechanism responsible for PFK1 protein upregulation [55]. These authors found that, after 48 h of chemically induced hypoxia in C6 glioma cells, PFK1 protein levels were upregulated with no significant change in the counterpart mRNA levels; this may explain the astrocytes' increased glycolytic capacity upon brain hypoxia [55].

6.2.3 Starvation

In the absence of nutrients, amino acids, growth factors and cytokines, which contribute to the activation of signal pathways related to cell survival and proliferation, the ISR facilitates cellular adaptation to stress conditions through the common target eIF2α [56]. Under starvation, there have been examples of leaky scanning and reinitiation events through which cells can make alternative protein products by selecting downstream initiation codons to better respond to the stress [56]. Reinitiation in ATF4, for example, is governed by the eIF2α pathway and also subjected to regulation by mRNA m^6A methylation. Zhou et al. demonstrated that m^6A in the 5′ UTR controls ribosome scanning and start codon selection [56].

IRES-mediated translation initiation is the most common alternative to canonical translation under starvation. One of the most widely used models to understand this mechanism in eukaryotes is the X-linked inhibitor of apoptosis (XIAP) mRNA [57]. When ternary complex (TC) is available, XIAP mRNA translation is maintained in a 5′-end m^7G cap-dependent mode; however, under serum deprivation, the XIAP IRES can initiate translation in an alternative eIF5B-dependent manner circumventing low TC numbers due to eIF2α phosphorylation [58]. Notably, not all cellular IRES use eIF5B-dependent mode of tRNA delivery during serum deprivation [27].

6.2.4 DNA Damage

Apaf-1 has a central role in DNA damage-induced apoptosis and its depletion contributes to malignant transformation [59]. As such Apaf-1 provides a good example of specialized DNA damage translation since the human Apaf-1 mRNA can initiate translation through an alternative mechanism, possibly involving an IRES [60, 61]. This internal ribosome entry site has been reported to require the assistance from a free 5′-end *in cis* [62], though it does not need m^7G cap recognition by eIF4E [60]. Under DNA damage conditions m^7G cap-binding factor eIF4E is often suppressed, but structured 5′ UTR regions such as IRES may mediate m^7G cap-independent, 5′-end-dependent translation initiation, which leads to preferential translation of some mRNAs like Apaf-1 [60]. A group of mRNAs including 53BP1, HIF1α, BRCA-1, and GADD45a, has also been shown to be more actively translated in response to DNA damage in breast cancer cells, through a selective eIF4G1-dependent process and with reduced dependence on eIF4E [63].

6.2.5 Heat Shock

Although we can find living organisms in a wide range of temperatures (from the freezing point of water, or below, to 113 °C) [64], each of them has

adapted to a certain optimal growth temperature. Heat above such temperatures becomes a major stressor, with often temperatures only moderately above the optimum growth temperature already causing a significant barrier to survival [65]. Heat stress can cause protein unfolding, entanglement, and unspecific aggregation [65]. During cellular heat shock response, a class of molecular chaperones, the heat shock proteins (Hsp), is upregulated in response to protein misfolding [65]. Heat shock goes beyond the unfolding of individual proteins, causing deleterious effects on the internal organisation of the cell, such as defects of the cytoskeleton [65].

Translation of Hsp70, a stress-induced molecular chaperone that also modulates tumour cell responses to cytotoxic agents and inhibits apoptosis [66], is mediated through internal ribosome entry, most likely due to increased m^6A modifications in its mRNA [67]. The increasing number of m^6A-containing transcripts results from the exposure to different cellular stresses that drive a widespread redistribution of m^6A [67]. A recent study by Coots et al. established the effect of m^6A in the 5′ UTR translation initiation [68]. They showed that when eIF4F-dependent translation is impaired, cells use the m^6A-dependent mode of translation initiation. They identified the ATP-binding cassette subfamily F member 1 (ABCF1) as a critical mediator of m^6A-dependent translation. This protein acts as an alternative to recruit the ternary complex during non-canonical translation as it can interact with eIF2 and ribosomes, thus playing a critical role in mRNA translation initiation under stress conditions [69, 70]. The HSP70 5′ UTR has also been shown to drive m^7G cap-independent translation via an IRES structure [71]; however, it is not yet known whether/how both features cooperate to enhance translation of heat shock-responsive proteins. Translation of the BiP protein was also found to be enhanced by continuous heat stress that activates an IRES-dependent translation [72]. This suggests that the IRES-dependent mechanism of translation initiation can be used by cells subjected to heat shock, being critical to cell survival and proliferation under stress [72].

Recently, the switch to activate a specialized ribosome for alternative translation of stress response genes was shown to be regulated by an alternative translation initiation process itself [73]: Mitochondrial ribosome protein-encoding MRPL18 mRNA was shown to translate into a shorter isoform from a downstream non-canonical CUG initiation codon following exposure to heat shock and phosphorylation of eIF2α. The shorter isoform is translated in frame but lacks the mitochondrial targeting signal in the N-terminus and is localized to the cytoplasm where it integrates – not the usual mitoribosome but now – the cytosolic 80S ribosomes promoting the specific synthesis of stress-proteins such as Hsp70.

6.2.6 Oxidative Stress

Under oxidative stress, NRF2, a master regulator of the oxidative stress response, is also translationally induced through an IRES [74–76]. NRF2 IRES-dependent translation is enhanced due to stimulation of an IRES element present within its 5′ UTR by La autoantigen IRES transacting factor (ITAF) binding [76]. Translation of some transcription factors is also mediated by IRES elements upon oxidative and genotoxic stress, such as p53, the octamer-binding protein 4 (OCT4) whose translation is stimulated by H_2O_2 treatment in breast cancer and liver carcinoma cells [77], and runt-related transcription factor 2 (RUNX2), whose translation is stimulated by mitomycin C [78]. BiP is also induced during oxidative stress through a mechanism that involves an alternative, less usual procedure for translation initiation, the usage of a non-canonical initiation codon, UUG, in a uORF [79]. The UUG-initiated uORF in the 5′ UTR of BiP and eIF2A were shown to be necessary for BiP expression during oxidative stress. At the same time, the uORFs generate peptides that could serve as major histocompatibility complex class I ligands.

6.3 Mechanisms of Alternative Translation Initiation

6.3.1 Translation Initiation by Internal Ribosome Entry

Internal ribosome entry site (IRES)-mediated translation is an additional and alternative mode of translation initiation used by some transcripts, which can be regulated independently of the canonical system. IRES-mediated translation has been extensively investigated. It has been estimated that at least 10–15% of cellular mRNAs can be translated by an IRES-dependent mechanism, in which the 40S ribosomal subunits bind the transcript internally, not at the 5′-end, often in a cap-independent manner and with a different requisite of translation factors [80]. We subdivided here this alternative internally initiated translation into two large groups: one more comprehensively studied, involving a structured RNA region usually just referred to as IRES; and a second more recently identified mechanism for internal entry that involves a methylated RNA site (Fig. 6.1). To note that it is still currently unknown how often actually these two mechanisms might overlap.

6.3.1.1 Internal Ribosome Entry @ Structured RNA Regions

According to a recent systematic screen for IRES-mediated translation activity, about 10% of all human 5′ UTRs have the potential to be IRES-translated [81], and these can be present in the coding region as well [82]. This translation initiation mechanism allows cells to cope with environmental changes affecting their viability, and thus must be essential for cellular life itself. In order to understand how central and elementary the mechanisms that govern IRES function are, Colussi et al. investigated if IRES could initiate translation in bacteria and saw that the IRES element could bind directly to both eukaryotic and bacterial ribosomes by occupying the space normally used by tRNAs [83]. Some IRES use dynamic RNA structures to target core and conserved, ancient domains of the translation machinery while circumventing organism-specific regulations to effectively initiate mRNA translation of specific transcripts in a large variety of cell types and cell conditions and with few requirements [83]. This is important because proteins with crucial roles in main cellular processes need backup regulation, their expression levels must be adjusted in response to external cues that impair the canonical mechanism of translation initiation. Indeed, alterations in their expression levels may account for many types of human diseases that arise in human population, including different types of cancer, and IRES-dependent translation initiation may play a decisive role in such processes.

From our current knowledge, most structured IRES described so far were identified in transcription factor mRNAs (21%), growth factor transcripts (15%), and in messages encoding transporters, receptors and channels (22%) [9]. FGF and VEGF families of proteins—growth factors of crucial importance to the development of specific tissues that play a significant role in promoting cell proliferation and differentiation, and in regulating cell survival—are translated via IRES elements present in the corresponding mRNAs [84–88]. As for transporters, receptors and channels, such as CAT-1, voltage-gated potassium channel and oestrogen receptor α, among others, play a critical role in signal transduction as they are main vehicles in cell-cell communication, which turns them into key elements to maintain cell homeostasis following environmental changes. Thus, alterations in their expression associate with changes in cellular function, which may lead to disease development and progression [84]. That is why transcripts encoding such proteins can be translated through an IRES-dependent mechanism that acts as a back-up tool when canonical translation initiation is impaired by environmental stress conditions, such as ER stress. Regarding transcription factors, they are fundamental in gene expression regulation, as they respond to quick changes in the environment in order to adapt their expression levels to a given context—c-MYC, HIF1α and p53 are good examples of transcription factors whose translation initiation is mediated by IRES elements [85, 86, 89].

6.3.1.2 Internal Ribosome Entry @ Methylated Regions

There are about 3–5 m^6A modifications per transcript [90]. Most m^6A modifications are located on the coding region and the 3′ UTR [91, 92]; however, Meyer et al. found recently that, when located in the 5′ UTR, such modifications could mediate translation initiation through internal ribosome entry [93] (Fig. 6.1). These authors have shown that m^6A modifications in the 5′ UTR act as ribosome engagement sites (MIRES) [93]. The ability of m^6A in the 5′ UTR to bind eIF3 is enough to recruit the 40S ribosomal subunit to initiate translation when eIF4E is not available to bind the 5′-end cap structure [93]. Although the mechanism that permits m^6A recognition by the translation machinery for subsequent m^7G-independent initiation is not yet completely understood, the significance of 5′ UTR m^6A residues has been observed in both ribosome profiling datasets and individual cellular mRNA analyses, such as the heat-shock protein 70 (Hsp70) [93, 94].

6.3.2 Non-AUG Translation

Another less common alternative translation process is the initiation at non-AUG codons (for a recent review see [95]) (Fig. 6.1). These translation starts are not errors but regulated events, leading to the production of stress response proteins as well as proteins involved in development. Impairment of this specialized translation process may lead to diseases such as cancer and neurodegeneration. It is increasingly clear that non-AUG translation is highly regulated by stress, signalling, translation factors and RNA structures and sequences. The most commonly used non-AUG codons are the near-cognate that differ by only one nucleotide, with CUG being most frequently used. Some proteins, like DAP5, are exclusively translated from a non-AUG (GUG) codon [96]. Interestingly, several proteins involved in activating alternative translation seem to be first translated from an alternative initiation codon: DAP5 regulates IRES-translation [97] and a CUG-translated shorter version of MRPL18 protein activates specialized translation of stress response genes by integrating the cytoplasmic 80S ribosome [73]. Non-AUG translation can be influenced by eIF2α phosphorylation during ISR, Kozak sequence, mRNA structures upstream or downstream or the expression of specific stress-induced translation factors like eIF2A.

6.4 Translation Initiation in Cancer

Oncogene activation and tumour suppressor gene inhibition are key events to the onset and development of cancer. Additionally, coding-independent mutations in regulatory elements, UTRs, splice sites and non-coding RNAs and synonymous mutations may also affect gene expression (reviewed in [98]). As any other stress situation, tumorigenesis includes backup mechanisms that allow tumour cells to cope with stress, such as those involved in stress-adaptive protein synthesis [48, 99–103]. Many transcripts relevant to cancer can initiate translation through non-canonical translation initiation mechanisms. Below we will briefly present a few well-known examples.

6.4.1 Non-AUG Translation in Cancer

Though IRES-dependent translation is the most widely studied, we will start by discussing initiation at non-AUG sites in cancer. Some of these sites have been shown to be regulated by eIF2A, which stimulates translation from non-AUG uORFs in cancer-related mRNAs that act to positively regulate the expression of their downstream ORFs [104]. eIF2A can initiate Leu-encoding codons at CUG and UUG by promoting the recruitment of LeutRNACUG for initiation [79, 105, 106]. Interestingly, mutation and inactivation (decreased expression) of the RNA helicase DDX3 in cancer leads to the formation of RNA structures in 5′ UTR downstream of non-AUG initiation codons, inducing their usage [107].

PTEN, a tumour suppressor gene that is frequently mutated in cancer, generates an alternative protein isoform, PTENβ, by using an upstream non-canonical AUU start codon [108]. PTENβ translation requires a favourable Kozak sequence and an evolutionarily conserved hairpin that is present 18 nucleotides downstream from the AUU. PTENβ negatively regulates rDNA transcription and cell proliferation and may have a role in tumour suppression [108].

FGF2 stimulates the growth and development of new blood vessels (angiogenesis) that contribute to the pathogenesis of cancer. In FGF2 mRNA, at least four upstream in-frame CUG codons were shown to generate longer isoforms that localized to the nucleus due to a nuclear localization signal present between the CUG and AUG codons [109, 110]. These isoforms could possibly affect cell growth and differentiation and one of them of 34 KDa enhanced survival in low serum conditions.

c-myc, a well-known protooncogene, uses an upstream in-frame CUG codon resulting in the production of a larger isoform with a distinct N-terminus that may contribute to the oncogenicity of c-myc, particularly in Burkitt's lymphoma [111]. As cell density increases, as during tumour formation and growth, the availability of amino acids, specifically methionine, becomes limiting, and translation initiation from the CUG is promoted [112].

6.4.2 Internal Ribosome Entry in Cancer

IRES elements within the 5′ UTR or coding region of transcripts encoding oncogenes, growth factors and proteins involved in the regulation of cell-cycle and programmed cell death, can mediate translation under stress situations triggered by the tumour's microenvironment, contributing to the survival of cancer cells [9, 99, 103, 113]. A common feature of these environments is the difficult access to oxygen. Indeed, cancer cells activate 4E-binding proteins (4E-BP) and inhibit the mTORC1 pathway in response to hypoxic conditions, but at the same time promote a switch to IRES-mediated translation thus maintaining tumour growth and angiogenesis [100, 102]. Large and advanced breast cancers were shown to overexpress 4E-BP and eIF4G and trigger m^7G-independent mRNA translation [102]. In inflammatory breast cancer, cells have adapted to a state of prolonged hypoxia and optimised the production of proteins required for tumour embolus survival and dissemination, a state promoted by high levels of eIF4GI protein coupled with a constitutively active 4E-BP1 [114]. This leads to higher rates of translation in IRES-containing mRNAs, namely VEGF and p120 catenin, which maintain high rates of angiogenesis, and membrane associated E-cadherin, respectively [115].

FGF (Fibroblast growth factors), such as *FGF1* and *FGF2*, are crucial for proliferation and differentiation of a wide variety of cells, and hence their translation has to be tightly regulated—some of them contain IRES elements within their 5′ UTRs, which allow cap-independent translation initiation [116, 117]. IRES-mediated regulation of FGF2 translation is considered a critical step in tumorigenesis, not only in solid tumours but also in multiple myeloma, which turns the FGF2 IRES into the non-cytotoxic primary molecular target of thalidomide, and therefore the preferred target of immunomodulatory drugs in multiple myeloma [114].

c-Myc IRES is also activated in multiple myeloma cells under thapsigangin- or tunicamycin-induced ER stress, or bortezomib (a myeloma therapeutic) treatment, thus maintaining c-Myc protein levels [118].

Sp1 (Specificity protein-1), a protein that is accumulated under hypoxic conditions in an IRES-dependent manner in lung tumour tissue, is another case of a protein whose expression is upregulated during tumorigenesis by activation of IRES-mediated translation, suggesting that translational regulation might contribute to the accumulation of Sp1 during tumorigenesis [119].

CAT-1 synthesis and sodium-coupled neutral amino acid transporter 2 (SNAT2), two amino acid transporters, is controlled by IRES under amino acid or glucose starvation [120, 121].

CAT-1 IRES-dependent translation is induced in tumour cells under glucose deprivation through phosphorylation of eIF2α by the transmembrane endoplasmic reticulum kinase (PERK) [121].

XIAP and sterol regulatory element-binding transcription factor 1 (SREBP-1) are translated via IRES during deprivation of growth factors in tumour cells, protecting them from apoptosis [122, 123]. IRES-dependent translation of these proteins allows the cell to survive under nutritional stress, which is an advantage for cancer cell continued existence [124]. Adding to this, XIAP expression is up-regulated under γ-irradiation through IRES-dependent translation, causing tumour cells to be resistant to radiotherapy [125, 126]. This agrees with the study by Holcik et al., in which they used RNA interference to inhibit XIAP, and saw that it enhances chemotherapeutic drug sensitivity and decreases myeloma cell survival [126].

β-catenin is also translated through internal ribosome entry in human ovarian cancer cells treated with paclitaxel (PTX), a chemotherapeutic drug used in the treatment of ovarian cancer—this then regulates the expression of downstream factors (c-Myc and cyclin D1), reducing PTX sensitivity [127].

c-Myc oncogenic transcription factor and Bcl2-associated athanogene 1 (BAG-1) are also regulated by IRES under oxidative and genotoxic stress and increase tumour cells' resistance to DNA damage-inducing drugs [128–131]. Bcl-2's expression in turn, and cIAP's are again enhanced by etoposide as well as arsenite *via* IRES-mediated translation [132, 133].

p53—a tumour suppressor, protooncogene [134] and transcriptional master regulator of the oxidative and genotoxic stress responses—is also translated through IRES-mediated processes [82, 135–138]. The p53 transcript contains IRES structures that control translation of the full-length (FL) p53 and the N-terminally truncated isoform Δ40p53 from the same mRNA [82, 135]. Several stress conditions that induce DNA damage as well as ER stress induce FLp53 or Δ40p53, respectively, *via* two different IRES structures, one in the 5′ UTR – for FL – and another in the 5′ coding region – controlling mostly Δ40p53 [138]. Furthermore, in response to doxorubicin, IRES-mediated translation of both p53 isoforms is stimulated by the ITAF polypyrimidine tract-binding protein (PTB), following PTB relocation from the nucleus to the cytoplasm [139]. Other ITAFs such as DAP5, Annexin A2, and PTB-associated Splicing Factor (PSF), have also been reported to control p53 IRES activity [140, 141]. Besides, identification of two other p53 ITAFs [translational control protein 80 (TCP80) and RNA helicase A (RHA)] that positively regulate p53 IRES activity, established a connection between IRES-mediated p53 translation and p53 tumour suppressive function in two breast cancer cell lines. Following DNA damage, the levels of TCP80 and RHA are extremely low and these two cell lines exhibited defective p53 induction and synthesis, since expression of both proteins was required to significantly increase p53 IRES activity [142, 143]. Cells devised a critical cellular response that counteracts cellular transformation—the oncogene-induced senescence (OIS)—which is characterized by cell cycle arrest and induction of p53, which prevents the proliferative potential of preneoplasic clones [144]. During OIS, there is a switch from canonical translation initiation to IRES-mediated translation, during which p53 IRES-dependent translation is promoted, providing a molecular barrier for cellular transformation [145].

In conclusion, and considering the aforementioned examples, it seems clear that the IRES-mediated translation of key regulators and pro-survival factors grant tumour cells enough tools for attaining resistance to chemotherapy and radiation [146]. On the other hand, the presence of IRES within transcripts coding tumour suppressor proteins can prevent cancer outbreak by maintaining the protein levels. Expression of some proteins is crucial to determine the cell fate under stress conditions—apoptosis or survival and proliferation. Thus, IRES-mediated translation is of key importance in the process of tumorigenesis. Furthermore, the IRES structures themselves and the cooperating ITAFs are vital targets for cancer treatment.

Acknowledgments Marco M Candeias was partially supported by grants PTDC/MED-ONC/32048/2017 and PTDC/BIMONC/4890/2014 from Fundação para a Ciência e a Tecnologia (FCT), by Grants-in-Aid 16K21111 and 18K07229 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, by Takeda Foundation and Astellas Foundation. Juliane Menezes is a posdoc fellow (SFRH/BPD/98360/2013) from FCT.

References

1. Erickson FL, Hannig EM (1996) Ligand interactions with eukaryotic translation initiation factor 2: role of the gamma-subunit. EMBO J 15:6311–6320
2. Gomez E, Mohammad SS, Pavitt GD (2002) Characterization of the minimal catalytic domain within eIF2B: the guanine-nucleotide exchange factor for translation initiation. EMBO J 21:5292–5301. https://doi.org/10.1093/emboj/cdf515
3. Pestova TV, Borukhov SI, Hellen CUT (1998) Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. Nature 394:854–859. https://doi.org/10.1038/29703
4. Valásek L, Nielsen KH, Hinnebusch AG (2002) Direct eIF2-eIF3 contact in the multifactor complex is important for translation initiation in vivo. EMBO J 21:5886–5898
5. Olsen DS, Savner EM, Mathew A et al (2003) Domains of eIF1A that mediate binding to eIF2, eIF3 and eIF5B and promote ternary complex recruitment in vivo. EMBO J 22:193–204. https://doi.org/10.1093/emboj/cdg030
6. Majumdar R (2003) Mammalian translation initiation factor eIF1 functions with eIF1A and eIF3 in the formation of a stable 40 S preinitiation complex. J Biol Chem 278:6580–6587. https://doi.org/10.1074/jbc.M210357200
7. Kolupaeva VG (2005) Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association. RNA 11:470–486. https://doi.org/10.1261/rna.7215305
8. Pestova TV, Lorsch JR, Hellen CUT (2007) The mechanisms of translation initiation in eukaryotes. In: Mathews M, Sonenberg N, Hershey JWB (eds) Translational control in biology and medicine. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
9. Lacerda R, Menezes J, Romão L (2017) More than just scanning: the importance of cap-independent mRNA translation initiation for cellular stress response and cancer. Cell Mol Life Sci 74:1659–1680. https://doi.org/10.1007/s00018-016-2428-2
10. Feoktistova K, Tuvshintogs E, Do A, Fraser CS (2013) Human eIF4E promotes mRNA restructuring by stimulating eIF4A helicase activity. Proc Natl Acad Sci 110:13339–13344. https://doi.org/10.1073/pnas.1303781110
11. Villa N, Do A, Hershey JWB, Fraser CS (2013) Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome. J Biol Chem 288:32932–32940. https://doi.org/10.1074/jbc.M113.517011
12. Kozak M (1989) The scanning model for translation: an update. J Cell Biol 108:229–241
13. Kozak M (2002) Pushing the limits of the scanning mechanism for initiation of translation. Gene 299:1–34
14. Kozak M (1995) Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. Proc Natl Acad Sci U S A 92:2662–2666
15. Dikstein R (2012) Transcription and translation in a package deal: the TISU paradigm. Gene 491:1–4. https://doi.org/10.1016/j.gene.2011.09.013
16. Yueh A, Schneider RJ (1996) Selective translation initiation by ribosome jumping in adenovirus-infected and heat-shocked cells. Genes Dev 10:1557–1567
17. Pestova TV, Kolupaeva VG (2002) The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. Genes Dev 16:2906–2922. https://doi.org/10.1101/gad.1020902
18. Kapp LD, Lorsch JR (2004) The molecular mechanics of eukaryotic translation. Annu Rev Biochem 73:657–704. https://doi.org/10.1146/annurev.biochem.73.030403.080419
19. De La Cruz J, Iost I, Kressler D, Linder P (1997) The p20 and Ded1 proteins have antagonistic roles in eIF4E-dependent translation in Saccharomyces cerevisiae. Proc Natl Acad Sci 94:5201–5206
20. Hinnebusch AG (2011) Molecular mechanism of scanning and start codon selection in eukaryotes. Microbiol Mol Biol Rev 75:434–467. https://doi.org/10.1128/MMBR.00008-11
21. Kuhle B, Ficner R (2014) eIF5B employs a novel domain release mechanism to catalyze ribosomal subunit joining. EMBO J 33:1177–1191. https://doi.org/10.1002/embj.201387344
22. Lee JH, Pestova TV, Shin B-S et al (2002) Initiation factor eIF5B catalyzes second GTP-dependent step in eukaryotic translation initiation. Proc Natl Acad Sci U S A 99:16689–16694. https://doi.org/10.1073/pnas.262569399
23. Shin B-S, Maag D, Roll-Mecak A et al (2002) Uncoupling of initiation factor eIF5B/IF2 GTPase and translational activities by mutations that lower ribosome affinity. Cell 111:1015–1025
24. Jackson RJ, Hellen CU, Pestova TV (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol 11:113–127. https://doi.org/10.1038/nrm2838
25. Ron D (2002) Translational control in the endoplasmic reticulum stress response. J Clin Invest 110:1383–1388. https://doi.org/10.1172/JCI200216784

26. Pakos-Zebrucka K, Koryga I, Mnich K et al (2016) The integrated stress response. EMBO Rep 17:1374–1395. https://doi.org/10.15252/embr.201642195
27. Holcik M (2015) Could the eIF2α-independent translation be the Achilles heel of cancer? Front Oncol 5:264. https://doi.org/10.3389/fonc.2015.00264
28. Rzymski T, Milani M, Pike L et al (2010) Regulation of autophagy by ATF4 in response to severe hypoxia. Oncogene 29:4424–4435. https://doi.org/10.1038/onc.2010.191
29. Roussou I, Thireos G, Hauge BM (1988) Transcriptional-translational regulatory circuit in Saccharomyces cerevisiae which involves the GCN4 transcriptional activator and the GCN2 protein kinase. Mol Cell Biol 8:2132–2139
30. Lu L, Han A-P, Chen J-J (2001) Translation initiation control by heme-regulated eukaryotic initiation factor 2alpha kinase in erythroid cells under cytoplasmic stresses. Mol Cell Biol 21:7971–7980. https://doi.org/10.1128/MCB.21.23.7971-7980.2001
31. Taylor DR, Lee SB, Romano PR et al (1996) Autophosphorylation sites participate in the activation of the double-stranded-RNA-activated protein kinase PKR. Mol Cell Biol 16:6295–6302
32. Harding HP, Zhang Y, Ron D (1999) Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. Nature 397:271–274. https://doi.org/10.1038/16729
33. Wu S, Hu Y, Wang J-L et al (2002) Ultraviolet light inhibits translation through activation of the unfolded protein response kinase PERK in the lumen of the endoplasmic reticulum. J Biol Chem 277:18077–18083. https://doi.org/10.1074/jbc.M110164200
34. Yerlikaya A, Kimball SR, Stanley BA (2008) Phosphorylation of eIF2alpha in response to 26S proteasome inhibition is mediated by the haem-regulated inhibitor (HRI) kinase. Biochem J 412:579–588. https://doi.org/10.1042/BJ20080324
35. Denoyelle C, Abou-Rjaily G, Bezrookove V et al (2006) Anti-oncogenic role of the endoplasmic reticulum differentially activated by mutations in the MAPK pathway. Nat Cell Biol 8:1053–1063. https://doi.org/10.1038/ncb1471
36. Liu B, Qian S-B (2014) Translational reprogramming in stress response. Wiley Interdiscip Rev RNA 5:301–305. https://doi.org/10.1002/wrna.1212
37. Lu PD, Harding HP, Ron D (2004) Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. J Cell Biol 167:27
38. Nair S, Xu C, Shen G et al (2007) Toxicogenomics of endoplasmic reticulum stress inducer tunicamycin in the small intestine and liver of Nrf2 knockout and C57BL/6J mice. Toxicol Lett 168:21–39. https://doi.org/10.1016/j.toxlet.2006.10.012
39. Sano R, Reed JC (2013) ER stress-induced cell death mechanisms. Biochim Biophys Acta, Mol Cell Res 1833:3460–3470. https://doi.org/10.1016/J.BBAMCR.2013.06.028
40. Bravo R, Parra V, Gatica D et al (2013) Endoplasmic reticulum and the unfolded protein response: dynamics and metabolic integration. Int Rev Cell Mol Biol 301:215–290. https://doi.org/10.1016/B978-0-12-407704-1.00005-1
41. Bertolotti A, Zhang Y, Hendershot LM et al (2000) Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. Nat Cell Biol 2:326–332. https://doi.org/10.1038/35014014
42. Okamura K, Kimata Y, Higashio H et al (2000) Dissociation of Kar2p/BiP from an ER sensory molecule, Ire1p, triggers the unfolded protein response in yeast. Biochem Biophys Res Commun 279:445–450. https://doi.org/10.1006/bbrc.2000.3987
43. Wu J, He G-T, Zhang W-J et al (2016) IRE1α signaling pathways involved in mammalian cell fate determination. Cell Physiol Biochem 38:847–858. https://doi.org/10.1159/000443039
44. Guan B-J, Krokowski D, Majumder M et al (2014) Translational control during endoplasmic reticulum stress beyond phosphorylation of the translation initiation factor eIF2α. J Biol Chem 289:12593–12611. https://doi.org/10.1074/jbc.M113.543215
45. Harding HP, Novoa I, Zhang Y et al (2000) Regulated translation initiation controls stress-induced gene expression in mammalian cells. Mol Cell 6:1099–1108. https://doi.org/10.1016/S1097-2765(00)00108-8
46. Chan CP, Kok KH, Tang HMV et al (2013) Internal ribosome entry site-mediated translational regulation of ATF4 splice variant in mammalian unfolded protein response. Biochim Biophys Acta, Mol Cell Res 1833:2165–2175. https://doi.org/10.1016/j.bbamcr.2013.05.002
47. Warnakulasuriyarachchi D, Cerquozzi S, Cheung HH, Holcík M (2004) Translational induction of the inhibitor of apoptosis protein HIAP2 during endoplasmic reticulum stress attenuates cell death and is mediated via an inducible internal ribosome entry site element∗. https://doi.org/10.1074/jbc.M308737200
48. Lewis SM, Cerquozzi S, Graber TE et al (2007) The eIF4G homolog DAP5/p97 supports the translation of select mRNAs during endoplasmic reticulum stress. Nucleic Acids Res 36:168–178. https://doi.org/10.1093/nar/gkm1007
49. Yoffe Y, David M, Kalaora R et al (2016) Cap-independent translation by DAP5 controls cell fate decisions in human embryonic stem cells. Genes Dev 30:1991–2004. https://doi.org/10.1101/gad.285239.116
50. Henis-Korenblit S, Shani G, Sines T et al (2002) The caspase-cleaved DAP5 protein supports internal ribosome entry site-mediated translation of death proteins. Proc Natl Acad Sci USA 99:5400–5405. https://doi.org/10.1073/pnas.082102499
51. Sridharan S, Varghese R, Venkatraj V, Datta A (2017) Hypoxia stress response pathways: modeling and targeted therapy. IEEE J Biomed Health

Inform 21:875–885. https://doi.org/10.1109/JBHI.2016.2559460
52. Young RM, Wang S-J, Gordan JD et al (2008) Hypoxia-mediated selective mRNA translation by an internal ribosome entry site-independent mechanism. J Biol Chem 283:16309–16319. https://doi.org/10.1074/jbc.M710079200
53. Shi D-Y, Xie F-Z, Zhai C, et al (2009) The role of cellular oxidative stress in regulating glycolysis energy metabolism in hepatoma cells. https://doi.org/10.1186/1476-4598-8-32
54. Wheaton WW, Chandel NS (2011) Hypoxia. 2. Hypoxia regulates cellular metabolism. Am J Physiol Cell Physiol 300:C385–C393. https://doi.org/10.1152/ajpcell.00485.2010
55. Ismail R, Ul Hussain M (2017) The up regulation of phosphofructokinase1 (PFK1) protein during chemically induced hypoxia is mediated by the hypoxia-responsive internal ribosome entry site (IRES) element, present in its 5′ untranslated region. Biochimie 139:38–45. https://doi.org/10.1016/J.BIOCHI.2017.05.012
56. Zhou J, Wan J, Shu XE et al (2018) N6-Methyladenosine guides mRNA alternative translation during integrated stress response. Mol Cell 69:636–647.e7. https://doi.org/10.1016/j.molcel.2018.01.019
57. Jackson RJ (1991) mRNA translation. Initiation without an end. Nature 353:14–15. https://doi.org/10.1038/353014a0
58. Thakor N, Holcik M (2012) IRES-mediated translation of cellular messenger RNA operates in eIF2 – independent manner during stress. Nucleic Acids Res 40:541–552. https://doi.org/10.1093/nar/gkr701
59. Soengas MS, Capodieci P, Polsky D et al (2001) Inactivation of the apoptosis effector Apaf-1 in malignant melanoma. Nature 409:207–211. https://doi.org/10.1038/35051606
60. Andreev DE, Dmitriev SE, Terenin IM, Shatsky IN (2013) Cap-independent translation initiation of Apaf-1 mRNA based on a scanning mechanism is determined by some features of the secondary structure of its 5′ untranslated region. Biochemistry 78:157–165. https://doi.org/10.1134/S0006297913020041
61. Coldwell MJ, Mitchell SA, Stoneley M et al (2000) Initiation of Apaf-1 translation by internal ribosome entry. Oncogene 19:899–905. https://doi.org/10.1038/sj.onc.1203407
62. Terenin IM, Andreev DE, Dmitriev SE, Shatsky IN (2013) A novel mechanism of eukaryotic translation initiation that is neither m7G-cap-, nor IRES-dependent. Nucleic Acids Res 41:1807–1816. https://doi.org/10.1093/nar/gks1282
63. Badura M, Braunstein S, Zavadil J, Schneider RJ (2012) DNA damage and eIF4G1 in breast cancer cells reprogram translation for survival and DNA repair mRNAs. Proc Natl Acad Sci USA 109:18767–18772. https://doi.org/10.1073/pnas.1203853109
64. Stetter KO (2006) Hyperthermophiles in the history of life. Philos Trans R Soc B Biol Sci 361:1837–1843. https://doi.org/10.1098/rstb.2006.1907
65. Richter K, Haslbeck M, Buchner J (2010) Molecular cell review the heat shock response: life on the verge of death. https://doi.org/10.1016/j.molcel.2010.10.006
66. Mosser DD, Morimoto RI (2004) Molecular chaperones and the stress of oncogenesis. Oncogene 23:2907–2918. https://doi.org/10.1038/sj.onc.1207529
67. Meyer KD, Patil DP, Zhou J et al (2015) 5′ UTR m(6)a promotes cap-independent translation. Cell 163:999–1010. https://doi.org/10.1016/j.cell.2015.10.012
68. Coots RA, Liu X-M, Mao Y et al (2017) M 6 a facilitates eIF4F-independent mRNA translation. Mol Cell 68:504–514.e7. https://doi.org/10.1016/j.molcel.2017.10.002
69. Merrick WC, Pavitt GD (2018) Protein synthesis initiation in eukaryotic cells. Cold Spring Harb Perspect Biol 10:a033092. https://doi.org/10.1101/cshperspect.a033092
70. Wek RC (2018) Role of eIF2α kinases in translational control and adaptation to cellular stress. Cold Spring Harb Perspect Biol 10:a032870. https://doi.org/10.1101/cshperspect.a032870
71. Sun J, Conn CS, Han Y et al (2011) PI3K-mTORC1 attenuates stress response by inhibiting cap-independent Hsp70 translation. J Biol Chem 286:6791–6800. https://doi.org/10.1074/jbc.M110.172882
72. Kim YK, Jang SK (2002) Continuous heat shock enhances translational initiation directed by internal ribosomal entry site. Biochem Biophys Res Commun 297:224–231. https://doi.org/10.1016/S0006-291X(02)02154-X
73. Zhang X, Gao X, Coots RA et al (2015) Translational control of the cytosolic stress response by mitochondrial ribosomal protein L18. Nat Struct Mol Biol 22:404–410. https://doi.org/10.1038/nsmb.3010
74. Li W, Thakor N, Xu EY et al (2010) An internal ribosomal entry site mediates redox-sensitive translation of Nrf2. Nucleic Acids Res 38:778–788. https://doi.org/10.1093/nar/gkp1048
75. Shay KP, Michels AJ, Li W et al (2012) Cap-independent Nrf2 translation is part of a lipoic acid-stimulated detoxification stress response. Biochim Biophys Acta 1823:1102–1109. https://doi.org/10.1016/j.bbamcr.2012.04.002
76. Zhang J, Dinh TN, Kappeler K et al (2012) La autoantigen mediates oxidant induced de novo Nrf2 protein translation. Mol Cell Proteomics 11:M111.015032-M111.015032. https://doi.org/10.1074/mcp.M111.015032
77. Wang X, Zhao Y, Xiao Z et al (2009) Alternative translation of OCT4 by an internal ribosome entry site and its novel function in stress response. Stem Cells 27:1265–1275. https://doi.org/10.1002/stem.58

78. Xiao Z-S, Simpson LG, Quarles LD (2003) IRES-dependent translational control of Cbfa1/Runx2 expression. J Cell Biochem 88:493–505. https://doi.org/10.1002/jcb.10375
79. Starck SR, Tsai JC, Chen K et al (2016) Translation from the 5′ untranslated region shapes the integrated stress response. Science 351:aad3867. https://doi.org/10.1126/science.aad3867
80. Spriggs KA, Stoneley M, Bushell M, Willis AE (2008) Re-programming of translation following cell stress allows IRES-mediated translation to predominate. Biol Cell 100:27–38. https://doi.org/10.1042/BC20070098
81. Weingarten-Gabbay S, Elias-Kirma S, Nir R, et al (2016) Comparative genetics: systematic discovery of cap-independent translation sequences in human and viral genomes. Science (80–):351. https://doi.org/10.1126/science.aad4939
82. Candeias MM, Powell DJ, Roubalova E et al (2006) Expression of p53 and p53/47 are controlled by alternative mechanisms of messenger RNA translation initiation. Oncogene 25:6936–6947. https://doi.org/10.1038/sj.onc.1209996
83. Colussi TM, Costantino DA, Zhu J et al (2015) Initiation of translation in bacteria by a structured eukaryotic IRES RNA. Nature 519:110–113. https://doi.org/10.1038/nature14219
84. Nakayama K (2009) Cellular signal transduction of the hypoxia response. J Biochem 146:757–765. https://doi.org/10.1093/jb/mvp167
85. Brocato J, Chervona Y, Costa M (2014) Molecular responses to hypoxia-inducible factor 1 and beyond. Mol Pharmacol 85:651–657. https://doi.org/10.1124/mol.113.089623
86. Kress TR, Sabò A, Amati B (2015) MYC: connecting selective transcriptional control to global RNA production. Nat Rev Cancer 15:593–607. https://doi.org/10.1038/nrc3984
87. Masoud GN, Li W (2015) HIF-1α pathway: role, regulation and intervention for cancer therapy. Acta Pharm Sin B 5:378–389. https://doi.org/10.1016/j.apsb.2015.05.007
88. Rohban S, Campaner S (2015) Myc induced replicative stress response: how to cope with it and exploit it. Biochim Biophys Acta 1849:517–524. https://doi.org/10.1016/j.bbagrm.2014.04.008
89. Grover R, Candeias MM, Fahraeus R, Das S (2009) p53 and little brother p53/47: linking IRES activities with protein functions. Oncogene 28:2766–2772. https://doi.org/10.1038/onc.2009.138
90. Peer E, Moshitch-Moshkovitz S, Rechavi G, Dominissini D (2018) The epitranscriptome in translation regulation. Cold Spring Harb Perspect Biol:a032623. https://doi.org/10.1101/cshperspect.a032623
91. Dominissini D, Moshitch-Moshkovitz S, Schwartz S et al (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 485:201–206. https://doi.org/10.1038/nature11112
92. Meyer KD, Saletore Y, Zumbo P et al (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell 149:1635–1646. https://doi.org/10.1016/j.cell.2012.05.003
93. Meyer KD, Patil DP, Zhou J et al (2015) 5′ UTR m6A promotes cap-independent translation. Cell 163:999–1010. https://doi.org/10.1016/j.cell.2015.10.012
94. Zhou J, Wan J, Gao X et al (2015) Dynamic m(6)A mRNA methylation directs translational control of heat shock response. Nature 526:591–594. https://doi.org/10.1038/nature15377
95. Kearse MG, Wilusz JE (2017) Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes Dev 31:1717–1731. https://doi.org/10.1101/gad.305250.117
96. Takahashi K, Maruyama M, Tokuzawa Y et al (2005) Evolutionarily conserved non-AUG translation initiation in NAT1/p97/DAP5 (EIF4G2). Genomics 85:360–371. https://doi.org/10.1016/j.ygeno.2004.11.012
97. Liberman N, Gandin V, Svitkin YV et al (2015) DAP5 associates with eIF2β and eIF4AI to promote Internal Ribosome Entry Site driven translation. Nucleic Acids Res 43:3764–3775. https://doi.org/10.1093/nar/gkv205
98. Diederichs S, Bartsch L, Berkmann JC et al (2016) The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations. EMBO Mol Med 8:442–457. https://doi.org/10.15252/emmm.201506055
99. Holcik M, Sonenberg N (2005) Translational control in stress and apoptosis. Nat Rev Mol Cell Biol 6:318–327. https://doi.org/10.1038/nrm1618
100. Blais JD, Addison CL, Edge R et al (2006) Perk-dependent translational regulation promotes tumor cell adaptation and angiogenesis in response to hypoxic stress. Mol Cell Biol 26:9517–9532. https://doi.org/10.1128/MCB.01145-06
101. Gaccioli F, Huang CC, Wang C et al (2006) Amino acid starvation induces the SNAT2 neutral amino acid transporter by a mechanism that involves eukaryotic initiation factor 2alpha phosphorylation and cap-independent translation. J Biol Chem 281:17929–17940. https://doi.org/10.1074/jbc.M600341200
102. Braunstein S, Karpisheva K, Pola C et al (2007) A hypoxia-controlled cap-dependent to cap-independent translation switch in breast Cancer. Mol Cell 28:501–512. https://doi.org/10.1016/j.molcel.2007.10.019
103. Topisirovic I, Sonenberg N (2011) mRNA translation and energy metabolism in cancer: the role of the MAPK and mTORC1 pathways. Cold Spring Harb Symp 76:355–368
104. Sendoel A, Dunn JG, Rodriguez EH et al (2017) Translation from unconventional 5′ start sites drives

tumour initiation. Nature 541:494–499. https://doi.org/10.1038/nature21036
105. Starck SR, Jiang V, Pavon-Eternod M et al (2012) Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. Science (80-) 336:1719–1723. https://doi.org/10.1126/science.1220270
106. Liang H, He S, Yang J et al (2014) PTENα, a PTEN isoform translated through alternative initiation, regulates mitochondrial function and energy metabolism. Cell Metab 19:836–848. https://doi.org/10.1016/j.cmet.2014.03.023
107. Guenther U-P, Weinberg DE, Zubradt MM et al (2018) The helicase Ded1p controls use of near-cognate translation initiation codons in 5′ UTRs. Nature 559:130–134. https://doi.org/10.1038/s41586-018-0258-0
108. Liang H, Chen X, Yin Q et al (2017) PTENβ is an alternatively translated isoform of PTEN that regulates rDNA transcription. Nat Commun 8:14771. https://doi.org/10.1038/ncomms14771
109. Bugler B, Amalric F, Prats H (1991) Alternative initiation of translation determines cytoplasmic or nuclear localization of basic fibroblast growth factor. Mol Cell Biol 11:573–577
110. Arnaud E, Touriol C, Boutonnet C et al (1999) A new 34-kilodalton isoform of human fibroblast growth factor 2 is cap dependently synthesized by using a non-AUG start codon and behaves as a survival factor. Mol Cell Biol 19:505–514
111. Hann SR, King MW, Bentley DL et al (1988) A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. Cell 52:185–195
112. Hann SR, Sloan-Brown K, Spotts GD (1992) Translational activation of the non-AUG-initiated c-myc 1 protein at high cell densities due to methionine deprivation. Genes Dev 6:1229–1240
113. Sonenberg N, Hinnebusch AG (2007) New modes of translational control in development, behavior, and disease. Mol Cell 28:721–729. https://doi.org/10.1016/j.molcel.2007.11.018
114. Silvera D, Schneider RJ (2009) Inflammatory breast cancer cells are constitutively adapted to hypoxia. Cell Cycle 8:3091
115. Lien I-C, Horng L-Y, Hsu P-L et al (2014) Internal ribosome entry site of bFGF is the target of thalidomide for IMiDs development in multiple myeloma. Genes Cancer 5:127
116. Vagner S, Gensac MC, Maret A et al (1995) Alternative translation of human fibroblast growth factor 2 mRNA occurs by internal entry of ribosomes. Mol Cell Biol 15:35–44
117. Martineau Y, Le Bec C, Monbrun L et al (2004) Internal ribosome entry site structural motifs conserved among mammalian fibroblast growth factor 1 alternatively spliced mRNAs. Mol Cell Biol 24:7622–7635. https://doi.org/10.1128/MCB.24.17.7622-7635.2004
118. Shi Y, Yang Y, Hoang B et al (2016) Therapeutic potential of targeting IRES-dependent c-myc translation in multiple myeloma cells during ER stress. Oncogene 35:1015–1024. https://doi.org/10.1038/onc.2015.156
119. Yeh SH, Yang WB, Gean PW et al (2011) Translational and transcriptional control of Sp1 against ischaemia through a hydrogen peroxide-activated internal ribosomal entry site pathway. Nucleic Acids Res 39:5412–5423. https://doi.org/10.1093/nar/gkr161
120. Fernandez J, Yaman I, Mishra R et al (2001) Internal ribosome entry site-mediated translation of a mammalian mRNA is regulated by amino acid availability. J Biol Chem 276:12285–12291. https://doi.org/10.1074/jbc.M009714200
121. Fernandez J, Bode B, Koromilas A et al (2002) Translation mediated by the internal ribosome entry site of the cat-1 mRNA is regulated by glucose availability in a PERK kinase-dependent manner. J Biol Chem 277:11780–11787. https://doi.org/10.1074/jbc.M110778200
122. Damiano F, Alemanno S, Gnoni GV, Siculella L (2010) Translational control of the sterol-regulatory transcription factor SREBP-1 mRNA in response to serum starvation or ER stress is mediated by an internal ribosome entry site. Biochem J 429:603–612
123. Riley A, Jordan LE, Holcik M (2010) Distinct 5′ UTRs regulate XIAP expression under normal growth conditions and during cellular stress. Nucleic Acids Res 38:4665–4674. https://doi.org/10.1093/nar/gkq241
124. Liu T, Zhang H, Xiong J et al (2015) Inhibition of MDM2 homodimerization by XIAP IRES stabilizes MDM2, influencing cancer cell survival. Mol Cancer 14. https://doi.org/10.1186/s12943-015-0334-0
125. Holcik M, Lefebvre C, Yeh C et al (1999) A new internal-ribosome-entry-site motif potentiates XIAP-mediated cytoprotection. Nat Cell Biol 1:190–192. https://doi.org/10.1038/11109
126. Holcik M, Sonenberg N, Korneluk RG (2000) Internal ribosome initiation of translation and the control of cell death. Trends Genet 16:469–473
127. Fu Q, Chen Z, Gong X et al (2015) β-Catenin expression is regulated by an IRES-dependent mechanism and stimulated by paclitaxel in human ovarian cancer cells. Biochem Biophys Res Commun 461:21–27. https://doi.org/10.1016/j.bbrc.2015.03.161
128. Yang X, Hao Y, Ferenczy A et al (1999) Overexpression of anti-apoptotic gene BAG-1 in human cervical cancer. Exp Cell Res 247:200–207. https://doi.org/10.1006/excr.1998.4349
129. Subkhankulova T, Mitchell SA, Willis AE (2001) Internal ribosome entry segment-mediated initiation of c-Myc protein synthesis following genotoxic stress. Biochem J 359:183
130. Dobbyn HC, Hill K, Hamilton TL et al (2007) Regulation of BAG-1 IRES-mediated translation following chemotoxic stress. Oncogene 27:1167–1174

131. Ott G, Rosenwald A, Campo E (2013) Understanding MYC-driven aggressive B-cell lymphomas: pathogenesis and classification. Blood 122:3884–3891. https://doi.org/10.1182/blood-2013-05-498329
132. Sherrill KW, Byrd MP, Van Eden ME, Lloyd RE (2004) BCL-2 translation is mediated via internal ribosome entry during cell stress. J Biol Chem 279:29066–29074. https://doi.org/10.1074/jbc.M402727200
133. Van Eden ME (2004) Demonstrating internal ribosome entry sites in eukaryotic mRNAs using stringent RNA test procedures. RNA 10:720–730. https://doi.org/10.1261/rna.5225204
134. Candeias MM, Hagiwara M, Matsuda M (2016) Cancer-specific mutations in p53 induce the translation of Delta160p53 promoting tumorigenesis. EMBO Rep 17:1542–1551. https://doi.org/10.15252/embr.201541956
135. Ray PS, Grover R, Das S (2006) Two internal ribosome entry sites mediate the translation of p53 isoforms. EMBO Rep 7:404–410. https://doi.org/7400623 [pii] https://doi.org/10.1038/sj.embor.7400623
136. Khan D, Katoch A, Das A et al (2015) Reversible induction of translational isoforms of p53 in glucose deprivation. Cell Death Differ 22:1203–1218. https://doi.org/10.1038/cdd.2014.220
137. Grover R, Candeias MM, Fåhraeus R, Das S (2009) p53 and little brother p53/47: linking IRES activities with protein functions. Oncogene 28:2766–2772. https://doi.org/10.1038/onc.2009.138
138. Candeias MM (2011) The can and can't dos of p53 RNA. Biochimie 93:1962–1965. https://doi.org/10.1016/j.biochi.2011.06.010
139. Grover R, Ray PS, Das S (2008) Polypyrimidine tract binding protein regulates IRES-mediated translation of p53 isoforms. Cell Cycle 7:2189–2198. https://doi.org/6271 [pii]
140. Sharathchandra A, Lal R, Khan D, Das S (2012) Annexin A2 and PSF proteins interact with p53 IRES and regulate translation of p53 mRNA. RNA Biol 9:1429–1439. https://doi.org/10.4161/rna.22707
141. Weingarten-Gabbay S, Khan D, Liberman N et al (2014) The translation initiation factor DAP5 promotes IRES-driven translation of p53 mRNA. Oncogene 33:611–618. https://doi.org/10.1038/onc.2012.626
142. Halaby M-J, Harris BRE, Miskimins WK, et al (2015) Deregulation of IRES-mediated p53 translation in cancer cells with defective p53 response to DNA damage. Mol Cell Biol MCB.00365-15. https://doi.org/10.1128/MCB.00365-15
143. Halaby M-J, Li Y, Harris BR, et al. Translational control protein 80 stimulates IRES-mediated translation of p53 mRNA in response to DNA damage. https://doi.org/10.1155/2015/708158
144. Serrano M, Lin AW, McCurrach ME et al (1997) Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. Cell 88:593–602
145. Bellodi C, Kopmar N, Ruggero D (2010) Deregulation of oncogene-induced senescence and p53 translational control in X-linked dyskeratosis congenita. EMBO J 29:1865–1876. https://doi.org/emboj201083 [pii]https://doi.org/10.1038/emboj.2010.83
146. Leprivier G, Rotblat B, Khan D et al (2015) Stress-mediated translational control in cancer cells. Biochim Biophys Acta Gene Regul Mech 1849:845–860. https://doi.org/10.1016/j.bbagrm.2014.11.002

RNA Therapeutics: How Far Have We Gone?

7

Maria Francisca Coutinho, Liliana Matos, Juliana Inês Santos, and Sandra Alves

Abstract

In recent years, the RNA molecule became one of the most promising targets for therapeutic intervention. Currently, a large number of RNA-based therapeutics are being investigated both at the basic research level and in late-stage clinical trials. Some of them are even already approved for treatment. RNA-based approaches can act at pre-mRNA level (by splicing modulation/correction using antisense oligonucleotides or U1snRNA vectors), at mRNA level (inhibiting gene expression by siRNAs and antisense oligonucleotides) or at DNA level (by editing mutated sequences through the use of CRISPR/Cas). Other RNA approaches include the delivery of *in vitro* transcribed (IVT) mRNA or the use of oligonucleotides aptamers. Here we review these approaches and their translation into clinics trying to give a brief overview also on the difficulties to its application as well as the research that is being done to overcome them.

Keywords

Antisense oligonucleotides · Aptamers · CRISPR-Cas gene editing · Modified mRNA replacement therapy · siRNA-based drugs · U1snRNA vectors

Abbreviations

2′-F	2′-fluoro
2′-MOE	2′-O-methoxyethyl
2′-O-CH2Py(4)	2′-O-methyl-4-pyrimidine
2′-OMe	2′-O-methyl
AADC	Aromatic L-amino Acid Decarboxylase
AAVs	Adeno-associated Viruses;
AGO	Argonaute
AMD	Age-related Macular Degeneration
AONs	Antisense Oligonucleotides
ASGPR	Asialoglycoprotein receptor
ASGR	Asialoglycoprotein Receptor
BBB	Blood Brain Barrier
BMD	Becker Muscular Dystrophy
CNS	Central Nervous System
CPPs	Cell-penetrating Peptides
CRISPR	Clustered Regulatory Interspaced Short Palindromic Repeats

M. F. Coutinho · L. Matos · J. I. Santos · S. Alves (✉)
Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge, Porto, Portugal
e-mail: sandra.alves@insa.min-saude.pt

L. Romão (ed.), *The mRNA Metabolism in Human Disease*, Advances in Experimental Medicine and Biology 1157, https://doi.org/10.1007/978-3-030-19966-1_7

CRISPR-Cas 9	CRISPR-associated protein 9 (Cas9)
CSF	Cerebrospinal Fluid
DLin-MC3-DMA	Anionic lipid dilinol eylmethyl-4-dimethyl-aminoburyate
DMD	Duchenne Muscular Dystrophy
dsRNAs	double-stranded RNAs
ExSpeU1s	Exon-Specific U1 snRNAs
FH	Familial Hypercholesterolemia
GalNAc	N-acetylgalactosamine
GPCRs	G Protein-coupled Receptors
HCV	Hepatitis C Virus
hFVII	Human Factor VII
hTTRA	Hereditary Transthyretin Amyloidosis
I2S	Iduronate 2-sulfatase
IVT	*in vitro* transcribed
LDL-C	Low-density Lipoprotein Cholesterol
LNAs	Locked Nucleic Acids
LNPs	Lipid-based Nanoparticles
miRNAs	microRNAs
mNIS+7	Modified Neurologic Impairment Score +7
mRNA	messenger RNA
NDA	New Drug Application
PCK9	Proprotein Convertase Subtilisin/Kexin Type 9
PD	Pharmacodynamics
PILs	Pegylated immunoliposomes
PIWI	P-element induced wimpy testis
PK	Pharmacokinetics
PMOs	Phosphoroamidate Morpholino Oligomers
PNAs	Peptide Nucleic Acids
PS	Phosphorothioate
RIS	RNA-induced silencing complex
RNAi	RNA interference
RNP	Ribonucleoprotein
RTK	Human Receptor Tyrosine Kinase
SELEX	Systematic Evolution of Ligands by Exponential enrichment
sgRNA	single guide RNA
shRNA	short hairpin RNA
siRNAs	small interference RNAs
SLNs	Solid Lipid Nanoparticles
Sm	Smith antigen
SMA	Spinal Muscular Atrophy
SNALPs	Nucleic-acid-lipid-particles
SSOs	Splice Switching Oligonucleotides
TLRs	Toll-like Receptors
TTR	Transthyretin
U.S. FDA	U.S. Food and Drug Administration

7.1 Introduction

The RNA molecule has traditionally been viewed as an intermediate between DNA and protein. Recently though, this reductive view has been abandoned as more classes and functions of RNA have been discovered as well as therapeutic applications involving this molecule are being developed. RNA therapeutics can either mimic or antagonize the endogenous RNA functions and have several advantages. They can act even on targets that were previously "undraggable" and, most importantly, they are easy to design, cost effective, stable and easy to combine with other drugs presenting also low immunogenicity. Despite these advantages, the use of RNAs as drugs requires the overcoming of two major obstacles: the poor pharmacological properties of RNA, which is rapidly degraded by RNases and the difficulties in its delivery to the target organs and tissues. In this chapter we present the major RNA-based therapeutics currently under research, discussing the challenges to their translation into the clinic and the recent advances in delivery strategies. RNA tools such as ribozymes, riboswitches and SINE-UP strategy are no less important but will not be discussed in this chapter.

7.2 Antisense Oligonucleotides

7.2.1 Brief Overview

Antisense oligonucleotides (AONs) are short synthetic oligonucleotides that bind to RNA through standard Watson-Crick base pairing and can modulate the function of their target RNA [1, 2]. AONs can function in various ways (Fig. 7.1). For example, AONs can mediate targeted gene knockdown through the recruitment of endogenous RNAse H to degrade mRNA at sites of DNA:RNA hybridization caused by AON binding (Fig. 7.1a). They can also be designed to bind to translation initiation sites on mRNAs in cytosol to block translation (Fig. 7.1b). Another approach uses single-stranded AONs to modulate miRNAs expression; these AONs directly bind target miRNAs to inhibit their function (anti-miRs), and thus depress their target gene (Fig. 7.1c). Moreover, AONs can also be used to modulate pre-mRNA splicing in the target gene bypassing the disease-causing mutation (Fig. 7.1d) [3, 4]. These AONs are designated Splice Switching Oligonucleotides (SSOs) and are single stranded 15–25 nucleotides long, which direct pre-mRNA splicing to a new pathway by binding sequence elements and sterically blocking access to the transcript by the spliceosome and other splicing factors [1, 5–7].

The modification of gene expression, using a synthetic single stranded DNA, resulting in inhibition of mRNA translation was demonstrated for the first time by Paterson and colleagues in 1977 in a cell-free system [8]. Almost a year later, Zamecnik and Stephenson showed that in chicken fibroblast tissue culture containing Rous Sarcoma virus, the addition of a synthetic 13-mer oligonucleotide complementary to the 3′ end of the virus, could inhibit its replication and the subsequent transformation of fibroblasts into sarcoma cells [9]. Since then, remarkable progress

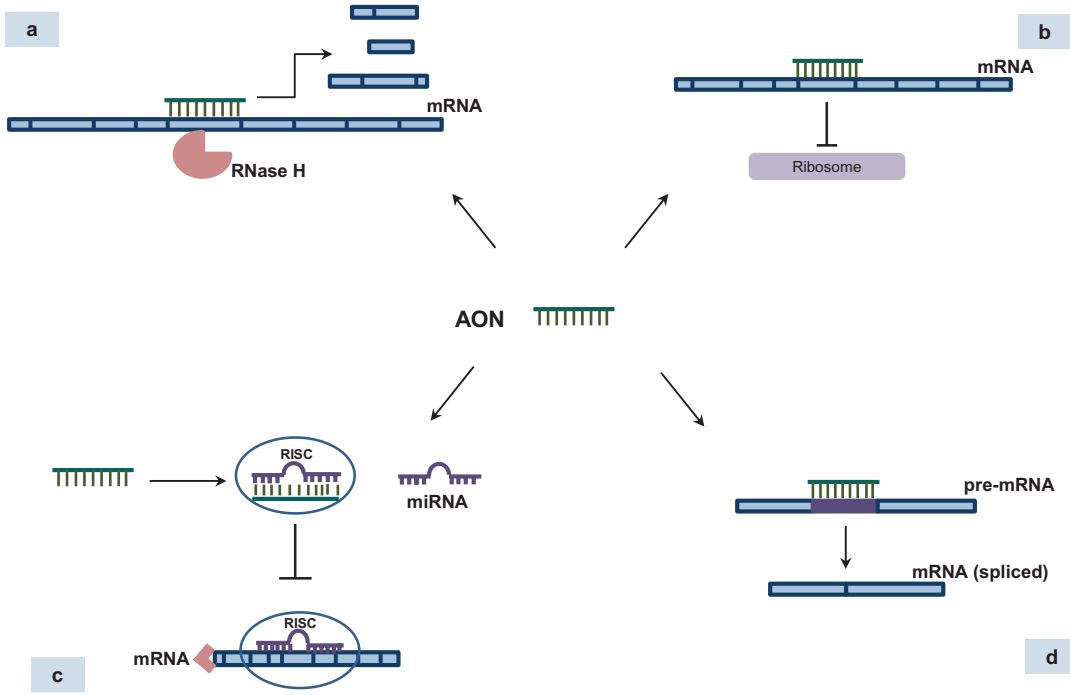

Fig. 7.1 Antisense mechanisms of RNA-based drugs. Antisense oligonucleotides (AONs) impact in gene expression through four different mechanisms: (**a**) RNase H-mediated mRNA degradation; (**b**) steric block of ribosome binding; (**c**) complementary binding to target microRNAs (miRNAs) in order to inhibit their function (antagomirs); and (**d**) splicing modulation

has been made in oligonucleotide drug development and currently, antisense technology is a powerful tool that can be used for target validation and to correct or alter RNA expression for therapeutic benefit [1, 10–12].

Initially, AONs were just synthetic unmodified DNA or RNA molecules, which despite delivering some promising results, would prove to be quite ineffective in biological systems due to their susceptibility to degradation by nucleases, poor affinity for their target mRNA, multiple off-target effects, inability to cross the cell membrane given their negative charge, and weak binding to plasma proteins, leading to rapid clearance by the kidney [1, 3, 13]. Therefore, a wide variety of chemically modified analogues of nucleotides have been developed since then. These chemical modifications were made in the oligonucleotides, generating three categories of AONs, commonly known as generations, with different chemical and pharmacological properties.

7.2.2 Antisense Oligonucleotides Chemistry

Several characteristics need to be fulfilled for the clinical application of antisense oligonucleotides. First, the sequence of the antisense oligonucleotide must be specific enough to avoid off-targets. Other important aspects to be taken into account are the chemistry of the AON and its resistance to degradation by nucleases in order not only to maintain the integrity of the molecule but also to ensure that it is present in an amount, which is sufficient for a true efficacy. In addition, ideal AON should have good pharmacokinetic (PK) and pharmacodynamic (PD) properties and, above all, should not be toxic. Finally, and a very fundamental thing to check, is whether the designed AONs are possible to deliver to target tissues or organs. To try to cope with these desired AON properties, several chemical modifications have been made to the backbone, ribose sugar moiety or nucleobase components, which have a profound effect on the enhanced stability, binding strength and specificity to the target RNA sequence [13] (Fig. 7.2).

The first generation of AONs is characterized by alterations in the backbone, the most common being the phosphorothioate (PS) backbone, accomplished by the replacement of one of the non-bridging oxygen atoms by a sulphur atom. AONs bearing PS linkages are compatible with recruitment of RNase H, which cleaves the target of AONs. This modification allows for an improved nuclease resistance, as well as strong binding to plasma proteins, reducing renal clearance, but still presents poor binding affinity, low specificity and poor cellular uptake [1, 7, 10, 13–15]. Despite these disadvantages, PS oligonucleotides are still the most commonly used AONs and were the first antisense-based drug approved for clinical use in 1998 with fomivirsen (Vitravene®) used for repression of cytomegalovirus mRNA translation [16]. It gained U.S. FDA (U.S. Food and Drug Administration) approval for intraocular treatment of cytomegalovirus retinitis in immunosuppressed patients in 1998 [16] and was discontinued later due to commercial considerations.

In order to surpass the downsides of the first generation oligonucleotides, a second generation was developed through modifications at the 2′ position of the ribose. The most widely studied second generation AONs are 2′-O-methyl (2′-OMe) and 2′-O-Methoxyethyl (2′-MOE), which present higher nuclease resistance and higher affinity for the target RNA, while also reducing non-specific protein binding and toxicity [7, 10, 13, 15]. These second generation AONs, however, do not support RNase H-mediated cleavage of the target mRNA, which impairs their usage for purposes of gene downregulation [1, 14, 17]. This limitation has been minimized with the development of "gapmer" structures where 2′ sugar-modified residues are present on either side of a central "gap" region comprising 8–10 PS-modified nucleotides. The external sugar modified residues thus increase affinity and nuclease resistance, while the internal "gap" region allows RNase H-mediated cleavage of the target RNA [1, 6, 18, 19].

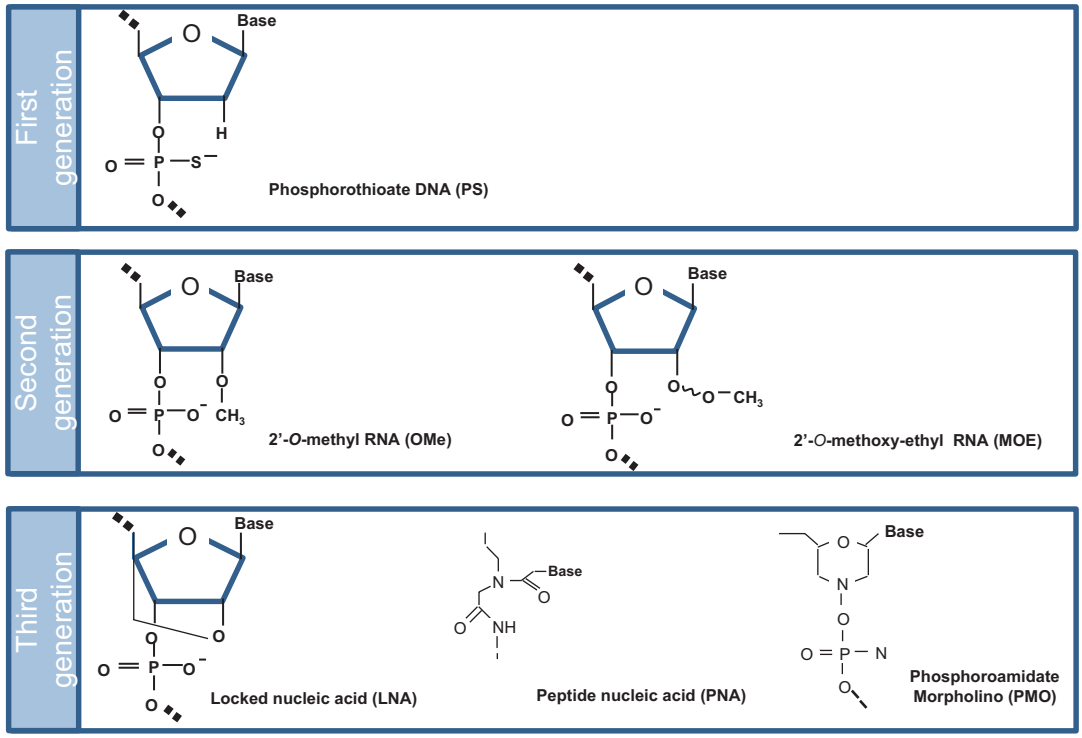

Fig. 7.2 Chemical modifications of antisense oligonucleotides. First generation antisense oligonucleotides (AONs) are characterized by phosphorothioate (PS) backbone; second generation AONs contain a methyl or methoxyethyl group at the 2′ position of the ribose; finally, third generation AONs are characterized by modifications of the furanose ring structure

Finally, the third generation of oligonucleotides is characterized by modifications of the furanose ring of the nucleotide, with the most common being peptide nucleic acids (PNAs), locked nucleic acids (LNAs) and phosphoroamidate morpholino oligomers (PMOs) [12, 14]. These modifications further increase nuclease and protease resistance, target affinity, specificity and *in vivo* stability of antisense drugs, while reduce non-specific interactions with proteins [1, 13, 14, 20, 21]. Nevertheless, PNAs and PMOs present poor cellular uptake, low water solubility and are rapidly cleared from the blood due to their uncharged nature [1, 10, 20, 22] whereas LNAs appear to generate higher toxicity than other chemical modifications, questioning their safety for therapeutic applications [1, 20].

7.2.3 Recent Successful Applications of Antisense Oligonucleotides

The therapeutic application of AONs is very promising. A huge amount of preclinical data has been produced in recent years and many studies have even undergone clinical trials (Table 7.1). Of those, four drugs with different AON chemistries and treatment targets reached, or almost reached clinical practice [12] (Fig. 7.3). One of them is Mipomersen (Kynamro®; Genzyme) that was approved by the U.S. FDA in 2013 for the treatment of familial hypercholesterolemia (FH). Mipomersen is a gapmer of 20 nucleotides and has a sequence complementary to a segment of the Apo b-100 mRNA. Its binding creates a

Table 7.1 Summary table of clinical trials for Antisense Oligonucleotides (AONs)

Chemistry	Drug	Target/organ	Dose/route	Indication	Key observations	References
Commercialized						
2′-H (PS backbone)	Fomivirsen	HCMV *UL122*, Eye	330 μg/eye once every 4 weeks, ITV	CMV retinitis	(1) Effective for CMV retinitis in AIDS patients	[23–25]
2′-MOE (PS backbone)	Mipomersen	*APOB*, Liver	200 mg once weekly, SC	HoFH	(1) Reduction of apoB100 containing lipoprotein particles; (2) Target-related ALT elevations; (3) Clinically meaning full tolerability issues	[26–29]
2′-MOE, fully modified (PS backbone)	Nusinersen	*SMN2* (Intron 7), CNS	12 mg once every 4 months, IT	SMA	(1) Correction of SMN2 splicing defect; (2) Two phase 3 studies terminated early for significant benefit; (3) No drug related adverse events	[30–33]
PMO	Eteplirsen	*DMD* (Exon 51), Muscle	30 mg/kg once weekly, IV	DMD	(1) Increased production of dystrophin; (2) Interpretation of results controversial	[34, 35]
Phase 3						
2′-H	Alicaforsen	*ICAM1*, Colon	240 mg once daily, Enema	Chronic pouchitis	Phase 2 findings included reduction in Pouchitis Disease Activity Index and endoscopy subscore	NCT02525523 [36]
2′-H	Mongersen	*SMAD7*, Intestine	160 mg once daily, Oral	Crohn's disease	Phase 3 failed to demonstrate benefit in patients with Crohn's Disease	NCT02596893 [37] NCT02601300
2′-MOE	Inotersen	*TTR*, Liver	300 mg once weekly, SC	Hereditary ATTR	Phase 3 findings included (1) significant benefit in inotersen-treated patients in both neuropathic disease progression and quality of life measures compared to placebo-treated patients, (2) reduction of mutant and normal TTR, (3) thrombocytopenia (1 fatal bleeding event) and (4) renal events	NCT01737398 [38] (Ionis Pharmaceuticals, press release, Nov 2, 2017)
2′-MOE	Volanesorsen	*APOC3*, Liver	300 mg once weekly, SC	FCS	Phase 3 findings included (1) consistent reduction of triglycerides, (2) mean reduction in triglycerides of 1712 mg/dLin FCS patients, (3) reduced abdominal pain and pancreatitis, and (4) FCS-related thrombocytopenia	NCT02211209 [39–42]

PMO	SRP-4045	*DMD* (Exon 45), Muscle	30 mg/kg once weekly, IV	DMD	Not published	NCT02500381
PMO	Golodirsen (SRP-4053)	*DMD* (Exon 53), Muscle	30 mg/kg once weekly, IV	DMD	On clinical hold due to serious adverse event	NCT02500381
Phase 2						
2′-MOE	IONIS-FXI$_{Rx}$/BAY 2306001	*F11*, Liver	100–300 mg once weekly, SC	Clotting disorders	Phase 2 findings included (1) reduction of Factor XI protein, and (2) reduction of thrombotic events without increase in bleeding	NCT02553889 NCT01713361 [43]
2′-MOE	IONIS-PTP1B$_{Rx}$	*PTPN1*, Liver	200 mg once weekly, SC	T2D	Phase 2 findings included (1) reduction of HbA1c, (2) improved leptin and diponectin levels, and (3) decreased body weight	NCT00455598
2′-MOE	IONIS-GCCR$_{Rx}$	*NR3C1*, Liver	60–420 mg once weekly, SC	T2D	Phase 1 findings included (1) improvement in lipid profile, and (2) attenuation of dexamethasone-induced hepatic insulin resistance	NCT01968265 [44]
2’-MOE	IONIS-GCGR$_{Rx}$	*GCGR*, Liver	50–200 mg once weekly, SC	T2D	Phase 2 findings included (1) attenuation of glucagon-induced increase in blood glucose levels, and (2) reduction of HbA1c	NCT02583919 [45]
2′-MOE	IONIS-FGF$_{R4}$	*FGFR4*, Liver	100–200 mg once weekly, SC	Obesity	Not published	NCT02463240
2′-MOE	IONIS-HTT$_{Rx}$	*HTT*, CNS	10–120 mg once every 4 weeks, IT	Huntington’s disease	In progress	NCT02519036
2′-MOE	IONIS-DGAT2$_{Rx}$	*DGAT2*, Liver	once weekly, SC	NASH	In progress	NCT03334214
2′-MOE	Apatorsen	*HSPB1*, Tumor Cells	200–1000 mg once weekly, IV	Cancer	Phase 1 findings included decrease in tumor markers and decline in CTCs	NCT01454089 [46]
2′-MOE	ATL1102	*ITGA4*, Immune Cells	200 mg twice weekly, SC	MS	Phase 2 findings included reduction in new active lesions	[47]
2′-MOE	Atesidorsen/ ATL1103	*GHR*, Liver	6 mg/kg twice weekly, SC	Acromegaly	Not published	ACTRN12615000289516

(continued)

Table 7.1 (continued)

Chemistry	Drug	Target/organ	Dose/route	Indication	Key observations	References
2′-MOE	IONIS-PKK$_{Rx}$	*KLKB1*, Liver	200 mg once weekly, SC	HAE, chronic migraine	In progress	NCT03108469
2′-MOE	IONIS-HBV$_{Rx}$	*HBV S*, Liver	SC	HBV, chronic atypical	In progress	NCT02981602
cET	AZD9150/IONISSTAT3-2.5$_{Rx}$	*STAT3*, Cancer and Stromal Cells	2–4 mg/kg once weekly, IV	Cancer	Phase 1 findings included (1) reduction of STAT3, (2) reduction in serum IL6, and (3) reduction in tumor burden	NCT02549651 NCT01563302 [48]
cET	AZD5312/ IONISAR-2.5$_{Rx}$	*AR*, Cancer Cells	150–1150 mg once weekly, IV	Prostate cancer	Phase 1 findings included declines in PSA and circulating tumor cells in some patients	NCT02144051 [49]
LNA	Miravirsen	*MIR122*, Liver	3–7 mg/kg once weekly, SC	HCV	Phase 2 findings included inhibition of miR-122 function.	NCT01200420 [50, 51]
2′-MOE, GalNAc	IONIS-APO(a)-L$_{Rx}$	*LPA*, Liver	10–40 mg once weekly, SC	CVD	Phase 1 findings included reduction of Lp(a)	NCT03070782 [52]
2′-MOE, GalNAc	IONIS-ANGPTL3-L$_{Rx}$	*ANGPTL3*, Liver	10–60 mg once weekly, SC	Dyslipidemias	Phase 1 findings included reduction of ANGPTL3, LDL cholesterol and triglycerides	NCT02709850 [53]
2′-MOE, GalNAc	IONIS-APOCIII-L$_{Rx}$	*APOC3*, Liver	once weekly, SC	CVD	In progress	NCT02900027
2′-MOE, GalNAc	GSK3389404/ IONIS-HBV-L$_{Rx}$	*HBV S*, Liver	30–120 mg single dose/once weekly, SC	Chronic HBV	In progress	NCT03020745
Undisclosed, GalNac	AZD4076/RG-125	*MIR107*, Liver	SC	Diabetic NASH	In progress	NCT02826525
Undisclosed	RG-012	*MIR21*, Kidney	110–220 mg once weekly, SC	Alport syndrome	In progress	NCT02855268
Phase 1						
2′-MOE	Eluforsen	Lung/ respiratory epithelium	Intranasal (three times weekly for 4 weeks)	Cystic fibrosis	Phase findings included (1) eluforsen was well tolerated (2) improved CFTR function	NCT02564354; [54]
2′-MOE	BIIB067/ IONISSOD1$_{Rx}$	*SOD1*, CNS	IT	Familial ALS	In progress	NCT02623699

LNA	MRG-106	*MIR155*, Cancer Cells	75–900 mg once weekly, ITM/SC/IV	Hematological malignancies	Phase findings included (1) improvements in cutaneous lesions, and (2) transcriptional changes consistent with target activity	NCT02580552 [55]
LNA	ISTH0036	*TGFB2*, Eye	6.75–225 μg single dose, IVT	Glaucoma	Phase 1 findings included (1) dose-response trend observed in post-operative intraocular pressure, and (2) no adverse events	NCT02406833 [56]
PO, 2′-OMe, ENA	DS-5141b	*DMD* (Exon 45), Muscle	0.1–6.0 mg/kg once weekly, SC	DMD	In progress	NCT02667483
2′-MOE, GalNAc	IONIS-FB-L_{Rx}	*CFB*, Liver	10–40 mg once every 2 weeks, SC	Ocular disease	Phase 1 findings included (1) dose-dependent reduction in factor B levels accompanied by similar reduction in factor B function and complement split factor Bb, and (2) no drug related adverse events	ACTRN12616000335493 [57]
2′-MOE, GalNAc	IONIS-AGT-L_{Rx}	*AGT*, Liver	once weekly, SC	Treatment resistant Hypertension	In progress	NCT03101878
2′-MOE, GalNAc	IONIS-PKK-L_{Rx}	*KLKB1*, Liver	SC	HAE, chronic migraine	In progress	NCT03263507
Discontinued						
2′-OMe	Drisapersen	*DMD* (Exon 51), Muscle	6 mg/kg once weekly, SC	DMD	Rejected by FDA	[58, 59] (Biomarin Pharmaceutical, press release, May 31, 2016)
2′-MOE	Custirsen	*CLU*, Tumor Cells	640 mg once weekly, IV	Prostate cancer and NSCLC	Failure to meet primary endpoints in phase 3 trials	NCT01578655 NCT01630733 [60] (OncoGeneX, press release, Nov 10, 2016)
2′-MOE	IONIS-APO(a)$_{Rx}$	*LPA*, Liver	300 mg once weekly, SC	CVD	Replaced with GalNac conjugate	[61]
2′-MOE	ISIS 388626	*SLC5A2*, Kidney	50–200 mg once weekly, SC	T2D	Availability of small-molecule inhibitors of SGLT2	[62]

(continued)

Table 7.1 (continued)

Chemistry	Drug	Target/organ	Dose/route	Indication	Key observations	References
2′-MOE	ISIS 333611	*SOD1*, CNS	0.15–3.0 mg single dose, IT	Familial ALS	Replaced by more potent compound	[63]
2′-MOE	ISIS 104838	*TNF*, Immune Cells	0.1–6 mg/kg IV, 200 mg once weekly, SC	Inflammatory disease	Inadequate activity	[64]
2′-MOE	ISIS 113715	*PTPN1*, Liver	100–600 mg once weekly, SC	T2D	Replaced by more potent compound	
cET, 2′-MOE	IONIS-DMPK$_{2.5Rx}$	*DMPK*, Muscle	100–600 mg once weekly, SC	DM1	Inadequate activity	NCT02312011 [65]
LNA	EZN-4176	*AR*, Cancer Cells	0.5–10 mg/kg once weekly, IV	Prostate cancer	ALT elevations	NCT01337518 [66]
Undisclosed, GalNac	RG-101	*MIR122*, Liver	SC	HCV	Cases of hyperbilirubinemia	EudraCT 2016-002069-77 (Regulus, press release, Jun 12, 2017)

Abbreviations: **Chemistry:** *2′ – H* 2′- deoxy, *2′ – MOE* 2′ –O- methoxy ethyl, *PMO* phosphorodiamidate morpholino oligomer, *cET* (S)- constrained ethyl, *LNA* locked nucleic acid, *GalNAc* N-Acetilgalactosamine, ENA, 2′ –O, 4′ –C- ethylene-bridged nucleic acid. **Target/organ:** *HCMV UL122* Human cytomegalovirus with IE-2, *APOB* apolipoprotein B, *SMN2* survival of motor neuron 2, *CNS* central nervous system, *DMD* dystrophin, *ICAM1* intercellular adhesion molecule, *SMAD7* SMAD family member, *TTR* transthyretin, *APOC3* apolipoprotein 3, *PTPN1* protein tyrosine phosphatase, non-receptor type1, *NR3C1* nuclear receptor subfamily 3, group C member 1 (glucocorticoid receptor), *GCGR* glucagon receptor, *FGFR4* fibroblast growth factor receptor 4, *HTT* huntingtin, *DGAT* diacylglycerol O-acyltransferase 2, *HSPB1* heat shock protein family B (small) member 1 (heat shock protein27), *ITGA4* integrin subunit alpha 4, *GHR* growth hormone receptor, *KLKB1* kallikrein B1, *HBV S* hepatitis surface antigen (HBsAg), *STAT3* signal transducer and activator of transcription 3, *AR* androgen receptor, *MIR* microRNA, *LPA* lipoprotein (a), *ANGPTL3* angiopoietin like 3, *SOD1* superoxide dismutase 1, *TGFB2* transforming growth factor beta 2, *CFB* complement factor B, *AGT* angiotensinogen, *CLU* clusterin, Solute Carrier Family 5 Member 2, *TNF* tumor necrosis factor alpha, *DMPK* DM1 protein kinase. **Dose/route:** *IVT* intravitreal, *SC* subcutaneous, *IT* intrathecal, *IV* intravenous. **Indication:** *CMV* cytomegalovirus, *HoFH* homozygous familial hypercholesterolemia, *SMA* spinal muscular atrophy, *ATTR* transthyretin amyloidosis, *FCS* familial chylomicronemia, *T2D* type 2 diabetes, *NASH* non-alcoholic steatohepatitis, *MS* multiple sclerosis, *HAE* hereditary angiodema, *HCV* hepatitis C virus, *CVD* cardiovascular disease, *ALS* amyotrophic lateral sclerosis, *NSCLC* non-small cell lung cancer, *DM1* myotonic dystrophy type 1. **Key observations:** *ALT* alanine aminotransferase, *HbA1c* hemoglobin A1c, *CTC* circulating tumor cells, *PSA* prostate specific antigen, *FDA* food and drug administration, *SGLT2* sodium-glucose co-transporter 2, *ALT* alanine aminotransferase

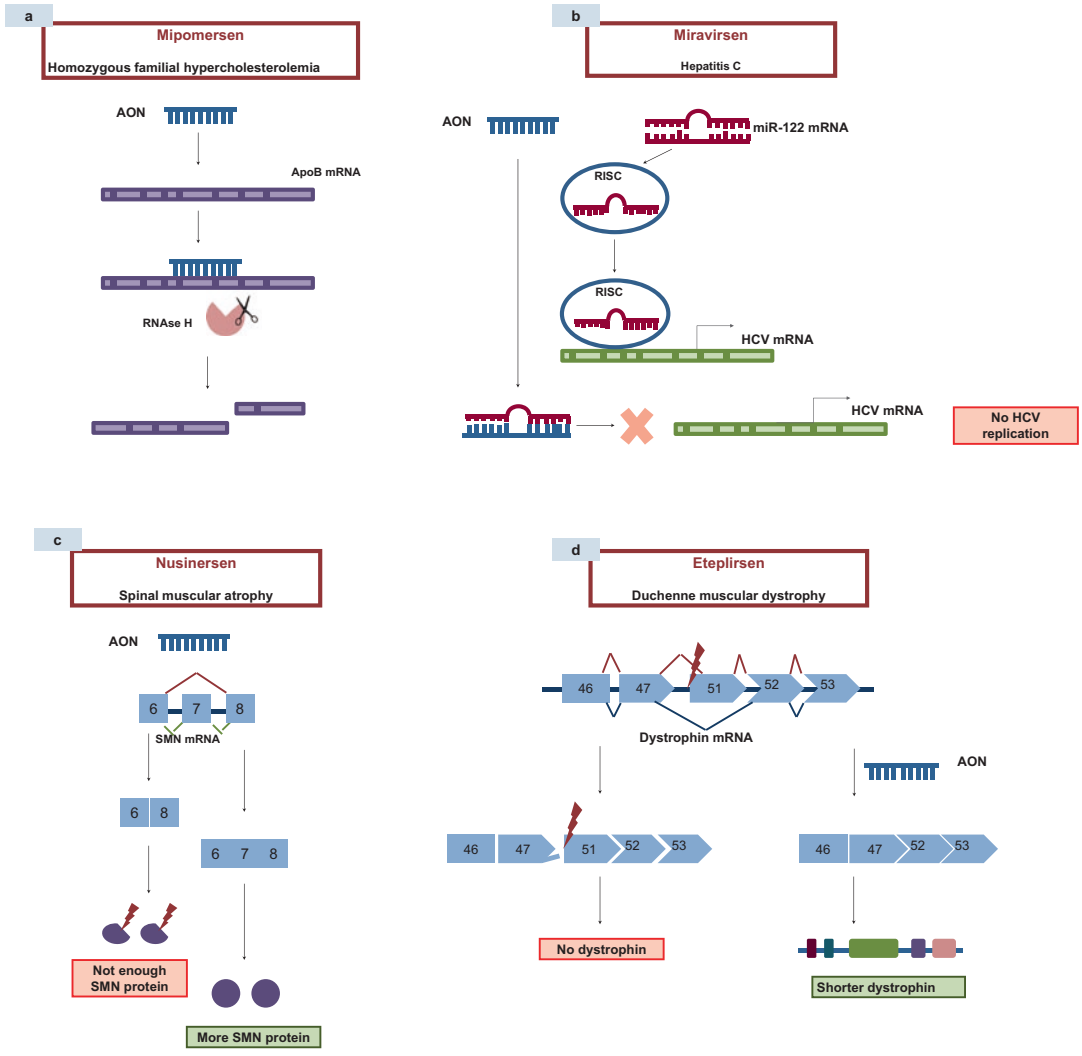

Fig. 7.3 Antisense drugs for clinical practice. Currently, four drugs with different AON chemistries and mechanisms of action have either obtained U.S. FDA approval (Mipomersen, Nusinersen and Eteplirsen) and reached clinical practice, or are seeking accelerated approval soon, with significant pre-clinical data supporting their rapid translation into clinic (Miraversen). (**a**) Mipomersen (Kynamro®; Genzyme), a 2′-MOE-modified AON approved by the U.S. FDA in 2013 for the treatment of familial hypercholesterolemia (FH); (**b**) Miravirsen (SPC3649, Santaris Pharma), an antagomir to miR-122, seeking approval for hepatitis C treatment; (**c**) Nusinersen (Spinraza®, Biogen), a fully modified 2′-MOE AON, approved by U.S. FDA in 2016 for the treatment of spinal muscular atrophy (SMA) and, (**d**) Eterplirsen® (EXONDYS 51TM, Sarepta), a PMO approved by the U.S. FDA in 2016 for the use in Duchenne muscular dystrophy (lightning symbol means the existence of a pathogenic alteration)

DNA:RNA hybrid that is substrate for the enzyme RNase H thus inducing the cleavage of the human Apo b-100 mRNA. The drug has a PS backbone, with 2′-MOE-modified ends, which when compared with earlier antisense technologies, provides greater biological stability and higher affinity to the target mRNA [12, 67]. When administered subcutaneously at a dose of 200 mg per week, it was shown to reduce ApoB-100 production and low-density lipoprotein cholesterol (LDL-C) in a dose-dependent fashion [68]. In general, the results achieved with Mipomersen

point to its efficacy, safety, and tolerability, demonstrating its suitability for use in the target patient population, and providing a tangible tool for use in the management of FH and severe hypercholesterolemia [11, 12, 68]. However, mild-to-moderate injection site reactions, flu-like symptoms and hepatic effects (despite transient and generally reversible) limited its utilization and therefore its commercial success [11].

AONs complementary to mature miRNAs (antagomirs) are also being developed to counteract miRNAs implicated in disease pathogenesis. An example is Miravirsen (SPC3649, Santaris Pharma), an antagomir to miR-122, a liver-specific microRNA that the hepatitis C virus (HCV) requires for replication. Miravirsen is designed to recognize and sequester miR-122, making it unavailable to HCV. As a result, viral replication is inhibited, and the level of HCV infection is reduced [12]. Positive results were observed in a phase II study. The use of Miravirsen in patients with chronic HCV genotype 1 infection showed prolonged dose-dependent reductions in HCV RNA levels without evidence of viral resistance [50]. The updated results revealed no long-term safety issues among 27 Miravirsen-treated patients [51, 69]. Moreover, there was a prolonged decrease in miR-122 plasma levels in patients dosed with Miravirsen but the plasma levels of other miRNAs were not significantly affected by antagonizing miR-122 [69, 70].

The above examples use AONs to alter gene expression, either directly or indirectly, to change disease progression. Another precise method to alter gene expression is to manipulate pre-mRNA splicing using SSOs. This is the case of Nusinersen (Spinraza®, Biogen), a fully modified 2′-MOE AON, approved by U.S. FDA in 2016 for the treatment of spinal muscular atrophy (SMA). SMA is an autosomal recessive neuromuscular disease caused by progressive loss of alfa-motor neurons in the anterior horn of the spinal cord [12, 71]. The most severe form, infant onset or type 1, is the most common one, representing 50% of all SMA cases. Type 2 is less severe, but also very debilitating. These infants never walk, and as they grow the disease progresses and patients begin to lose the capacity of lift even their arms. In humans, a parolog gene of SMA exists, the *SMN2* gene that differs of the *SMN1* by 5–11 nucleotides. However, in the majority of the *SMN2* transcripts the exon 7 is lacking, resulting in a truncated protein, which is rapidly degraded [72]. Nusinersen induces the inclusion of exon 7 in the *SMN2* mRNA by targeting and blocking an intron 7 internal splice site. This action increases SMN protein production, thus improving its function [73]. Intrathecal injection of Nusinersen (every 4 months) allows therapeutic delivery directly into the cerebrospinal fluid (CSF) bathing the spinal cord, the site of motor neuron degeneration and, substantially prolonging survival of type 1 infants, while also resulting in improvements in all measures evaluated [32]. Similar benefit was demonstrated in patients with later onset type 2 SMA [33]. More remarkable, treatment of type 1 pre-symptomatic infants with Nusinersen has been demonstrated to result, in many cases, in achievement of motor milestones at the age expected for healthy infants. Moreover, 92% of the infants treated prior to the development of symptoms were able to sit without support, a milestone never achieved by a type 1 SMA infant before Nusinersen treatment was introduced and 50% were able to walk without support [74].

Another SSO, already in the market is Eterplirsen® (EXONDYS 51™, Sarepta) that was approved by the U.S. FDA in 2016 for use in Duchenne muscular dystrophy (DMD) patients, a severe, childhood-onset disease that results mostly from deletions within the dystrophin gene. DMD is a progressive, neuromuscular disease, occurring mainly in males (1 in 3500–5000 males born worldwide) [75, 76]. It is caused by an absence of the protein dystrophy, a membrane-associated protein that forms a network with sarcolemmal glycoproteins by linking the cytoskeleton actin in muscle fibers within the first few extracellular matrix [77], which results in altered myocyte integrity, muscle wasting and relentlessly progressive weakness. Becker muscular dystrophy (BMD) is a milder disease

caused by dystrophin truncations (due to "in frame" deletions) rather than its absence. A viable strategy for generating truncated, but functional, dystrophin protein involves the skipping of exons to correct DMD-linked mutations (which includes 83% of mutations in DMD) [78]. This can reduce the severity of the disease and produce a milder phenotype, similar to that of BMD. Eteplirsen, the first PMO drug ever approved, binds to the exon/intron splice site at the beginning of exon 51, resulting in its skipping, giving origin to an in frame transcript, which prevents the unwanted degradation of the mutant transcripts by the nonsense-mediated mRNA decay (NMD) pathway, and allowing the production of an internally deleted but functional dystrophin protein [79]. Eterplirsen is applicable for approximately 14% of patients with DMD mutations. It is administered via intravenous infusion and was found to be well tolerated, with no adverse effects, in several clinical trials [79]. In addition, over 3 years of follow-up, Eterplirsen-treated patients showed a slower rate of decline in ambulation assessed by the 6-min walk test compared to untreated matched historical controls from two DMD natural history cohorts: the Leuven Neuromuscular Reference Center (LNMRC) and the Italian Telethon registry [80]. Previously, the ability of Eteplirsen to induce expression of dystrophin had been demonstrated by an observed increase of dystrophin-positive fibers in skeletal muscle of DMD patients [35]. Recently, Kinane and coworkers compared the pulmonary function data from DMD patients, who received Eteplirsen in studies 201/202 (included 12 patients treated with Eteplirsen over 5 years) with the natural history data published. This study verified that the deterioration of respiratory muscle function with Eteplirsen treatment as measured by forced vital capacity was half of that seen in natural history. Maximum expiratory pressure and maximum inspiratory pressure also declined more slowly in Eteplirsen treated patients compared to natural history, thus demonstrating its potential to preserve respiratory function in patients with DMD [81].

7.2.4 Antisense Oligonucleotides Delivery

One of the major issues for the use of AONs for therapeutic purposes is the efficient delivery to their target site. AONs need to reach the target tissue and, once there, they must reach the appropriate intracellular compartment [2, 7]. Parenteral injection, such as intravenous infusion or subcutaneous injection is the main method at the moment of delivery of PS modified single-stranded AONs formulated in a simple saline solution [15, 21]. However, even though AON activity has been observed in many tissues such as lung, stomach, bladder, and heart, AONs predominantly accumulate in liver, kidney, bone marrow, adipocytes, and lymph nodes [21]. Therefore, delivery problems must be considered in terms of sets of barriers to movement of AON within the body. Tissue barriers to delivery include the vascular endothelial barrier, first-pass renal excretion (which strongly affects PK and bio-distribution of AONs), and the blood brain barrier (BBB) that AONs cannot cross due to their size and charge, limiting their access to the central nervous system (CNS), in the case of CNS diseases. The one exception to this is intrathecal injection of single-stranded AONs with specific chemical modifications into the CSF, which allows AONs into the CNS [21].

Two main strategies are being developed to improve AON delivery: viral and non-viral delivery. Despite viral vectors are efficient systems for the delivery of genetic material and for the capability to infect a large number of cell types, they also showed some constraints, such as immunogenicity, tumorogenicity risks, limited loading capacity and scaling-up problems [7]. However, adeno-associated viruses (AAVs), which are non-integrative vectors and therefore present a low risk of genomic insertions, have been used in *in vitro* cells and in animal models to efficiently deliver AONs sequences embedded into modified snRNA systems (modified U1 snRNAs and modified U7 snRNAs). Indeed, promising therapeutic results were obtained with this strategy to induce exon-skipping in diseases like Leber Congenital

Amaurosis [82] and DMD [83–87] and exon-inclusion in SMA [88, 89] (for a more extended review on this subject see [90]). Non-viral delivery also represents a good alternative and the conjugation of free AONs with non-viral delivery carriers can be achieved by different strategies. One option is to conjugate AONs (or their carrier) to a ligand that interacts selectively with a cell surface receptor. Ideally, one such receptor should be expressed only in the tissue to be targeted. Additionally, it should also be abundantly expressed, rapidly and extensively internalized, and have a high affinity to its ligands, so that they become readily available. Some receptors used to target AONs include integrins, G protein-coupled receptors (GPCRs), human receptor tyrosine kinase (RTK), scavenger receptors, asialoglycoprotein receptor (ASGR), Toll-like receptors (TLRs) and folate receptor [91–97], as reviewed in [98]. However, receptors fulfilling all the above-referred criteria are not available for the majority of tissues and, once the AONs reach the cell surface of the target cell, they must be internalized by endocytosis and packed into a vesicle, termed endosome. Then, the endosomes fuse with lysosomes, organelles rich in hydrolases, which ultimately degrade a high portion of the internalized AONs. In fact, one strategy to improve delivery of AONs is the use of short cell-penetrating peptides (CPPs), sequences of short cationic and/or amphipathic peptides (fewer than 30 amino acids) that translocate small drugs/cargo across cell membranes. CPPs are attached to their cargo through covalent linkages or through the formation of noncovalent nanoparticle complexes [99] that can promote uptake of macromoleclues via endocytosis. CPPs covalently conjugated to AONs were already used for therapeutic purposes in DMD [100–105] Myotonic Dystrophy type I [106] and SMA [107, 108]. However, CPPs have a limited endosomal escape and overcoming the rate-limiting step of endosomal escape into the cytoplasm remains a major challenge to their successful use. Several studies tried to overcome this and some relevant results have been obtained. For instance, specific synthetic endosomal escape domains (EEDs) significantly enhanced cytoplasmic delivery in the absence of cytotoxicity [109] and a CPP-adaptor system capable of efficient intracellular delivery was also recently developed [110]. Another possibility is to incorporate AONs into nanoparticles (NPs) that based on their size and materials, will determine the AON biodistribution and interaction. In fact, the progress of nanotechnology has provided several nanosystems with the aim to increase the drug targeting efficacy. The most common types used for drug delivery are solid lipid nanoparticles (SLNs), polymer nanoparticles, lipid-based nanoparticles (LNPs) and carbon-based nanomaterials [110]. For example, cationic core-shell NPs named T1 and ZM2 (a type of polymer nanoparticles) were used to conjugate AONs for exon skipping application in preclinical studies in DMD mice [110, 111]. As these obstacles of delivery are overcome, the advantages of antisense technology will warrant that antisense oligonucleotide therapeutics will be one of the most promising clinical approaches to genetic diseases in the future.

7.3 U1 snRNA-Mediated Therapy

7.3.1 Brief Overview

Since its discovery in the early days of splicing research, U1 snRNA has been recognized as a crucial player in the first stages of the splicing process [112–114]. U1 snRNA is a 164 nucleotides long molecule with a well-defined structure consisting of four stem-loops, which primarily exerts its function in the form of a ribonucleoprotein (RNP) complex (termed U1 snRNP) containing seven Smith antigen (Sm) proteins and three U1-specific proteins U1A, U1C and U1-70K [115] (Fig. 7.4a). It is now well-established that U1 snRNP initiates spliceosome assembly by binding to the 5′ splice donor site (ss) through base pairing between the single stranded 5′ tail of the U1 snRNA molecule and the moderately conserved stretch of nucleotides at the 5′ss (CAG/GURAGU; R-purine) marking the exon-intron boundary [116]. However, not all base pairs at different 5′ss positions are equally important, and their contribution to splicing

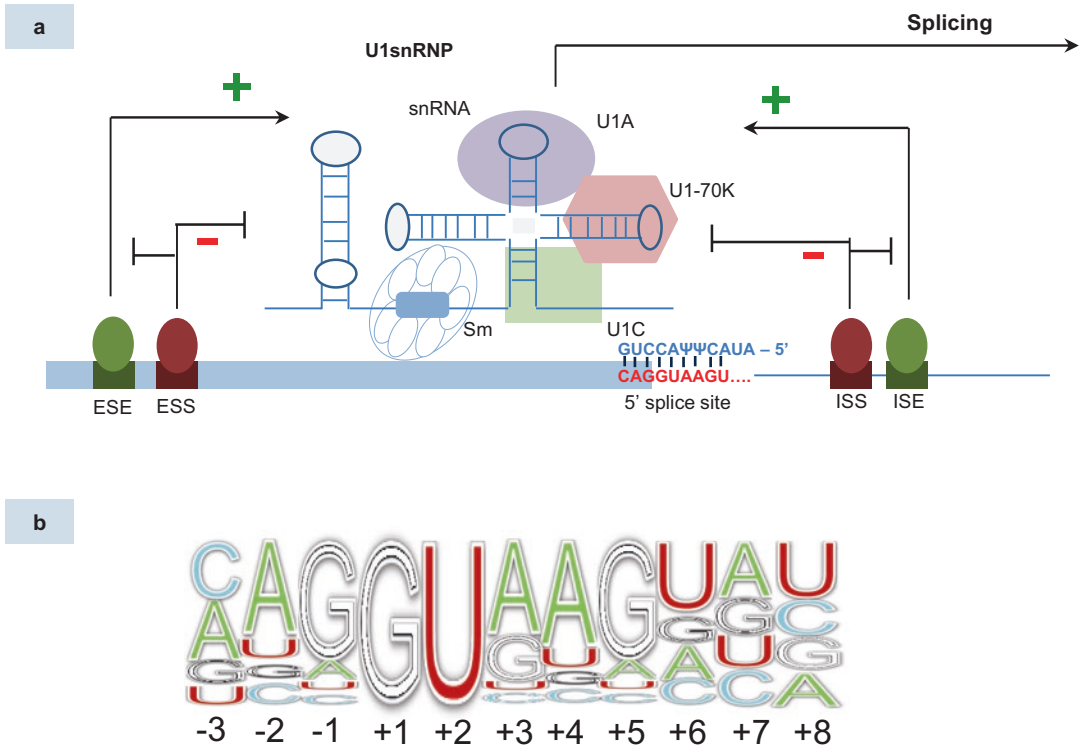

Fig. 7.4 Role of U1 small nuclear ribonucleoproteins (snRNPs) in splicing. (**a**) The 5′ end of U1 snRNA base pairs to the 5′ splice site (ss) in order to define a functional splice donor site. This process is positively and negatively modulated by different splicing factors, which bind to exonic and intronic splicing enhancer and silencer motifs (ESE, ISE, ESS and ISS, respectively). (**b**) The 5′ss motif. The height of each nucleotide corresponds to its conservation at the corresponding positon (−3 to −1 are exonic positions, while +1 to +8 correspond to intronic positions). The most conserved 5′ss positions are +1 and +2, which determine the 5′ss subtype: the GU subtype accounts for ~99% of 5'ss. Minoritary subtypes have a mismatch to U1 at either +1 or +2 and include the GC and the very rare AU 5′ss. The next most conserved 5′ ss positions are -1G and +5G, which form strong G-C base pairs with U1 through three hydrogen bonds. Consensus nucleotides −2A, +3A, +4A, and + 6U are also conserved but have a lesser although important contribution to 5'ss strength because their base pairing to U1 involves only the formation of two hydrogen bonds. The 5'ss positions +7 and +8 do not exhibit substantial conservation in humans, yet several lines of evidence indicate that these positions can base-pair to U1 and contribute to splicing

roughly correlates with their conservation (Fig. 7.4b). In the 9 nucleotides consensus sequence the most conserved 5′ss positions lie at the first two intronic nucleotides (+1 and +2), and the sequence GU at these positions accounts for ~99% of all 5′ss. The next most conserved 5′ss positions are −1G and +5G, which form strong G-C base pairing with U1 [117]. Once the donor site does not always conform to the consensus sequence, but can instead have a degenerate pattern feature, it is understandable that many other additional elements such as splicing silencer and enhancer motifs, the presence of alternative splice sites, secondary structures and regulatory proteins can influence the splice site selection (Fig. 7.4a) [117, 118].

U1 snRNA is classically known for its role in pre-mRNA splicing events. However, the finding that U1 snRNA levels far exceed other spliceosomal associated snRNA levels led to the notion that it may have additional roles in the cell apart from splicing regulation [115, 119]. Indeed, emerging evidence suggests that U1 snRNA plays a key role in transcription initiation and in the protection of pre-mRNAs from degradation, as also has a regulatory function in the 3′-end for-

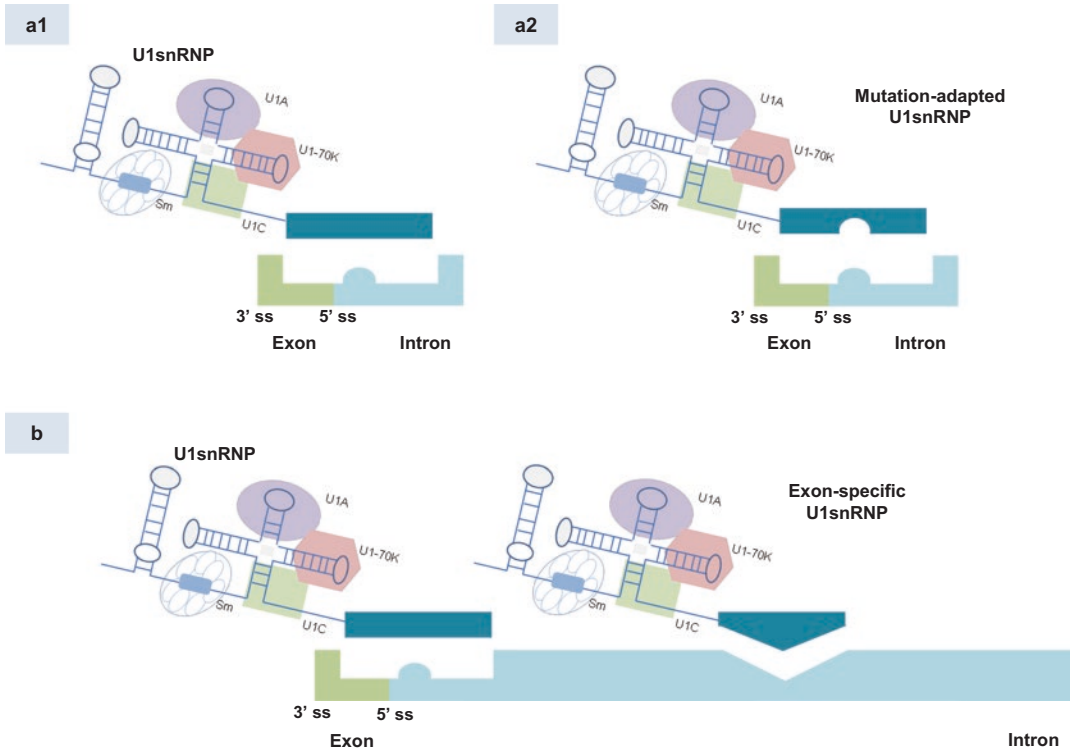

Fig. 7.5 U1 snRNA-mediated therapy for mutations affecting 5′ splice site (ss). (**a1**) A wild-type endogenous U1 small nuclear ribonucleoprotein (snRNP) does not bind to 5′ss due to the presence of a 5′ss mutation. (**a2**) An exogenous U1 snRNP (first generation particle) is modified in 5′ tail with a compensatory alteration (semi-circle) that allows for base-pairing with the mutated 5′ss and the restoration of exon recognition and inclusion. (**b**) The presence of a 5′ss mutation does not allow the correct 5′ss recognition by U1 snRNP but an exon-specific U1 snRNP (second generation particle) with an engineered 5′ tail which binds a downstream non-conserved intronic region can activate the mutated 5′ss, through a mechanism not yet fully understood, allowing the correct exon recognition and inclusion

mation, protecting pre-mRNA transcripts against premature polyadenylation and contributing to the regulation of alternative polyadenylation [115, 119–121].

Splicing mutations at the 5′ss, which are frequent among defects that cause human disease, compromise U1 snRNA binding and can prevent spliceosome assembly and subsequent splicing, which results in exon skipping, intron retention or activation of cryptic splice sites [117, 122]. The most deleterious mutations at a 5′ss are those affecting the nearly invariant GU dinucleotide at the positions +1 and +2. For the remaining nine positions the effects on splicing are less understood. Indeed, nucleotide substitutions in the less conserved positions can cause splicing defects in several but not all 5′ss, suggesting that this 5′ss positions and/or the general context define at what level splicing is changed [117, 123].

7.3.2 Two Generations of Engineered U1 snRNAs to Correct Splicing Defects

As donor splice site mutations disrupt the complementarity of the donor site with the endogenous U1 snRNA, restoring the complementarity through engineered modification of the U1 snRNA represents a valuable approach (Fig. 7.5a1, a2). In fact, in the mid-80s, Zhuang and Weiner [124] demonstrated for the first time that modified U1 snRNAs were able to suppress 5′ss mutations. Since then, the physiological role

of the U1 snRNA to promote exon inclusion in the presence of 5′ss mutations affecting different positions of the donor site, has been extensively exploited as a possible therapy for numerous diseases. Significant correction levels have been achieved for mutations located in less conserved 5′ss positions in diseases like Neurofibromatosis type 1 [125], Coagulation factor VII deficiency [126, 127], Retinitis pigmentosa [128, 129], Propionic acidemia [130], Phenylketonuria [131] and Bardet-Biedl syndrome [132, 133]. Furthermore, recent studies have also demonstrated the feasibility of this approach in the more conserved GU region. In fact, partial correction of splicing defects caused by mutations in the +1 position of 5′ss was observed, not only in a Fanconi anemia case [134], but also in Sanfilippo C disease patient cells [135]. Also, for a mutation in the +2 position causing Hemophilia B, the treatment with a modified U1 snRNA led to an increase in the proportion of correct transcripts (~20%) [136]. In general, though, results of U1 snRNA therapeutic approaches can vary depending on the nature of the mutation and on the overall genomic context.

Until now, modified U1s effects *in vivo* were only addressed in two studies. In the first one, Balestra and co-workers [137] showed the rescue of the expression of a splicing-defective human factor VII (hFVII) mutant by a mutation-adapted U1 snRNA which improved hFVII circulating levels in mice, highlighting the potential of this strategy as a therapy for FVII coagulation deficiency. In the second study, Lee et al. [138] demonstrated the therapeutic effect of a mutation-adapted U1 snRNA in a knock-in mouse model of Aromatic L-amino acid decarboxylase (AADC) deficiency.

In common with other rescue strategies based on targeting RNA by complementarity, modified U1 snRNAs have to deal with potential off-target effects. This is particularly dangerous for modified U1 snRNAs with only one base change from the natural U1 snRNA, which might activate normally silent cryptic donor splice sites and induce aberrant splicing in other genes [139]. The consequences of such unwanted side reactions are hard to predict and depend on the function of the spliced transcript. However, the binding site sequence screening and mapping against the human genome to rule out sequence homologies should extensively decrease nonspecific events although their total exclusion cannot be guaranteed [140]. Therefore, experimental analysis should be performed whenever possible to test the effect of the U1 treatment on non-target transcripts. In mutation-adapted U1 snRNA *in vitro* approaches to correct 5′ splicing defects in Retinitis pigmentosa [129] and Bardet-Biedl syndrome [133], this type of test was performed and no missplicing events were found in the non-target transcripts. Also, in the *in vivo* U1 snRNA therapeutic strategy for AADC deficiency, the treatment was well tolerated and no toxic effects were seen within the study period [138]. However, in the *in vivo* study for hFVII deficiency [137], the authors observed hepatotoxicity, most probably caused by the binding of the engineered U1 to similar consensus 5′ss in other genes.

It was previously shown that U1 snRNAs do not necessarily have to bind at the 5′ss to promote exon definition. Some atypical 5′ss are recognized by U1 snRNA shifted by one nucleotide [141] and U1 snRNAs complementary to intronic sequences downstream of the 5′ss were originally reported to enhance the recognition of 5′ss in model gene systems [142, 143]. Given this, to reduce the possible interaction of modified U1 snRNAs with non-target 5′ss, a second generation of engineered U1s called Exon-Specific U1 snRNAs (ExSpeU1s) was developed. The ExSpeU1s have engineered 5′ tails that direct their loading into non-conserved intronic regions downstream of the 5′ss of a specific exon, and are expected to improve specificity and reduce potential off-target events [139, 144] (Fig. 7.5b). In different cellular models (i.e. minigene assays, patient's cells or iPSC's), a number of ExSpeU1s has been successfully applied, allowing an efficient rescue of exon skipping caused by various types of splicing mutations in Hemophilia B [144, 145], Cystic Fibrosis [144], SMA [144, 146, 147], Fanconi anemia [148] and Netherton syndrome [149]. The ExSpeU1 strategy has also been investigated in mouse models. For SMA, Dal Mas et al. [146], reported that AAV-mediated

delivery of ExSpeU1 corrects splicing, increasing the inclusion of *SMN2* exon 7 in different tissues/organs. In another study, Rogalska and colleagues [150] created a mouse expressing a particular ExSpeU1 and, after crossing it with a severely affected SMA mouse, observed increased inclusion of the missing exon, followed by SMN protein production and increased mice lifespan. Possible gene expression side effects were also addressed and, from a panel of 12,414 analysed genes, only 12 had altered expression after treatment.

ExSpeU1 molecules have also successfully rescued splicing in a transgenic mouse model of Familial Dysautonomia, a rare genetic disease with no treatment [151]. For Hemophilia B, another ExSpeU1 was explored in mice expressing two natural *F9* splicing defective variants at 5′ss or 3′ss, and efficiently rescued human *F9* splicing in liver resulting in an increase of the target protein and coagulation activity [152]. This study, as the pivotal one developed by Fernandez Alanis et al. with the ExSpeU1 strategy [144], interestingly showed that a single ExSpeU1 can be used to correct exon-skipping mutations at the consensus 5′ss (apart from canonical GU dinucleotide), the polypyrimidine tract, and even at exonic regulatory elements, thus extending the applicability of ExSpeU1s to panels of mutations and cohorts of patients with the same genetic disorder.

Despite the promising results obtained with ExSpeU1s in different studies, its precise mechanism of action for splicing correction is not totally clear. Pagani and co-workers demonstrated that ExSpeU1s are assembled as U1-like particles and that their splicing rescue activity is dependent on the U1-70K protein and on the loop IV structure of the U1 snRNA; not on the recruitment of endogenous U1 snRNP to the upstream 5′ss [150]. This may indicate that ExSpeU1s promote correct exon recognition through the recruitment of splicing factors that subsequently activate the mutated 5′ss [144, 146, 149–152]. However, it is important to stress that the splicing stimulator activity of ExSpeU1s was also responsible for the activation of a cryptic 5′ss in an approach attempted to correct a splicing defect causing Intrahepatic Cholestasis, which resulted in the production of an additional splice transcript with intron retention [153].

Globally, both mutation-adapted U1 snRNA and exon-specific U1 snRNA constitute a novel therapeutic strategy to correct splicing defects associated to defective exon definition in several human disorders. Once the U1 snRNA-mediated approaches act at pre-mRNA level, they have the main advantage of maintaining the regulated expression of the targeted gene in the normal chromosomal context [139, 154]. Also, given that the U1 snRNA gene used for splicing rescue includes promoter and regulatory sequences, it has the capability of guaranteeing long term correction of the genetic defect [139]. Despite these advantages, U1 snRNA-mediated therapies may also face some problems such as the presence of off-target effects and low efficacies. Therefore, in a near future, it will be imperative not only to develop a specific method or tool to search for off-target effects, but also to adjust the expression levels of U1 snRNA therapeutic particles in preclinical *in vivo* studies [154].

7.3.3 Engineered U1 snRNAs Delivery

U1 snRNA-mediated therapies also have to deal with the challenge of an efficient delivery to a target tissue. In the *in vivo* studies already developed, one of the most successful gene therapy systems available nowadays – AAV vectors – has been chosen as the method for U1 snRNA-engineered particles delivery into mice [137, 138, 146, 150, 151]. AAV vectors allow a highly efficient delivery to various tissues following systemic injection, even though dependent on the viral serotype used [90, 155]. Also, the low packaging capacity of AAV vectors is quite adequate for U1 snRNA-based approaches given the small cassette size to package [90]. However, despite the modifications that have been introduced in viruses, the potential for antiviral immunity and phenotoxicity of the transgene are still major limitations to the use of viral vectors for therapy.

Possible alternatives to viruses are liposomes and nanoparticle delivery [155].

Among the several RNA tools enabling the rescue of splicing, both mutation-adapted U1 snRNA and ExSpeU1 snRNA therapeutic strategies have already shown their efficacy to repair different types of splicing defects at least in animal models of disease. Still, further developments will be necessary for this therapeutic approach to be translated to human trials.

7.4 siRNA-Based Drugs

7.4.1 Brief Overview

The last decade of the twentieth century has also witnessed the discovery of a new mechanism of gene regulation whose therapeutic potential is still being unveiled: RNA interference (RNAi). Interestingly, the first experimental observation of this mechanism came up from a failed genetic experiment aimed at developing more attractive petunia flowers. In fact, in 1990, Jorgensen and co-workers attempted to genetically engineer flower pigmentation genes, to be inserted into the target plant genome. To their surprise, however, instead of generating more colorful flowers, they ended up producing a generation of plants that had virtually lost all pigmentation, thus becoming white. This observation prompted additional studies to check the expression levels of endogenous genes involved in the natural pigmentation biosynthetic pathway and most of them were strongly reduced. Thus, a concept of co-suppression, whereby sequence-related genes could negatively regulate each other, was born [156, 157]. Still, little was known on its underlying mechanism. The first major breakthrough came from the pivotal studies by Andrew Fire and Craig Mello. By introducing various forms of long RNA molecules into *C. elegans*, their team observed that those with a double-stranded presentation (double stranded RNAs, dsRNAs) were the actual inducers of the silencing phenomenon, which was then coined RNAi [158]. Thus, the work by Fire and Mello, which earned them the 2006 Nobel Prize of Medicine, has not only represented a major advance in the understanding of RNAi basic mechanism, but also provided a simple and reproducible method by which long dsRNAs could be used to induce specific gene silencing in lower organisms commonly used in genetic research, such as *C. elegans* [159] and *D. melanogaster* [160]. In the meantime, other teams kept their focus in plant systems, aiming at a better understanding of the role that RNAi and additional silencing processes assume in plant homeostasis. Soon it became clear that gene silencing operating at the RNA level has roles in adaptative protection against viruses [161], genome defense against mobile DNA elements [162, 163] and developmental regulation of gene expression (reviewed in [164]). A second component of RNA silencing, in addition to dsRNAs, was then identified and coined short interfering RNAs, which resulted from the processing of dsRNAs into 21–26 nt counterparts [165]. Interestingly, those short interfering RNA molecules could be sorted into two classes depending on their size, and soon it became clear that each of those classes assumed different functions. The long ones (24–26 nt) were dispensable for sequence specific mRNA degradation, but essential for systemic silencing and methylation of homologous DNA [164]. Another interesting contribution to the deeper understanding of the overall RNA silencing process came from a work of Cogoni and co-workers, who described a new biological function for RNA silencing in *Neurospora* called quelling, which can be activated upon the introduction of transgenic DNA. These authors observed that quelling targets preferentially transgenes arranged in large tandem arrays and its effectors are also short interfering RNAs [166], reviewed in [167]. Altogether, these works unveiled an unexpected complexity in the RNA silencing process in plants, prompting additional studies to check whether the same would also apply in animals. By this time, however, no one foresaw that the RNAi mechanism would also work in mammalian systems because long dsRNAs were already known to induce a strong interferon response. The first demonstrations that RNAi also works in humans came from the work of two

independent groups in Germany, one operating at the Max Planck Institute, and the other at the University of Bayreuth, and a third one in the United States, operating at the NIH, Bethesda. The team at the Max Plank Institute showed that synthetic versions of short dsRNA molecules were able to trigger a strong gene silencing effect in mammalian cells without inducing the interferon response. Moreover, they tested a series of design features for those short dsRNAs including length, blunt or sticky ends and chemical modifications, finding that structurally defined 21–23 base-pair small RNAs, with 2 nucleotide unpaired overhangs at the 3′ ends, were the most efficient mediators of RNAi [168, 169]. This fundamental work was published in Nature in 2001, and became the scientific content for a key patent in the field called "Tuschl II". In parallel, the NIH team came up with another demonstration that synthetic siRNAs can induce gene-specific inhibition of expression in *C. elegans* and in cell lines from humans and mice. They did it by systematically comparing the level of gene expression decrease caused by siRNAs *versus* that caused by single stranded AONs [170]. Their work, published in PNAS, was another step to open a path toward the use of siRNAs as a reverse genetic and therapeutic tool in mammalian cells, as the authors themselves have stated. Around the same time, at the University of Bayreuth, Kreutzer and Limmer had also reasoned that short fragments of dsRNA would putatively mediate a RNAi response similar to the one originally described by Fire and Mello and, even though their findings have never been published, they did file key patents around the discovery. Additional studies on the subject ended up unveiling the endogenous RNA silencing pathway that was being fed by small dsRNAs, from now on called small interfering RNAs (siRNAs). It also became evident that the same pathway is also able to process microRNAs (miRNAs), as previously seen (Fig. 7.1c). Here, we will focus solely on the RNAi process triggered by siRNAs.

The siRNA pathway starts with the cytoplasmic cleavage of long dsRNAs by an enzyme called Dicer. As a result, short dsRNA duplexes are formed. Then, those dsRNAs are incorporated into the RNA-induced silencing complex (RISC), where the strands are separated, and one strand guides RISC to the complementary region of target mRNA (Fig. 7.6). The heart of RISC is the Argonaute (AGO) proteins. In humans there are 8 AGO proteins, 4 from AGO clade (AGO1-4) and 4 from P-element induced wimpy testis (PIWI) clade (PIWI1-4; [171]). Still, not all AGO proteins are cleavage competent. In fact, AGO2 is the sole executer that accomplishes siRNA-induced silencing. Thus, whenever the siRNA strand loaded into RISC has complete sequence complementarity with its target mRNA sequence, it triggers site-specific mRNA cleavage, which ultimately results in a reduced expression of that mRNA and of the target protein (Fig. 7.4; reviewed in [172]). This exact same process can also be induced by direct exogenous supply of synthetic siRNAs. Over the years, a series of empirical and rational guidelines started accumulating from the analysis of hundreds of functional siRNAs. There are now a number of guidelines one should follow in order to design an effective siRNA, which have been well reviewed elsewhere [173]. There are also many websites and companies that either offer reliable methods for the design of effective siRNAs or even design them on demand. Because of their small size, the chemical synthesis of siRNAs is relatively easy and nowadays, several companies offer them delivered in ready-to-transfect format. This is, therefore, a simple, easy-to-handle RNAi-effector for virtually every lab need.

Since the half-life of siRNA is short, an alternative RNAi-effector molecule has also been developed: short hairpin RNAs (shRNAs), which are not directly transfected into their target cells. Instead, shRNAs are transcribed in the nucleus from an exogenous DNA expression vector bearing a palindromic sequence with a spacer in between, whose transcript folds into a short dsRNA with a terminal loop. The shRNA transcript is processed by Drosha, an RNase III endonuclease. The resulting pre-shRNA is exported to the cytoplasm, where it can then be processed by another RNase III, called Dicer, and incorporated into RISC, thus triggering the same RNAi process previously described (reviewed in [172]). In

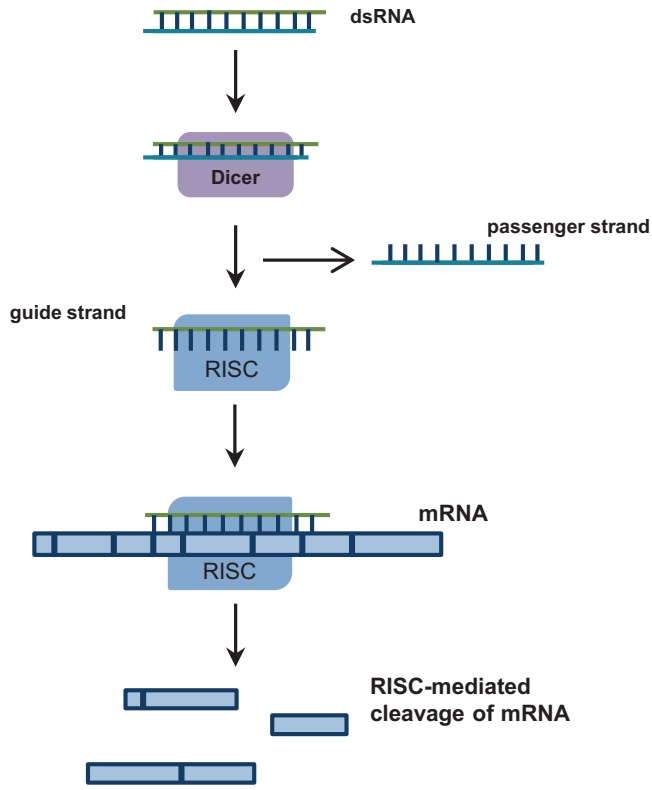

Fig. 7.6 The RNA interference (RNAi) mechanism. Entry of double-stranded RNA (dsRNA) into eukaryotic cells results in targeted RNA-induced silencing complex (RISC)-mediated cleavage of messenger RNA (mRNA) through activation of the endogenous RNAi mechanism: dsRNAs are recognized and cleaved into shorter fragments by Dicer, and subsequently loaded into a multiprotein conglomerate called RISC, which facilitates the separation of the two RNA strands. Once the double-stranded RNA is separated, one strand gets degraded while the other associated with RISC acts as a template for RISC-mediated cleavage of complementary RNA, thus reducing protein translation

general, shRNAs are harder to complex/internalize. Still, by delivering DNA instead of the effector RNA molecule, they take advantage of the cell's transcription machinery to produce specific shRNA transcripts, they allow for high potency sustainable effects using low-copy numbers. One such approach results in less off-target effects, putatively ensuring greater safety. Additionally, a shRNA expression vector does also cost less than the bulk manufacturing of siRNAs (reviewed in [174]).

Once all cells have the RNAi machinery and, in principle, any gene can be knocked down, soon siRNAs became invaluable tools in the lab, enabling the easy genetic knockdown of any sequence. RNAi was rapidly exploited as a tool to promote unbiased genome-wide screening to search for relevant genes involved in specific biological processes, first in invertebrate cells [170, 175–177] and latter in mammalian cells [178–182]. In fact, this knockdown technique provides a valuable tool for the functional annotation of mammalian genes [183, 184], for the creation of knockout animals [185] and for the identification of new drug targets (reviewed in [186]), but these are far from being the major application of this technology. In fact, RNAi has been regarded as one of the major breakthroughs in the field of molecular medicine, and its potential as a therapeutic effector has been largely tested over the last decades.

The need to optimize the technique and take it from bench to clinic is also prompting extra research efforts to gain a deeper understanding of

the overall RNAi mechanism. For example, in order to function, siRNAs need to escape to the cytosol, where the RISC works. Thus, release from the endosome is an important barrier. Understanding the mechanism(s) that promotes and limits endosomal release may help to optimize this limiting step. This remains, though, a major area of investigation for all nucleic acid therapeutics [11].

7.4.2 Recent Successful Applications of siRNA-Based Drugs

Being a naturally occurring post-transcriptional gene silencing process, this mechanism has several advantages when compared to other AON technologies, and that recognition triggered major investments in RNAi-based drug development by large pharmaceutical and biotechnological companies [187]. The potential of siRNA therapeutics was first demonstrated by Song and co-workers 15 years ago, when injection of *Fas* siRNAs protected mice from autoimmune hepatitis. Fas-mediated apoptosis is implicated in a broad spectrum of liver diseases, where inhibiting hepatocyte death can be life-saving. These authors investigated the silencing effect of siRNA duplexes targeting the gene encoding the Fas receptor (*Fas*), to protect mice from liver failure and fibrosis in two models of autoimmune hepatitis. Intravenous injection of *Fas* siRNA specifically reduced both *Fas* mRNA and Fas protein expression levels in mouse hepatocytes, and the effects persisted without diminution for 10 days [188]. This pioneer work has not only shown that siRNA-directed Fas silencing could work *in vivo* and be of therapeutic effect for preventing and/or treating acute and chronic liver injury [188], but also provided the proof-of-principle on the potential of the overall RNAi technology to treat or prevent disease (reviewed in [189]). Since then, drug development has been rapid, with siRNAs facing virtually the same obstacles as AONs. Fortunately, some of the AON strategies could be adapted to siRNA therapeutics, thus accelerating siRNA preclinical drug development and clinical evaluation. In general, RNAi clinical trials are progressing well. Clinical Phase I and II studies of siRNA therapeutics have demonstrated potent (as high as 98%) and persistent (lasting for weeks) gene knockdown effects, especially in liver, with some signs of clinical improvement and without unacceptable toxicity (reviewed in [189]). There are also several trials in Phase III development (Table 7.2; reviewed in [11]).

Early this year Alnylam has announced U.S. FDA acceptance of New Drug Application (NDA) and Priority Review Status for Patisiran, an investigational RNAi therapeutic for the treatment of hereditary transthyretin amyloidosis (hATTR) [225]. Almost at the same time, the company presented new clinical results from the APOLLO Phase III study of this drug at the 16th International Symposium on Amyloidosis. The APOLLO Phase III trial was a randomized, double-blind, placebo-controlled, global study designed to evaluate the efficacy and safety of Patisiran in hATTR amyloidosis patients with polyneuropathy. The primary endpoint of the study was the change from baseline in modified Neurologic Impairment Score +7 (mNIS+7) relative to placebo at 18 months. According to the general manager of the transthyretin (TTR) program at Alnylam, "*the clinical results presented further highlight the robust profile of Patisiran and provide evidence supporting Patisiran as a potentially transformative treatment approach for patients with hATTR amyloidosis*". Also the results obtained in the cardiac subpopulation, which corresponded to approximately 50% of the patients enrolled in the APOLLO study, revealed significant improvements in measures of cardiomyopathy, the leading cause of death in patients with hATTR amyloidosis, relative to placebo [201]. Finally, in August 2018, the drug got its U.S. FDA approval, and is now commercialized under the designation Onpattro™ [226].

Hopefully, the approval of the first RNAi therapeutic will pave the way for approval of other targets (reviewed in [227]), especially if we take into account that there are several other siRNA drugs under evaluation, which have recently advanced for phase III development (Table 7.2; reviewed in [11]). The most relevant examples

Table 7.2 Summary table of clinical trials for Small interfering RNAs (siRNAs)

Chemistry	Drug	Target/organ	Dose/route	Indication	Key observations	References
Phase 3						
PS, 2′-OMe, 2′-F (LNP)	Patisiran (ALN-TTR02)	*TTR*, Liver	0.3 mg/kg once every 3 weeks, IV (Pretreatment w/ glucocorticoids, antihistamines and analgesics)	Hereditary ATTR	Phase 3 findings included (1) significant improvement with patisiran compared to placebo in neuropathy and quality of life metrics, (2) infusion-related reactions, and (3) peripheral edema	NCT01960348 (Coelho [190, 191]) (Alnylam Pharmaceuticals, press release, Sept 20, 2017)
2′-MOE (None)	QPI-1002 (15NP)	*TP53*, Kidney	0.5–10 mg/kg, IV	Delayed graft function; AKI following cardiac surgery	Phase 2 findings in renal transplant patients included improved (1) dialysis-free survival time, and (2) reduced number dialysis in first 30 days	NCT02610296 NCT02610283 [192, 193]
Unmodified (None)	SYL1001	*TRPV1*, Eye	11.25 g/mL/drop, Eye drops	Ocular pain; dry eye syndrome	Phase 2 findings included analgesic trend that was not significant compared to placebo	NCT03108664 [194]
2′-MOE (None)	QPI-1007	*CASP2*, Eye	1.5 mg, IVT	Acute NAION, acute primary angle closure glaucoma	Phase 1 findings included conjunctival hemorrhage	NCT02341560 [195, 196]
PS, 2′-OMe, 2′-F, GalNAc (None)	Fitusiran (ALN-AT3SC)	*SERPINC1*, Liver	80 mg once monthly, SC	Severe Hemophilia A or B	Phase 1/2 open-label and open-label extension findings in patients included (1) mean 80% target reduction and reduced bleeding, and (2) 1 death due to thromboembolic event	NCT03417245 NCT03417102 NCT02035605 NCT02554773 [197, 198]
PS, 2′-OMe, 2′-F,GalNAc (None)	Inclisiran (ALN-PCSSC)	*PCSK9*, Liver	300 mg once every 3–6 months, SC	ASCVD, HC, HoFH, HeFH	Phase 2 findings included (1) mean 53% reduction of LDL-C, (2) injection site reactions, and (3) ALT elevations	NCT03400800 NCT03399370 NCT03397121 [199, 200]

(continued)

Table 7.2 (continued)

Chemistry	Drug	Target/organ	Dose/route	Indication	Key observations	References
PS, 2′-OMe, 2′-F, GalNAc (None)	Givosiran, (ALN-AS1)	*ALAS1*, Liver	2.5 mg/kg once monthly, SC	Acute hepatic porphyrias	Phase 1 findings in ASHE and AIP patients included (1) dose-dependent reduction of target, (2) 63% reduction in annualized AIP attack rate, (3) injection site reactions, and (4) myalgia	NCT03338816 [201], press release, Nov 7, 2017) NCT02452372
Phase 2						
Undisclosed (PLGA polymer)	siG12D-LODER	mutant *KRAS*, Tumor	0.025-3 mg, Surgical implant	Pancreatic ductal adenocarcinoma	Phase 2 findings included (1) suggestion of improved median survival, and (2) cholangitis and neutropenia	NCT01676259 NCT01188785 [202, 203]
Unmodified (None)	Bamosiran (SYL 040012)	*ADRB2*, Eye	300-900 mg/eye/day, Eye drops	Ocular hypertension; open angle glaucoma	Phase 1/2 findings included (1) reduction of intraocular pressure, and (2) no drug related adverse events	NCT02250612 [204, 205]
Undisclosed (LNP)	TKM-080301 (TKM-PLK1)	*PLK1*, Liver	0.15-0.9 mg/kg, IV	Solid tumors with liver involvement; neuroendocrine tumors; adrenocortical carcinoma	Phase 1/2 findings included (1) some evidence of antitumor activity, (2) infusion reaction-like inflammatory response, (3) ALT elevations, and (4) acute respiratory failure	NCT01437007 NCT01262235 NCT02191878 [206]
Undisclosed (None)	RXI-109	*CTGF*, Skin/ Eye	1-10 mg, ID/IVT	Hypertrophic scars; keloids, and subretinal fibrosis	Phase 1 findings included modest target reduction	NCT02246465 NCT02079168 NCT02599064 [207]
2′-OMe (cationic lipoplex)	Atu027	*PKN3*, Vascular endothelium	0.001-0.336 mg/kg, IV (administered w/ gemcitabine)	Advanced solid tumors; pancreatic ductal carcinoma	Phase 1/2 findings included possible anti-tumor effects	NCT00938574 NCT01808638
2′-OMe (None)	PF-655	*DDIT4*, Eye	0.05-3 mg/eye, IVT	AMD	Phase 1/2 findings included (1) no evidence of benefit in AMD patients, and (2) increased intraocular pressure	NCT01445899 [208]

PS, 2′-OMe, 2′-F, GalNAc	Cemdisiran (ALN-CC5)	*C5*, Liver	600 mg once every 4 weeks, SC	Complement mediated diseases (aHUS, PNH)	Open-label phase 1 findings included (1) sustained control of hemolysis with up to 67% reduction in eculizumab dose, (2) 99% reduction of C5, (3) injection site reactions, and (4) 1/6 (17%) Gr 3 LFT	NCT03303313 NCT02352493 [209]
PS, 2′-OMe, 2′-F, GalNAc (None)	Lumisiran (ALN-GO1)	*HAO1*, Liver	1.0 mg/kg, SC	1st Hyperoxaluria type 1	Phase 1 findings in healthy volunteers included (1) increased plasma and urine glycolate, and (2) injection site pain	NCT02706886 [210]
3 siRNAs Undisclosed (LNP)	ARB-1467 (TKM-HBV)	*HBV*, Liver	2 mg/mL, IV in combination w/ nucleoside analogues (Pretreatment w/ medication to mitigate infusion-related reactions)	HBV infection	Phase 1 findings in patients included dose-dependent additive HBsAg reduction	NCT02631096 [211]
Phase 1						
PS, 2′-OMe, 2′-F, GalNAc (None)	ALN-TTR$_{SC}$02	*TTR*, Liver	5-300 mg, SC	Hereditary ATTR	Phase 1 findings in healthy volunteers included (1) dose dependent reduction in plasma TTR	NCT02797847 (Alnylam Pharmaceuticals, press release Dec 16, 2016)
Undisclosed (LNP, vitamin A)	ND-L02-s0201	*SERPINH1* Liver	0.03-0.6 mg/kg/wk, IV (Pretreat w/ antihistamines)	Hepatic fibrosis	Phase 1b findings included (1) no doselimiting toxicities, and (2) infusion reactions.	NCT02227459 [212]
Undisclosed (siRNAelectroporated PBMCs)	APN401	*CBLB*, Tumors	5-50 3 10^5 cells /kg, IV	Melanoma, pancreatic cancer, renal cell cancer	(1) Immune modulation shown by in vitro increase in IFN-g and IL-2 with CD3/CD28 stimulation, and (2) infusion reaction like adverse events	NCT02166255 NCT03087591 [213, 214]

(continued)

Table 7.2 (continued)

Chemistry	Drug	Target/organ	Dose/route	Indication	Key observations	References
Undisclosed (liposome)	siRNA-EphA2-DOPC	*EPHA2*, Tumors	IV	Advanced cancers	Not published	NCT01591356
2 siRNAs, Undisclosed (LNP)	ALN-VSP02	KIF11 and *VEGFA*, Liver/Tumors	0.1-1.5 mg/kg, IV (Pretreatment w/ glucocorticoids, antihistamines, and analgesics)	Advanced solid tumors with liver involvement	Phase 1/2 findings included (1) evidence of antitumor activity, and (2) infusion reactions, dose-limiting thrombocytopenia, and complement split product Bb increase	NCT01158079 [215, 216]
Undisclosed (LNP)	ALN-PCS02	*PCSK9*, Liver	0.4 mg /kg, IV (Pretreatment w/ glucocorticoids, antihistamines, and analgesics)	HC	Phase 1 findings included (1) reductions in PCSK9 of up to 84% and LDL cholesterol plasma levels of up to 50% after a single IV dose, and (2) infusion reactions	NCT01437059 [217]
Undisclosed (None)	TD101	*KRT6A*, Nails/Skin	0.1-17 mg, Intra-lesion	PC	Limited data from one patient	NCT00716014 [218]
Undisclosed (gold NP)	NU-0129	*BCL2L12*, Cancer	IV	Glioblastoma multiforme	Not published	NCT03020017
Discontinued						
Unmodified (None)	Bevasiranib (Cand5)	*VEGFA*, Eye	IVT	Macular degeneration, diabetic macular edema	Terminated in Phase 3 due to lack of efficacy.	NCT00499590 (Health, press release Mar 2009)
Unmodified (None)	AGN211745	*FLT1*, Eye	0.100-1 mg, IVT	AMD w/choroidal neo-vascularization	Terminated after Phase 2 due to lack of efficacy	NCT00395057
Undisclosed, GalNAc (None)	ALN-HBV	*HBV*, Liver	0.1-3.0 mg/kg, SC	Non-cirrhotic chronic HBV	Replaced by more potente compound	NCT02826018

2′-OMe, 2′-F, GalNAc (None)	Revusiran (ALN-TTRSC)	*TTR*, Liver	500 mg once weekly, SC	TTR-mediated cardiac amyloidosis	Terminated due to imbalance of mortality in the drug arm compared to placebo, adverse events included ISRs and LFT elevations	NCT02319005 [219, 220]
Unmodified (SNALP)	PRO-040201 (TKM-ApoB)	*APOB*, Liver	IV	HC	Terminated after phase 1due to infusion reactions	NCT00927459 [221]
Undisclosed (RONDEL™ cyclodextrin insert polymer, adamantane-PEG-transferrin)	CALAA-01	*RRM2*, Tumors	3-30 mg/m^2, IV	Solid tumors	Terminated in phase 1due to dose-limiting toxicities, including infusion reaction-like adverse events, elevated creatinine and LFTs, and dose-dependent thrombocytopenia	NCT00689065 [222]
Undisclosed (LNP)	DCR-MYC	*MYC*, Tumors	0.1-0.85 mg/kg, IV	Multiple cancers	No evidence of clinical activity	NCT02110563 NCT02314052 (Dicerna Pharmaceuticals, press release Sept 26, 2016)
Undisclosed (LNP)	DCR-PH1	*AGXT*, Liver	0.005-0.05 mg/kg, IV	Primary hyperoxaluria type 1		NCT02795325
PS, 2′-OMe, 2′-F, GalNAc (None)	ALN-AAT	*SERPINA1*, Liver	0.1-6.0 mg/kg, SC	Alpha-1 Antitrypsin deficiency	Terminated after phase 1/2 due to LFT elevations.	NCT02503683
Undisclosed, cholesterol (MLPCDM-GalNAc)	ARC-520	*HBV*, Liver	4 mg/kg, IV (Pretreatment w/ antihistamine and acetaminophen)	Chronic HBV	Terminated due to toxicities in NHPs, but had favorable tolerability in clinic	NCT02349126 (Arrowhead Pharmaceuticals, press release Nov 29, 2016)
Undisclosed, cholesterol (MLPCDM-GalNAc)	ARC-521	*HBV*, Liver	0.6-6.0 mg/kg, IV (Pretreatment w/ antihistamine and acetaminophen)	Chronic HBV	Terminated due to toxicities in NHPs, but had favorable tolerability in clinic.	NCT02797522 [223] (Arrowhead Pharmaceuticals, press release Nov 29, 2016)

(continued)

Table 7.2 (continued)

Chemistry	Drug	Target/organ	Dose/route	Indication	Key observations	References
UNA, cholesterol (MLP-CDM-GalNAc)	ARC-AAT	*SERPINA1*, Liver	0.38-8.0 mg/kg, IV (Pretreatment not disclosed, but assumed same as other ARC drugs)	Alpha-1 antitrypsin deficiency	Terminated due to toxicities in NHPs, but had favorable tolerability in clinic and dose-dependent reductions in sérum AAT up to 90%	NCT02363946 [224] (Arrowhead Pharmaceuticals, press release Nov 29, 2016)
Undisclosed (LNP)	ALN-TTR01	*TTR*, Liver	1 mg/kg, IV (Pretreatment w/ glucocorticoids, anti-histamines and analgesics)	Hereditary ATTR	Terminated due to greater potency of patisiran (ALN-TTR02)	NCT01148953 [190]

Abbreviations: **Chemistry:** *PS* phosphorothioate, *2′ –Ome* 2′ –methoxy, *2′ – F* 2′ – fluoro, *2′ – MOE* 2′ –O- methoxy ethyl, *GalNAc* N-Acetilgalactosamine, *PLGA* poly(lactic-co-glycolic) acid, *LNP* lipid nanoparticle, *NP* nanoparticle, *PEG* polyethylene glycol, *SNALPs* stable nucleic acid-lipid paricles, *MLP* melittin-like protein, *CDM* carboxy-dimethylmaleamide, *UNA* unlocked nucleic acid. **Target/ organ:** *TTR* transthyretin, *TP53* tumor protein 53, *TRPV1* transient receptor potential cation chanel subfamily V member 1, *CASP2* caspase2, *SERPINC1* serpin family C member 1, *PCSK9* proprotein convertase subtilisin/kexin type 9, *ALAS1* 5′-aminolevulinate synthase 1, *KRAS* KRAS proto-oncogene, *ADRB2* adrenoreceptor beta 2, *PLK1* Serine/threonine-protein kinase, *CTGF* connective tissue growth factor, *PKN3* protein kinase N3, *DDIT4* DNA damage inducible transcript 4, *C5* complement C5, *HAO1* hydroxyacid oxidase 1, *HBV* Hepatitis B Virus, *SERPINH1* serpin family H member 1, *CBLB* Cbl proto-oncogene B, *EPHA2* ephrin receptor A2, *KIFF11* kinesin family member 11, *VEGFA* vascular endothelial growth factor, *KRT6A* keratin 6A, *BCL2L 12* BCL2 like 12, *FLT1* fms related tyrosine kinase 1, *APOB* apolipoprotein B, *RRM2* ribonucleotide reductase regulatory subunit M2, *MYC* MYC proto-oncogene, *AGXT* alanine-glyoxylate and serine-pyruvate aminotransferase, *SERPINA1* serpin family A member 1. **Dose/route:** *IV* intravenous, *IVT* intravitreal, *SC* subcutaneous, *ID* intradermal. **Indication:** *ATTR* transthyretin amyloidosis, *AKI* acute kidney injury, *NAION* non-arteritic anterior ischemic optic neuropathy, *ASCVD* atherosclerotic cardiovascular disease, *HC* hypercholesterolemia, *HoFH* homozygous familial hypercholesterolemia, *HeFH* heterozygous familial hypercholesterolemia, *AMD* age-related macular degeneration, *aHUS* atypical hemolytic-uremic syndrome, *PNH* paroxysmal nocturnal hemoglobinuria, *HBV* Hepatitis B Virus, *ATTR* transthyretin amyloidosis, *HC* hypercholesterolemia, *PC* pachyonychia congenital, *AMD* age-related macular degeneration. **Key observations:** *ALT* alanine aminotransferase, *ASHE* asymptomatic high excreters, *AIP* acute intermittent porphyria, *AMD* age-related macular degeneration, *C5* complement C5, *LFT* liver function tests, *TTR* transthyretin amyloidosis, *IFN-g* interferon gamma, *IL-2* interleukin 2, *PCSK9* Proprotein convertase subtilisin/kexin type 9, *LDL* Low-density lipoprotein, *ISR* injection site reaction, *NHP* non-human primate, *AAT* Alpha-1-antitrypsin

include Revusiran, which is a second siRNA drug under evaluation as a treatment option for patients with familial amyloid cardiomyopathy by reducing plasma TTR levels [220]; QPI-1002, which is being developed for the treatment of delayed graft function for kidney transplants [192]; Fitusiran, a potent siRNA drug under study for patients with hemophilia, which aims at ameliorating the disorder by reducing the plasma levels of anti-thrombin [198], and Inclisiran, which targets proprotein convertase subtilisin/kexin type 9 (PCK9) to reduce the risk of cardiovascular disease [200], as reviewed in [11].

7.4.3 Delivery of siRNA-Based Drugs

Intracellular delivery of double-stranded siRNAs is more challenging than delivery of single stranded AONs [186]. Still, it is also worth mentioning that, as suggested by the best estimates, only a few hundred cytosolic siRNAs per cell are needed for efficient and sustained gene knockdown [228, 229]. This happens because the guide strand of the siRNA remains stable within the RISC for weeks, even though it gets diluted with every cell division [230]. Thus, the same siRNA molecule can target multiple transcripts, knocking down gene expression in slowly dividing or non-dividing cells over the same period. Overall, as noticed by several authors, this actually contributes to turn the delivery obstacle into a less formidable one than that faced by other antisense mechanisms, which act on a one-to-one basis [230] (reviewed in [229]).

Still, knowing that the translation of siRNAs from the bench to the clinic would be hindered by their limited cellular uptake, low biological stability and unfavorable pharmacokinetics, the development of appropriate delivery methods became mandatory to proceed with preclinical studies. Therefore, different approaches have been (and are being) attempted to ensure safer and long-lasting delivery methods for siRNA-based drugs, for both systemic and targeted delivery. Most of these developments were made in parallel with siRNA drug development and only through the combined efforts of several independent teams were these drugs modified in ways that allowed their clinical evaluation, with the promising results highlighted in the previous section. Whatever the case, an effective delivery system must fulfill a series of criteria, which have already been listed by Tatiparti et al. amongst other author: be stable at the body temperature and pH variations, have an endocytosis promoting shape, cannot be toxic, must exhibit high siRNA loading abilities and have a size that avoids rapid renal and hepatic clearance [231]. In general, all the delivery systems developed for gene therapy may also be adapted for siRNA delivery [232].

7.4.4 Non-targeted Delivery

Early strategies for solving the dual problems of intracellular delivery and rapid excretion involved incorporating siRNAs into LNPs – smaller, more homogeneous analogues of lipoplexes used for laboratory transfection [233–235], (reviewed in [189]). LPNs were first shown to be effective in targeting the hepatitis B virus (HBV) in mice, where the LPN-formulated siRNA was given in 3 daily injections of 3 mg/Kg/day. This treatment regimen resulted in a decrease of HBV levels by 1–2 orders of magnitude [236], as reviewed in [237]. Nevertheless, these complexes (and other nanoparticle strategies for siRNA delivery) accumulate in the liver and other filtering organs, which limits their effectiveness in penetrating other tissues [235, 238] (reviewed in [189]). Furthermore, the administration of siRNAs with LNP delivery vehicles is quite pro-inflammatory. In fact, lipid-based vehicles can become entrapped in endosomes [237], where the Toll-like receptors (TLR) will recognize various moieties in dsRNAs, modified siRNAs or even from their degradation products [239], eliciting an undesirable innate inflammatory response. So, in most circumstances the siRNAs require pretreatment regimens including antihistamines, non-steroidal anti-inflammatories and even relatively high doses of glucocorticoids [190, 191, 240] (reviewed in [11]).

Still, recent developments by several independent teams have demonstrated the feasibility of systemic administration of either chemically modified or complexed siRNAs. In fact, even though unmodified siRNAs do not distribute broadly to tissues after systemic administration (reviewed in [11]), simple chemical modifications of the 2′-position of the ribose and substitution of phosphorothioate linkages, such as the ones described for AONs in the first section of this chapter (2′-OMe and 2′-MOE), and 2′-fluoro (2′-F), protect siRNAs from nuclease digestion, thus prolonging their half-lives, both in serum and other body fluids [236, 241] (reviewed in [189]). 2′-modifications can also prevent recognition by innate immune receptors by blocking the binding to TLR [242–244] and reduce off-target effects that could arise from the suppression of partially complementary sequences [245] (reviewed in [189]). As already referred, these modifications had been previously designed for use in AONs and not siRNAs and, even though they did show effective in improving stability, specificity and immunogenic properties, they do not improve potency. Recently, however, two novel, siRNA-optimized 2′-O modifications, were shown to increase *in vivo* activity of siRNAs, not only by increasing their potency but also their *in vivo* duration compared to their unmodified counterparts when delivered using LNPs: 2′-O-benzyl and 2′-O-methyl-4-pyrimidine (2′-O-CH2Py(4); [246]). Several teams have also been assessing different complexation methods with functional peptides [247] and/or different vectors: exosomes [248] including lipid nanocarriers such as pegylated immunoliposomes (PILs; [249]), stable-nucleic-acid-lipid-particles (SNALPs; [250]), or polyhydroxyalkanoate-based nanovehicles [251], reviewed in [252]. Also under consideration are siRNA delivery strategies that use viral particles. The viral delivery of siRNAs is composed of two main strategies: siRNAs are either chemically synthesized and loaded into a viral capsule, or they can be expressed from the DNA of a recombinant virus (reviewed in [253]).

Meanwhile, second generation LPNs were also developed. Constructed with the anionic lipid dilinol eylmethyl-4-dimethylaminoburyate (DLin-MC3-DMA), they mediate potent gene knockdown at reduced doses compared with first generation LNPs, while improving delivery [190, 217]. Partisiran (previously termed ALN-TTR02; ChemIDplus-Partisiran), for example, is exemplary of a minimally chemically modified siRNA delivered primarily to the liver in a second generation liposome formulation.

7.4.5 Targeted Delivery

Overall, there has been a huge progress over the last decade concerning not only non-targeted but also targeted delivery of siRNA drugs. In fact, siRNAs can also be targeted for uptake in selected tissues or cell types by taking advantage of high-affinity antibody or antibody fragments [254–256], aptamers (nucleic acids selected for high-affinity binding; [257–259] or receptor ligands [260–264], which bind to specific cell surface receptors and mediate cell-specific uptake. The targeting moieties can be either directly conjugated to siRNAs (bound non-covalently) or incorporated into LPNs or other nanoparticles (reviewed in [189]). In general, targeted uptake has the advantages of being effective at a lower dose while exhibiting lower toxicity, which may potentially occur from knockdown effects in unintended tissues. It is also a tool of great advantage for the treatment of non-systemic diseases. The easiest organ to target is the liver, which is a filtering organ that traps nanoparticles. It is also the primary site of synthesis of many circulating proteins. That is why it has been the target organ in most early clinical attempts at translating RNAi (reviewed in [186]). Furthermore, several diseases, which directly affect this organ may benefit from a straightforward liver targeting method. The most successful possibility under study includes a series of more drastic chemical modifications, where siRNAs have a trivalent N-acetylgalactosamine (GalNAc) moiety conjugated to the 3′ terminus of one of the strands [220, 264] (reviewed in [11]). GalNAc mediates hepatocyte uptake through the hepatocyte-

restricted asialoglycoprotein receptor (ASGPR), thus being a suitable mediator for whole-liver delivery [265] (reviewed in [189]). Uptake by this receptor is primarily through clathrin-dependent endocytosis [97]. Examples of GalNAc-modified siRNAs include Revusiran and Fitusiran (ChemIDplus-Revusiran; ChemIDplus-Fitusiran (reviewed in [11]). Also, local delivery to the CNS, a region that is difficult to deliver drugs to due to the BBB, is being addressed, with promising results. First preliminary evidence that *in vivo* downregulation of specific genes by RNAi could work at the CNS level came from studies in rats and mice using invasive local delivery methods (reviewed in [266]). Lately, however, evidence is accumulating on the successful brain delivery of si/shRNAs using specifically designed vectors and/or modifications that include the use of enzyme-sensitive LPNs [267], carbosilane dendrimers [268], cholesterol modifications [269], and recombinant fusion proteins [270]. Also, the pharmaceutical industry is investing in developing BBB-directed vectors. One of those examples is a family of vectors that take advantage on the existence of specific receptors and transport systems, which are highly expressed at the BBB to provide essential substances to brain cells. These vectors comprise a full-length protein (Melanotransferrin) and may be used to facilitate receptor mediated drug delivery into the brain to treat CNS disorders [271]. Recently, the application of this new peptide vector to siRNA and ongoing studies addressing the brain delivery of Iduronate 2-sulfatase (I2S) for the treatment of Hunter Syndrome, a rare X-linked lysosomal storage disorder, was discussed and its results in knockout mice were quite promising [272].

7.5 CRISPR-Cas Gene Editing

In addition to the most well known RNA-based therapeutics (antisense drugs and siRNA-based drugs) several other mechanisms of action are also potential strategies. Recently, a new gene editing technology, Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR) and the CRISPR-associated protein 9 (Cas9) (CRISPR-Cas 9) system, has received unprecedented acceptance in the scientific community for a variety of genetic applications (reviewed in [273]) (Fig. 7.7a). Even though this technology lies beyond the scope of this chapter, it does deserve some attention, as it may become a leading method for gene editing and even RNA-based therapeutics, in the long term.

Similarly to what had already happened with RNAi, the CRISPR-Cas system was not specifically developed as a method for gene editing. Instead, it is a naturally occurring prokaryotic immune defense strategy against non-self DNA based invasions (e.g., viruses, plasmids), which was recently discovered in bacteria and archaea [273–276], and latter adapted for bench applications [277, 278]. Also like RNAi, the specificity of CRISPR-Cas relies on the antisense pairing of RNAs (here termed single guide RNA, sgRNA) to specific genes but instead of binding directly to RNA, sgRNAs bind to chromosomal DNA. Another relevant difference between the RNAi and CRISPR technologies has to do with the transiency of their effect. In fact, unlike siRNAs, sgRNAs induce stable changes in gene expression, which are invaluable for *in vivo* gene screening. Thus, genomic targeting through CRISPR-Cas creates indels that can be adapted for stable eukaryotic genome engineering, namely Cas-mediated gene knockdown (reviewed in [186, 273]). In general, the application of CRISPR/Cas9 for DNA editing as well as for mammalian gene editing was established in the 2012–2013 period and, in just 3 years, this technique has revolutionized the entire gene editing field. Currently, CRISPR-Cas gene knockdown in zygotes provides a fast method for the development of different animal models, when compared to homologous recombination. Nevertheless, it does hold a series of drawbacks and raises a number of concerns, particularly when its therapeutic potential is considered. In fact, since this technique has the ability to modify the genome, its ethical and safe concerns are enormous. Furthermore (and like every other

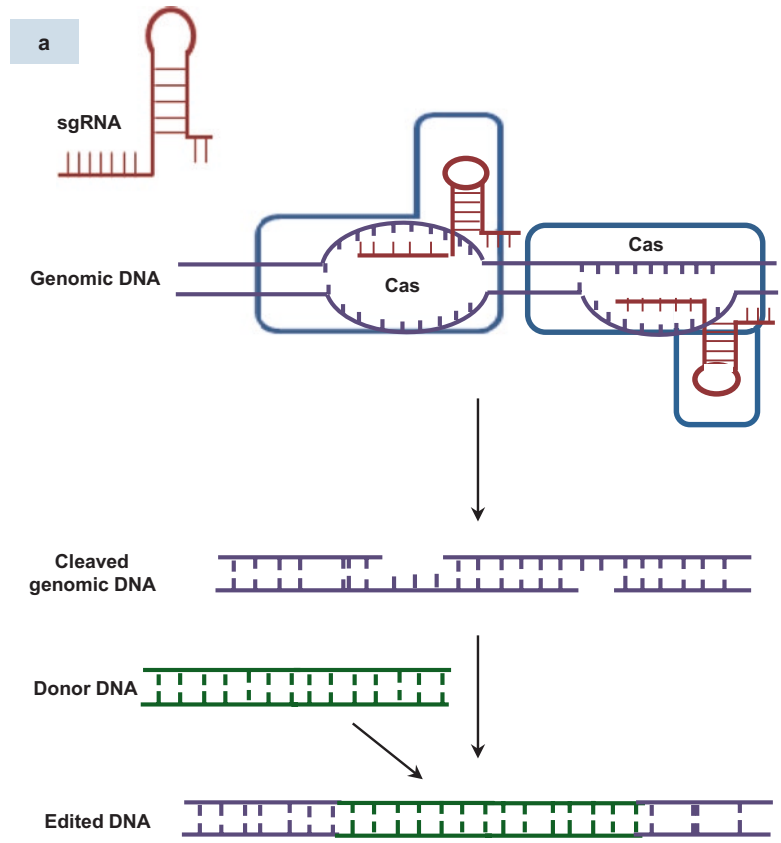

Fig. 7.7 Additional RNA-based drug mechanisms. **(a)** CRISPR/Cas9: CRISPR/Cas9 induces double strand breaks (DBS) when targeted to a specific genomic site by an appropriate guide RNA (sgRNA). This property may be used for mutation correction, by adding a donor DNA sequence that has homologous overlaps to the DBS (typically 100–1000 bp of overlap is used), thus promoting homologous repair of the cleaved genomic DNA; **(b)** Modified mRNAs: This approach consists in introducing chemically modified, stabilized mRNAs into cells to be translated to protein. Once internalized, sense-RNA drugs can be used for transient *in vivo* transcription (IVT) of mRNAs to replace mutated proteins or for vaccination without the risk of genomic alteration; **(c)** Aptamers: Aptamers take advantage of their selection for high-affinity binding to molecular ligands, often in the nanomolar or subnanomolar range. They can be compared to nucleic acid antibodies, having many of the advantages of conventional protein antibodies. They can be either agonists or antagonists, linked for bifunctional targeting and conjugated to other RNAs, small-molecule drugs, toxins or peptides. However, unless modified, they are rapidly excreted and do not activate immune functions, as other antibodies do

antisense technology), CRISPR-Cas holds potential for both on- and off-target effects. Moreover, by creating double-stranded DNA breaks, the Cas endonuclease can also lead to oncogenic gene translocations and trigger a DNA damage response, ultimately causing cell-cycle arrest or even cell death. Finally, depending on the repair pathway which is activated, gene editing may be imprecise [186]. Still, it must be noticed that in the 5 years following the publication of the method, several improvements to reduce off-target effects and provide a better control of the whole mechanism, while enhancing its efficiency have been developed and reported (reviewed in [186, 279]). It should also be mentioned that CRISPR/Cas9 is certainly a more versatile technique than RNAi, has it may not only induce indels but also repress or activate gene expression and cause both heritable and non-heritable genomic changes [280–

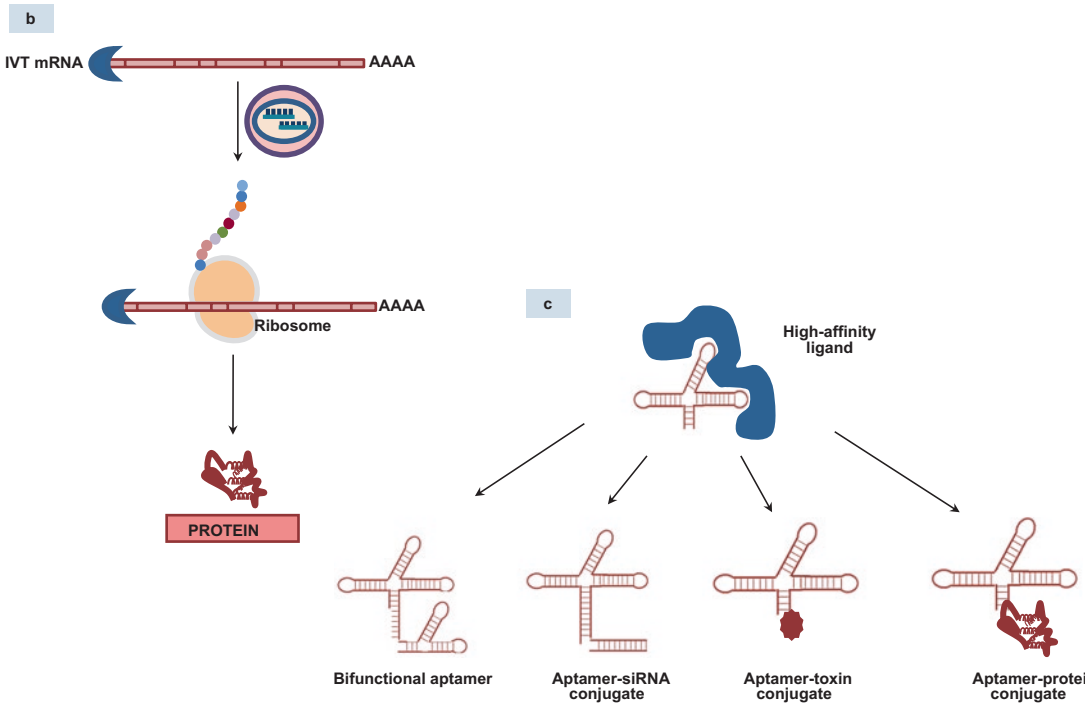

Fig. 7.7 (continued)

282]. In fact, CRISPR/Cas9 can be adapted to upregulate gene expression in different ways: the first, most obvious approach consists in using this technology to stably introduce constitutive active promoter elements to a gene, thus stably enhancing its expression. Alternatively, a modified Cas9 (fused to a transcriptor activator protein) may be targeted to any gene of interest, driving a transient enhancement in gene expression known as CRISPR activation (CRISPRa; [273]). In addition, the emergence of newer gene editing tools such as the Cpf1 enzyme, which is a single RNA-guided endonuclease, will eventually strengthen the portfolio of applications that may be achieved by CRISPR mediated genome engineering [283]. Therefore, it is becoming clear that clinical CRISPR-Cas studies will also be a trend in research over the next years, starting with *ex vivo* editing of differentiated cells, which may then be infused into patients. Furthermore, as other authors have already stated, it will also greatly benefit from the accumulated knowledge on other non-RNA-based gene editing tools, such as zinc-finger nucleases and on the delivery methods previously developed for AON- and siRNA-based drugs.

7.6 Messenger RNA as a Novel Therapeutic Approach

Another RNA-based approach is to introduce chemically modified stabilized mRNAs into cells, where those exogenous mRNAs will eventually be translated to protein (Fig. 7.7b). In fact, *in vitro* transcribed (IVT) mRNA has recently come into focus as a potential new drug class to deliver genetic information. Such synthetic mRNAs can be engineered to transiently express proteins by structurally resembling natural mRNAs [186, 284]. One advantage of mRNA-based therapy over viral gene delivery is that mRNA does not transit to the nucleus, thereby mitigating insertional mutagenesis risks. Moreover, mRNA provides transient, half-life-

dependent protein expression, while avoiding constitutive gene activation and maintaining dose responsiveness. Because of these advantages, IVT mRNA treatment is an emerging class of therapy, with multiple mRNA-based cancer immunotherapies and vaccines currently in clinical trials [284–286]. However, the fact that IVT mRNA, despite its strong resemblance to naturally occurring mRNA, can be recognized by the innate immune system may play an important part in its applicability. For vaccination approaches, the inflammatory cytokine production resulting from mRNA-induced immune stimulation might add to the effectiveness of the evoked immune response. For non-immunotherapy approaches, however, the story is different and so far, cancer immunotherapy is the only field in which mRNA based therapeutics have reached clinical trials [287–294]. Nevertheless, the potential of IVT mRNA is currently being explored for a variety of applications, ranging from inherited or acquired disorders to regenerative medicine, all of which remain at the preclinical stage [295, 296]. In fact, an increasing number of preclinical studies has evaluated mRNA-based therapy for a wide range of diseases such as surfactant B deficiency, myocardial infarction [297] sensory nerve disorders [298], fulminant hepatitis [299] hemophilia B [300, 301], congenital lung disease [302], cancer [303], liver and lung fibrosis [304], and methylmalonic acidemia [305]. However, the main hurdle in implementation of mRNA for therapeutics, the systemic delivery of mRNA molecules to target cells, remains a challenge. Better understanding of the factors that determine translational efficiency as well as RNA recognition by innate immune receptors, has improved the intracellular stability and functionality of mRNA transfected to cells. Still, when aiming to harness mRNA molecules for gene therapy purposes, this progress was insufficient. The need for mRNA protection from degradation in extracellular compartments, as well as for enabling its entry to the cell, has raised the demand for suitable delivery platforms [285, 295]. A possible solution for this challenge relies in the rapidly evolving field of nucleic acid-loaded NPs. In fact, the progress in the field of NPs-mediated RNAi-based therapy, has led to similar development of nanocarriers for mRNA. Particularly, the widely investigated family of LNPs was proposed to be such appropriate mRNA nanocarriers [296, 305, 306]. Moreover, the use of polyplex nanomicelles has also been explored [298, 299]. In order to achieve high efficacy *in vivo* some IVT mRNA specific formulation adjustments should be done in a near future. These adjustments are more important when systemic administration is required. Moreover, in order to expand the variety of mRNA-based therapies, cell specific targeted delivery systems are also needed especially in diseases involving a certain organ, which is inaccessible by standard LNPs, as well as in many types of solid tumors [296]. In conclusion, innovative design of nanocarriers for IVT mRNAs delivery will help to increase their potential and turn them into a valid therapeutic approach.

7.7 Aptamer-Based Drugs

Another potential class of RNA therapeutics are oligonucleotide aptamers (see Fig. 7.7c). The term aptamer comes from the Latin word "*aptus*", which means "to fix", as a clear reference to the lock and key relationship of aptamers and their targets [307, 308].

Aptamers are short (20–70 bases) single stranded oligonucleotides (ssRNA/ssDNA), which bind to their targets through 3D conformational complementarities with high affinity and specificity. Unlike the previously referred strategies, aptamers can be tailored selectively against a variety of targets, from nucleotides to amino acids, proteins, small molecules or even live cells [309]. Still, proteins are the major targets in aptamer research (reviewed in [310]). Oligonucleotide aptamers have affinity and specificity capacities, which are comparable to those of monoclonal antibodies, whilst having minimal immunogenicity, high production, low cost and high stability. These oligonucleotides can be selected trough an *in vivo* process called Systematic Evolution of Ligands by Exponential

enrichment (SELEX), which dates back to 1990. This method was originally described and performed by Szostak and Gold [307, 308]. The whole process starts with the synthesis of a screening library formed by a large number of randomly combinatorial ssDNA and/or ssRNAs. Each one of those random ssDNA/ssRNAs has one conserved sequence at each end. That sequence allows primer binding and amplification. The random library is then incubated with the target proteins, under proper conditions. Then, through a partition step, the sequences that had bind to target proteins are separated from those that did not bind. In the third step, the binding sequences are eluted and amplified with primers complementary to their conserved sequences, either by PCR (for ssDNA) or RT-PCR (for ssRNA). All these steps form a single SELEX cycle. This selection process is then repeated for about 7–20 rounds of incubation, partitioning and amplification. Ultimately, this results in the identification of a small number of binding sequences with high affinity and specificity for further processing and optimization. Generally, the binding sequences are then transformed into bacteria (*E. coli*) for further sequencing and characterization (reviewed in [310]). Naturally, in the post-SELEX process, the synthesized aptamers (as every other AON) can be chemically modified for therapeutic purposes, to stabilize and protect them against nucleases *in vivo*. Recent advances in SELEX technology, with the introduction of chemically modified bases and the use of deep sequencing to analyze enriched RNAs in early rounds of selection, have greatly reduced the time needed and the likelihood of identifying high-affinity aptamers (reviewed in [186]).

Over approximately 10 years, starting in 2005, when the first aptamer Pegaptanib (PubChem, Pegaptanib) was approved for wet age-related macular degeneration (AMD) therapy by U.S. FDA, oligonucleotide aptamers were growing more and more popular. Until 2016, when the last estimates were published online, there had been over 900 aptamers developed against various targets for diagnostic and therapeutic purposes [311]. Nevertheless, drug development of aptamers is currently not very active, with big pharmacological companies being much more focused on the technologies reviewed in the previous sections of this chapter. Still, it is worth mentioning that these oligonucleotides could substitute for some applications of therapeutic antibodies, with lower risk of developing immunological responses. They could also be used for targeted intracellular delivery of other molecules, including RNA-based drugs.

7.8 Conclusion

Over the last decades, an exceptional increase on the understanding of the versatile roles of RNAs has sparked the development of new classes of RNA-based drugs. Therapeutic RNA-based applications are emerging, in different fields, from inherited genetic diseases, oncology, viral infections and diabetes to neurological, cardiovascular, bone-related and ocular diseases. Over the last years in particular, much effort has been focused on the development of RNA-based therapeutics. Currently, even though there are a number of RNA-based therapeutic strategies, which may be attempted in order to either correct or modulate gene expression, there has been a clear prevalence of studies focused on splicing modification and gene expression inhibition using different types of AONs. Actually, the first AON-based drugs were recently approved, closely followed by the first siRNA-based therapeutic drug, which was approved last year. Still, there is a strong need to optimize the delivery steps of RNA-based technologies and to improve the drug-like properties of therapeutic nucleic acids. Expanding the range of targeted cells and tissues will require the development of robust strategies for cytosolic delivery, thus overcoming the two major hurdles of getting across the plasma membrane and out of the endosome.

In conclusion, as the first generation of nucleic acid therapeutics become drugs, the barrier for investing in RNA-based therapeutics will be lowered, and more resources will become available for exploring other mechanisms of action for RNA-based drugs apart from splicing modulation and single-gene knockdown. As already

pointed out by other authors, the flexibility of RNA design should allow for the facile construction of potent multifunctional drugs that have more than one mode of action and disrupt multiple targets. One such multifunctional drug may hold the promise of substituting for drug cocktails in a future not so distant. There is also the largely unexplored potential of targeting other RNA species and disrupting their functions. Therefore, in the near future, RNA-based drugs may become an increasing component of the pharmacopoeia, greatly expanding the universe of druggable targets and providing affordable treatment options for previously untreatable diseases. Ultimately, this kind of drugs may hold potential to actually cure genetic diseases [186].

References

1. Bennett CF, Swayze EE (2010) RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. Annu Rev Pharmacol Toxicol 50:259–293
2. Godfrey C et al (2017) Delivery is key: lessons learnt from developing splice-switching antisense therapies. EMBO Mol Med 9(5):545–557
3. McClorey G, Wood MJ (2015) An overview of the clinical application of antisense oligonucleotides for RNA-targeting therapies. Curr Opin Pharmacol 24:52–58
4. Rossor AM, Reilly MM, Sleigh JN (2018) Antisense oligonucleotides and other genetic therapies made simple. Pract Neurol 18(2):126–131
5. Bauman JA et al (2010) Anti-tumor activity of splice-switching oligonucleotides. Nucleic Acids Res 38(22):8348–8356
6. Kole R, Krainer AR, Altman S (2012) RNA therapeutics: beyond RNA interference and antisense oligonucleotides. Nat Rev Drug Discov 11(2):125–140
7. Sardone V et al (2017) Antisense oligonucleotide-based therapy for neuromuscular disease. Molecules 22(4):E563
8. Paterson BM, Roberts BE, Kuff EL (1977) Structural gene identification and mapping by DNA-mRNA hybrid-arrested cell-free translation. Proc Natl Acad Sci USA 74(10):4370–4374
9. Zamecnik PC, Stephenson ML (1978) Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. Proc Natl Acad Sci USA 75(1):280–284
10. Schoch KM, Miller TM (2017) Antisense oligonucleotides: translation from mouse models to human neurodegenerative diseases. Neuron 94(6):1056–1070
11. Crooke ST et al (2018) RNA-targeted therapeutics. Cell Metab 27(4):714–739
12. Oberemok VV et al (2018) A half-century history of applications of antisense oligonucleotides in medicine, agriculture and forestry: we should continue the journey. Molecules 23(6):1302
13. Khvorova A, Watts JK (2017) The chemical evolution of oligonucleotide therapies of clinical utility. Nat Biotechnol 35(3):238–248
14. Chan JH, Lim S, Wong WS (2006) Antisense oligonucleotides: from design to therapeutic application. Clin Exp Pharmacol Physiol 33(5–6):533–540
15. Chery J (2016) RNA therapeutics: RNAi and antisense mechanisms and clinical applications. Postdoc J 4(7):35–50
16. Roehr B (1998) Fomivirsen approved for CMV retinitis. J Int Assoc Phys AIDS Care 4(10):14–16
17. Kurreck J (2003) Antisense technologies. Improvement through novel chemical modifications. Eur J Biochem 270(8):1628–1644
18. Rigo F, Seth P, Bennett C (2014) Antisense oligonucleotide-based therapies for diseases caused by pre-mRNA processing defects. In: Yeo G (ed) Systems biology of RNA binding proteins, advances in experimental medicine and biology. Springer, New York, pp 303–352
19. Monia BP et al (1993) Evaluation of 2′-modified oligonucleotides containing 2′-deoxy gaps as antisense inhibitors of gene expression. J Biol Chem 268(19):14514–14522
20. Evers MM, Toonen LJ, van Roon-Mom WM (2015) Antisense oligonucleotides in therapy for neurodegenerative disorders. Adv Drug Deliv Rev 87:90–103
21. Geary RS et al (2015) Pharmacokinetics, biodistribution and cell uptake of antisense oligonucleotides. Adv Drug Deliv Rev 87:46–51
22. Aartsma-Rus A (2012) Overview on AON design. Methods Mol Biol 867:117–129
23. Vitravene Study G (2002) A randomized controlled clinical trial of intravitreous fomivirsen for treatment of newly diagnosed peripheral cytomegalovirus retinitis in patients with AIDS. Am J Ophthalmol 133(4):467–474
24. Vitravene Study G (2002) Randomized dose-comparison studies of intravitreous fomivirsen for treatment of cytomegalovirus retinitis that has reactivated or is persistently active despite other therapies in patients with AIDS. Am J Ophthalmol 133(4):475–483
25. Vitravene Study G (2002) Safety of intravitreous fomivirsen for treatment of cytomegalovirus retinitis in patients with AIDS. Am J Ophthalmol 133(4):484–498
26. Duell PB et al (2016) Long-term mipomersen treatment is associated with a reduction in cardiovascular events in patients with familial hypercholesterolemia. J Clin Lipidol 10(4):1011–1021
27. Raal FJ et al (2010) Mipomersen, an apolipoprotein B synthesis inhibitor, for lowering of LDL cholesterol concentrations in patients with homozygous

familial hypercholesterolaemia: a randomised, double-blind, placebo-controlled trial. Lancet 375(9719):998–1006
28. Raal FJ et al (2016) Pediatric experience with mipomersen as adjunctive therapy for homozygous familial hypercholesterolemia. J Clin Lipidol 10(4):860–869
29. Santos RD et al (2015) Mipomersen, an antisense oligonucleotide to apolipoprotein B-100, reduces lipoprotein(a) in various populations with hypercholesterolemia: results of 4 phase III trials. Arterioscler Thromb Vasc Biol 35(3):689–699
30. Chiriboga CA et al (2016) Results from a phase 1 study of nusinersen (ISIS-SMN(Rx)) in children with spinal muscular atrophy. Neurology 86(10):890–897
31. Finkel RS et al (2016) Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. Lancet 388(10063):3017–3026
32. Finkel RS et al (2017) Nusinersen versus sham control in infantile-onset spinal muscular atrophy. N Engl J Med 377(18):1723–1732
33. Mercuri E et al (2018) Nusinersen versus sham control in later-onset spinal muscular atrophy. N Engl J Med 378(7):625–635
34. Cirak S et al (2011) Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study. Lancet 378(9791):595–605
35. Mendell JR et al (2013) Eteplirsen for the treatment of Duchenne muscular dystrophy. Ann Neurol 74(5):637–647
36. Miner P et al (2004) An enema formulation of alicaforsen, an antisense inhibitor of intercellular adhesion molecule-1, in the treatment of chronic, unremitting pouchitis. Aliment Pharmacol Ther 19(3):281–286
37. Monteleone G et al (2015) Mongersen, an oral SMAD7 antisense oligonucleotide, and Crohn's disease. N Engl J Med 372(12):1104–1113
38. Ackermann EJ et al (2016) Suppressing transthyretin production in mice, monkeys and humans using 2nd-generation antisense oligonucleotides. Amyloid 23(3):148–157
39. Digenio A et al (2016) Antisense-mediated lowering of plasma apolipoprotein C-III by Volanesorsen improves dyslipidemia and insulin sensitivity in type 2 diabetes. Diabetes Care 39(8):1408–1415
40. Gaudet D et al (2017) The APPROACH study: a randomized, double-blind, placebo-controlled, phase 3 study of volanesorsen administered subcutaneously to patients with familial chylomicronemia syndrome (FCS). J Clin Lipidol 11:814–815
41. Gaudet D et al (2014) Targeting APOC3 in the familial chylomicronemia syndrome. N Engl J Med 371(23):2200–2206
42. Gaudet D et al (2015) Antisense inhibition of apolipoprotein C-III in patients with hypertriglyceridemia. N Engl J Med 373(5):438–447
43. Buller HR et al (2015) Factor XI antisense oligonucleotide for prevention of venous thrombosis. N Engl J Med 372(3):232–240
44. Bhanot S et al (2013) ISIS-GCCRRX, a novel glucocorticoid (GC) receptor antisense drug reduces cholesterol and triglycerides and attenuates dexamethasone induced hepatic insulin resistance without systemic GC antagonism in normal subjects. Diabetologia 56:S282
45. Morgan E, et al (2014) ISIS-GCGRRx, an antisense glucagon receptor antagonist, caused rapid, robust, and sustained improvements in glycemic control without changes in BW, BP, lipids, or hypoglycemia in T2DM patients on stable metformin therapy. Diabetes 56(11):2183–2195
46. Chi KN et al (2016) A phase I dose-escalation study of apatorsen (OGX-427), an antisense inhibitor targeting heat shock protein 27 (Hsp27), in patients with castration-resistant prostate cancer and other advanced cancers. Ann Oncol 27(6):1116–1122
47. Limmroth V et al (2014) CD49d antisense drug ATL1102 reduces disease activity in patients with relapsing-remitting MS. Neurology 83(20):1780–1788
48. Hong D et al (2015) AZD9150, a next-generation antisense oligonucleotide inhibitor of STAT3 with early evidence of clinical activity in lymphoma and lung cancer. Sci Transl Med 7(314):314ra185
49. Chowdhury S et al (2016) A phase I dose escalation, safety and pharmacokinetic (PK) study of AZD5312 (IONISARRx), a first-in-class Generation 2.5 antisense oligonucleotide targeting the androgen receptor (AR). Eur J Cancer 69:S145
50. Janssen HL et al (2013) Treatment of HCV infection by targeting microRNA. N Engl J Med 368(18):1685–1694
51. van der Ree MH et al (2014) Long-term safety and efficacy of microRNA-targeted therapy in chronic hepatitis C patients. Antivir Res 111:53–59
52. Viney NJ et al (2016) Antisense oligonucleotides targeting apolipoprotein(a) in people with raised lipoprotein(a): two randomised, double-blind, placebo-controlled, dose-ranging trials. Lancet 388(10057):2239–2253
53. Graham MJ et al (2017) Cardiovascular and metabolic effects of ANGPTL3 antisense oligonucleotides. N Engl J Med 377(3):222–232
54. Sermet-Gaudelus I et al (2018) Antisense oligonucleotide eluforsen improves CFTR function in F508del cystic fibrosis. J Cyst Fibros S1569–1993(18)30914–7
55. Querfeld C et al (2017) Ph 1 trial evaluating MRG-106, a microrna-155 inhibitor, administered by intratumoral, subcutaneous, or intravenous delivery in cutaneous T-cell lymphoma (CTCL) patients. Hematol Oncol 35:275
56. Pfeiffer N et al (2017) First-in-human phase I study of ISTH0036, an antisense oligonucleotide selectively targeting transforming growth factor beta 2 (TGF-β2), in subjects with open-angle glaucoma

undergoing glaucoma filtration surgery. PLoS One 12(11):e0188899
57. McCaleb M et al (2017) Systemic pharmacodynamic efficacy of a complement factor B antisense oligonucleotide in preclinical and phase 1 clinical studies. Invest Ophthalmol Vis Sci 58:1952
58. Goemans NM et al (2016) Long-term efficacy, safety, and pharmacokinetics of drisapersen in Duchenne muscular dystrophy: results from an open-label extension study. PLoS One 11(9):e0161955
59. Voit T et al (2014) Safety and efficacy of drisapersen for the treatment of Duchenne muscular dystrophy (DEMAND II): an exploratory, randomised, placebo-controlled phase 2 study. Lancet Neurol 13(10):987–996
60. Saad F et al (2011) Randomized phase II trial of Custirsen (OGX-011) in combination with docetaxel or mitoxantrone as second-line therapy in patients with metastatic castrate-resistant prostate cancer progressing after first-line docetaxel: CUOG trial P-06c. Clin Cancer Res 17(17):5765–5773
61. Tsimikas S et al (2015) Antisense therapy targeting apolipoprotein(a): a randomised, double-blind, placebo-controlled phase 1 study. Lancet 386(10002):1472–1483
62. van Meer L et al (2016) Renal effects of antisense-mediated inhibition of SGLT2. J Pharmacol Exp Ther 359(2):280–289
63. Miller TM et al (2013) An antisense oligonucleotide against SOD1 delivered intrathecally for patients with SOD1 familial amyotrophic lateral sclerosis: a phase 1, randomised, first-in-man study. Lancet Neurol 12(5):435–442
64. Sewell KL et al (2002) Phase I trial of ISIS 104838, a 2′-methoxyethyl modified antisense oligonucleotide targeting tumor necrosis factor-alpha. J Pharmacol Exp Ther 303(3):1334–1343
65. Mignon L et al (2016) ISIS-DMPKRx in healthy volunteers: a placebo-controlled, randomized, single ascending-dose phase 1 Study. Neurology 86(16 Supplement):P3.166
66. Bianchini D et al (2013) First-in-human Phase I study of EZN-4176, a locked nucleic acid antisense oligonucleotide to exon 4 of the androgen receptor mRNA in patients with castration-resistant prostate cancer. Br J Cancer 109(10):2579–2586
67. Toth PP (2011) Antisense therapy and emerging applications for the management of dyslipidemia. J Clin Lipidol 5(6):441–449
68. Visser ME et al (2012) Mipomersen, an apolipoprotein B synthesis inhibitor, lowers low-density lipoprotein cholesterol in high-risk statin-intolerant patients: a randomized, double-blind, placebo-controlled trial. Eur Heart J 33(9):1142–1149
69. Nguyen DD, Chang S (2017) Development of novel therapeutic agents by inhibition of oncogenic microRNAs. Int J Mol Sci 19(1):E65
70. van der Ree MH et al (2016) Miravirsen dosing in chronic hepatitis C patients results in decreased microRNA-122 levels without affecting other microRNAs in plasma. Aliment Pharmacol Ther 43(1):102–113
71. Hua Y et al (2010) Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. Genes Dev 24(15):1634–1644
72. Lefebvre S et al (1997) Correlation between severity and SMN protein level in spinal muscular atrophy. Nat Genet 16(3):265–269
73. Hua Y et al (2008) Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. Am J Hum Genet 82(4):834–848
74. De Vivo DC et al (2017) Interim efficacy and safety results from the phase 2 NURTURE study evaluating nusinersen in presymptomatic infants with spinal muscular atrophy. Neurology 88:S46.003.75
75. Mendell JR et al (2012) Evidence-based path to newborn screening for Duchenne muscular dystrophy. Ann Neurol 71(3):304–313
76. Emery AE (1991) Population frequencies of inherited neuromuscular diseases–a world survey. Neuromuscul Disord 1(1):19–29
77. Ervasti JM (2007) Dystrophin, its interactions with other proteins, and implications for muscular dystrophy. Biochim Biophys Acta 1772(2):108–117
78. Kole R, Krieg AM (2015) Exon skipping therapy for Duchenne muscular dystrophy. Adv Drug Deliv Rev 87:104–107
79. Lim KR, Maruyama R, Yokota T (2017) Eteplirsen in the treatment of Duchenne muscular dystrophy. Drug Des Devel Ther 11:533–545
80. Mendell JR et al (2016) Longitudinal effect of eteplirsen versus historical control on ambulation in Duchenne muscular dystrophy. Ann Neurol 79(2):257–271
81. Kinane TB et al (2018) Long-term pulmonary function in Duchenne muscular dystrophy: comparison of Eteplirsen-treated patients to natural history. J Neuromuscul Dis 5(1):47–58
82. Garanto A et al (2016) In vitro and in vivo rescue of aberrant splicing in CEP290-associated LCA by antisense oligonucleotide delivery. Hum Mol Genet 25(12):2552–2563
83. Denti MA et al (2006) Body-wide gene therapy of Duchenne muscular dystrophy in the mdx mouse model. Proc Natl Acad Sci USA 103(10):3758–3763
84. Denti MA et al (2008) Long-term benefit of adeno-associated virus/antisense-mediated exon skipping in dystrophic mice. Hum Gene Ther 19(6):601–608
85. Goyenvalle A et al (2009) Enhanced exon-skipping induced by U7 snRNA carrying a splicing silencer sequence: promising tool for DMD therapy. Mol Ther 17(7):1234–1240
86. Goyenvalle A et al (2012) Engineering multiple U7snRNA constructs to induce single and multiexon-skipping for Duchenne muscular dystrophy. Mol Ther 20(6):1212–1221
87. Vulin A et al (2012) Muscle function recovery in golden retriever muscular dystrophy after AAV1-U7 exon skipping. Mol Ther 20(11):2120–2133

88. Geib T, Hertel KJ (2009) Restoration of full-length SMN promoted by adenoviral vectors expressing RNA antisense oligonucleotides embedded in U7 snRNAs. PLoS One 4(12):e8204
89. Odermatt P et al (2016) Somatic therapy of a mouse SMA model with a U7 snRNA gene correcting SMN2 splicing. Mol Ther 24(10):1797–1805
90. Imbert M, Dias-Florencio G, Goyenvalle A (2017) Viral vector-mediated antisense therapy for genetic diseases. Genes (Basel) 8(2):E51
91. Krause DS, Van Etten RA (2005) Tyrosine kinases as targets for cancer therapy. N Engl J Med 353(2):172–187
92. Cornelio DB, Roesler R, Schwartsmann G (2007) Gastrin-releasing peptide receptor as a molecular target in experimental anticancer therapy. Ann Oncol 18(9):1457–1466
93. Low PS, Kularatne SA (2009) Folate-targeted therapeutic and imaging agents for cancer. Curr Opin Chem Biol 13(3):256–262
94. McGettrick AF, O'Neill LA (2010) Localisation and trafficking of Toll-like receptors: an important mode of regulation. Curr Opin Immunol 22(1):20–27
95. Millard M, Odde S, Neamati N (2011) Integrin targeted therapeutics. Theranostics 1:154–188
96. Canton J, Neculai D, Grinstein S (2013) Scavenger receptors in homeostasis and immunity. Nat Rev Immunol 13(9):621–634
97. D'Souza AA, Devarajan PV (2015) Asialoglycoprotein receptor mediated hepatocyte targeting – strategies and applications. J Control Release 203:126–139
98. Juliano RL (2016) The delivery of therapeutic oligonucleotides. Nucleic Acids Res 44(14):6518–6548
99. McClorey G, Banerjee S (2018) Cell-penetrating peptides to enhance delivery of oligonucleotide-based therapeutics. Biomedicine 6(2):E51
100. Yin H et al (2008) Cell-penetrating peptide-conjugated antisense oligonucleotides restore systemic muscle and cardiac dystrophin expression and function. Hum Mol Genet 17(24):3909–3918
101. Yin H et al (2009) A fusion peptide directs enhanced systemic dystrophin exon skipping and functional restoration in dystrophin-deficient mdx mice. Hum Mol Genet 28(4):699
102. Wu B et al (2008) Effective rescue of dystrophin improves cardiac function in dystrophin-deficient mice by a modified morpholino oligomer. Proc Natl Acad Sci USA 105(39):14814–14819
103. Betts C et al (2012) Pip6-PMO, a new generation of peptide-oligonucleotide conjugates with improved cardiac exon skipping activity for DMD treatment. Mol Ther Nucleic Acids 1:e38
104. Gao X et al (2014) Effective dystrophin restoration by a novel muscle-homing peptide-morpholino conjugate in dystrophin-deficient mdx mice. Mol Ther 22(7):1333–1341
105. Jirka SMG et al (2014) Peptide conjugation of 20-O-methyl phosphorothioate antisense oligonucleotides enhances cardiac uptake and exon skipping in mdx mice. Nucleic Acid Ther 24:25–36
106. Leger AJ et al (2013) Systemic delivery of a Peptide-linked morpholino oligonucleotide neutralizes mutant RNA toxicity in a mouse model of myotonic dystrophy. Nucleic Acid Ther 23(2):109–117
107. Hammond SM et al (2016) Systemic peptide-mediated oligonucleotide therapy improves long-term survival in spinal muscular atrophy. Proc Natl Acad Sci USA 113(39):10962–10967
108. Shabanpoor F et al (2017) Identification of a peptide for systemic brain delivery of a morpholino oligonucleotide in mouse models of spinal muscular atrophy. Nucleic Acid Ther 27(3):130–143
109. Lönn P et al (2016) Enhancing endosomal escape for intracellular delivery of macromolecular biologic therapeutics. Sci Rep 6:32301
110. Salerno JC et al (2016) Novel cell-penetrating peptide-adaptors effect intracellular delivery and endosomal escape of protein cargos. J Cell Sci 129(12):2473–2474
111. Falzarano MS, Passarelli C, Ferlini A (2014) Nanoparticle delivery of antisense oligonucleotides and their application in the exon skipping strategy for Duchenne muscular dystrophy. Nucleic Acid Ther 24(1):87–100
112. Lerner MR et al (1980) Are snRNPs involved in splicing? Nature 283(5743):220–224
113. Rogers J, Wall R (1980) A mechanism for RNA splicing. Proc Natl Acad Sci USA 77(4):1877–1879
114. Mount SM et al (1983) The U1 small nuclear RNA-protein complex selectively binds a 5′ splice site in vitro. Cell 33(2):509–518
115. West S (2012) The increasing functional repertoire of U1 snRNA. Biochem Soc Trans 40(4):846–849
116. Buratti E, Baralle D (2010) Novel roles of U1 snRNP in alternative splicing regulation. RNA Biol 7(4):412–419
117. Roca X, Krainer AR, Eperon IC (2013) Pick one, but be quick: 5′ splice sites and the problems of too many choices. Genes Dev 27(2):129–144
118. Raponi M, Baralle D (2008) Can donor splice site recognition occur without the involvement of U1 snRNP? Biochem Soc Trans 36.(Pt 3:548–550
119. Guiro J, O'Reilly D (2015) Insights into the U1 small nuclear ribonucleoprotein complex superfamily. Wiley Interdiscip Rev RNA 6(1):79–92
120. Spraggon L, Cartegni L (2013) U1 snRNP-dependent suppression of polyadenylation: physiological role and therapeutic opportunities in Cancer. Int J Cell Biol 2013:846510
121. Valadkhan S, Gunawardane LS (2013) Role of small nuclear RNAs in eukaryotic gene expression. Essays Biochem 54:79–90
122. Buratti E et al (2007) Aberrant 5′ splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. Nucleic Acids Res 35(13):4250–4263

123. Roca X et al (2008) Features of 5′-splice-site efficiency derived from disease-causing mutations and comparative genomics. Genome Res 18(1):77–87
124. Zhuang Y, Weiner AM (1986) A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. Cell 46(6):827–835
125. Baralle M et al (2003) Identification of a mutation that perturbs NF1 agene splicing using genomic DNA samples and a minigene assay. J Med Genet 40(3):220–222
126. Pinotti M et al (2008) U1-snRNA-mediated rescue of mRNA processing in severe factor VII deficiency. Blood 111(5):2681–2684
127. Pinotti M et al (2009) Rescue of coagulation factor VII function by the U1+5A snRNA. Blood 113(25):6461–6464
128. Tanner G et al (2009) Therapeutic strategy to rescue mutation-induced exon skipping in rhodopsin by adaptation of U1 snRNA. Hum Mutat 30(2):255–263
129. Glaus E et al (2011) Gene therapeutic approach using mutation-adapted U1 snRNA to correct a RPGR splice defect in patient-derived cells. Mol Ther 19(5):936–941
130. Sánchez-Alcudia R et al (2011) Overexpression of adapted U1snRNA in patients' cells to correct a 5′ splice site mutation in propionic acidemia. Mol Genet Metab 102(2):134–138
131. Martínez-Pizarro A et al (2018) Intronic PAH gene mutations cause a splicing defect by a novel mechanism involving U1snRNP binding downstream of the 5′ splice site. Plos Genet 14:e1007360
132. Schmid F et al (2011) U1 snRNA-mediated gene therapeutic correction of splice defects caused by an exceptionally mild BBS mutation. Hum Mutat 32(7):815–824
133. Schmid F et al (2013) A gene therapeutic approach to correct splice defects with modified U1 and U6 snRNPs. Hum Gene Ther 24(1):97–104
134. Hartmann L et al (2010) Correct mRNA processing at a mutant TT splice donor in FANCC ameliorates the clinical phenotype in patients and is enhanced by delivery of suppressor U1 snRNAs. Am J Hum Genet 87(4):480–493
135. Matos L et al (2014) Therapeutic strategies based on modified U1 snRNAs and chaperones for Sanfilippo C splicing mutations. Orphanet J Rare Dis 9:180
136. Scalet D et al (2019) Disease-causing variants of the conserved +2T of 5′ splice sites can be rescued by engineered U1snRNAs. Hum Mutat 40(1):48–52
137. Balestra D et al (2014) An engineered U1 small nuclear RNA rescues splicing defective coagulation F7 gene expression in mice. J Thromb Haemost 12(2):177–185
138. Lee NC et al (2016) Mutation-adapted U1 snRNA corrects a splicing error of the dopa decarboxylase gene. Hum Mol Genet 25(23):5142–5147
139. Pinotti M et al (2011) RNA-based therapeutic approaches for coagulation factor deficiencies. J Thromb Haemost 9(11):2143–2152
140. Wally V, Murauer EM, Bauer JW (2012) Spliceosome-mediated trans-splicing: the therapeutic cut and paste. J Invest Dermatol 132(8):1959–1966
141. Roca X, Krainer AR (2009) Recognition of atypical 5′ splice sites by shifted base-pairing to U1 snRNA. Nat Struct Mol Biol 16(2):176–182
142. Cohen JB et al (1994) Suppression of mammalian 5′ splice-site defects by U1 small nuclear RNAs from a distance. Proc Natl Acad Sci USA 91(22):10470–10474
143. Hwang DY, Cohen JB (1997) U1 small nuclear RNA-promoted exon selection requires a minimal distance between the position of U1 binding and the 3′ splice site across the exon. Mol Cell Biol 17(12):7099–7107
144. Fernandez Alanis E et al (2012) An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. Hum Mol Genet 21(11):2389–2398
145. Tajnik M et al (2016) Molecular basis and therapeutic strategies to rescue factor IX variants that affect splicing and protein function. PLoS Genet 12(5):e1006082
146. Dal Mas A et al (2015) Exon-specific U1s correct SPINK5 exon 11 skipping caused by a synonymous substitution that affects a bifunctional splicing regulatory element. Hum Mutat 36(5):504–512
147. Nizzardo M et al (2015) Spinal muscular atrophy phenotype is ameliorated in human motor neurons by SMN increase via different novel RNA therapeutic approaches. Sci Rep 5:p. 11746
148. Mattioli C et al (2014) Unusual splice site mutations disrupt FANCA exon 8 definition. Biochim Biophys Acta 1842(7):1052–1058
149. Dal Mas A et al (2015) Improvement of SMN2 pre-mRNA processing mediated by exon-specific U1 small nuclear RNA. Am J Hum Genet 96(1):93–103
150. Rogalska ME et al (2016) Therapeutic activity of modified U1 core spliceosomal particles. Nat Commun 7:11168
151. Donadon I et al (2018) Exon-specific U1 snRNAs improve ELP1 exon 20 definition and rescue ELP1 protein expression in a familial dysautonomia mouse model. Hum Mol Genet 27(14):2466–2476
152. Balestra D et al (2016) An exon-specific U1snRNA induces a robust factor IX activity in mice expressing multiple human FIX splicing mutants. Mol Ther Nucleic Acids 5(10):e370
153. van der Woerd WL et al (2015) Analysis of aberrant pre-messenger RNA splicing resulting from mutations in ATP8B1 and efficient in vitro rescue by adapted U1 small nuclear RNA. Hepatology 61(4):1382–1391
154. Hwu WL, Lee YM, Lee NC (2017) Gene therapy with modified U1 small nuclear RNA. Expert Rev Endocrinol Metab 12(3):171–175

155. Suñé-Pou M et al (2017) Targeting splicing in the treatment of human disease. Genes (Basel) 8(3):E87
156. Napoli C, Lemieux C, Jorgensen R (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. Plant Cell 2(4):279–289
157. van der Krol AR et al (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell 2(4):291–299
158. Fire A et al (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391(6669):806–811
159. Montgomery MK, Xu S, Fire A (1998) RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans. Proc Natl Acad Sci USA 95(26):15502–15507
160. Kennerdell JR, Carthew RW (1998) Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell 95(7):1017–1026
161. Voinnet O (2001) RNA silencing as a plant immune system against viruses. Trends Genet 17(8):449–459
162. Ketting RF et al (1999) Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell 99(2):133–141
163. Wu-Scharf D et al (2000) Transgene and transposon silencing in Chlamydomonas reinhardtii by a DEAH-box RNA helicase. Science 290(5494):1159–1162
164. Hamilton A et al (2002) Two classes of short interfering RNA in RNA silencing. EMBO J 21(17):4671–4679
165. Hamilton AJ, Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286(5441):950–952
166. Cogoni C et al (1996) Transgene silencing of the al-1 gene in vegetative cells of Neurospora is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. EMBO J 15(12):3153–3163
167. Cecere G, Cogoni C (2009) Quelling targets the rDNA locus and functions in rDNA copy number control. BMC Microbiol 9:44
168. Elbashir SM et al (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411(6836):494–498
169. Elbashir SM, Lendeckel W, Tuschl T (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev 15(2):188–200
170. Caplen NJ et al (2000) dsRNA-mediated gene silencing in cultured Drosophila cells: a tissue culture model for the analysis of RNA interference. Gene 252(1–2):95–105
171. Hutvagner G, Simard MJ (2008) Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol 9(1):22–32
172. Deng Y et al (2014) Therapeutic potentials of gene silencing by RNA interference: principles, challenges, and new strategies. Gene 538(2):217–227
173. Sandy P, Ventura A, Jacks T (2005) Mammalian RNAi: a practical guide. BioTechniques 39(2):215–224
174. Aagaard L, Rossi JJ (2007) RNAi therapeutics: principles, prospects and challenges. Adv Drug Deliv Rev 59(2–3):75–86
175. Fraser AG et al (2000) Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. Nature 408(6810):325–330
176. Lum L et al (2003) Identification of Hedgehog pathway components by RNAi in Drosophila cultured cells. Science 299(5615):2039–2045
177. Boutros M et al (2004) Genome-wide RNAi analysis of growth and viability in Drosophila cells. Science 303(5659):832–835
178. Paddison PJ et al (2004) A resource for large-scale RNA-interference-based screens in mammals. Nature 428(6981):427–431
179. Silva JM et al (2008) Profiling essential genes in human mammary cells by multiplex RNAi screening. Science 319(5863):617–620
180. Schlabach MR et al (2008) Cancer proliferation gene discovery through functional genomics. Science 319(5863):620–624
181. Brass AL et al (2008) Identification of host proteins required for HIV infection through a functional genomic screen. Science 319(5865):921–926
182. Luo J et al (2009) A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. Cell 137(5):835–848
183. Mittal V (2004) Improving the efficiency of RNA interference in mammals. Nat Rev Genet 5(5):355–365
184. Hannon GJ, Rossi JJ (2004) Unlocking the potential of the human genome with RNA interference. Nature 431(7006):371–378
185. Rytlewski JA, Beronja S (2015) RNAi in the mouse: rapid and affordable gene function studies in a vertebrate system. Wiley Interdiscip Rev Dev Biol 4(1):45–57
186. Lieberman J (2018) Tapping the RNA world for therapeutics. Nat Struct Mol Biol 25(5):357–364
187. Hayden EC (2014) RNA interference rebooted. Nature 508(7497):443
188. Song E et al (2003) RNA interference targeting Fas protects mice from fulminant hepatitis. Nat Med 9(3):347–351
189. Wittrup A, Lieberman J (2015) Knocking down disease: a progress report on siRNA therapeutics. Nat Rev Genet 16(9):543–552
190. Coelho T et al (2013) Safety and efficacy of RNAi therapy for transthyretin amyloidosis. N Engl J Med 369(9):819–829
191. Suhr OB et al (2015) Efficacy and safety of patisiran for familial amyloidotic polyneuropathy: a phase II multi-dose study. Orphanet J Rare Dis 10:109

192. Demirjian S et al (2017) Safety and tolerability Study of an intravenously administered small interfering ribonucleic acid (siRNA) post on-pump cardiothoracic surgery in patients at risk of acute kidney injury. Kidney Int Rep 2(5):836–843
193. Peddi V, Ratner L, Cooper M, Gaber O, Feng S, Tso P, Bowers V, Naraghi R, Budde K, Polinsky M et al (2014) Treatment with QPI-1002, a short interfering (SI) RNA for the prophylaxis of delayed graft function. Transplantation 98:153
194. Benitez-Del-Castillo JM et al (2016) Safety and efficacy clinical trials for SYL1001, a novel short interfering RNA for the treatment of dry eye disease. Invest Ophthalmol Vis Sci 57(14):6447–6454
195. Antoszyk A, Katz B, Singh R, Gurses-Ozden R, Erlich S, Rothenstein D, Sharon N, Hodge J, Levin L, Miller N et al (2013) A phase I open label, dose escalation trial of QPI-1007 delivered by a single intravitreal (IVT) injection to subjects with low visual acuity and acute non-arteritic anterior ischemic optic neuropathy (NAION). Invest Ophthalmol Vis Sci 54(4575):4575
196. Solano EC et al (2014) Toxicological and pharmacokinetic properties of QPI-1007, a chemically modified synthetic siRNA targeting caspase 2 mRNA, following intravitreal injection. Nucleic Acid Ther 24(4):258–266
197. Pasi KJ et al (2016) Fitusiran, an investigational RNAi therapeutic targeting antithrombin for the treatment of hemophilia: updated results from phase 1 and phase 2 extension studies in patients with inhibitors. Blood 128:1397
198. Ragni, et al (2016) Fitusiran, an investigational RNAi therapeutic targeting antithrombin for the treatment of hemophilia: updated results from a phase 1 and phase 1/2 extension study in patients without inhibitors. Blood 128:2572
199. Fitzgerald K et al (2017) A highly durable RNAi therapeutic inhibitor of PCSK9. N Engl J Med 376(1):41–51
200. Ray KK et al (2017) Inclisiran in patients at high cardiovascular risk with elevated LDL cholesterol. N Engl J Med 376(15):1430–1440
201. Drugs.com. Alnylam announces FDA acceptance of New Drug Application (NDA) and priority review status for patisiran, an investigational RNAi therapeutic for the treatment of hereditary ATTR (hATTR) amyloidosis. 2018 10 Aug 2018. Available from: https://www.drugs.com/nda/patisiran_180201.html
202. Golan T et al (2015) RNAi therapy targeting KRAS in combination with chemotherapy for locally advanced pancreatic cancer patients. Oncotarget 6(27):24560–24570
203. Zorde Khvalevsky E et al (2013) Mutant KRAS is a druggable target for pancreatic cancer. Proc Natl Acad Sci USA 110(51):20723–20728
204. Gonzalez, et al (2014) Phase 2 of bamosiran (SYL040012), a novel RNAi based compound for the treatment of increased intraocular pressure associated to glaucoma. Invest Ophthalmol Vis Sci 55(13):564
205. Moreno-Montanes J et al (2014) Phase I clinical trial of SYL040012, a small interfering RNA targeting beta-adrenergic receptor 2, for lowering intraocular pressure. Mol Ther 22(1):226–232
206. Demuere et al (2016) A phase I/II study of TKM-080301, a PLK1-targeted RNAi in patients with adrenocortical cancer (ACC). J Clin Oncol 34:2547
207. Libertine, et al. (2015) Update on phase 2 clinical trial results of RXI-109 treatment to reduce the formation of hypertrophic dermal scars. J Am Acad Dermatol 72:AB273
208. Nguyen QD et al (2012) Dose-ranging evaluation of intravitreal siRNA PF-04523655 for diabetic macular edema (the DEGAS study). Invest Ophthalmol Vis Sci 53(12):7666–7674
209. Hill, et al (2016) A subcutaneously administered investigational RNAi Therapeutic (ALN-CC5) targeting complement C5 for treatment of PNH and complement-mediated diseases: preliminary phase 1/2 study results in patients with PNH. Blood 128:3891
210. Hulton, et al. (2016) A phase 1/2 trial of ALN-GO1, an investigational RNAi therapeutic for primary hyperoxaluria type 1 (PH1). Pediatr Nephrol 31:1763
211. Sterinu-Cercel et al (2017) A phase 2a study evaluating the multi-dose activity of ARB-1467 in HBeAg-Positive and –negative virally suppressed subjects with hepatitis B. J Hepatol p 66:S688–S689
212. Lawitz E et al (2015) Safety, pharmacokinetics, and biologic activity of ND-L02-s0201, a novel targeted lipid-nanoparticle to deliver HSP47 siRNA for the treatment of patients with advanced liver fibrosis: interim results from clinical phase 1b/2 studies. S Hepatology 62(909A):S688–S689
213. Salzberg et al (2016) Adoptive cellular immunotherapy with APN401, autologous Cbl-b-silenced peripheral blood mononuclear cells, in patients with solid tumors. J Clin Oncol 34:e14541
214. Triozzi, et al. (2015) Phase I clinical trial of adoptive cellular immunotherapy with APN401 in patients with solid tumors. Immun Cancer 3(Suppl 2):175
215. Alsina M et al (2012) Open-label extension study of the RNAi therapeutic ALN-VSP02 in cancer patients responding to therapy. J Clin Oncol 30(4):3062
216. Cervantes et al (2011) Phase I dose escalation study of ALN-VSP02, a novel RNAi therapeutic for solid tumors with liver involvement. J Clin Oncol 29:3025
217. Fitzgerald K et al (2014) Effect of an RNA interference drug on the synthesis of proprotein convertase subtilisin/kexin type 9 (PCSK9) and the concentration of serum LDL cholesterol in healthy volunteers: a randomised, single-blind, placebo-controlled, phase 1 trial. Lancet 383(9911):60–68
218. Leachman SA et al (2010) First-in-human mutation-targeted siRNA phase Ib trial of an inherited skin disorder. Mol Ther 18(2):442–446
219. Gillmore, et al (2016) Phase 2 open-label extension study of revusiran, an investigational RNAi thera-

peutic for the treatment of patients with transthyretin amyloidosis with cardiomyopathy: updated interim results. 2016: Abstract book, international symposium of amyloidosis, p PA84
220. Zimmermann TS et al (2017) Clinical proof of concept for a novel hepatocyte-targeting GalNAc-siRNA conjugate. Mol Ther 25(1):71–78
221. Xu CF et al (2015) Targeting glucose uptake of glioma cells by siRNA delivery with polymer nanoparticle. J Control Release 213:e23–e24
222. Zuckerman JE et al (2014) Correlating animal and human phase Ia/Ib clinical data with CALAA-01, a targeted, polymer-based nanoparticle containing siRNA. Proc Natl Acad Sci USA 111(31):11449–11454
223. Yuen et al (2017) Prolonged RNA interference therapy with ARC-520 Injection in treatment nai¨ve, HBeAg positive and negative patients with chronic HBV results in significant reductions of HBs antigen. J Hepatol 66:S27
224. Turner, et al. (2017) Hepatic targeted RNA interference provides deep and prolonged knockdown of alpha-1 antitrypsin levels in ZZ patients. J. Hepatol 66:S92
225. Drugs.com. Alnylam presents new clinical results from the APOLLO phase 3 study of patisiran at the 16th international symposium on amyloidosis. 2018 10 Aug 2018. Available from: https://www.drugs.com/clinical_trials/alnylam-presents-new-clinical-results-apollo-phase-3-study-patisiran-16th-international-symposium-17805.html
226. Drugs.com (2018) Available from: https://www.drugs.com/history/onpattro.html
227. Bobbin ML, Rossi JJ (2016) RNA interference (RNAi)-based therapeutics: delivering on the promise? Annu Rev Pharmacol Toxicol 56:103–122
228. Gilleron J et al (2013) Image-based analysis of lipid nanoparticle-mediated siRNA delivery, intracellular trafficking and endosomal escape. Nat Biotechnol 31(7):638–646
229. Wittrup A et al (2015) Visualizing lipid-formulated siRNA release from endosomes and target gene knockdown. Nat Biotechnol 33(8):870–876
230. Bartlett DW, Davis ME (2006) Insights into the kinetics of siRNA-mediated gene silencing from live-cell and live-animal bioluminescent imaging. Nucleic Acids Res 34(1):322–333
231. Tatiparti K et al (2017) siRNA delivery strategies: a comprehensive review of recent developments. Nanomaterials (Basel) 7(4):322–33
232. Han L, Tang C, Yin C (2014) Oral delivery of shRNA and siRNA via multifunctional polymeric nanoparticles for synergistic cancer therapy. Biomaterials 35(15):4589–4600
233. Akinc A et al (2008) A combinatorial library of lipid-like materials for delivery of RNAi therapeutics. Nat Biotechnol 26(5):561–569
234. Semple SC et al (2010) Rational design of cationic lipids for siRNA delivery. Nat Biotechnol 28(2):172–176
235. Shi B et al (2011) Biodistribution of small interfering RNA at the organ and cellular levels after lipid nanoparticle-mediated delivery. J Histochem Cytochem 59(8):727–740
236. Morrissey DV et al (2005) Activity of stabilized short interfering RNA in a mouse model of hepatitis B virus replication. Hepatology 41(6):1349–1356
237. Gooding M et al (2012) siRNA delivery: from lipids to cell-penetrating peptides and their mimics. Chem Biol Drug Des 80(6):787–809
238. Akinc A et al (2010) Targeted delivery of RNAi therapeutics with endogenous and exogenous ligand-based mechanisms. Mol Ther 18(7):1357–1364
239. Czech MP, Aouadi M, Tesz GJ (2011) RNAi-based therapeutic strategies for metabolic disease. Nat Rev Endocrinol 7(8):473–484
240. Robbins M, Judge A, MacLachlan I (2009) siRNA and innate immunity. Oligonucleotides 19(2):89–102
241. Chiu YL, Rana TM (2003) siRNA function in RNAi: a chemical modification analysis. RNA 9(9):1034–1048
242. Morrissey DV et al (2005) Potent and persistent in vivo anti-HBV activity of chemically modified siRNAs. Nat Biotechnol 23(8):1002–1007
243. Judge AD et al (2006) Design of noninflammatory synthetic siRNA mediating potent gene silencing in vivo. Mol Ther 13(3):494–505
244. Judge AD et al (2009) Confirming the RNAi-mediated mechanism of action of siRNA-based cancer therapeutics in mice. J Clin Invest 119(3):661–673
245. Jackson AL et al (2006) Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. RNA 12(7):1197–1205
246. Kenski DM et al (2012) siRNA-optimized modifications for enhanced in vivo activity. Mol Ther Nucleic Acids 1:e5
247. Tai W, Gao X (2017) Functional peptides for siRNA delivery. Adv Drug Deliv Rev 110-111:157–168
248. Ha D, Yang N, Nadithe V (2016) Exosomes as therapeutic drug carriers and delivery vehicles across biological membranes: current perspectives and future challenges. Acta Pharm Sin B 6(4):287–296
249. Boado RJ (2005) RNA interference and nonviral targeted gene therapy of experimental brain cancer. NeuroRx 2(1):139–150
250. Rossi JJ (2006) RNAi therapeutics: SNALPing siRNAs in vivo. Gene Ther 13(7):583–584
251. Li Z, Loh XJ (2016) Recent advances of using polyhydroxyalkanoate-based nanovehicles as therapeutic delivery carriers. Nanomed Nanobiotechnol 9(3):e1429
252. Coutinho MF et al (2016) Genetic substrate reduction therapy: a promising approach for lysosomal storage disorders. Diseases 4(4):1429
253. Buduru S et al (2018) RNA interference: new mechanistic and biochemical insights with application in oral cancer therapy. Int J Nanomedicine 13:3397–3409

254. Song E et al (2005) Antibody mediated in vivo delivery of small interfering RNAs via cell-surface receptors. Nat Biotechnol 23(6):709–717
255. Peer D et al (2007) Selective gene silencing in activated leukocytes by targeting siRNAs to the integrin lymphocyte function-associated antigen-1. Proc Natl Acad Sci USA 104(10):4095–4100
256. Peer D et al (2008) Systemic leukocyte-directed siRNA delivery revealing cyclin D1 as an anti-inflammatory target. Science 319(5863):627–630
257. McNamara JO 2nd et al (2006) Cell type-specific delivery of siRNAs with aptamer-siRNA chimeras. Nat Biotechnol 24(8):1005–1015
258. Berezhnoy A et al (2014) Aptamer-targeted inhibition of mTOR in T cells enhances antitumor immunity. J Clin Invest 124(1):188–197
259. Wheeler LA et al (2011) Inhibition of HIV transmission in human cervicovaginal explants and humanized mice using CD4 aptamer-siRNA chimeras. J Clin Invest 121(6):2401–2412
260. Soutschek J et al (2004) Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. Nature 432(7014):173–178
261. Kortylewski M et al (2009) In vivo delivery of siRNA to immune cells by conjugation to a TLR9 agonist enhances antitumor immune responses. Nat Biotechnol 27(10):925–932
262. Davis ME et al (2010) Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. Nature 464(7291):1067–1070
263. Wong SC et al (2012) Co-injection of a targeted, reversibly masked endosomolytic polymer dramatically improves the efficacy of cholesterol-conjugated small interfering RNAs in vivo. Nucleic Acid Ther 22(6):380–390
264. Nair JK et al (2014) Multivalent N-acetylgalactosamine-conjugated siRNA localizes in hepatocytes and elicits robust RNAi-mediated gene silencing. J Am Chem Soc 136(49):16958–16961
265. Manoharan M(2014) GalNAc-siRNA with enhanced stabilization chemistry: ESC-GalNAc-siRNA. Available from: http://www.alnylam.com/web/assets/ALNY-ESC-GalNAc-siRNA-TIDES-May2014-Capella.pdf
266. Roberts TC et al (2016) Synthetic SiRNA delivery: progress and prospects. Methods Mol Biol 1364:291–310
267. Bruun J et al (2015) Investigation of enzyme-sensitive lipid nanoparticles for delivery of siRNA to blood-brain barrier and glioma cells. Int J Nanomedicine 10:5995–6008
268. Serramia MJ et al (2015) In vivo delivery of siRNA to the brain by carbosilane dendrimer. J Control Release 200:60–70
269. Bruck J et al (2015) Cholesterol modification of p40-specific small interfering RNA enables therapeutic targeting of dendritic cells. J Immunol 195(5):2216–2223
270. Haroon MM et al (2016) A designed recombinant fusion protein for targeted delivery of siRNA to the mouse brain. J Control Release 228:120–131
271. Gabathuler R (2012) New protein vectors for physiological transfer of therapeutic agents to the central nervous system. Biol Aujourdhui 206(3):191–203
272. Gabathuler R (2015) Using a peptide derived from transcend (mtf, p97) to deliver biologics to the cns using a physiologic pathway. Available from: http://www.brains4brain.eu/wp-content/uploads/2015/01/9th-B4B-Workshop-Scientific-Programme.pdf
273. Unniyampurath U, Pilankatta R, Krishnan MN (2016) RNA interference in the age of CRISPR: will CRISPR interfere with RNAi? Int J Mol Sci 17(3):291
274. Barrangou R et al (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315(5819):1709–1712
275. Sorek R, Kunin V, Hugenholtz P (2008) CRISPR–a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol 6(3):181–186
276. Grissa I, Vergnaud G, Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics 8:172
277. Mali P et al (2013) RNA-guided human genome engineering via Cas9. Science 339(6121):823–826
278. Cong L et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819–823
279. Cornu TI, Mussolino C, Cathomen T (2017) Refining strategies to translate genome editing to the clinic. Nat Med 23(4):415–423
280. Yin H et al (2014) Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. Nat Biotechnol 32(6):551–553
281. Nelson CE et al (2016) In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. Science 351(6271):403–407
282. Tabebordbar M et al (2016) In vivo gene editing in dystrophic mouse muscle and muscle stem cells. Science 351(6271):407–411
283. Zetsche B et al (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163(3):759–771
284. Sahin U, Kariko K, Tureci O (2014) mRNA-based therapeutics–developing a new class of drugs. Nat Rev Drug Discov 13(10):759–780
285. Stanton MG (2018) Current status of messenger RNA delivery systems. Nucleic Acid Ther 28(3):158–165
286. Pardi N et al (2018) mRNA vaccines – a new era in vaccinology. Nat Rev Drug Discov 17(4):261–279
287. Heiser A et al (2002) Autologous dendritic cells transfected with prostate-specific antigen RNA

stimulate CTL responses against metastatic prostate tumors. J Clin Invest 109(3):409–417
288. Morse MA et al (2002) The feasibility and safety of immunotherapy with dendritic cells loaded with CEA mRNA following neoadjuvant chemoradiotherapy and resection of pancreatic cancer. Int J Gastrointest Cancer 32(1):1–6
289. Morse MA et al (2003) Immunotherapy with autologous, human dendritic cells transfected with carcinoembryonic antigen mRNA. Cancer Investig 21(3):341–349
290. Weide B et al (2009) Direct injection of protamine-protected mRNA: results of a phase 1/2 vaccination trial in metastatic melanoma patients. J Immunother 32(5):498–507
291. Rittig SM et al (2011) Intradermal vaccinations with RNA coding for TAA generate CD8+ and CD4+ immune responses and induce clinical benefit in vaccinated patients. Mol Ther 19(5):990–999
292. Kübler, et al. (2011) Final analysis of a phase I/IIa study with CV9103, an intradermally administered prostate cancer immunotherapy based on self-adjuvanted mRNA. J Clin Oncol 29:4535
293. Sebastian M et al (2012) Messenger RNA vaccination and B cell responses in NSCLC patients. J Clin Oncol 30(15_suppl):2573–2573
294. Wilgenhof S et al (2013) A phase IB study on intravenous synthetic mRNA electroporated dendritic cell immunotherapy in pretreated advanced melanoma patients. Ann Oncol 24(10):2686–2693
295. Devoldere J et al (2016) Evading innate immunity in nonviral mRNA delivery: don't shoot the messenger. Drug Discov Today 21(1):11–25
296. Granot Y, Peer D (2017) Delivering the right message: challenges and opportunities in lipid nanoparticles-mediated modified mRNA therapeutics-An innate immune system standpoint. Semin Immunol 34:68–77
297. Zangi L et al (2013) Modified mRNA directs the fate of heart progenitor cells and induces vascular regeneration after myocardial infarction. Nat Biotechnol 31(10):898–907
298. Baba M et al (2015) Treatment of neurological disorders by introducing mRNA in vivo using polyplex nanomicelles. J Control Release 201:41–48
299. Matsui A et al (2015) Messenger RNA-based therapeutics for the treatment of apoptosis-associated diseases. Sci Rep 5:15810
300. DeRosa F et al (2016) Therapeutic efficacy in a hemophilia B model using a biosynthetic mRNA liver depot system. Gene Ther 23(10):699–707
301. Ramaswamy S et al (2017) Systemic delivery of factor IX messenger RNA for protein replacement therapy. Proc Natl Acad Sci USA 114(10):E1941–E1950
302. Kormann MS et al (2011) Expression of therapeutic proteins after delivery of chemically modified mRNA in mice. Nat Biotechnol 29(2):154–157
303. Wang Y et al (2013) Systemic delivery of modified mRNA encoding herpes simplex virus 1 thymidine kinase for targeted cancer gene therapy. Mol Ther 21(2):358–367
304. Schrom E et al (2017) Translation of angiotensin-converting enzyme 2 upon liver- and lung-targeted delivery of optimized chemically modified mRNA. Mol Ther Nucleic Acids 7:350–365
305. An D et al (2017) Systemic messenger RNA therapy as a treatment for methylmalonic acidemia. Cell Rep 21(12):3548–3558
306. Patel S et al (2017) Boosting intracellular delivery of lipid nanoparticle-encapsulated mRNA. Nano Lett 17(9):5711–5718
307. Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. Nature 346(6287):818–822
308. Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249(4968):505–510
309. Jayasena SD (1999) Aptamers: an emerging class of molecules that rival antibodies in diagnostics. Clin Chem 45(9):1628–1650
310. Yu Y et al (2016) Molecular selection, modification and development of therapeutic oligonucleotide aptamers. Int J Mol Sci 17(3):358
311. Cruz-Toledo J et al (2012) Aptamer Base: a collaborative knowledge base to describe aptamers and SELEX experiments. Database (Oxford) 2012:bas006

Index

L. Romão (ed.), *The mRNA Metabolism in Human Disease*, Advances in Experimental Medicine and Biology 1157, https://doi.org/10.1007/978-3-030-19966-1