Transcriptional and translational dynamics of the human heart

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat.)
im Fach Biologie

eingereicht an der

Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

von
M.Sc Valentin Schneider-Lunitz

(komm.) Präsident der Humboldt-Universität zu Berlin
Prof. Dr. Peter Frensch

Dekan der Lebenswissenschaftlichen Fakultät

der Humboldt-Universität zu Berlin

Prof. Dr. Dr. Christian Ulrichs

Gutachter:

1 Prof. Dr. Norbert Hübner
2 Prof. Dr. Thomas Sommer
3 Prof. Dr. Dieter Beule

Tag der mündlichen Prüfung: 01.07.2022

моим детям

Table of Contents

Table of Figures

Abstract

Gene expression has primarily been studied on transcriptional and protein levels, largely disregarding the extent of translational regulation that directly influences protein abundance.

To elucidate its role, I used ribosome profiling (Ribo-seq) data, obtained through ribosome profiling, to study translational regulation and identify novel translation events in 65 left ventricular samples of end-stage DCM patients and 15 non-DCM controls. This dataset helped dissect transcriptional and translational regulation between diseased and unaffected human hearts, revealing genes and processes purely under translational control. These would have remained undetected by only looking at the transcriptional level. Furthermore, I predicted novel cardiac proteins translated from long non-coding RNAs (lncRNAs) and circRNAs. Translated lncRNAs showed sequence conservation to primates, while only a few candidates showed positional and translation initiation conservation across mammals. Computational analysis of these evolutionary young proteins suggested involvement in diverse molecular processes with a particular enrichment for mitochondrial energy metabolism.

Finally, I identified RNA-binding proteins (RBPs) whose expression influences target mRNA abundance or translational efficiency (TE) rates. Among the RBPs that influence TE, we identified the muscle-specific splicing factor RBM20 whose switch in isoform production impacts target translational rates. For a subset of 21 RBPs, I have observed regulation on both quantitative traits, which resulted in different mechanistic basis expression control for independent sets of genes. Though the precise switch in RBP function is likely achieved by a combination of multiple factors, for three candidates we have observed a strong dependency on target length and 5’ UTR structure.

This work presents a catalogue of newly identified translation events and a quantitative approach to study translational regulation in the healthy and failing human heart.

Zusammenfassung

Die Genexpression wurde bisher hauptsächlich auf Transkriptions- und Proteinebene untersucht, wobei der Einfluss der Translation, die die Proteinhäufigkeit direkt beeinflusst, weitgehend außer Acht gelassen wurde.
Um diese Rolle besser zu verstehen, habe ich Ribosomen-Profiling-Daten (Ribo-seq) verwendet, um die Translationsregulation zu untersuchen und neue Translationsvorgänge in 65 linksventrikulären Proben von DCM-Patienten im Endstadium und 15 Nicht-DCM-Kontrollen zu identifizieren. Dieser Datensatz half dabei, die Transkriptions- und Translationsregulation zwischen erkrankten und nicht betroffenen menschlichen Herzen zu sezieren und enthüllte Gene und Prozesse, die rein unter Translationskontrolle stehen. Diese wären unentdeckt geblieben, wenn man nur die Transkriptionsebene betrachtet hätte. Darüber hinaus habe ich neue kardiale Proteine vorhergesagt, die von langen nicht-kodierenden RNAs (lncRNAs) und zirkulären RNAs (circRNAs) translatiert werden. Translatierte lncRNAs zeigten Sequenzkonservierung zu Primaten, während nur wenige davon Positions- und Translationsinitiationskonservierung über Säugetiere hinweg aufwiesen. Computergestützte Analysen dieser evolutionär jungen Proteine legten eine Beteiligung an verschiedenen molekularen Prozessen nahe, mit einer besonderen Anreicherung für den mitochondrialen Energiestoffwechsel.
Schließlich identifizierte ich RNA-bindende Proteine (RBPs), deren Expression die Menge der Ziel-mRNA oder die Frequenz der Translationseffizienz (TE) beeinflusst. Unter den RBPs, die die TE beeinflussen, haben wir den muskelspezifischen Spleißfaktor RBM20 identifiziert, dessen Wechsel in der Isoformproduktion die Translationsrate der gebundenen RNA steuert. Für eine Untergruppe von 21 RBPs habe ich die Regulation auf beiden quantitativen Merkmalen beobachtet, was zu einer unterschiedlichen mechanistischen Basis der Expressionskontrolle für unabhängige Gensätze führte. Obwohl die genaue Umschaltung der RBP-Funktion wahrscheinlich durch eine Kombination von mehreren Faktoren erreicht wird, haben wir für drei Kandidaten eine starke Abhängigkeit von der Zielgenlänge und der 5'-UTR-Struktur beobachtet.
Diese Arbeit präsentiert einen Katalog von neu identifizierten Translationsereignissen und einen quantitativen Ansatz zur Untersuchung der Translationsregulation im gesunden und kranken menschlichen Herzen.

Introduction

From the genome to the proteome

Deoxyribonucleic acid (DNA) is a double-stranded molecule made of four nitrogen-containing nucleobases; adenine (A), guanine (G), cytosine (C), and thymine (T). Its structure was revealed in 1953 by James Watson and Francis Crick, explaining how genetic information is stored and replicated (Watson & Crick, 1953b, 1953a). Five years later, this discovery was used to describe how genetic information is transferred from DNA via RNA to the protein, known as the “central dogma of molecular biology” (Crick, 1958) (**Figure 1**).

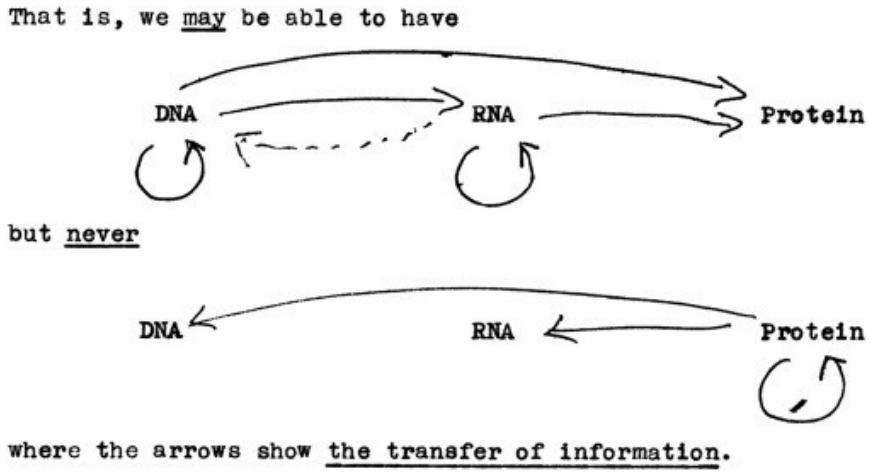


Figure 1: Drawing of the central dogma by Francis Crick in the article “On protein synthesis”.

The central dogma of molecular biology illustrates the flow of genetic information from DNA via RNA to protein in living organisms.

Although the central dogma of molecular biology is under debate (Camacho, 2021; Koonin, 2012; Pukkila, 2001), the flow of information is a complex crosstalk between multiple layers requiring interaction between all molecules besides the ones suggested in the central dogma. Among these, we find histone modification (Bannister & Kouzarides, 2011), post-transcriptional modifications such as pre-mRNA splicing (Wilkinson et al., 2020), transport (Köhler & Hurt, 2007) of RNA from the nucleus into the cytosol, RNA stability (Boo & Kim, 2020), translation of RNA into protein (Hershey et al., 2012), and post-translational modifications (Conibear, 2020). All these layers are under tight regulatory control, and errors can cause cardiovascular diseases

(Darling & Uversky, 2018), metabolic disorders (Morita et al., 2013), or neurodegenerative diseases (Fernandopulle et al., 2021). Along the way from pre-mRNA splicing to degradation RNA molecules are associated with different RNA-binding proteins (RBPs), position them as central regulators of gene expression.

Gene expression regulation captured by high-throughput sequencing

Molecular basis of gene expression regulation by RNA-binding proteins

RNA-binding proteins have been identified as regulators of gene expression across in all levels of gene expression. The life of a mRNAs transcripts starts with its transcription from the DNA by RNA Polymerase II (Pol II) in the nucleus (F. X. Chen et al., 2018). A premature messenger RNA is bound by RBPs that recruit the spliceosome to execute splicing. Alternative splicing essentially joins selected exons by removing intronic regions in-between (Hang et al., 2015). This process which results alternative coding mRNA transcripts is executed by a RBP group called splicing factor (Brody & Abelson, 1985; Frendewey & Keller, 1985; Grabowski et al., 1985; Shi, 2017). Each mature mRNAs is again bound by RBPs that orchestrate the export into the cytoplasm through nuclear pore complexes (NPC) and protect the 5'cap and the 3' poly(A) tails (Carmody & Wente, 2009). In the cytoplasm, RBPs are responsible for the highly sensitive mRNA translation which starts with the recognition of the 5' cap, formation of the ribosome and the actual translation (R. F. Harvey et al., 2018; Moore & von Lindern, 2018) which will be more extensively introduced downstream in the coming section. For purpose of translation, individual mRNAs are interconnected with stability and decay by RBPs and other factors. Hereby, 5'cap and the poly(A) tail acts as not only as regulators of mRNA translation but also prevent degradation of mRNA by exosomes (Díaz-Muñoz & Turner, 2018). Degradation is induced by deadenylation of the poly(A) site by polyA ribonuclease (Webster et al., 2019) or decapping by DCP2 (D'Lima et al., 2017; Erickson et al., 2015). Messenger RNAs lacking 5'cap site are susceptible to 5'-3'-exonuclease degradation. Stable transcripts on the hand are bound by PABP and needs to be displaced before initiation of mRNA decay (Schoenberg & Maquat, 2012).

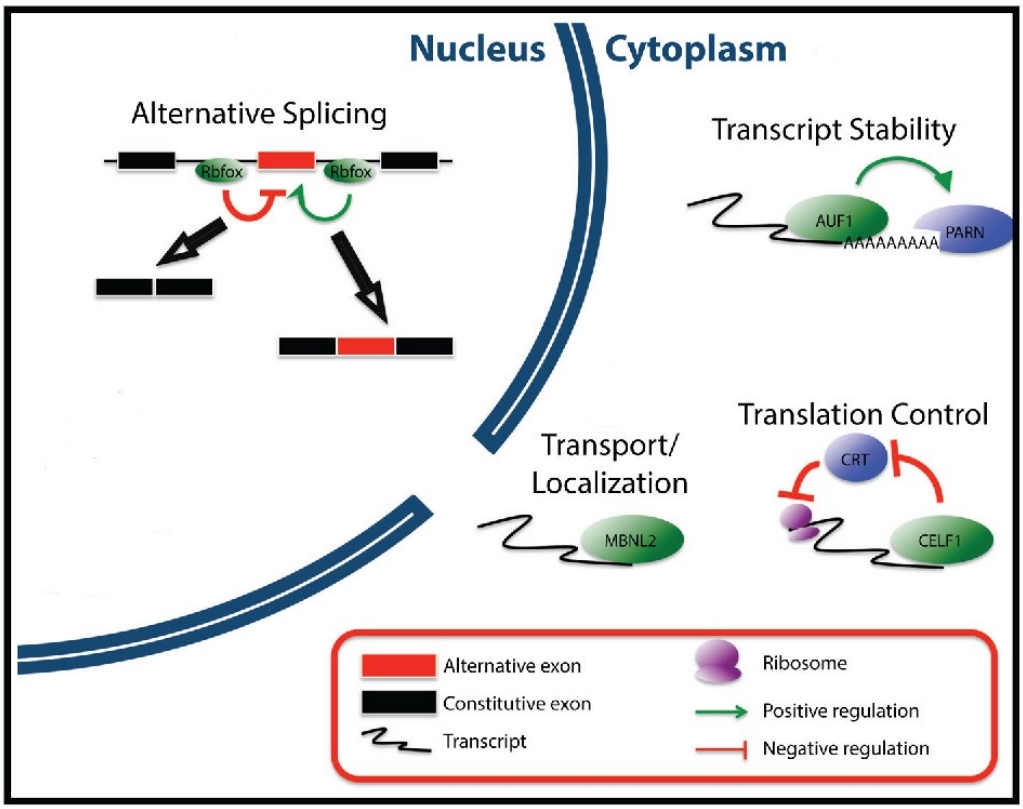


Figure 2: Schematic overview of REP-mediated post-transcriptional regulation.

RBPs are involved in multiple mechanisms of gene expression regulation. The cover alternative splicing of premature mRNA which is located in the nucleus, transport of the spliced transcript into the cytoplasm where it is stabilized or translated into proteins. All these steps and many more are under tight control of RBPs.

Adapted from Blech-Hermoni at al. 2013

RNA-protein interaction by CLIP-seq

Interaction of RBPs with their mRNA targets can be measured by UV crosslinking assays (CLIP, HITS-CLIP, iCLIP, PAR-CLIP, or eCLIP (Hafner et al., 2010; König et al., 2010; Licatalosi et al., 2008; Ule et al., 2003; Van Nostrand et al., 2016) at nucleotide resolution. This method covalently binds RNA to its attached protein. Later, specific antibodies against the protein are used to immunoprecipitate the RBP-RNA complexes, followed by RNA purification and high-throughput sequencing. After mapping the sequenced reads to the transcriptome, the exact RBP-RNA interaction sites can be defined and used for downstream analysis (e.g. target or motif prediction). Several studies were able to identify binding motifs for splicing factors, such as

NOVA (Ule et al., 2006), FOX (C. Zhang et al., 2008), RBM20 (Maatz et al., 2014), which gave insight into how alternatively spliced exons can influence target gene expression. However, genome-wide annotation of binding sites does not explain the precise regulation of alternative splicing. Furthermore, binding sites are frequently located distant to the alternatively spliced exons (Ule et al., 2006). Short motifs that frequently occur across the whole genome give rise to false-positive signals resulting in strong signals without any functional relevance (Witten & Ule, 2011). A combination of different methods, as has been applied for RBM20 (Maatz et al., 2014), will help annotate full splicing maps and the affected targets.

Transcriptome analysis using RNA-seq

The development of direct cDNA sequencing, termed RNA-seq (RNA sequencing), has several advantages over previous methods, including whole-transcriptome coverage and the possibility of quantifying or identifying novel splicing events (Nagalakshmi et al., 2008). The main idea is to convert a population of RNA (total or poly(A)+) in cDNA fragments with attached adaptors to both ends followed by high-throughput sequencing. The cDNA libraries can be sequenced with several platforms such as the HiSeq4000 by Illumina®, Applied Biosystems SOLiD v4 or Roche 454 Life Science® (Zhong Wang et al., 2009). Until 2019, over 95% of the published datasets in the Short Read Archive (SRA) were generated using Illumina-based platforms (Stark et al., 2019). Besides identification and quantification of whole transcriptomes, RNA-seq provides single-nucleotide resolution, high dynamic range during quantification, reveals variations (SNPs) (Cloonan et al., 2008) in the transcribed regions, has low background noise and high reproducibility (Nagalakshmi et al., 2008).

The RNA-seq data analysis can be subdivided into four main steps. After sequencing the fragments/reads, FASTQ files including raw reads are usually generated (Cock et al., 2009). In the first step, the raw reads are mapped to a known reference transcriptome or genome using a variety of available alignment tools, including TopHat (D. Kim et al., 2013), STAR (Dobin et al., 2013), or HISAT (D. Kim et al., 2015). To account for reads spanning exon boundaries, previously mentioned tools can perform spliced alignments allowing gaps within reads during the mapping procedure.

In case transcriptome annotations are incomplete or not available, reads aligned to the genome can be used to generate *de novo* transcriptome assemblies. Tools like StringTie (Pertea et al., 2015) or SOAPdenovo-Trans (Xie et al., 2014) use genome-guided and genome-independent strategies to generate splicing graph from RNA-seq reads that are used to assemble novel transcripts.

In the next step of RNA-seq analysis, mapped reads are assigned to gene features (e.g. exons) to determine their abundance. Quantification tools like RSEM (B. Li & Dewey, 2011), MMSeq (Turro et al., 2011), or HTSeq (Anders et al., 2015) can be used for this purpose. All of them use different quantification strategies and suffer from distinct issues. HTSeq by default discards many multi-mapping reads or those overlapping several exons. This bias may result in lower expression quantification of homologous or overlapping genes. Decision on the preferred quantification tool may have a considerable impact on the final quantification as has been shown previously (Robert & Watson, 2015; Williams et al., 2017), making it the most critical step in RNA-seq analysis procedure with a significant impact on all downstream analysis steps.

After quantification of the RNA-seq data, poorly expressed genes or transcripts are filtered out and the remaining ones are normalized to account for differences in sample size, expression, and technical bias (Risso et al., 2011, 2014; Wagner et al., 2012). Average genes expression is frequently determined by TPM (transcripts per kilobase million) or RPKM/FPKM (reads/fragments per kilobase of exon per million reads/fragments mapped).

Both provide counts per length of transcript (kb) per million reads mapped, though RPKM/FPKM allow gene comparison only within a sample but not between samples making TPM the more flexible count normalization method.

Normalization methods that account for different sets of expressed genes need to be carefully chosen (X. Li et al., 2017). Among the available tools, edgeR (Robinson et al., 2009) and DESeq2 (Love et al., 2014) emerged as robust tools applying various normalization methods, properly correcting such effects. While DESeq2 divides counts by sample-specific size factors which are calculated by median ratio of gene counts relative to geometric mean, EdgeR uses a weighted trimmed mean of the log expression ratios between samples.

Finally, the filtered and normalized expression counts can be used to identify differentially expressed genes or transcripts between two or more conditions. Differential behaviour is estimated using normalized read counts followed by evaluation of generalized linear models. Tools such as DESeq2, edgeR, or limma+voom (Law et al., 2014) successfully apply this method and achieve comparable results (Conesa et al., 2016).

All the methods presented above are frequently used and constantly undergo extensive improvements. Still, identification of transcribed transcripts and their quantification by RNA-seq alone is a poor measurement since much more is happens at the translational level(Schafer, Adami, et al., 2015).

Translational regulation

Fundamentals of eukaryotic translation

Protein synthesis is an essential step to transfer genetic information from RNA into a functional protein. The active translation is ultimately mediated by ribosomes, complex macromolecules consisting of assembled proteins and small RNAs. A complete eukaryotic 80S ribosome can be divided into a large (60S) and a small (40S) subunit. The synthesis of a protein during RNA translation can be subdivided into three parts: translation initiation, elongation, termination, and termination; in addition, recycling can be also considered as an additional step. The process of translation initiation in eukaryotes can be subdivided into multiple stages (**Figure 3**) making the complete model of eukaryotic translation rather enormous and reviewed elsewhere (Dever et al., 2018; Jackson et al., 2010, 2012). A short overview is given in the next section.

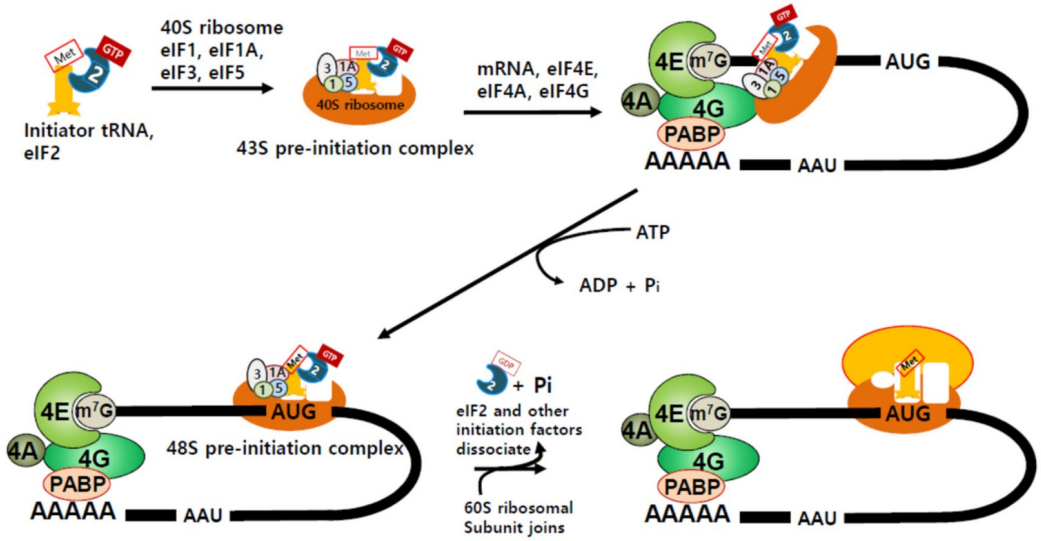


Figure 3: The canonical pathway of eukaryotic translation initiation (H. J. Kim, 2019).

The translation initiation phase can be treated as a scanning process searching for the start codon “AUG” within a 5’ untranslated region (5’ UTR). The scanning starts with the binding of the 40S-Meth-tRNA$_i$ to the 5’-terminus of an m^{7}G-capped mRNA by eIF4F complex, made of the cap-binding protein eIF4E, the RNA helicase eIF4A, and the scaffolding protein eIF4G. To ensure accessibility of the ribosome to the potentially folded 5’ leading structure of the RNA, eIF4A, enhanced by eIF4B and eIF4H helicases, unwind the structured RNA. The small 40S ribosomal subunit recruitment is facilitated by eIF4G and eIF4 (Jackson et al., 2010; Preiss &

Hentze, 2003). The pre-initiation complex (PIC), small 40S ribosomal subunit along with eIFs, and the Met-tRNA$_i$ scanning are executed until a translation initiation codon is found. The mere "AUG" sequence is frequently not enough to initiate active translation. However, it is commonly situated within a Kozak sequence "GCC(**A/G**)CCAUGG" (M. Kozak, 1986; Marilyn Kozak, 1999). Alternative start codons (e.g., CUG or GUG) can also efficiently initiate translation in favourable contexts (Diaz de Arce et al., 2018). Complete motif analysis of sequence requirements for translation initiation at non-AUG start codons. Upon recognition of the appropriate translation start site by the Met-tRNA, the complex transitions into a closed conformation called the 48S complex. The large 60S ribosomal subunit joins the smaller 48S complex forming the final 80S ribosome complex. At this point, several initiation factors are released, among them eIF1A and eIF5B-GDP (Jackson et al., 2010) finishing the first stage of protein translation. Cap-independent translation initiation via IRES structures has been observed for some cellular and many viral RNA long ago (J Pelletier et al., 1988), expanding the canonical proteome (Y. Yang & Wang, 2019).

Translation elongation can be partitioned into three substeps. First, the amino-acyl tRNA complementary to the following RNA codon is placed in the A-site of the ribosome, and a peptidyl-transferase catalyzes the peptide connection between the amino acids in the P- and A-sites. Afterwards, the ribosomes move to the tRNA to the E-site, which frees the A-site making it available for the next tRNA (J. Chen et al., 2012). To ensure the precise translation, elongation factors (EF1 and EF2) interact with tRNA and the ribosome, providing continuous proofreading during each cycle. It has been estimated that a single ribosome adds around six amino acids to the polypeptide in a second (Ingolia et al., 2011). The translational efficiency is modulated during the elongation phase by multiple ribosomes bound to the RNA.

Translation terminates with the recognition of one of three possible stop codons (TAG, TGA and TAA). Upon recognition of the stop codon, release factors trigger hydrolysis of the polypeptide chain, releasing the complete protein from the ribosome (Capecchi, 1967; Hellen, 2018).

Finally, the recycling step is initiated by ABCE1 a protein that binds to the release factor eRF1 on the post-termination complex (post-TC), causing a split of the ribosome into 60S and 40S ribosomal subunit, eRF1, and ABCE1. The initiation factors eIF1, eIF1A, eIF3 and eIF3j release the deacylated tRNA and messenger RNA (mRNA) from the 40S ribosomal subunit (Hellen, 2018).

Cis and *trans* acting translational regulation by RNA-binding proteins.

Regulation of the canonical translation machinery and protein synthesis is modulated via phosphorylation or cleavage of the translation initiation or elongation factors (Bushell et al., 2006; Hinnebusch, 2014). This regulatory mechanism usually occurs under pathophysiological stress such as viral infection, temperature changes, starvation, or hypoxia (Spriggs et al., 2010). The initiation can be reduced via phosphorylation of eIF2, causing a reduction of other factors; and dephosphorylation of 4E-BP inhibiting eIF4E formation. Besides that, translation can be controlled at the step of translation elongation (Richter & Coller, 2015), the phosphorylation status of elongation factors eEFs (Leprivier et al., 2013), cell differentiation (Gingold et al., 2014), and cold stress (Bastide et al., 2017).

Besides the numerous RBPs that are part of the canonical translation machinery acting in *cis* on the active translation, many known RBPs are distant (*trans*) regulators of protein synthesis. They frequently repress or stimulate the translation of selected subsets of target mRNAs. Translation can, in principle, occur at different cytosolic locations (cytosol, endoplasmic reticulum, or specialized districts in different cell types). Even though translational regulation can happen at various stages (initiation, elongation, termination), translation initiation is considered as the central, rate-limiting stage in protein synthesis and is thus tightly controlled.

The regulation is achieved via interaction with mRNA structural elements such as 5' and 3' UTRs or coding regions (CDS), directly regulating cytoplasmic transcript abundance as a response to external signalling (R. F. Harvey et al., 2018). Unlike global protein synthesis control that is mainly regulated via phosphorylation, *trans*-acting factors that are recruited to 5' and 3' UTRs have a significant impact on mRNA metabolism, including polyadenylation, nuclear export, localization, stability, and translation (Martin & Ephrussi, 2009). Interaction is exerted between RNA motifs and RNA-binding domains. Among the most frequently used motifs are terminal oligopyrimidine (5'TOP) motifs, and internal ribosome entry segments (IRES). At the same time, 5'TOP motifs have been identified in all ribosomal proteins and thus put them in a central position of translational regulation.

TOP-containing mRNA has the TOP motif positioned after the 7-methylguanosine triphosphate (m7 GTP) cap. It consists of an invariable C-residue followed by a 4-15 base pyrimidine tract of equally divided Cs and Us (Meyuhas & Kahan, 2015). Additionally, a GC-rich region immediately after the TOP motif has been observed, suggesting a coordinated requirement for translational control (D Avni et al., 1994; Dror Avni et al., 1996).

To date, studies identified 93 mRNAs containing TOP motifs, including 79/80 ribosomal proteins. Among them translation initiation factors (eIF3E, eIF3F, eIF4B, and eIF3H), all five translation elongation factors (EEF1A1, EEF1B2, EEF1D, EEF1G, and EEF2), and other

proteins required for translation (R. F. Harvey et al., 2018). All mRNA containing TOP motifs are translationally upregulated upon an increase of growth factors or insulin (Patursky-Polischuk et al., 2014) but undergo translational repression upon cellular stress (Fonseca et al., 2015; Miloslavski et al., 2014). Nevertheless, the precise composition of *trans*-acting factors and how they interact with each other is unknown. Several candidates such as LARP1, CNBP, AUF1, and TAIR/1 have been shown to play an important role in their regulation, and their localization in stress granules suggests a potential localization interaction (Cockman et al., 2020).

Translational regulation by 5' UTR features

As discussed above, multiple factors are required to initiate translation, including pre-initiation complex, mRNA to translate, and precise recognition of the AUG within the Kozak region. All these factors are under precise control and allow intervention during translation initiation. Along with m7G cap, 5' leader sequences harbour other regulatory features that have an effect on translational, such as sequence length, GC-content, or the presence of complex secondary structures.

The complexity of the secondary structure is determined by three sequence properties, including its length, nucleotide composition, and structural arrangement. Longer sequences are often more stable because of frequent stacking, hydrogen bonds, and higher GC content (Trotta, 2014). Complex secondary structures require specific RBPs such as DDX3X to unwind the structured sequence and make it accessible to initiation factors (Shen & Pelletier, 2020). Secondary structures do not always require to be unfolded but can also act as alternative translation start sites. Internal ribosome entry site (IRES) are among the regions that can provide alternative translation initiation sites (TIS), independent of the m7G-cap structure. Its complex self-complementary structure has been shown effective in the recruitment of ribosomes, initiation factors, and other RBPs crucial for translation (W. Yang et al., 2019). Although frequently present in viral RNAs (Martinez-Salas et al., 2018; Jerry Pelletier & Sonenberg, 1988), IRES sites have been found in numerous human transcripts (Y. Yang & Wang, 2019). Furthermore, multiple studies have focused on alternative translation initiation sites (TIS) upstream of the main open reading frame (ORF). Upstream ORFs (uORFs) are abundant regulatory elements present in 40-50% of all mRNAs (Somers et al., 2013). Although our understanding of its function is incomplete, uORF translation has been shown to cause alleviation of re-initiation efficiency (Hinnebusch, 2005; Marilyn Kozak, 2001) or even repression (Vattem & Wek, 2004) of the downstream main coding (CDS). With the introduction of ribosome profiling, it is now possible to quantitatively access uORF and main CDS translation rates on a

genome-wide scale, allowing dissection of the individual regulatory context. Furthermore, it has been shown that only a small fraction of uORFs are conserved across vertebrates (Chew et al., 2016), and therefore only limited protein evidence in mass spectrometry (Menschaert et al., 2013). Suggesting the act of translation itself and not the protein product as a regulatory mechanism on the main ORF.

The coding capacity of the non-coding genome

Translation, as measured by ribosome profiling, and its regulation adds another layer of complexity on gene expression which is of significant importance but a bit underappreciated.
After all, translational regulation is not only much more complex but also much more is translated than previously anticipated (Karakas & Ozpolat, 2021; Mudge et al., 2021). Advances using the ribosome profiling technique (Ingolia et al., 2009; Schafer, Adami, et al., 2015) and subsequent development of analysis tools (Calviello et al., 2016; Fields et al., 2015; Mackowiak et al., 2015; Raj et al., 2016) led to the detection of thousands of actively translated sequences by ribosomes. Surprisingly, several studies have shown in human and mouse (D. M. Anderson et al., 2015; Mackowiak et al., 2015; Y. Zhang et al., 2018), other species (Chew et al., 2013; Ruiz-Orera et al., 2014), yeast (Ingolia et al., 2009) and plants (Hellens et al., 2016) that a large fraction of long noncoding RNAs (lncRNAs) host small open reading frames (sORFs) with strong ribosome signals (**Figure 4**) and with significant three-nucleotide periodicity, indicative of active translation.
A noncoding RNA longer than 200 nucleotides (nt) is termed as long noncoding RNA (lncRNA), which lacks species conservation or open reading frames longer than 100 amino acids (aa). Long noncoding RNAs have been shown to act as gene expression regulators at (D. M. Anderson et al., 2015; Mackowiak et al., 2015; Y. Zhang et al., 2018) different layers, including epigenetics (Pandya-Jones et al., 2020), transcription (Latos et al., 2012), post-transcription (Miller & Olivas, 2011), or translation (Carlevaro-Fita et al., 2016), and even protein abundance (K. C. Wang & Chang, 2011).

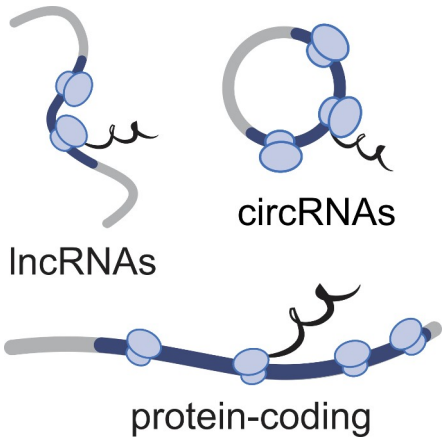


Figure 4: Schematic translation of the coding and non-coding transcriptome.

These discoveries suggest that a significant fraction of lncRNAs are translated and expand the annotated translatome. Though it has to be figured out how many of these proteins (called “microproteins”) result in functional products or whether the act of translation plays a regulatory role. Among the microproteins that have been functionally characterized over the last years, few are muscle-specific and have been shown to regulate SERCA activity (K. M. Anderson et al., 2016; Matsumoto et al., 2017; Nelson et al., 2016). For example, the muscle-specific 34 amino acids long peptide ”DWORF” localizes to the sarcoplasmic reticulum (SR) membrane where it enhances SERCA activity by replacing its inhibitors (Nelson et al., 2016). Intriguingly, for some other translated lncRNAs, a noncoding role has been previously shown, suggesting a dual function for some of them. These discoveries hint towards distinct functions encoded by one gene, still for each candidate wet-lab experiments need to show whether the lncRNA function is purely non-coding or if the microprotein makes the function. Particularly challenging will be experiments that confirm both non-coding and coding function.

Besides lncRNAs, circRNAs are another class of noncoding RNAs produced by a non-canonical splicing event termed back-splicing, which connects downstream splice donor site covalently with the upstream splice-acceptor site. The majority of circRNAs are transcribed from protein-coding genes via RNA polymerase II but lack, contrary to canonical transcripts, a 5’-cap and 3’-polyadenylation (poly-A) site (Qu et al., 2017). CircRNAs generally localize to the cytoplasm. Multiple studies have shown that circRNAs are an abundant and conserved class of RNA widely expressed across complex tissue-, cell type- or stage-specific mode (Hanan et al., 2017; Venø et al., 2015; You et al., 2015). Even though considered noncoding, several works have shown for selected circRNAs their ability to encode proteins (Granados-Riveron & Aquino-Jarquin,

2016; Legnini et al., 2017; Pamudurti et al., 2017), suggesting a yet unprecedented potential role of circRNAs.

Conservation of the noncoding genome

Traditional approaches assessed ORF conservation across species to identify protein-coding genes with preserved function, from human down to bacteria. Automated pipelines used this conservation approach to identify genes with long (≥ 100aa) sequences which inherently show more robust conservation than short (< 100aa) sequences (Couso & Patraquim, 2017). Even though many lncRNAs show similar features to protein-coding mRNAs, including transcription by polymerase II (Pol II) with 5'-cap and 3'-polyadenylation tail and accumulation in the cytosol (Van Heesch et al., 2014), signatures of conservation are largely absent in the non-coding genome, discarding transcripts with shorter ORFs (50 - 110 codons) by automatic annotation pipelines which require a translated gene to encode a protein longer than 100 amino acids and show nucleotide conservation across species. Short proteins with poor species conservation might have evolved recently, indicating that many of these proteins would be evolutionary young (Ruiz-Orera et al., 2018). However, there is also rising evidence that the act of translation can be conserved and contribute to the stability of a transcript, independent of the conservation (H. Zhang et al., 2018). Fields and colleagues have suggested that homologous lncRNAs might undergo translation across different lineages without the presence of a constrained amino acid sequence, and only the translation initiation start site would be conserved across further species. Even though not present across many species, conservation might be visible to a subset like hominids which discriminated primates from other mammals.

Dilated cardiomyopathy and heart failure

All the previously described steps is fundamental in molecular biology of which some parts are well studied and others yet poor understood. Dysregulation of these can results in severe diseases (Tahmasebi et al., 2018). Of particular interest are disruption causing cardiomyopathies.

Cardiomyopathies are disorders of the heart muscle with mechanical or electrical dysfunction resulting in distinct phenotypes, including dilated cardiomyopathy (DCM) and hypertrophic cardiomyopathy (HCM), arrhythmogenic cardiomyopathy (ACM), and left ventricular noncompaction cardiomyopathy (LVNC) (**Figure 5**). Dilated cardiomyopathy is the most

frequent type defined by left- or biventricular dilatation and systolic dysfunction in the absence of hypertension, valvular disease, coronary artery disease, or congenital heart disease (McKenna et al., 2017). DCM can be classified into a genetic and nongenetic form (Elliott et al., 2008), showing severe cardiac dysfunction that can lead to heart failure and arrhythmias (Richardson et al., 1996), making it the primary cause of heart transplantation. Besides left- or biventricular enlargement, resulting in increased weight up to 1kg, the walls are relatively thin, showing decreased contractility and impaired ventricular function (Elliott et al., 2008). The 2013 published population-based study estimated a prevalence of DCM to be > 1 per 250 individuals (Hershberger et al., 2013). The mortality rate associated with cardiomyopathy is 5.9 per 100,000 individuals (Lozano et al., 2012), with a survival rate of 50% after five years of diagnosis since patients develop heart failure or arrhythmias over time (Luk et al., 2009). DCM-related heart failure arises due to pump failure (70%), while the remaining fraction encompasses sudden cardiac death from arrhythmias (Dries et al., 1999).

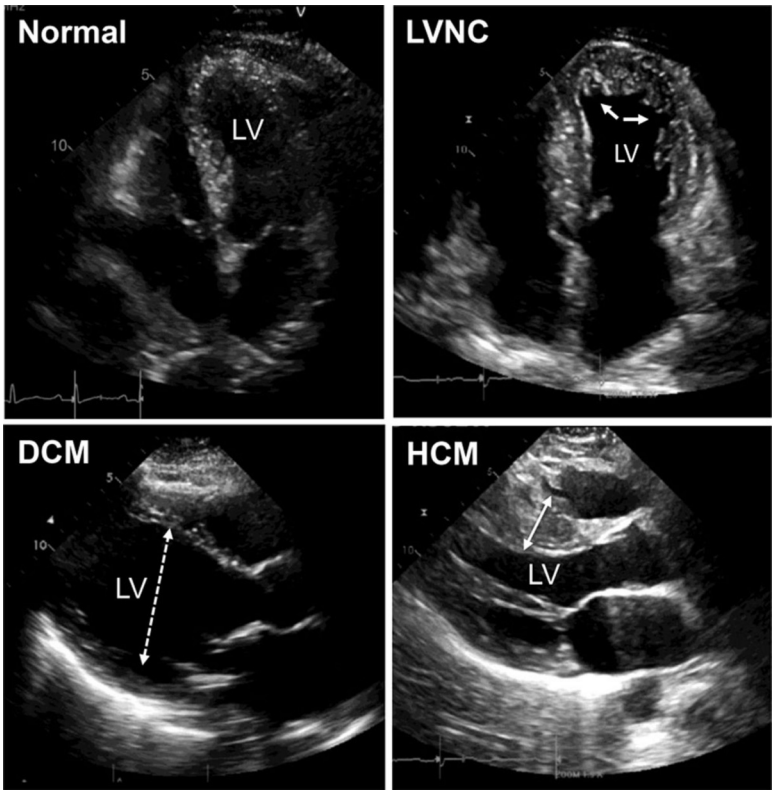


Figure 5: Echocardiography of different cardiomyopathies.

Left ventricular noncompaction cardiomyopathy (LVNC) on the right upper part. Dilated cardiomyopathy (DCM) at the bottom left and Hypertrophic cardiomyopathy (HCM). (McNally & Mestroni, 2017)

The risk of developing heart failure is ~30% higher in black individuals with an increased mortality rate of 1.2 - 1.5 than white individuals (Coughlin et al., 1994; Dries et al., 1999). Furthermore, DCM is diagnosed more often in men than in women (Nieminen et al., 2008).
Among the nongenetic factors associated with DCM are alcohol abuse, drugs, viral infection, or autoimmune diseases, still around 30% of cases have a familial and thus genetic pathogenesis (Hershberger et al., 2013). To date, more than 50 genes have been associated with DCM with different onset and progression of the disease (Raju et al., 2011). Affected genes are involved in various biological processes, such as nuclear-localized splicing and transcription, translation or encoded components of the sarcomere, membrane-scaffolding proteins (Brauch et al., 2009; Harakalova et al., 2015; Herman et al., 2012; Kamisago et al., 2000). Thus, evaluation of DCM patients should encompass detailed family history and their clinical screening combined with next-generation sequencing methods such as whole-exome sequencing to identify the disease variant and the affected gene(s).
The most common genes known to cause DCM are *TTN* (Gerull et al., 2002), *LNMA* (Fatkin et al., 1999), *MYH7*, *TNNT2* (Kamisago et al., 2000), and *RBM20* (Brauch et al., 2009; Maatz et al., 2014). Interestingly, distinct mutations within the same gene may cause either dilated or hypertrophic cardiomyopathy suggesting a shared molecular basis across cardiomyopathies.

Material and Methods

Note: All wet-lab experiments were done by colleagues and collaborators inside and outside of the MDC. Their work will be summarized shortly in this thesis. For a detailed description of the wet-lab methods, see references in the corresponding section. Those are mainly the following publications "*The Translational Landscape of the Human Heart*" *(DOI: 10.1016/j.cell.2019.05.010)* and "*Widespread Translational Control of Fibrosis in the Human Heart by RNA-Binding Proteins*" *(DOI: 10.1161/CIRCULATIONAHA.119.039596)*.

Generation of data sets

Human primary tissue

Human left ventricle heart tissue from 65 DCM patients was collected during left ventricular assistance device (LVAD) implantation or from explanted hearts, whereas 15 unaffected non-DCM controls came from unused donor hearts.

Institute	Condition
Cardiovascular Research Center (London, England) (n = 41)	DCM
UMCU (Utrecht, Netherlands) (n = 4)	DCM
HDZ-NRW (Bad Oeynhausen, Germany) (n = 20)	DCM
UMC (Utrecht, Netherland) (n = 4)	non-DCM
Sydney heart Bank (Sydney, Australia) (n = 10)	non-DCM
DHZB (Berlin, Germany) (n = 1)	non-DCM

The cohort consists of 22 females and 58 males whose age (mean = 43.6; SD = 15.48 years) and gender did not correlate with the condition of the patients (Fisher's exact test; p-value = 0.28) or the first two principal components (both mRNA-seq or Ribo-seq; Student's *t*-test for association with first two principal components; p-value = 0.13 - 0.35). For this study, power calculation was not applied, in order to be as comprehensive as possible in detecting all translation events.

Additionally, human kidney (n = 6) and liver (n = 6) tissue samples were obtained to validate potential translation events among long noncoding RNA genes. Tissue collection and processing have been described in the corresponding publication (van Heesch et al., 2019). Human kidney and liver samples were obtained under ethical approval (Germany; ethical approval EA1/134/12) for heart and (Japan; ethical approval H19-057-437) for kidney and (Germany; ethical approval PV4081) for the liver.

Human iPSC-CM and fibroblast cell culture

Human iPSC-derived cardiomyocytes (iPSC-CM) were differentiated as described previously (Burridge2014) by the Pluripotent Stem Cell facility at MDC Berlin. Libraries were generated and used for mRNA-seq, ribosome profiling, and shotgun MS. *XIST* expression was used for gender identification (females were defined by high *XIST* mRNA abundance). Differentiation and processing of the human primary cardiac fibroblasts (male donors) and iPSC-CM have been described previously (Chothani, Schäfer, et al., 2019; van Heesch et al., 2019).

Animal models

Mouse left ventricular heart tissue was obtained from 10-week-old C57BL/6 wild-type males. Rat left ventricular samples from previously described and acquired from the same source (Schafer, Adami, et al., 2015). These include 6-week-old inbred BN-Lx (n = 5) and SHR (n = 5) males. Tissue processing was described in the main publications previously (Schafer, Adami, et al., 2015; van Heesch et al., 2019)

Processing and alignment of datasets

Ribosome profiling (Ribo-seq) and matching mRNA-seq

Ribosome profiling was performed on human primary left ventricles (n = 80), human kidney (n = 6), human liver (n = 6), human iPSC-CM (n = 2), human primary cardiac fibroblasts (n = 20), and mouse left ventricles (n = 6) according to the TrueSeq Ribo Profile which was previously optimized for tissue samples (Schafer, Adami, et al., 2015).

The protocol was initially used to generate Ribo-seq rat left ventricle data from BN-*Lx* and SHR animals (available at the European Nucleotide Archive (ENA) under accession number PRJEB7498). In this work the dataset was used to validate translation events and their conservation between human and rodents. Ribosome profiling on the human iPSCM-CM cells was performed using the TruSeq Ribo Profile (Mammalian) Library Prep Kit (Illumina, San Diego, CA; USA) following the TruSeq Ribo Profile protocol. For all samples, library size distribution was verified on the Bioanalyzer 2100, multiplexed and sequenced on an Illumina HiSeq 2500, generating single end 1 × 51nt long reads.

To avoid preprocessing biases, samples were processed in batches of maximum 30 units.

For the human heart Ribo-seq libraries, an average library depth of 115M (min. 56M, max. 232M) raw reads was achieved.

Along with ribosome profiling, poly(A)-purified mRNA-seq libraries were generated from high-quality RNA (average RNA Integrity Number (RIN) of 8.1 (human left ventricle), 9.1 (rat), and 7.9 (mouse)). Rat RNA was extracted from the same tissue used for ribosome profiling. RNA-seq libraries were prepared in batches of 48 samples according to the TruSeq Stranded Total RNA and mRNA Reference Guides.

Libraries were multiplexed and sequenced on an Illumina HiSeq 4000, producing paired 2 × 101nt reads. TotRNA-seq was generated for the human heart but not for human kidney and liver tissues. For the human heart samples, an average library depth of 83M (min. 59M, max.

118M) and 82M (min. 52M, max. 205M) raw reads was achieved for the mRNA-seq and totRNA-seq, respectively.

Preprocessing of sequenced reads

While RNA-seq adapters were removed at the demultiplexing stage with software provided along with the sequencer, Ribo-seq reads are shorter than 51nt. Because of this, the sequencing reached out into the 3' adapters, which needs to be manually removed. It is essential to mention that reads that do not contain a 3' adapter should be discarded as they might come from different sources of interaction and not from ribosome-protected fragments during active translation. Ribo-seq read adapters were manually removed using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit) with the following command:

Unix Code:

```
fastx_clipper -a AGATCGGAAGAGCACACGTCT -l 20 -c -n -v | fastx_trimmer -f 1 -z -o .trimmed/<sampleID>/trimmed.fastq.gz
```

Parameters:

-a Adapter sequence used during library preparation.
-l Discard sequences shorter than N nucleotides; Default is 5.
-c Discard non-clipped sequences. This option needs to be used for Ribo-Seq but not for RNA-seq libraries.
-n Keep sequences with unknown nucleotides.
-v Verbose - report number of sequences.
-f First base to keep.
-z Compress output.

The resulting FASTQ files consist of the actual bases called during sequencing and the corresponding Phred quality score. The score is defined as the probability that the base-calling is wrong, e.g. $-10*\log_{10}(0.0001) = 30$. Reads showing on average Phred score above 25 are considered to be of good quality.
To obtain a set of fragments corresponding to autosomal and sex chromosomes, reads from each sample were mapped to the mitochondrial RNA, ribosomal RNA, and tRNA sequences of the respective species, and the unmapped fraction was kept for further processing. Due to

differences in average mRNA-seq and Ribo-seq read length, full-length mRNA-seq reads were trimmed with FASTX-Toolkit to 29-mers (average length of Ribo-seq reads) to be able to establish a comparative analysis between mRNA-seq and Ribo-seq datasets and escape any mapping and/or quantification bias:

Unix Code:

fastx_trimmer -f 1 -l 29 -z -o trim_dir/outname.fastq.gz

Parameters:

-f First base to keep.
-l Last base to keep.
-z Compress output with GZIP.
-o FASTA/Q output file.

Most recent genomes from mitochondria and ribosomal RNA for many species, including human, mouse and rat are available at the Ensemble database (Howe et al., 2021).

Initial

Read Alignment

At first, human, rat and mouse full-length 2 × 101nt mRNA-seq reads were mapped to the respective reference genomes (GRCh38.p10/hg38, Rnor6.0/rn6 and GRCm38.p5/mm38) using STAR v.2.5.2b (Dobin et al., 2013).

STAR mapper was selected because of its high accuracy and mapping speed. The mapping procedure is subdivided into two steps. In the first step, STAR searches for every read the longest sequence that matches the reference genome, called Maximal Mappable Prefixes (MMPs). The first MMP is called *seed1*. STAR will then search only for the unmapped fraction to match the reference genome or the next MMP, which is then called *seed2* and so on. This strategy, combined with the intelligent division of the genome and its storage in a suffix array, makes STAR faster than other algorithms that first try to map the whole read before splitting it into parts. In the second step, STAR stitches the seeds to generate a complete read by first clustering the seeds based on proximity and then based on the best alignment for the read, considering mismatches, indels, or gaps.

For the actual mapping of the mRNA-seq samples, we used a 2-pass mode that allows novel exon splice junction detection. STAR mapping was done using the following modified parameters:
--outSAMtype BAM SortedByCoordinate,
--outFilterMismatchNmax 6,
--outFilterMultimapNmax 20,
--alignSJDBoverhangMin 3,
--outFilterType BySJout,
--alignSJoverhangMin 10
--outSAMattributes All

For proper *de novo* assembly, an XS tag for each mapped read in each mRNA-seq BAM file is required and was added using "*tagXSstrandedData.awk*" script, provided by the STAR authors and made available via the GitHub development platform at: https://github.com/alexdobin/STAR/tree/master/extras/scripts.

De novo transcriptome assembly

To capture the complete transcriptome, including all cardiac expressed isoforms and unannotated lncRNAs, the mRNA-seq BAM files were used to create sample-specific *de novo* transcriptomes using StringTie v1.3.3 (Pertea et al., 2015). All novel transcriptomes were merged into one consensus transcriptome GTF file for each species, using the StingTie-merge option, guided by the canonical reference annotation (Ensembl release 87). To avoid the assembly of undesired mapping artefacts or short RNAs, a novel transcript was required to be longer than 200nt and to have a minimum expression level of FPKM ≥ 1. FPKM stands for Fragments Per Kilobase of transcript per Million mapped reads (paired-end).
In the next step, we filtered undesired novel transcripts coming from mono exonic genes that lack strand information, known non-polyadenylated transcripts (e.g. snRNA), newly annotated genes with no unique mapping reads, and transcript isoforms overlapping neighbouring genes on the same strand. In annotated genes, newly assigned StringTie IDs were manually replaced by the respective reference Ensembl IDs.
We denoted all antisense transcripts (AS), long intergenic (lincRNA), and processed transcripts as lncRNAs. Due to differences in lncRNA annotation across species, we took a particular effort to capture the complete set of already annotated lncRNAs and small protein-coding genes (longest ORF < 100aa) in each of the analyzed species. The protein-coding genes with short

ORFs were of particular interest since they are largely annotated as lncRNA in other species. It happened that a small protein-coding in human is a lncRNA in mouse and *vice versa*. Using the UCSC Batch Coordinate Conversion (LiftOver) utility (Kuhn et al., 2013), we cross-referenced annotated lncRNAs and small protein-coding genes with newly predicted genes, keeping potential orthologous lncRNAs and removing novel genes without a match in any of the other species.

In the human heart, the presented method added 117 potential novel lncRNAs, 978 and 224 novel isoforms of already annotated lncRNAs and small protein-coding genes, respectively. The newly detected candidate lncRNAs and isoforms of annotated genes were included in the respective species GTF and used as a reference for downstream analysis.

Second alignment

The extended GTF with a complete set of cardiac lncRNAs (annotated + novel) was used to map the trimmed and cleaned mRNA-seq, and Ribo-seq reads with STAR v2.5.2b. As all transcripts isoforms of interest were incorporated in the novel annotation, *de novo* splice junction detection was disabled to improve mapping precision. Due to the short read length for both the trimmed mRNA-seq and Ribo-seq, only two mismatches were allowed. Since ribosomal footprint span a range of 20 - 35nt in length, with the majority being 29nt long, we defined the starting point of the search at half of the read length with the option *-seedSearchStartLmaxOverLread 0.5*. Otherwise, both mRNA-seq and Ribo-seq reads were mapped with the same settings.

Identification of novel translation events

Detection of active translation with RiboTaper

In the next step, we used RiboTaper v1.3 (Calviello et al., 2016) to identify active translation across ribosome profiling data with default settings. The principle behind RiboTaper is to use Fourier transformation to determine the 3nt movement of the ribosome along with the open reading frame. RiboTaper additionally applies a *de novo* ORF discovery to capture the entire translatome, only considering ORFs with canonical start and stop codons. To ensure high periodicity, we kept reads length for which 70% of the reads matched the primary reading frame of the ORF; otherwise, the settings were not changed. All ORFs were required to be at least

eight amino acids long, showing evidence of uniquely mapping reads and have at least 21 P-sites. The resulting list was further filtered, requiring a gene to be considered as translated if the average mRNA FPKM ≥ 1 (transcription) and detected as translated in at least 10/80 in human, 3/10 in rat, and 2/6 in mouse heart samples. Each ORF was required to have an identical translation termination codon in at least 5/80 in human, 2/10 in rat, and 2/6 in mouse heart samples to avoid spurious signals. Besides canonical ORFs and short ORFs (sORFs) from lncRNAs, we were able to capture up- and downstream ORFs. Upstream ORF can be upstream of the main ORF or partially overlap it but on a different reading frame.

Translation of known functional lncRNAs

To quantify the fraction of translated lncRNAs with previously published noncoding function, we manually curated a database of 324 functionally characterized lncRNAs based on public databases (Amaral et al., 2011; Gray et al., 2013; Quek et al., 2015) and extensive literature search. We see 61/324 (18.82%) lncRNAs expressed in human, rat, or mouse hearts, and 32 were detected as translated by RiboTaper in the heart (27 in human, 1 in rat, 7 in mouse). Furthermore, we see translation evidence for 28 translated lncRNAs in human primary cardiac fibroblast and 17 in iPSC-CM, of which ten, respectively two have not been detected as translated in the human heart. Altogether, we were able to collect 42 translated, functionally characterized lncRNAs in the heart.

Identification of microprotein features

To better understand where microproteins localize in the cell and hypothesize about potential functions based on localization, we looked into embedded amino acid sequence features such as N-terminal regions that frequently define peptide localization. In the first step, we predicted microprotein localization using TargetP v1.1 (Emanuelsson et al., 2000) and DeepLoc v1.0 (Almagro Armenteros et al., 2017). To reduce false-positive signals, we omitted plant-specific chloroplasts as possible localization. SignalP v4.1 (Petersen et al., 2011) and TMHMM v2.0c (Krogh et al., 2001) were applied to predict signal peptides and transmembrane helices respectively across new amino acid sequences that target proteins into, or across membranes.

circRNA detection and translation

CircRNA detection in the human left ventricle was performed on the rRNA depleted totRNA-seq samples, since these datasets contain can capture nonpolyadenylated fragments; mRNA-seq was used as a negative control, particularly for detecting polyadenylated transcripts, e.g. shuffled exons mimicking backsplice junction signs. In the first step, reads were mapped with BWA-MEM (H. Li & Durbin, 2010) to the human genome (GRCh38.p10/hg38) with the following parameters: -t4, -L 3,3 -E 3,3 -k 14 -T 1. The actual circRNA detection was done using find_circ2 (v1.2 https://github.com/rajewsky-lab/find_circ2). The identified circRNAs were filtered for the ones with at least two unique reads in at least 10/80 samples, but 50 junction overlapping reads in total. Additionally, they had to be of an exonic length of 50 - 10,000nt, coming from a single gene of autosomes or sex chromosomes and absent in mRNA-seq.

To avoid false-positive circRNAs backsplice junctions arising from *trans*-splicing and/or exon shuffling in polyadenylated transcripts, we additionally applied a ratio cutoff of 100:1 for presence in totRNA-seq vs. mRNA-seq data. This led to the exclusion of 320/324 circRNAs that were detected in both datasets. Finally, we tagged circRNAs as good quality candidates by requiring a 15-fold higher support than warning flags, resulting in 85% of 8,878 good quality circRNAs.

CircRNA translation was identified by remapping Ribo-seq reads that could not be mapped to the human transcriptome and genome, to the backsplice junctions using Bowtie2 (v2.0.6) (Langmead & Salzberg, 2012). For this, 40 nucleotides of splice junction flanking exons were extracted, reversely combined, and used as novel reference circRNA transcriptome to map the Ribo-seq reads. Requiring an overlap of at least 9nt on either side of the backsplice junction, we mapped 1,298 (766 unique) Ribo-seq reads to 508 circRNAs. Applying stringent criteria, we identified 40 circRNAs encoded by 39 genes that whose backsplice junctions were covered by at least three unique and five total Ribo-seq reads and thus considered as potentially translated in the human heart. To confirm Ribo-seq mapping specificity and that mapping is an actual signal and did not arise by chance, we simulated ~3.8 million circRNAs backsplice junctions derived from random combinations of translated cardiac exons removing backsplice junctions predicted by find_circ2 on the actual dataset in both mRNA-seq and totRNA-seq. The remaining junctions were randomly subsampled into 10,000 sets each of 8,878 simulated junctions, preserving the circRNA length distribution observed in the actual circRNA dataset. We again used the unmapped fraction Ribo-seq reads (i.e. reads cannot be mapped to the linear transcriptome and genome) to map them against each of these 10,000 sets requiring overlap between 1 - 15nt to check if any other potential thresholds would results in a significant outcome or confirm that every test set has significantly lower mappability that the actual circRNA dataset.

Quantitative analysis of the cardiac transcriptome and translatome

Gene expression quantification and differential expression

Gene expression quantification for both mRNA-seq and Ribo-seq samples was done by counting reads that map to the coding sequence (CDS) region of all translated genes and on full exons of untranslated genes using HTSeq-count (Anders et al., 2015). Full-length mRNA-seq counts were used to identify transcribed genes with an average expression of FPKM ≥ 1 across all heart samples.
Finally, we removed translated genes that showed zero counts in the final quantification by HTSeq-count (both trimmed mRNA-seq and Ribo-seq).). This also removed overlapping genes with full CDS overlap that could not be quantified unambiguously.

mRNA-seq:
htseq-count -f bam -r pos --stranded=reverse -t exon -i gene_id human_mRNA.bam Homo_sapiens.GRCh38.87_custom.gtf > /path/out/human_mRNA_counts.txt

Ribo-seq:
htseq-count -f bam -r pos --stranded=yes -t exon -i gene_id human_ribo.bam Homo_sapiens.GRCh38.87_custom.gtf > /path/out/human_ribo_counts.txt

Parameters:
-f File type
-r Sorting of the file.
-c Discard non-clipped sequences. This option needs to be used for Ribo-Seq but not for RNA-seq libraries
-n Keep sequences with unknown nucleotides.
-v Verbose; report number of sequences.
-f First base to keep.
-z Compress output.

Our setup aimed to achieve an unbiased comparison of transcriptional and translational levels and disease conditions. For this, we integrated raw counts of transcribed and translated (n = 11,387) genes of both mRNA-seq and Ribo-seq, using reads mapping to CDS regions. For mRNA-seq reads, single-end reads trimmed to 29-mers (average ribosomal footprint size) were

used. Finally, joined sample size normalization and differential expression analysis between DCM and control using DESeq2 v1.12.4 (Love et al., 2014) was applied.

R code:

```
pool_sizeFactor ←
 estimateSizeFactorsForMatrix(cbind(raw_counts_polyA,raw_counts_ribo))
mrna_sizeFactor ← pool_sizeFactor[1:ncol(raw_counts_polyA)]
rpf_sizeFactor ← pool_sizeFactor[(ncol(raw_counts_polyA)+1):(ncol(raw_counts_polyA)
+ncol(raw_counts_ribo))]
```

A gene was considered differentially expressed if it met the significance threshold of $p_{adj} \leq 0.05$ and a fold change (FC) ≤ 1/1.2 (0.83) or ≥ 1*1.2.

The results were used to identify gene expression level changes from transcription to translation, revealing a hidden layer of regulation that has, e.g. a basis at the transcriptional level. For this a delta fold change of mRNA-seq FC (DCM vs. Control) and Ribo-seq FC (DCM vs. Control) was calculated.

Translational efficiency was calculated on mRNA-seq vs. Ribo-seq ratio for each gene across all samples. Alternative methods to estimate TE such as group-specific TE or ΔTE (Chothani, Adami, et al., 2019; W. Li et al., 2017; Z. Xiao et al., 2016; Zhong et al., 2017), were not suitable for calculating the RBP-driven regulation because they provide finalized group-specific ratios.

Coregulation analysis

We calculated Spearman correlations using DESeq2-normalized counts of pairwise complete observations in different sets of genes: (i) all translated genes, (ii) all differentially transcribed and translated genes, (iii) lncRNA-mRNA antisense pairs, and (iv) RBP target gene pairs, for a better understanding of how genes are coregulated. Only those with detected expression (mean FPKM ≥ 1) and translation in at least 20 samples were considered. Gene pairs that showed a significant ($p_{adj} \leq 0.05$) correlation after correction for multiple testing (Benjamini & Hochberg, 1995) were retained for downstream analysis. For the RBP target gene pairs correlations, we used expressions levels of the RBP (as measured by Ribo-seq) and either target gene mRNA-seq abundance or translational efficiency.

To test for significant differences between two correlation values, correlation coefficients were Fisher Z transformed (Fisher Z-Transformation or Fisher r to z transformation) (Fisher, 1915) and compared to a normal distribution, allowing for statistical comparison.

To study global (or subsets of genes) coregulation, the computed correlation matrix was used to elucidate the Euclidean distance followed by hierarchical clustering. Cluster visualization was done using heatmap.2 from the gplots v3.0.1, the modified heatmap.3 (https://github.com/obigriffith/biostar-tutorials/tree/master/Heatmaps) or pheatmap v1.0.12 R package.

To increase the interpretability of gene-gene clusters of differentially expressed genes (as measured by transcription, translation, or both) and their relationship to each other without losing information, we performed principal component analysis (PCA). Analysis was done on translationally downregulated (delta Log2FC < 1/1.2), translationally upregulated (delta Log2FC > 1.2) and transcriptionally regulated (delta Log2FC > 1/1.2 and < 1.2) genes together with average transcription and translation log2FC values between DCM and control samples.

Gene Ontology (GO) enrichment analysis

We assigned biological function to selected gene sets of interest via GO enrichment analysis using gProfiler v0.6.4 (Reimand et al., 2016) archive revision 1741 (Ensembl 90, Ensembl Genomes 38).

Because of the time difference of the two presented works in this thesis, GO enrichment analysis for the RBP section was performed using gProfiler2 v0.1.9 (archive revision fof4439, (Raudvere et al., 2019). For all GO enrichment analyses, translated genes were used as a custom background set. Only GO terms of 20 - 500 genes were considered to avoid general terms that frequently show strong significance due to beneficial input and term size ratio.

Gene specificity across tissues and cell types

GTEX v6 data was used to define cardiac and/or muscle-specific expression of translated lncRNAs by requiring the mean expression level of a gene in the left ventricle or atrial appendage (cardiac-specific) to be 12-fold higher than the mean expression in other GTEx tissues or the mean expression in the left ventricle, atrial appendage and skeletal muscle (muscle-specific) to be 10-fold higher than the mean expression in other GTEx tissues. We further used ribosome profiling information to validate the fraction of translated lncRNAs in the human kidney, liver primary cardiac fibroblasts (Chothani, Schäfer, et al., 2019), and hiPSC-derived cardiomyocytes (21 days old). RNA-binding protein (RBP) tissue expression was

defined as ubiquitously expressed if present in more than 30 samples with mean TPM ≥ 1 or high abundance with mean TPM ≥ 10.

Conservation analysis of translated lncRNAs

To understand whether potentially translated lncRNAs show any patterns of conservation, we first investigated nucleotide conservation of the detected sORFs using a PhyloCSF (Lin et al., 2011) based pipeline (Mackowiak et al., 2015) containing alignments of 49 species. LncRNA sORFs with a PhyloCSF score > 10 were considered as potentially conserved across vertebrates. Some of the tested candidates were completely lacking alignment information and thus could not be tested for conservation. These candidates might be translated from human-specific regions that are absent in other species completely, or from fast-evolving regions that could no be aligned to other genomes due to a high sequence divergence. Positional conservation was determined using LiftOver by converting the genomic lncRNA coordinates between human (hg38), rat (rn6), and mouse (mm10) genomes and requiring a gene to show relative localization to homologous neighbouring protein-coding genes where possible and without sense-overlap with annotated genes. Important to mention the poor annotation of the rat transcriptome with multiple protein coding genes annotated as lncRNAs and vice versa. Positional conservation analysis should be therefore reduced on translated transcript no longer than 150 amino acid.

Additionally, we required a sense-antisense gene pair to be in 5kb neighbourhood. Appropriate chain files for genomic conversion between human, rat and mouse are available UCSC Genome Browser repository (Navarro Gonzalez et al., 2021). Translation initiation site (TIS) conservation was also defined reciprocally between human, rat, and mouse as previously described (Fields et al., 2015). For this, we extracted TIS of translated lncRNAs coordinates and converted their coordinates using LiftOver to the other two species. Conservation was assumed when the conversion was successful in at least one species within a window of 9nt up- or downstream of converted TIS.

Computational analysis of RBPs-targets

Note: The following methods are part of the second paper which investigates multifunctionality of RNA-binding proteins.

Identification of CLIP-seq targets

ENCODE database (Davis et al., 2018) was used to download eCLIP-seq data for 150 RBPs for two different cell lines, HepG2 (n = 103) and K562 (n = 120). RBP datasets consisted of eCLIP-seq peaks and BAM files of already mapped reads to the human reference genome (GRCh38.p10/hg38). For the identification of robust eCLIP peaks across replicates and different cell lines, we processed the datasets as suggested by van Nostrand and colleagues (Van Nostrand et al., 2020). We first quantified the coverage of each predicted peak for all input (mock) and immunoprecipitation (IP, antibody against RBP) BAM files using the BEDTools suite (Quinlan & Hall, 2010). The relative information content was defined as $IC = p_i \; x \; \log2 \; (p_i / q_i)$, where p and q are the sums of reads mapping to the peak in IP and negative control, respectively, and then further used to calculate the Irreproducible Discovery Rate (IDR) (Q. Li et al., 2011), a parameter indicating reproducible peaks across biological replicates. We considered peaks as significant and reproducible when meeting an IDR cutoff < 0.01, p-value ≤ $10^{-5,}$ and fold-enrichment (FC) > 8. For two or more significant peaks overlapping the same genomic region, the most significant was retained. To capture all RBP targets, non-overlapping peaks across cell lines were pooled into one final target set. For the disease-relevant and muscle-specific splicing repressor RBM20, whose targets were not part of the ENCODE dataset, peak regions were obtained directly from the authors (Maatz et al., 2014) and added to the peak table. As RBM20 peaks were derived from rat cardiomyocytes, we used the UCSC Batch Coordinate Conversion (LiftOver) utility to convert rat RBM20 HITS-CLIP targets to GRCh38.p10/hg38 coordinates. Finally, the peaks were added to the table containing ENCODE RBP targets. Overall, we obtained on average 4,300 CLIP-seq peaks that were further overlapped the linear transcriptome for each RBP experiment to identify genes that were supported at least one CLIP-seq peak and thus denoted as RBP targets.

Target gene enrichment

RBPs significantly correlating with at least ten CLIP targets were considered as potential regulators of mRNA abundance and/or TE levels. We leveraged the significance of the correlation by generating 100,000 random equally sized theoretical target sets out of all cardiac translated genes, as has been suggested before (Chothani, Schäfer, et al., 2019). For each of these sets, the number of correlating RBP targets (theoretical) was quantified and compared to the actual observation using an empirical test.

Empirical p-value = sum (theoretical targets > true RBP targets) / 100,000

Empirical p-values were corrected for multiple testing using the Benjamini-Hochberg approach. RBPs that showed a significantly ($p_{adj} \leq 0.05$) higher number of CLIP-derived than theoretical targets were considered as putative regulators of either mRNA abundance (n = 58) or TE (n = 37). Because of the fixed number of random test sets, the minimum empirical p-value, after correction for multiple testing, is 5.25×10^{-5}. Thus the p-value does not necessarily the strength of the significance. To overcome this gap, we calculated Glass' Delta (△) (Freeman et al., 1986) effect sizes, a measure of RBP target enrichment strength. Glass' Delta is defined as the difference between two sets divided by the standard deviation of the theoretical set.

Effect size = (*true RBP targets* - mean (*theoretical targets*)) / sd (*theoretical targets*)

Fibroblast replication cohort

To validate observed RBP-target associations in an independent cohort, we retrieved raw RNA-seq and Ribo-seq data from a set of 20 primary cardiac fibroblast cultures treated with TGF-beta (Chothani, Schäfer, et al., 2019). Raw data is available on the gene expression omnibus (GEO submission: GSE131112, GSE123018, GSE131111) repository. For consistency reasons, read preprocessing, mapping, gene expression quantification, and correlation analysis were done identically as in the human heart dataset (see 'Second alignment' and 'Coregulation analysis' subsections). To confirm RBP regulatory effects on their target genes and thus its replication in an independent cohort, we quantified the fraction of RBP-target correlations where the direction (positive or negative correlation) of regulation between RBP and its target genes was similar in the fibroblast and human heart cohort. Statistical analysis was assessed by running 10,000 permutations of the correlation coefficients in the fibroblast cohort and comparing its fraction to the observed directionality in both the actual and random cohorts.

Analysis of exon splicing

To understand whether RBM20 protein abundance could influence the TE ratios of its targets by modulating isoform ratios (exon in- or exclusion), we calculated the percentage spliced in (PSI) for all exons of all known and correlating RBM20 target genes (Maatz et al., 2014). For this, mapped human heart full-length (2 × 101nt) mRNA-seq reads were analyzed (Schafer, Miao, et al., 2015) to improve splice site coverage.

Network analysis

The Weighted Topological Overlap (wTO) (Gysi et al., 2018; Nowick et al., 2009) analysis was performed to generate an RBP-RBP network and identify master regulators of cardiac translation. For this, the target TE and RBP expression matrix was randomly resampled 400 times, each followed by a calculation of Spearman's correlation. The sampling analysis results were used to calculate a weight score, denoted as wTO. An interaction score with $p_{adj} \leq 0.05$ was considered significant and retained for visualization using Cytoscape (Shannon et al., 2003).

Minimum free energy in 5' UTRs

We set out to predict the 5' UTR secondary structure using RNAfold from the Vienna Package v2.4 (Lorenz et al., 2011). For each target transcript, the 5' UTR sequence was used to calculate the minimum free energy followed by length normalization to observe differences in UTR complexity between positively and negatively correlating target genes.

General remarks on statistical analysis

Statistical analysis and figure generation were done using R v3.6.2 (Team, 2016). All tools and methods used to generate the results are listed in the corresponding method sections. Statistical parameters such as "n", mean, standard deviation, and significance levels are named in figures and figure captions. The "p_{adj}" denotes the p-value corrected for multiple testing using the Benjamini-Hochberg approach (Benjamini & Hochberg, 1995).

Abbreviations

mRNA → messenger ribonucleic acid
lncRNA → long noncoding RNA
RBP → RNA-binding protein
TE → translational efficiency
TE-RBP → RBPs regulating target translational efficiency
mRNA-RBP → RBPs regulating target mRNA abundance

Ribo-seq → Ribosome profiling

mRNA-seq → messenger RNA sequencing

totRNA-seq → rRNA depleted total RNA sequencing

CLIP-seq → Cross-linking immunoprecipitation sequencing

ORF → Open Reading Frame

sORF → short Open Reading Frame

uORF → upstream Open Reading Frame

dORF → downstream Open Reading Frame

TIS → Translation Initiation Site

CDS → coding sequence

UTR → untranslated region

p_{adj} → adjusted p-value

DCM → Dilated Cardiomyopathy

Results

Disclaimer: Collaborators outside and inside of the MDC provided material, personal expertise, and statistical advice. J. Schulz and colleagues did all microprotein in vitro and in vivo analyses presented here.
F. Witte worked with me on the transcriptional and translational control in the failing human heart tissue section. S. v. Heesch and N. Hübner led the study's overall design and fusion of the wet-lab and computational work. Presented results and many more have been published recently (van Heesch et al. 2019). For the section of RNA-binding proteins (RBPs) driven translational regulation J. Ruiz-Orera, N. Hübner and S. v. Heesch contributed to the study design (https://doi.org/10.1101/2021.04.13.439465). Figures used in this thesis are directly extracted from the two papers mentioned above. The Results presented downstream cover two papers and even though both deal with translational activity in the heart, the text is subdivided in two parts.

Data generation and quality assessment

To understand the transcriptional and translational regulation in the failing human heart, we applied mRNA-seq, and ribosome profiling (Ribo-seq) (Ingolia et al., 2009) on human left ventricular biopsies of 65 DCM patients and 15 non-DCM controls (**Figure 6**). The DCM tissue was gathered from the left ventricular assist device (LVAD) core at the time of implant or transmural samples of explanted hearts. The non-DCM came largely from unused donor hearts.

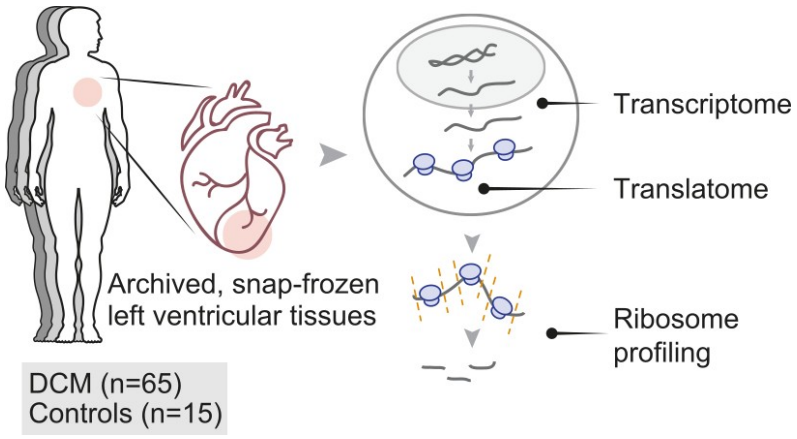


Figure 6: Study experiment design illustrating the experimental approach.

Furthermore, we generated transcriptional and translational information from 6-week old inbred male BN-*Lx* (n = 5) and SHR (n = 5) rats (Witte et al., 2020) and six mice left ventricular biopsies. This information was used in the downstream analysis to investigate differences in conservation and translational profiles between human and rodents, indicating potential conservation of translation events across mammals (see "*LncRNA translation across human tissues*" section). Reproducibility of translation events across tissues was verified additionally across six human kidney and liver samples each. Finally, we investigated the cardiac specificity of novel translation events on 20 primary cardiac fibroblast (Chothani, Schäfer, et al., 2019) and two iPSC-CM samples.

To ensure high-quality sequenced read fragments at the read mapping step, we first removed naturally occurring non-mRNA reads such as mitochondrial RNA, ribosomal RNA, and tRNA sequences (**Figure 7**) with an overall high mitochondrial fraction (mean 30.71% in mRNA-seq and mean 24.06% in Ribo-seq). Interestingly, within the Ribo-seq dataset, we capture on average 58.36% ribosomal RNA fragments owned by the nature of the ribosome profiling library preparation since it is with over 85% the most abundant cellular RNA (Boisvert et al., 2007).

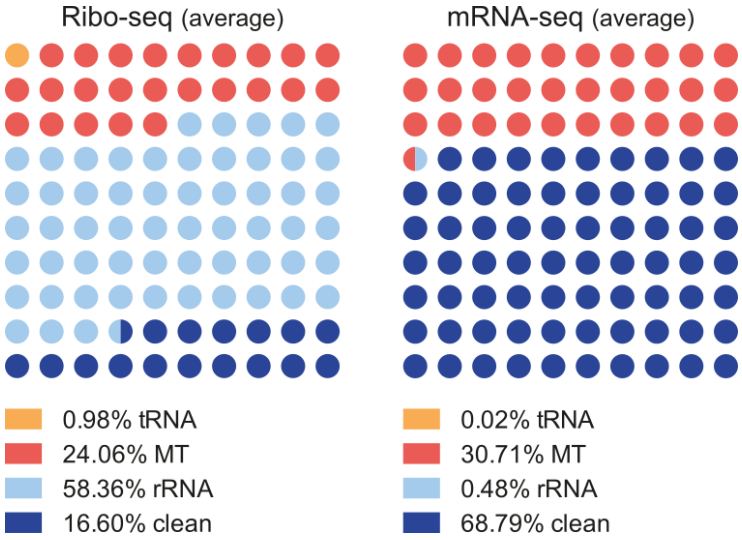


Figure 7: Fraction of different read groups.

Waffle plot displaying the fraction of raw sequence reads derived from tRNA, mitochondrial RNA, ribosomal RNA and canonical genes of Ribo-seq and mRNA-seq data averaged across all 80 human heart samples.

Actively translating ribosomes are known to show a footprint around 28 - 29nt (Ingolia, 2016) which we expect to see after the removal of adapter sequences. The majority of the human left ventricle ribosomal footprints match the expected size and fall almost exclusively

within coding regions (**Figure 8A** and **8B**), indicative of good quality. A small fraction of ribosomal footprints fall within 5' and 3' UTR regions, likely derived from translation initiation and/or elongation of alternative main, upstream or downstream ORFs (Barbosa et al., 2013; Orr et al., 2021).

Assessment of active translation is ensured by estimating the P-site position of every read, which is usually located at the 12th position of a ribosome protected fragment. A P-site is the second nucleotide of the codon and the binding site for the tRNA within the ribosome. Ribosomes periodic 3nt movement along the ORF represents active translation. Since ribosomes move codon by codon steps, their footprints tend to map at the same relative position for all reads.

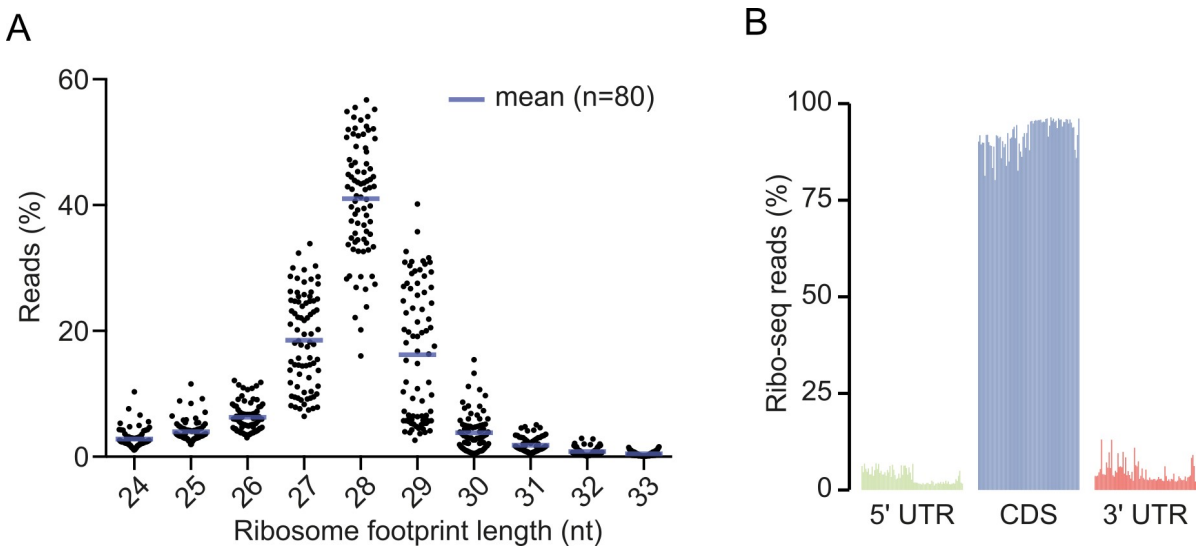


Figure 8: Ribo-seq read mapping results.

(A) *Dot plot showing the ribosomal footprint length across all 80 samples.* ***(B)*** *Bar plot showing the percentage of reads falling into coding sequence (CDS) or untranslated regions (5' and 3' UTR) of protein-coding genes.*

We can use this information to evaluate the periodicity among known protein-coding genes at sub-codon resolution. For this, we plotted the periodicity for the first 100nt of annotated protein-coding genes, showing high (> 70%) trinucleotide periodicity on annotated mRNA ORFs (**Figure 9**), hence indicating good data quality and periodicity.

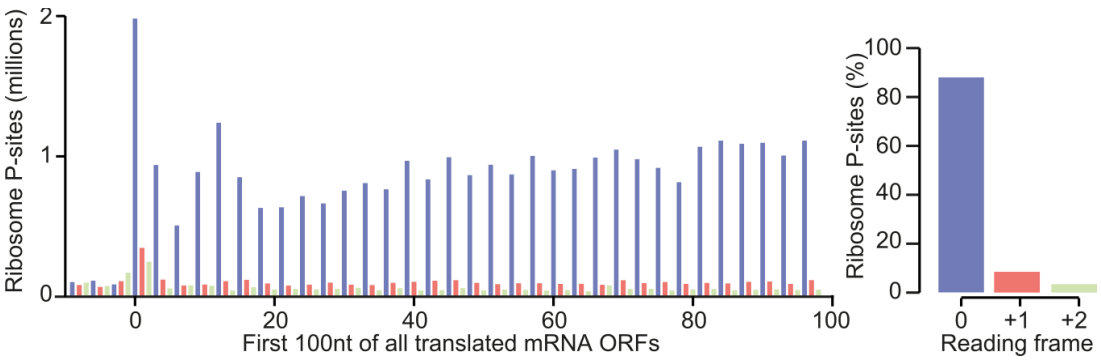


Figure 9: Ribosome footprint quality.

Bar plot showing the P-site position derived from ribosome footprints for the first 100nt of annotated ORFs (left) and its percentage matching the reading frame (right).

Detection of actively translating open reading frames (ORFs)

To capture the complete cardiac transcriptome, especially novel splice isoforms and unannotated lncRNAs, we generated *de novo* transcriptomes for human, mouse, and rat using mRNA-seq data. For this, we ran a *de novo* transcriptome assembly using StringTie (Pertea et al., 2015) on every sample and combined the resulting annotations in a meta GTF, preserving Ensembl gene nomenclature. Novel genes annotated as lncRNA in one of the other two species were annotated as those and added to the reference annotation/GTF. In the human heart, this analysis resulted in 117 potential novel and 978 novel isoforms of already annotated lncRNAs.
To enable read comparisons between distinct sequencing types, full-length mRNA-seq reads were trimmed (29nt) and mapped together with ribosome profiling (Ribo-seq) reads to the extended transcriptome assemblies. We used the resulting BAM files for an unsupervised open reading frame (ORF) detection using Ribo-seq data with RiboTaper v1.3. Hereby, translation events were identified using the 3nt periodicity of mapped reads on exonic sequences. This allows discrimination between ribosome occupancy and identification of active translation (**Figure 10**). Each ORF is tested for its 3nt periodicity using a multitaper approach that converts raw signal (time-domain) to its spectrum of fixed periodic components (frequency domain) using Fourier transformation (Calviello et al., 2016) across all three reading frames.

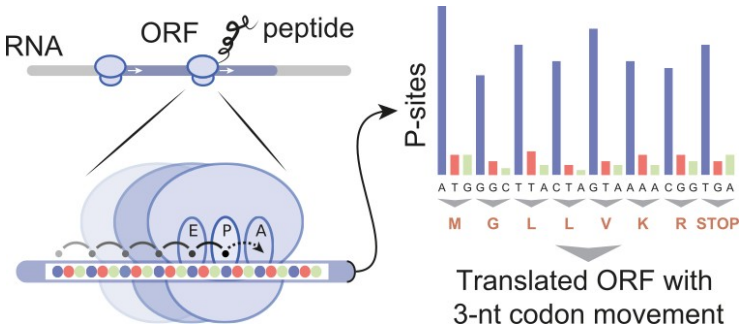


Figure 10: Schematic overview of RiboTapers ORF detection strategy.

Among the 12,631 transcribed genes in the human heart, 11,387 (91.2%) show active translation, the majority of which are located to the canonical ORFs (n = 20,763; 94.69%) (**Figure 11A**).

It is known that a gene can encode multiple ORFs by using different reading frames or position them in a non-overlapping way. Interestingly, we see translation evidence for 1,090 (919 genes) upstream ORFs (uORFs) and 74 (62 genes) downstream ORFs (dORFs). Upstream ORFs are known to impact the translation of the downstream positioned main coding sequence by modulating its biological sequence context (H. Zhang et al., 2019).

Interestingly, we see 339 short open reading frames (sORFs) -denoted as “microproteins” if translated- encoded by 169 lncRNAs (**Figure 11B**), suggesting a hidden layer of yet largely ignored gene sets that potentially encode proteins with biological function.

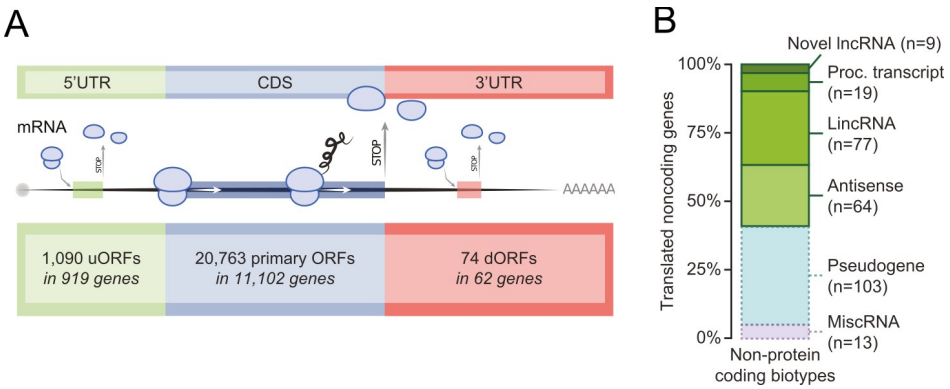


Figure 11: Ribo-seq reads fraction mapping to the distinct mRNA features and lncRNAs.

(A) *Number of translated distinct genes features, including upstream ORFs (uORFs), coding sequences (CDS) and downstream ORFs (dORFs).* ***(B)*** *Stacked barplot showing the fraction and number of translated noncoding genes. The green bars denote translated lncRNAs.*

LncRNA translation across human tissues

A lncRNA is defined as a gene that lacks a canonical open reading frame (> 100aa), longer than 200nt and with absence of sequence conservation across mammalian species (Ulitsky & Bartel, 2013). Our conservation and length-independent approach allows identification of all possible ORFs, including those smaller than 100aa whose function remain largely unknown (Makarewich & Olson, 2017) and need further validation *in vivo* and *in vitro*. Among 783 transcribed lncRNAs in the human heart, 169 (22%) are potentially translated into microproteins. Compared to protein-coding genes, translated lncRNAs are, as expected, almost nine times shorter (mean aa length 426 vs. 47.5) (**Figure 12A**).

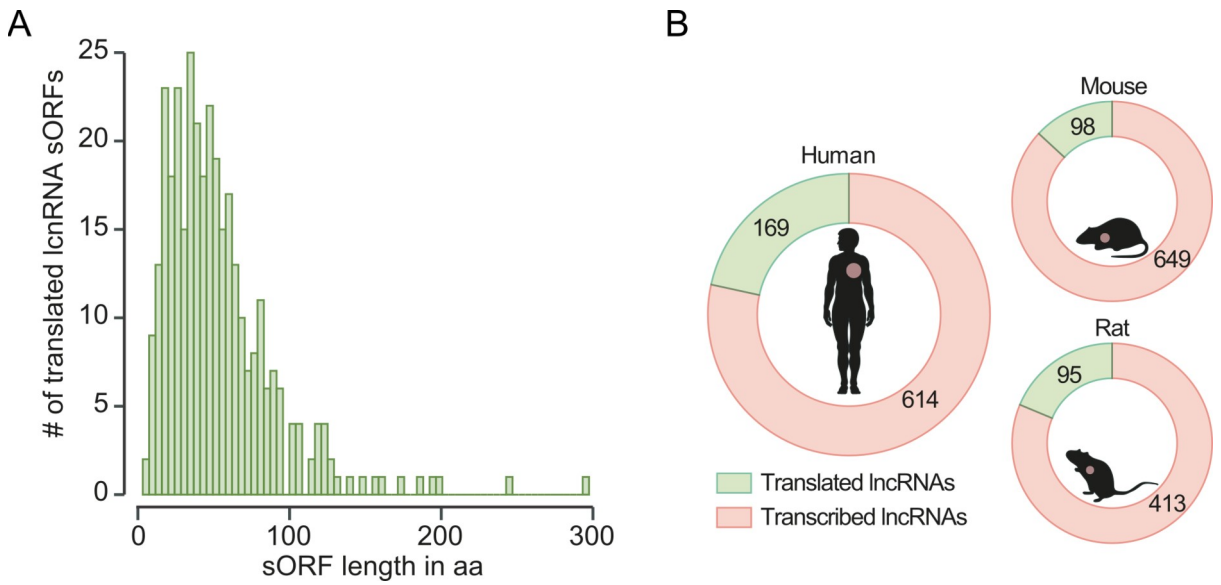


Figure 12: Translated lncRNAs in human and rodents.

(A) *Histogram showing the length distribution of translated lncRNA sORFs in human hearts.*
(B) *Doughnut chart with transcribed and translated lncRNAs in human, rat and mouse hearts.*

To understand to what extent the novel translation events are cardiac-specific or potentially undergo translation in other tissues, we set out to study distinct microprotein expression characteristics.

In the human heart, we detect translation for majority of known cardiac microproteins (longest ORFs < 100aa) (190/199; 95%) together with recently discovered microproteins proteins who at the stage of analysis were annotated as proteins, including DWORF (Nelson et al., 2016), SPAR (Matsumoto et al., 2017), and ALN (also known as C4orf3) (K. M. Anderson et al., 2016). Using public gene expression data provided by the GTEx consortium (GTEx Consortium, 2013) across 44 tissues, we observed that out of 169 translated lncRNAs, 16 are specifically expressed in heart or muscle tissue, suggesting a muscle-specific role for these translated elements. The rest of translated lncRNAs show a rather ubiquitous expression. In the GTEx dataset, 122 lncRNAs are expressed (RPKM > 1) in at least ten and 44 lncRNAs in all tissues. To understand whether translation of lncRNAs also occurs in other human tissues and to what extent our observations are reproducible, we generated translatomes for six liver and kidney tissues, each. Even though the tissues were collected independently, we found 71/169 (42%) and 116/169 (69%) lncRNAs expressed in the liver and kidney, respectively. Of these, 56 (liver) and 87 (kidney) are translated. Altogether 50 lncRNAs are translated in all three tissues (**Figure 13**).

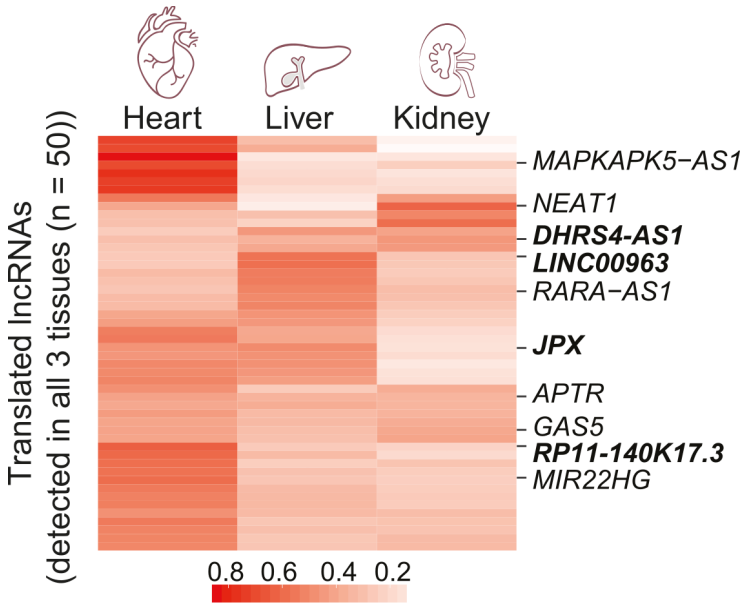


Figure 13: Ubiquitously translated lncRNAs.

Heatmap of scaled Ribo-seq expression (scaled DESeq2 normalized counts) values for 50 translated lncRNAs in human heart, liver and kidney tissue.

Furthermore, 72 (52 genes) ORFs had been previously identified in human cell lines and stored at the sORFs.org (Olexiouk et al., 2016, 2018), a public data base collecting all published translated lncRNAs across all possible organisms, tissues and cell lines. Our dataset underlines that lncRNAs possess the capability to produce novel microproteins potentially across several human tissues and cell lines.

To understand the extent of conservation among translated lncRNAs, we used a PhyloCSF based pipeline (Mackowiak et al., 2015) to check for amino acid conservation of their translated lncRNAs. For 79 nucleotide sequences of the translated lncRNAs could be aligned to hominid species (chimp, gorilla, and orangutan), 43 to mammals, and the remaining 16 are specific to human. Overall, conserved sORFs are mostly absent in rodents and mostly restricted to primates, suggesting that these proteins are evolutionary young (Ruiz-Orera & Albà, 2019) and with yet uncharacterised functions. Apart from assessing the conservation at the amino acid level, we used UCSC Batch Coordinate Conversion (LiftOver) utility to investigate the positional conservation of translated human lncRNAs in rodents. For 76 out of 169 lncRNAs, positional conservation was observed in the same relative orientation to their sense mRNAs (Ulitsky, 2016). To check for potential translation of lncRNAs in model organisms, we used in-house rat and mouse hearts transcriptomes and translatomes and discovered similar translation rates (13

- 22%) of expressed lncRNAs as in the human heart (**Figure 12B**). Among the 76 positionally conserved lncRNAs, 18 are also translated in rodents, and seven share the same translation initiation site. As denoted previously, translated lncRNAs are expressed across most human tissues and show translation evidence in both kidney and liver, suggesting a general translational behaviour that has not been studied yet.

Detection of microproteins in human hearts *in vitro* and *in vivo*

To validate the predicted translated sORFs *in vitro* and *in vivo,* we first performed *in vitro* translation (IVT) assays on 58 randomly selected translated lncRNAs. Forty-four (75%) of these produced stable proteins, and subsequent mutation of the start codon prevented translation (**Figure 14**), confirming presence and translation capability resulting in a detectable protein signal.

In vivo protein detection using mass spectrometry (MS) datasets of short proteins such as signalling proteins with low dynamic range and abundance is particularly challenging in cardiac tissue and can result in false-positive results (Bánfai et al., 2012; Bazzini, Johnstone, Christiano, MacKowiak, et al., 2014; Low et al., 2013; Mackowiak et al., 2015; Nesvizhskii, 2014; Omenn et al., 2017; Slavoff et al., 2013).

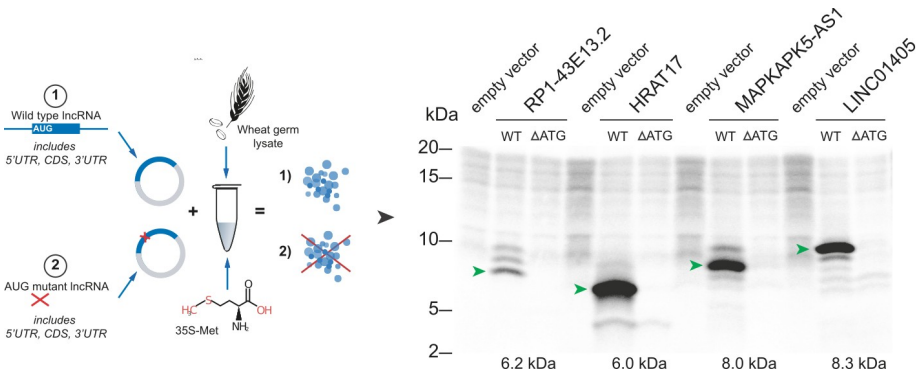


Figure 14: In vitro detection of translated lncRNAs.

Schematic of the in vitro translation (IVT) approach (left) and AUG mutagenesis of selected lncRNA sORFs (right). Predicted product molecular weight in kilodaltons (kDa) for each candidate shown on the bottom.

For the identification of microproteins *in vivo*, we used the deepest human heart shotgun MS dataset to date (Doll et al., 2017) together with the in-house generated iPSC-CM proteome,

resulting in the identification of 140/339 (41.29%) sORF encoded by 93/169 (55.02%) translated lncRNAs. For 28 microproteins, we see unique peptide evidence in one sample, while 100 microproteins are identified in more than one sample. To increase the true positive detection rate and overcome mass-specific issues such as detection of microproteins with low dynamic range, we next applied a high-throughput selected reaction monitoring (SRM) assay (Picotti et al., 2010). It uses synthetic signature peptides to detect the investigated candidate's exact fragmentation pattern, thus increasing the sensitivity and specificity of microprotein detection. Application of this method to five human hearts (two replicates each) resulted in the identification of 76 out of 137 randomly selected microproteins (55.4%) translated from 50/83 (60.2%) lncRNAs.

The presented *in vitro* and *in vivo* methods (IVT assays, shotgun MS and SRM) methods elucidate that it is necessary to combine approaches to validate the translation of microproteins that otherwise would have remained undetected.

Functional characterization of translated lncRNAs

Among the identified translated lncRNAs, 27 human and 5 mouse candidates have been previously shown to host a noncoding function, including *LINC-PINT* (also known as *lincRNA-Mkln1*) (Huarte et al., 2010), *JPX (Tian et al., 2010)*, *CRNDE (Graham et al., 2011)*, *NEAT1 (Clemson et al., 2009)*, *DANCR (Kretz et al., 2012)*, *BANCR (Flockhart et al., 2012)*, *GATA6-AS1* (also known as *lncGATA6*) (Zhu et al., 2018), and the heart-function related *myheart (Han et al., 2014)*, *chaer* (Zhihua Wang et al., 2016), *UPPERHAND* (also known as *UPH* or *HAND2-AS1*) (K. M. Anderson et al., 2016), *ZFAS1 (Y. Zhang et al., 2018)* and *TRDN-AS* (also known as *RP11-532N4.2*) (L. Zhang et al., 2018) (**Figure 15A** and **15B**). Of which *NEAT1, GATA6-AS1,* and *UPPERHAND* are positionally conserved and translated in human and rodent hearts.

Furthermore, we detect the microproteins from these lncRNAs *in vivo*. Interestingly, 22 functionally characterized lncRNAs are also translated in the kidney and liver, showing that lncRNAs likely have noncoding and coding functions.

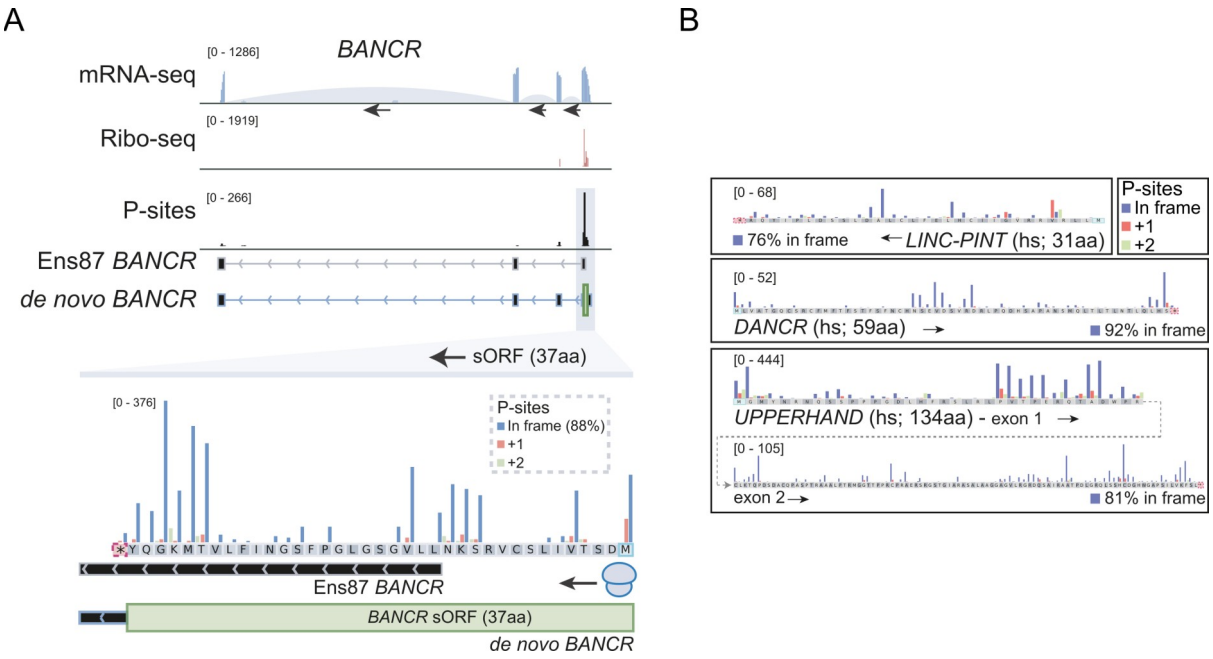


Figure 15: Genome browser examples of translated lncRNAs.

(A) *Genome browser example of the translated lncRNA BANCR (reverse strand), showing mRNA-seq Ribo-seq coverage across the whole gene followed by P-sites localization, canonical and de novo gene annotation. Zoom into P-sites localization of the sORF, which starts upstream of the canonical gene annotation. In-frame P-sites are colored blue.* ***(B)*** *Genome browser zooms into P-site tracks of three translated lncRNAs: LINC-PINT, DANCR and UPPERHAND (HAND2-AS1). In-frame P-sites are colored blue and red and green for +1 and +2 frames, respectively.*

Several of these lncRNAs are antisense to protein-coding genes known to be *cis*-regulatory (K. M. Anderson et al., 2016; Han et al., 2014; L. Zhang et al., 2018). To test for local coregulation of all translated lncRNAs, we extracted lncRNA-mRNA sense-antisense pairs and identified 18 with strong correlation (Spearman's rho: 0.52 - 0.76; $p = 3.3 \times 10^{-5}$ - 1.9×10^{-12}) (**Figure 16A**) of translated lncRNAs and protein-coding genes, including some cardiac transcription factors (*HAND2*, *TBX5,* and *GATA6*), and some important regulatory or structural cardiac genes (*CORIN, TRDN,* and *TNNI3*). We see a significant decrease in correlation during the translation except for *TRDN-TRDN-AS1* (Spearman's rho: 0.23 vs. 0.53; p = 0.0136) (**Figure 16B**). *TRDN-AS1* has been shown to *cis*-regulate cardiac and skeletal muscle triadin production (L. Zhang et al., 2018). The translational regulation we observe in the human heart suggests functional relatedness besides transcription.

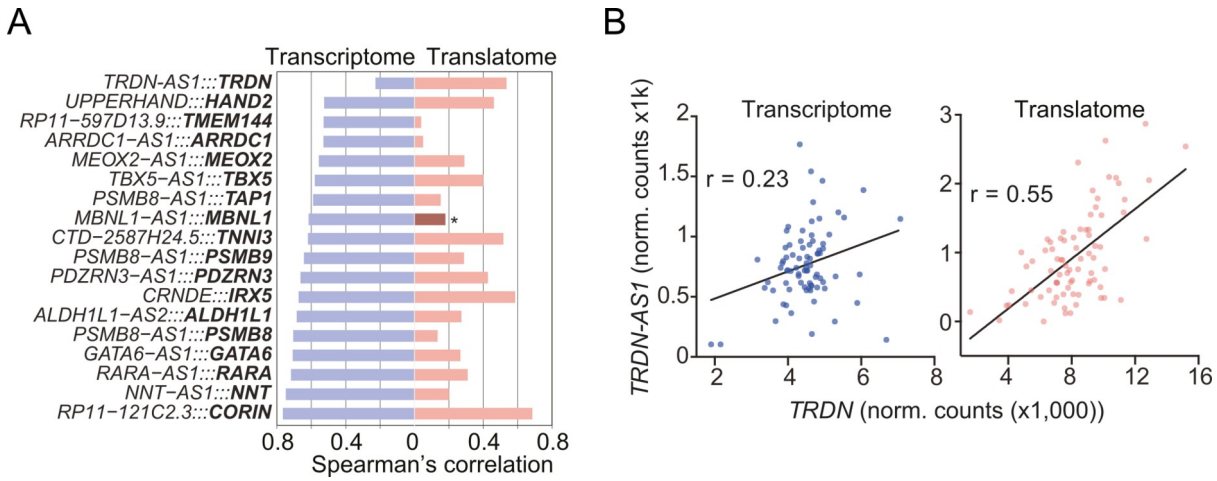


Figure 16: Sense-antisense correlation of translated lncRNAs and their sense mRNAs.

(A) *Bar pot showing translated lncRNA and sense mRNA Spearman rank correlation coefficients at both transcriptional (blue) and translational (red) level. "*" indicates a negative correlation between MBNL1-AS1 and MBNL1. Protein coding genes are shown in bold.* ***(B)*** *Scatterplots showing the correlation between TRDN and TRDN-AS1 at both transcription and translation across all 80 human heart samples.*

Apart the potential function resulting from coregulation to sense mRNA, genes involved in the same molecular pathway will likely show strong coregulation independent of the genomic localization (Lee et al. 2004; Pers et al. 2015). For this, we calculated genome-wide gene-gene correlation followed by hierarchical clustering. We identified a cluster significantly enriched in translated lncRNAs (93 out of 169; $p = 2.17 \times 10^{-15}$; Fisher's Exact Test) along with genes enriched in nuclear-encoded mitochondrial genes (GO:0005739 mitochondrion; $p_{adj} = 8.83 \times 10^{-149}$) (**Figure 17A**). Surprisingly among the 93 translated lncRNAs within the cluster, we observed three candidates (*RP11-140K17.3*, *PRR34-AS1*, *MIR4458HG*) that show a particularly strong correlation with genes involved in all five complexes responsible for oxidative phosphorylation (OXPHOS; KEGG:hsa00190; $p_{adj} = 6.43 \times 10^{-40}$) (**Figure 17B**), suggesting a role in mitochondria, particularly ATP production.

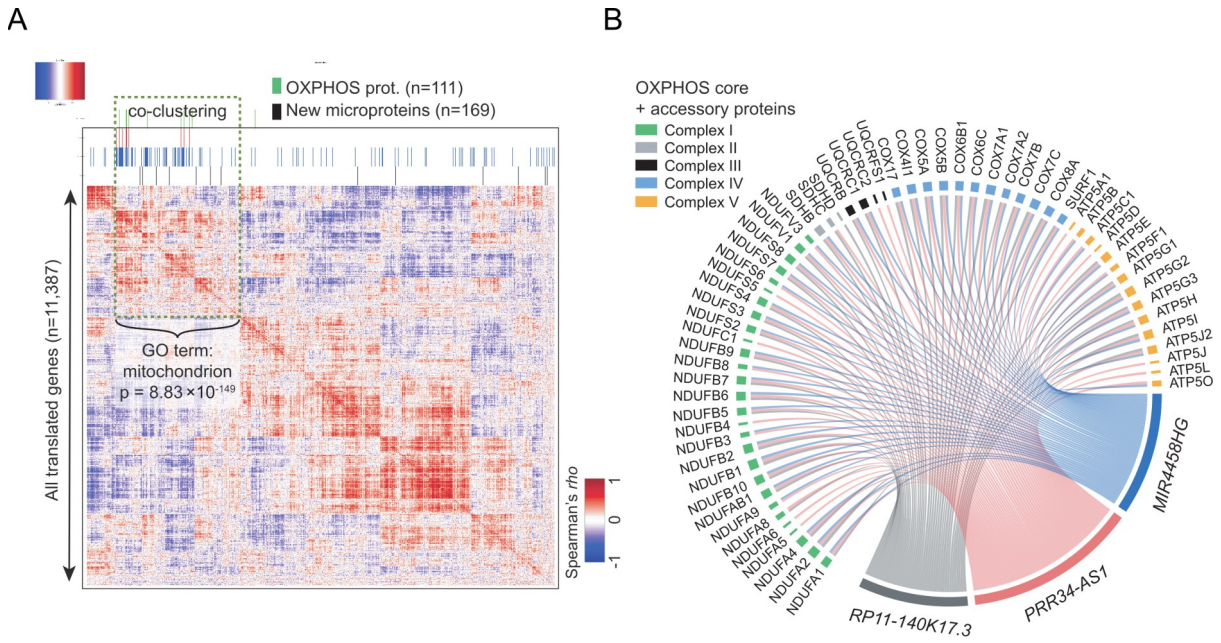


Figure 17: Translated lncRNAs correlate with OXPHOS genes.

(A) *Heatmap of all translated gene-gene correlations. The Black dotted box indicates co-clustering of translated lncRNAs (black bar) and genes encoding OXPHOS proteins (green bar)* ***(B)*** *Circos plot visualizing three translated lncRNAs and coregulated OXPHOS proteins. Each line represents a correlation coefficient ≥ 0.5.*

To determine individual microprotein functions, we applied a GO enrichment (Reimand et al. 2016) analysis on mRNAs that show a strong correlation (Spearman's rho ≥ 0.5) with a lncRNA across all samples. Correlating targets were checked for their biological role using GO enrichment analysis, indicating a shared molecular function of the translated lncRNAs.

Clustering of the resulting GO terms resulted in 42 translated lncRNAs with shared and distinct biological processes such as catabolic process, transcription, translation or DNA binding. The majority (22 lncRNAs) showed enrichment in mitochondria localized functions (**Figure 18A**). The genome-wide lncRNA-mRNA correlation analysis reveals the involvement of translated lncRNAs in mitochondrial processes and thus their localization.

An exciting example is PDZRN3-AS1, a lncRNA that encodes a microprotein which colocalizes to mitochondria and interacts explicitly with RMND1 in a immunoprecipitation (IP)-MS assay (**Figure 18B**). RMND1 is known to colocalize in the inner mitochondrial membrane and is required to translate OXPHOS subunits (Janer et al., 2015), making this protein a perfect candidate for follow up. For a subset of 18 translated lncRNAs, we verified mitochondrial colocalization using immunofluorescence (IF) staining upon overexpression in HeLa cells (see three selected examples; **Figure 18C**).

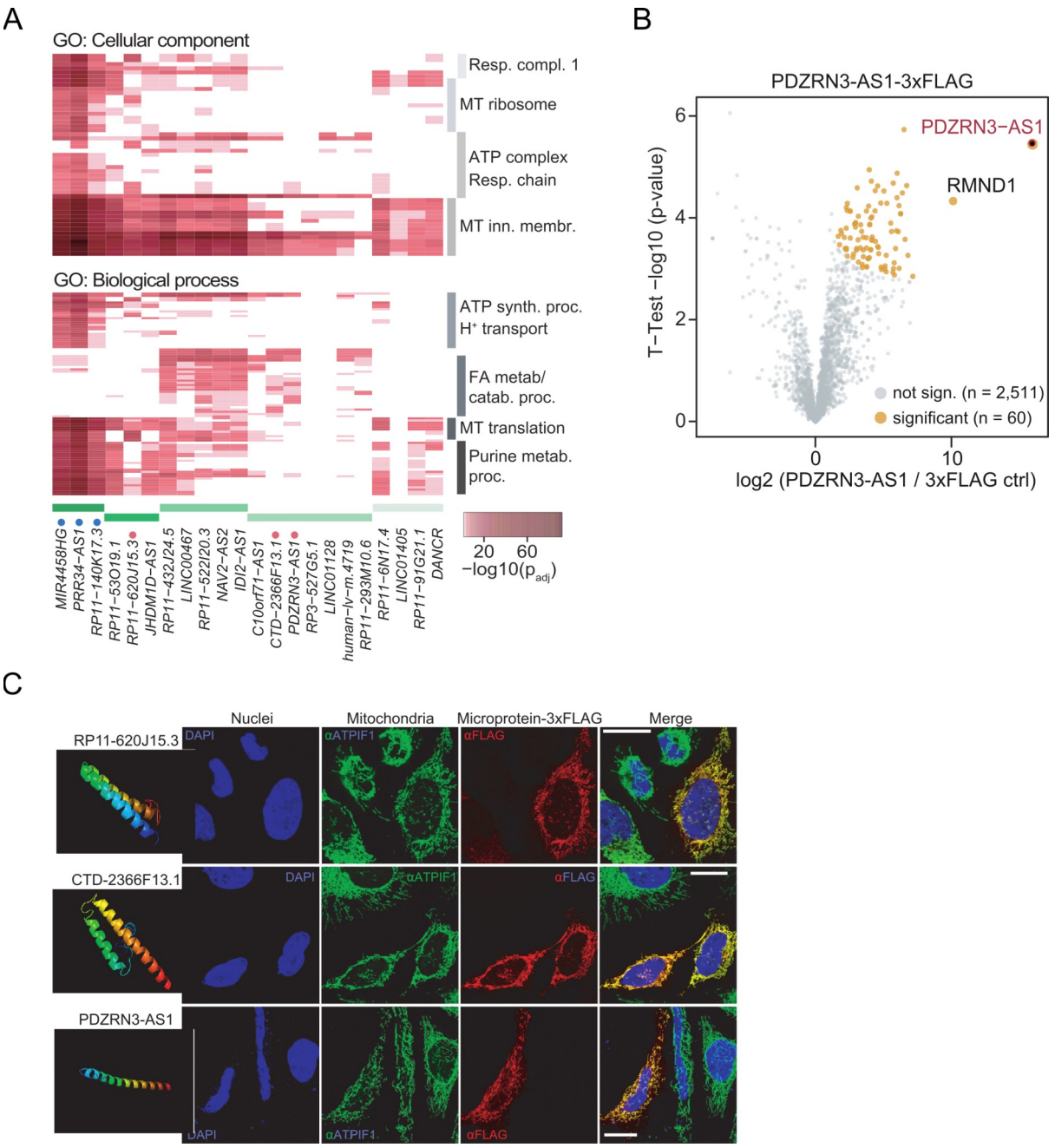


Figure 18: Biological function and localization of translated lncRNAs.

(A) *Heatmap of clustered GO terms for 22 translated lncRNA coregulated with mitochondrial processes.* ***(B)*** *Vulcano plot of immunoprecipitation (IP)-MS results for PDZRN3-AS1. Significant (two-sided T-test; padj ≤ 0.005) interactions are colored in yellow.* ***(C)*** *Immunofluorescence (IF) showing mitochondrial localization for three selected microproteins (right). Scale bars represent 20 µM. Predicted α-helical (left) 3D structures were modelled by I-Tasser.*

Another layer of computational methods that allow for molecular function prediction is the analysis of protein sequence features. Using SignalP (Petersen et al., 2011), we predicted the presence and location of a cleavage site for three microproteins suggesting a potential secretory function. Furthermore, TMHMM (Krogh et al., 2001) predicted for seven microproteins the presence of a transmembrane helix with a range of 13 - 38 amino acids. Finally, DeepLoc (Almagro Armenteros et al., 2017) predicted subcellular localization for all translated lncRNAs. Strikingly, 112 (33%) ORFs were predicted to have mitochondrial localization supporting our observation that many translated lncRNAs likely have a function in this cellular compartment.

Transcription and translation of cardiac circRNAs

Besides lncRNAs, circular RNAs (circRNAs) are a novel class in the noncoding family that has been shown to harbour the potential of being translated (Legnini et al., 2017; Pamudurti et al., 2017; Y. Y. Yang et al., 2017). CircRNAs are characterized by a closed-loop structure missing both the 5' cap and 3'polyA tail compared to mRNAs. Even though circRNAs are derived from linear precursor RNA transcribed by RNA polymerase II, they finally form a circle via the backsplice junction, requiring polyA independent sequencing techniques such as totRNA-seq data for circRNA detection.

To capture circRNAs in the human heart, we used totRNA-seq derived from the same cardiac tissue as the mRNA-seq and Ribo-seq data whereas mRNA-seq data was used as negative control. Even though both data sets showed similar amounts of mapped reads to the genome, backsplice junction reads were almost exclusively located to totRNA-seq (**Figure 19A**). Backsplice junctions within mRNA-seq data likely come from transcripts with exon shuffling (Gilbert, 1978; Smithers et al., 2019), repeats (Rigatti et al., 2004), or falsely aligned reads between homologues genes. Using find_circ2 we detected 9,198 circRNA of which 324 were present in both totRNA and mRNA data sets. To avoid false signals, we applied stringent threshold requiring a circRNA with mRNA evidence to show at least 100 times more totRNA reads (**Figure 19B**). Removing likely false circRNAs, we remained with 8,878 circRNAs encoded by 3,181 genes of which 1,324 (14,9%) showed 15 more find_circ2 warnings than in the remaining 7,554 (85,1) circRNAs (**Figure 19C**), indicating robust detection quality for the majority of the candidates.

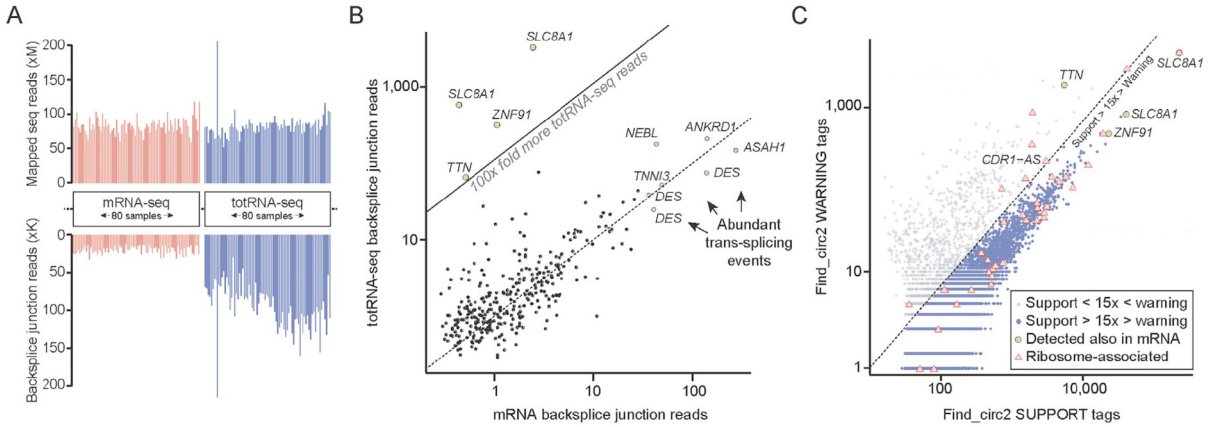


Figure 19: CircRNA identification statistics.

(A) *Abundance of mapped reads and reads matching backsplice junction as identified by find_circ2 in both mRNA-seq and totRNA-seq. The mRNA-seq data set is used as a negative control and thus should not contain backsplice junction reads.* ***(B)*** *Scatter plot of the number of backsplice junction reads for 324 circRNA identified by totRNA-seq and mRNA-seq data. CircRNA with 100-fold higher totRNA-seq reads are separated by a diagonal line. The other circRNAs with abundance in both totRNA-seq and mRNA-seq likely result from trans-splicing and/or exon shuffling and are removed from further downstream analysis.* ***(C)*** *Scatter plot of warning and support flags for all identified circRNAs. Ribosome-associated circRNAs are indicated as red triangles.*

A large fraction of the detected circRNAs (6,808; 76 %) overlaps with the publicly available datasets (Glažar et al., 2014; Khan et al., 2016) (**Figure 20A** and **20B**), showing that we can reproduce circRNAs but also find novel and potentially cardiac-specific candidates.
To identify ribosome-associated circRNAs, we first mapped ribosomal footprints that did not map to the linear transcriptome to the identified back-splice junctions. Requiring at least three unique and five reads covering the back-splice junction (**Figure 20A**), we identified 40 ribosome-associated circRNAs encoded by 39 genes, including the well-known microRNA sponge circCDR1-AS (Piwecka et al., 2017), circCFLAR (Papaioannou et al., 2020) as well as cardiac circRNAs such as circMYBPC3 (Tan et al., 2017), circRYR2 (Tan et al., 2017), and circSLC8A1(Holdt et al., 2018; Tan et al., 2017) (**Figure 20C**). To ensure that we look at circRNAs and not linear transcripts with similar features, we treated our circRNAs with RNase R. Their unique circular structure shows resistance and is thus a perfect target for validation (M. S. Xiao & Wilusz, 2019). For 16 out of 18 tested circRNAs, we see resistance to the nuclease, suggesting their existence and thus potential translation due to the previously mentioned ribosome association.

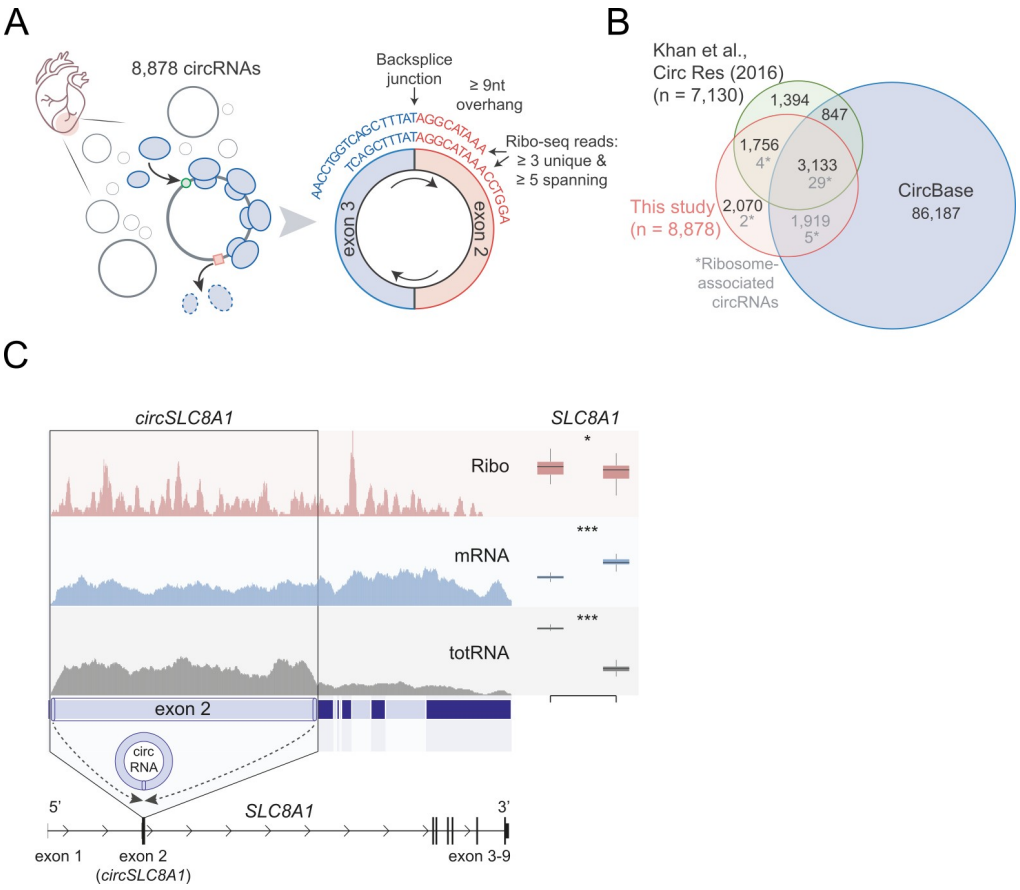


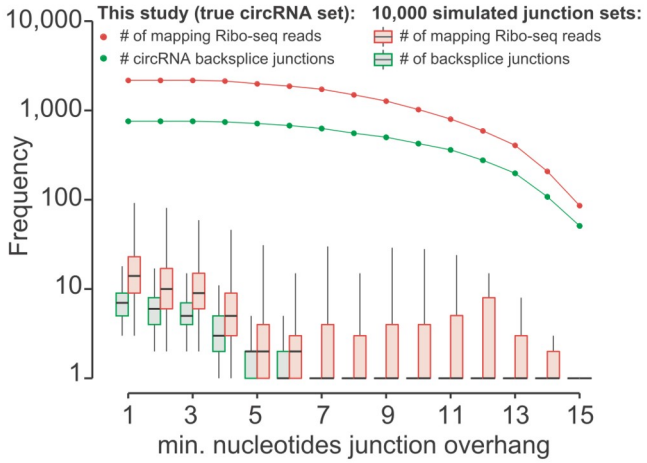
Figure 20: Identification of translated circRNAs.

(A) *Schematic visualization of ribosome-associated circRNAs detection at the back-splice junction site and corresponding thresholds.* ***(B)*** *Venn diagram showing a three-way overlap between the heart dataset, previously identified heart circRNAs (Khan et al., 2016) and circBase (Glažar et al., 2014).* ***(C)*** *Genome browser view of the translated circRNA SLC8A1. Ribo-seq, mRNA-seq and totRNA-seq read coverage for pooled SLC8A1 exons. Exon 2 forms the ribosome-associated circRNA with significantly increased totRNA abundance than the linear exons (right side).*

To confirm the specificity of the alignment, we mapped Ribo-seq reads to 10,000 simulated sets of size- and length-matched backsplice junctions. The resulting statistics showed a significantly lower mappability to the simulated backsplice junctions than to the observed set, confirming circRNA-ribosome specificity of the association ($p_{adj} = 1.5 \times 10^{-4}$) (**Figure 21A**). Five of the 40 candidates have already been reported as translated in human (Y. Y. Yang et al., 2017), otherwise we did not see any other translated circRNAs that have been previously described. Using the same heart mass-spectrometry dataset (Doll et al., 2017) as for the translated lncRNAs, we searched for potential peptides derived from backsplice junction regions and found unique peptide evidence for 6 out of 40 circRNAs (23%) (**Figure 21B**). These discoveries

suggest transcription and translation of circRNAs occurs in the human heart and warrants further investigation of their biological function. Though futures studies require careful analysis during circRNA detection and validation using mass spectrometry which may show evidence close to the noise level (Hansen, 2021).

A



B

Ribosome-associated circRNAs in the human heart

AC109642.1	ENGASE	PSD3	TECRL
ANKH	EXOC6B_1	PSEN1	TIMMDC1
ARHGAP10	EXOC6B_2	RELL1	TPST1
ASXL1	GSE1	RSRC1	TTN
AXL	GUCY1A2	RYR2	USP25
CDR1-AS	KIAA1429	SCARF1	VAMP3
CFLAR	MYBPC3	SLC8A1	XPO1
CSNK1G3	NAXD	SIPA1L1	YAF2
DHDDS	NUDC	SYMPK	YPEL2
DOCK6	PAPOLA	TBCD	ZFAND6

with mass-spec evidence
without mass-spec evidence

Figure 21: CircRNA detection sensitivity.

(A) *Combined dot and scatter plot showing mappability of ribosome footprints to simulated back-splice junctions (empirical* p_{adj} *= 1.5 x* 10^{-4}*).* ***(B)*** *Table of all ribosome-associated circRNAs. Blue colored circRNAs show mass-spec evidence.*

Transcriptional and translational regulation in human hearts

We can study quantitative transcriptional and translational regulation in human diseased and unaffected hearts with the presented dataset. For this, gene expression counts for both mRNA-seq and Ribo-seq datasets were combined and jointly normalized to achieve comparability. Normalized matrices were used for differential gene expression analysis with DESeq2 (Love et al., 2014) on both measured levels. We observed 2,660 and 2,648 genes with differential expression at transcriptional and translational level, respectively. For 964 genes, transcriptional differences are forwarded towards the translation (**Figure 22**), meaning they showed differences in transcription carried over to the translational level.

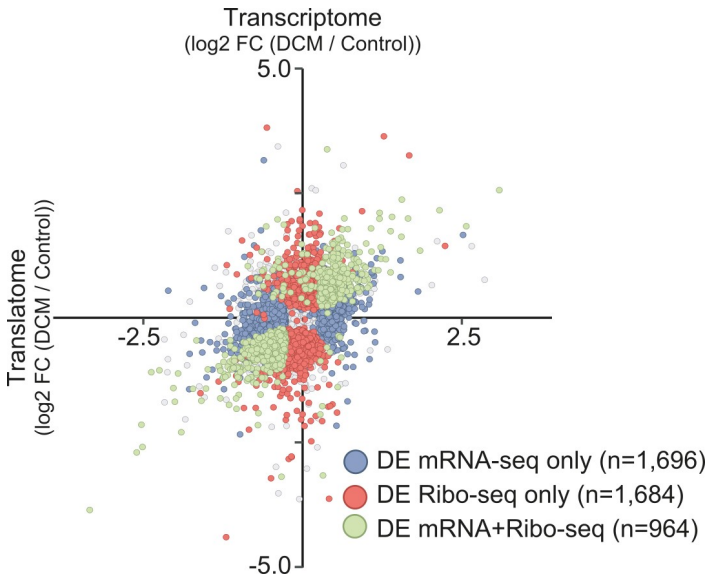


Figure 22: Differential expressed genes.

A Fold-change - fold-change (FC/FC) scatterplot for all differentially expressed genes between DCM and control samples.

We correlated translation levels of all 4,344 differentially expressed genes across all 80 samples to identify processes affected by the failing heart. The resulting correlation coefficients were hierarchically clustered and subdivided into 30 clusters, with 22 showing significant GO terms (The Gene Ontology Consortium, 2019). Each cluster contains genes involved in the process- and pathway-specific regulation of the human heart on a transcriptional or translational level (**Figure 23A**). They show enrichment in genes encoding proteins involved in transcription, translation, and immune response but also processes known to be hallmarks of the diseased heart, such as ECM, sarcomere or mitochondrion. It is important to mention that some cluster categories, such as ECM, are divided into five clusters, report distinct groups of genes with potentially different modes of regulation that warrant further investigation.

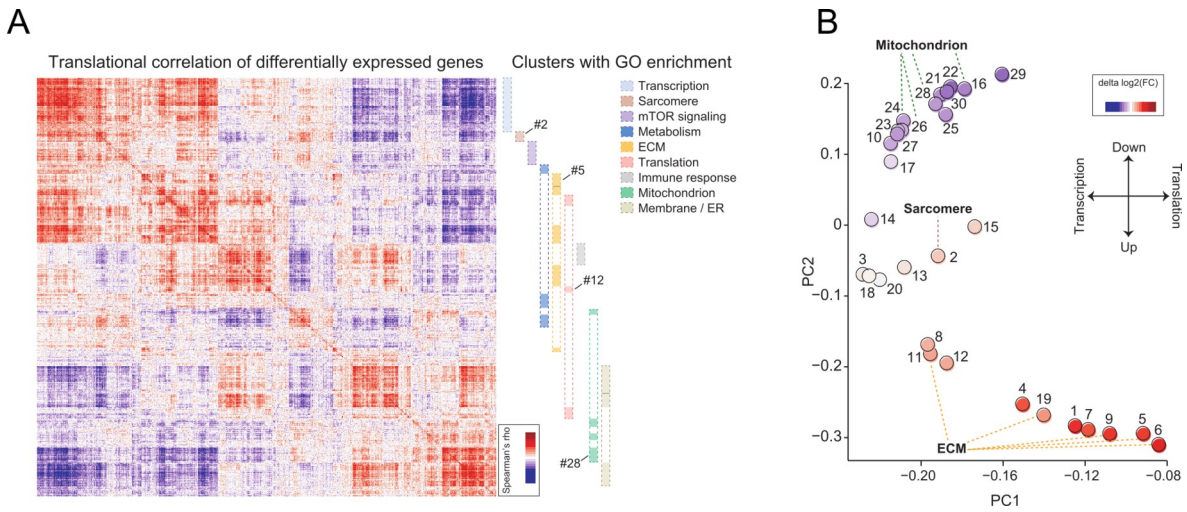


Figure 23: Process-specific regulation.

(A) *Heatmap with hierarchically clustered gene-gene correlation (Ribo-seq) values for all differentially expressed genes (left) and GO enrichment categories for individual clusters (right).* ***(B)*** *Scatter plot of the first two principal components which separates clusters from (A) into transcriptionally and translationally regulated groups.*

Principal-component analysis of transcriptionally and translationally differentially expressed genes together with average fold-change values revealed individual cluster expression contribution (**Figure 23B**). This separation divides clusters into gene sets that are largely up- or downregulated between DCM and unaffected controls or show transcriptional and/or translational regulation.

As mentioned before, the clusters show diverse modes of regulation and involvement in distinct biological processes. Among them, we see examples like four transcriptionally and translationally downregulated mitochondrial clusters, representing mitochondrial dysfunction in the failing heart (Okonko & Shah, 2014). Besides, we find transcriptional upregulation of the sarcomere cluster or translational upregulation of five extra-cellular matrices (ECM) clusters indicating excessive generation of ECM proteins during cardiac fibrosis (L. Li et al., 2018) (**Figure 24**).

Examples in **Figure 24** show little differential behaviour at the transcriptional level but a strong change at the level of translation and would have remained undetected otherwise.

The presented dataset reveals many process-specific pathways that are under translational control. Still, examples like mitochondrial function or sarcomere proteins that are under transcriptional control highlight the importance of both transcriptional and translational expression levels in the heart.

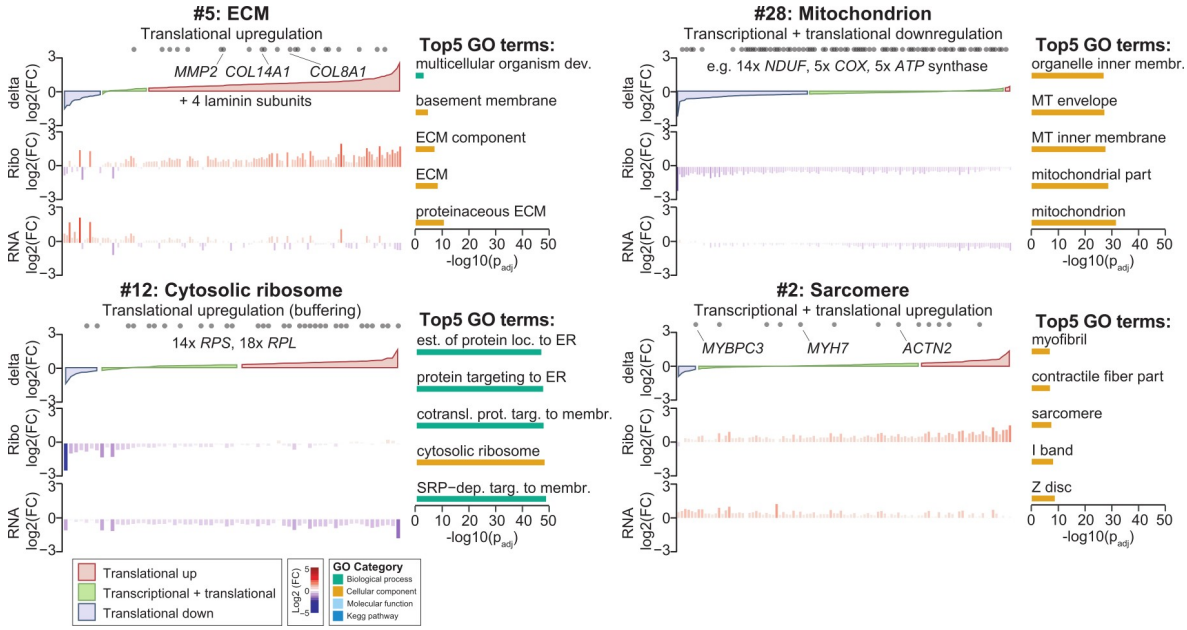


Figure 24: Coregulated gene clusters in the heart.

Bar plot of functionally coregulated clusters. Every bar represents a unique gene, and every sub-panel depicts fold-change (FC) at transcription, translation or the difference between these two. Dots show genes related to the GO terms shown at the top. On the right, the top 5 GO terms and the adjusted p-value for each example are given.

NOTE: The following part of the thesis uses the 80 human left ventricular heart samples from patients with dilated cardiomyopathy (DCM) and unaffected controls to study RBP regulatory roles in the human heart. Due to the time difference between the first and second part of the work, several analysis parts, such as GO enrichment analysis, were done with different release versions, denoted in the corresponding method sections.

RNA-binding protein abundance regulates target mRNA and translational efficiency levels

To understand the potential regulatory role of RBP levels (Ribo-seq) on mRNA abundance and translational efficiency (TE), we collected and reanalysed CLIP-seq derived protein-RNA interactions for 142 ubiquitously expressed RBPs (Van Nostrand et al., 2020) together with the muscle-specific splicing factor RBM20 (Maatz et al., 2014) (**Figure 25A**).

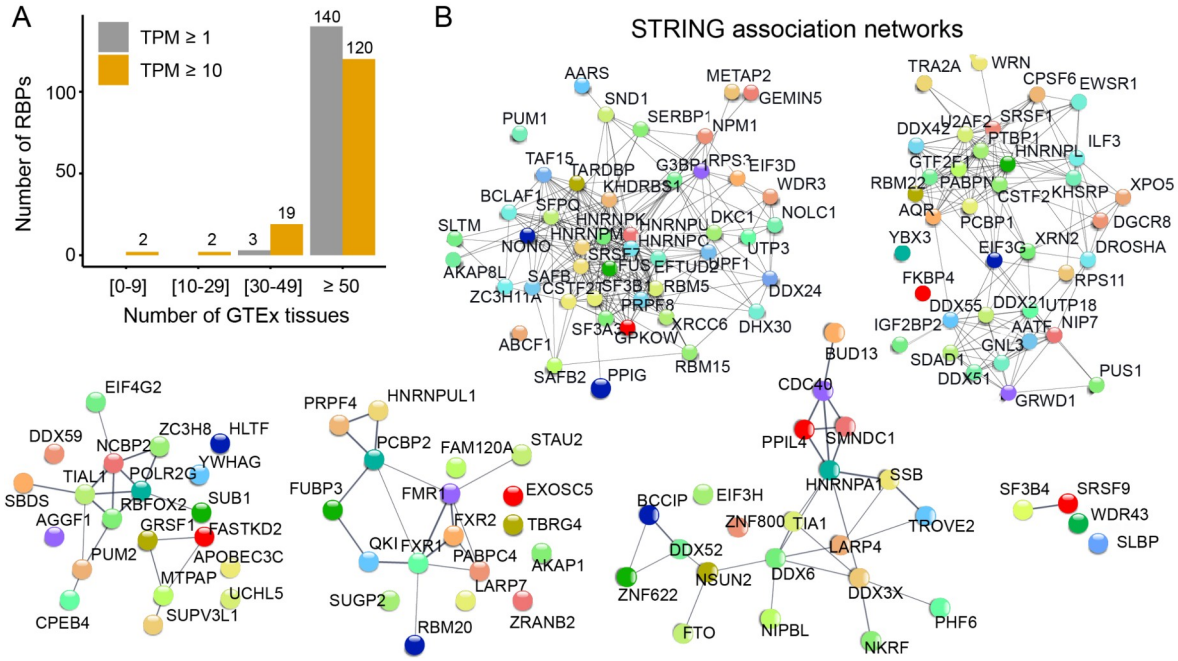


Figure 25: RBP abundance and known interactions.

(A) *Bar plot showing expression of 143 RBPs across all GTEx tissues.* ***(B)*** *STRING protein-protein interactions networks of six coregulated RBP clusters.*

In the first step, we calculated the correlation coefficients of RBPs protein abundance (as measured by Ribo-seq) against translational efficiency levels of all cardiac expressed genes (11,387) across all 80 human heart samples, followed by hierarchical clustering of the resulting correlation matrix. To gain knowledge about RBP with quantitative dependency to target gene translational control, we checked for already known and/or predicted protein-protein interaction among RBP clusters using the STRING database (Szklarczyk et al., 2019) (**Figure 25B**). Indeed, the majority of the proteins show already known or predicted interactions though for some RBPs interactions partners are missing.

In the next step, we calculated the correlation between RBP proteins levels with target mRNA abundance and TE and quantified the frequency of significant associations for each RBP. The association were statistically evaluated by generating 100,000 matched sets of randomly simulated mRNA targets as described previously (Chothani, Schäfer, et al., 2019). The analysis resulted in 58 RBPs that showed significant (empirical $p_{adj} \leq 0.05$) enrichment with target mRNA abundance association (hereafter called “mRNA-RBPs”) and 37 RBPs with target translational efficiency (hereafter called “TE-RBPs”) (**Figure 26A**). Among the significant mRNA-RBPs, we find the known tumor suppressor and splicing factor RBM5 (138 correlating targets; $p_{adj} = 2.83 \times 10^{-5}$; Glass’ $\Delta = 27.2$) (Sutherland et al., 2010), or the non-canonical translation initiation factor

EIF4G2, an RBP that correlates with target gene translational efficiencies (235 correlating targets; $p_{adj} = 5.26 \times 10^{-5}$; Glass' Δ = 6.3) (Liberman et al., 2009; Weber et al., 2021).

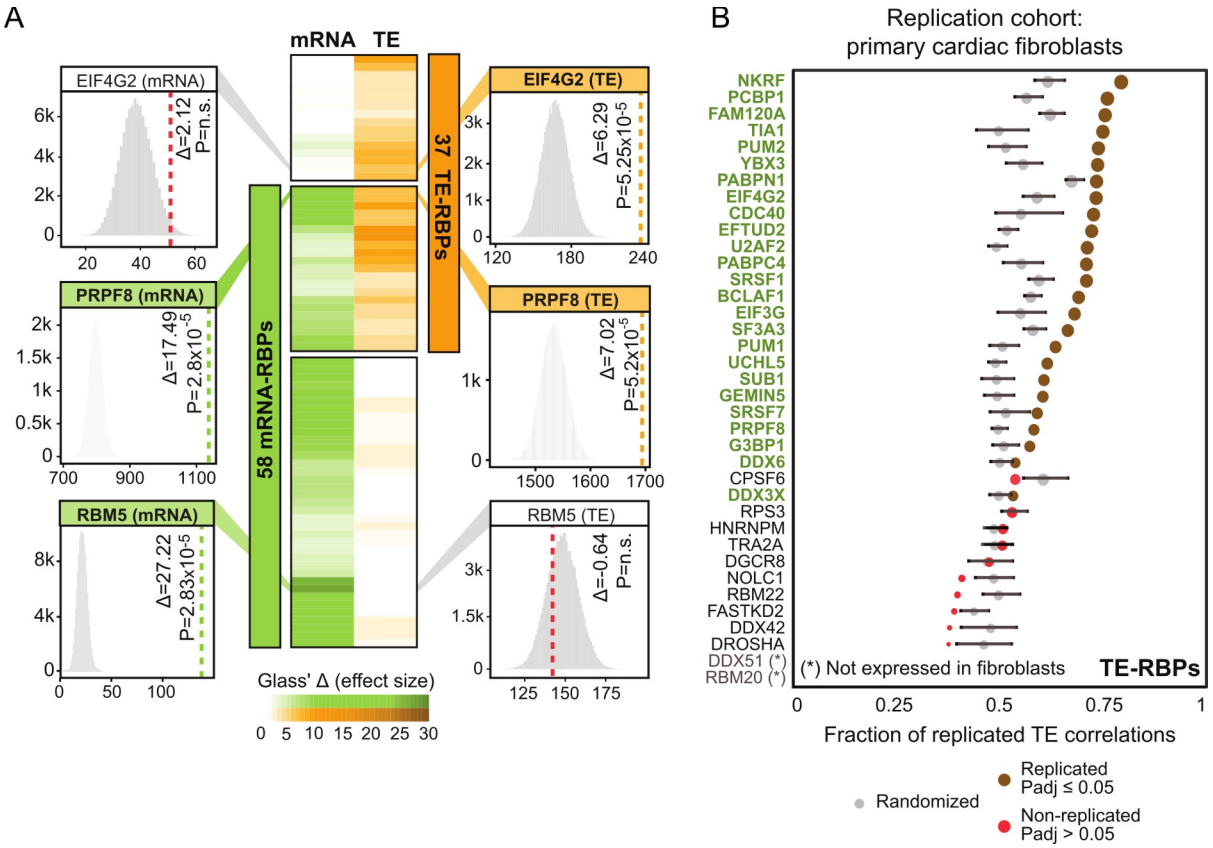


Figure 26: Ubiquitously expressed RBPs predict mRNA abundance and TE.

(A) *Two-column heatmap of Glass' delta scores quantifying the effect size of the RBP-target association in the human heart. For each category, only the significant values are shown. For three selected RBPs, histograms illustrate the theoretical distribution vs. actual observation and the significance of the comparison.* ***(B)*** *Dot plot showing the fraction of RBP-target correlations that can be replicated in an independent cohort of primary cardiac fibroblasts. For each RBP, the fraction replicated correlations on the replication cohort and randomized set is shown in brown (significant) or red (non-significant). The size of the dots indicates significance ($-\log_{10}(p_{adj})$). Grey dots indicate the average fraction of replicated correlations in the randomized set, and error bars show the mean value with standard deviation (SD).*

Reassuringly, 25 of 37 TE-RBP showed reproducible regulation of target translational efficiencies in an independent cohort of primary cardiac fibroblast translatome (n = 20; (Chothani, Schäfer, et al., 2019) (**Figure 26B**).

Bi-directional regulation of translation by known and unknown factors

To find RBPs that coordinately regulate target translational efficiency, we clustered significant correlations of all 37 TE-RBPs and their CLIP-seq targets. This divides RBPs and targets into two groups of opposite directions of regulation (**Figure 27A**), suggesting cooperative or competitive regulation. The two sets of target genes are enriched in mRNA metabolism (p_{adj} = 6.17×10^{-54}) and endoplasmic reticulum ($p_{adj} = 1.82 \times 10^{-7}$).

Example like the nuclear splicing factor U2AF2 and the protease UCHL5 show opposite effects on shared target genes (**Figure 27B**), including *MYL6* and *KPNA4,* whose effect on translation can be replicated independently in the cardiac fibroblast translatome cohort (**Figure 27C)**. For some RBPs (e.g. UCHL5 and FAM120A), interaction has not been described to our knowledge and are not visible in the STRING database (Szklarczyk et al., 2019), thus suggesting potential coordinated models of regulation and interaction that warrant further investigation but also as a resource for studying the impact of RBP abundance on the targets quantitative levels.

We used all 37 TE-RBPs to identify RBP master translational regulators and an internal hierarchy of how translation is coordinated. For this, we generated a weighted topological overlap (wTO) (Gysi et al., 2018; Nowick et al., 2009) network which assigned 5 out of 37 TE-RBPs including four splicing factors; HNRNPM, EFTUD2, U2AF2, SF3A3, and a ribosomal protein (RPS3) as central regulators (**Figure 28**).

Only RPS3 (Dong et al., 2017) and HNRNPM (Ainaoui et al., 2015; T.-M. Chen et al., 2019) have been previously associated with translational regulation but also splicing factors with yet unknown roles. The high number of splicing among the central translational regulators (14 out of 37; 38%) can be explained by the relatively high number of splicing factors in the starting RBPs (49 splicing RBPs out of 143 included RBPs; 34%) set included in this study and do not display any signs of enrichment.

The presented results suggest that ubiquitously expressed splicing factors are involved in translational regulation. Even though the same machinery is not known, this effect may be achieved by the splicing machinery or secondary effects such as the decision of which isoform is being produced.

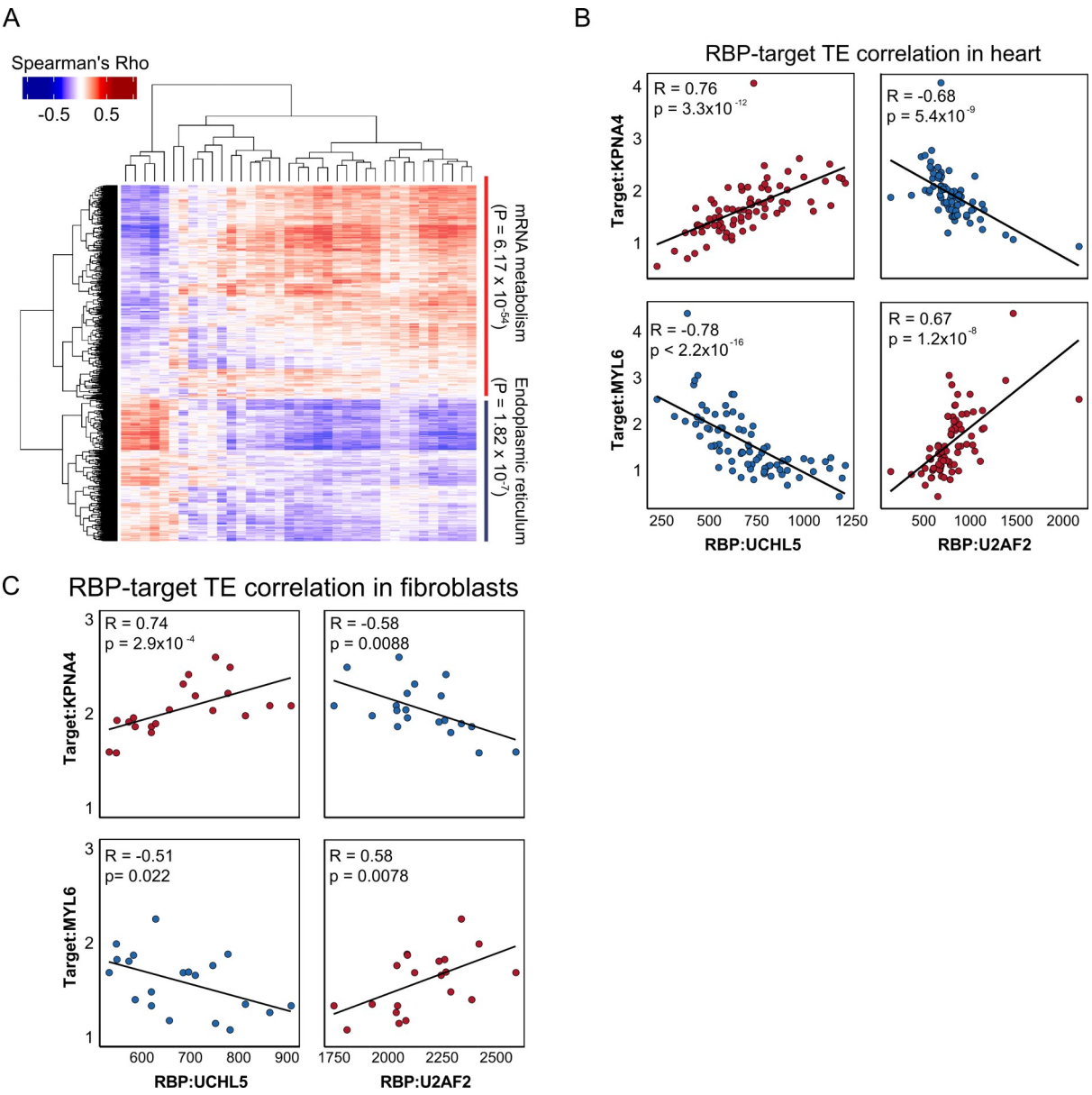


Figure 27: CLIP-seq identifies transitionally regulated targets.

(A) *Heatmap of hierarchically clustered translational efficiency correlations for all 37 TE-RBPs and their CLIP-seq targets. TE-RBP targets genes are separated into two groups with opposite TE effect whose genes show enrichment in mRNA metabolism and endoplasmic reticulum.* ***(B)*** *Scatter plots showing correlation of heart* ***(C)*** *and primary cardiac fibroblast between UCHL5 and U2AF2 and translational efficiencies of two shared target genes KPNA4 and MYL6. UCHL5 and U2AF2 with opposite effects on target genes, indicating competitive regulation reproducible in an independent cohort. Spearman's correlation coefficients and adjusted p-values are displayed.*

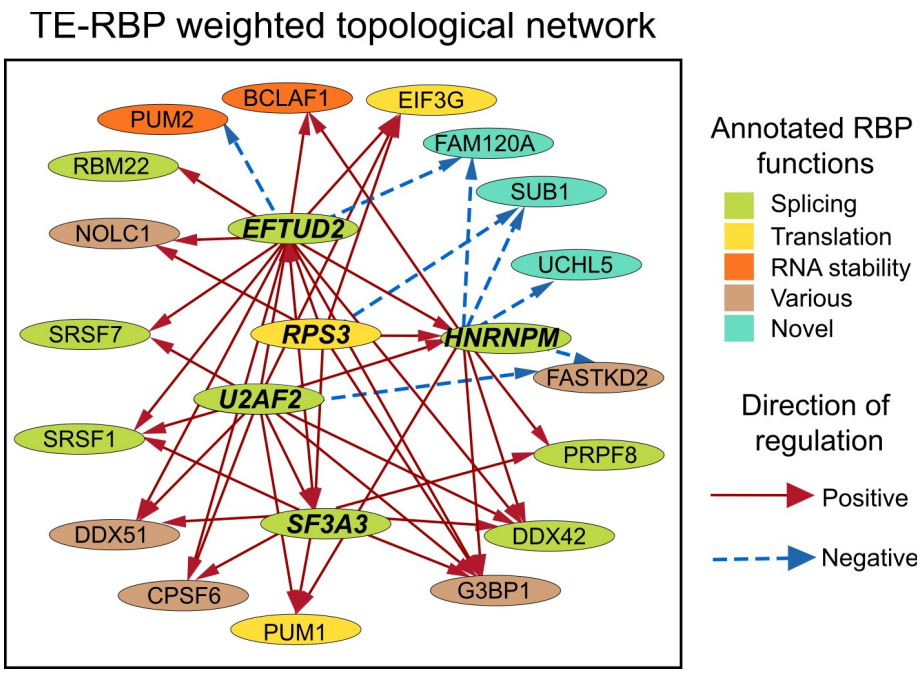


Figure 28: Network of coregulated RBPs.

RBP-RBP network based on weighted topological overlap (wTO) scores with centrally positioned 5 RBPs (EFTUD2, RPS3, U2AF2, HNRNPM, SF3A3) that show robust correlation to at least five target genes.

RBM20 dependent isoform production correlates with translational efficiency

Besides the splicing factors involved in the potentially central role of target regulation, we identified 27 out of 37 TE-RBPs splicing factors that have not been associated with translational regulation yet. Among these 27 splicing factors, our attention was drawn by the disease-relevant and muscle-specific splicing factor RBM20, whose expression significantly correlated with TE of 163 (out of 561 total evaluated targets; Glass' Δ = 7) experimentally validated target genes (**Figure 29A**) but showed no significant impact on mRNA abundance.

RBM20 showed a positive correlation to target genes that primarily consist of sarcomere genes, including *TTN* (Guo et al., 2012; Herman et al., 2012) and *TNNI3K* (Theis et al., 2014; Xi et al., 2015), showing enrichment for muscle function processes (GO:0003012; $p_{adj} \leq 5.97 \times 10^{-16}$) (**Figure 29A** and **29B**). To explore the possible relation between RB20 splicing and the efficiency of mRNA translation, we associated exon splicing rates of known target genes with mRNA translational efficiency. For 66 out of 163 (40.10%) translationally regulated RBM20 target genes, we find percent spliced in (PSI) rates correlating with RBM20

abundance. Among them, well-studied gene *TTN,* whose N2BA-specific I-band specific exons splicing rates negatively correlate with overall TTN TE (**Figure 29C**).

Meaning that I-band specific inclusion has a negative effect, and their removal, such as executed by RBM20, positively affects the overall TTN translational efficiency. We observed this isoform-specific decrease in translational levels previously (van Heesch et al., 2019). The analysis presented here provides a mechanistic explanation of this TTN N2BA specific property which might be a general condition for other sarcomere proteins as well.

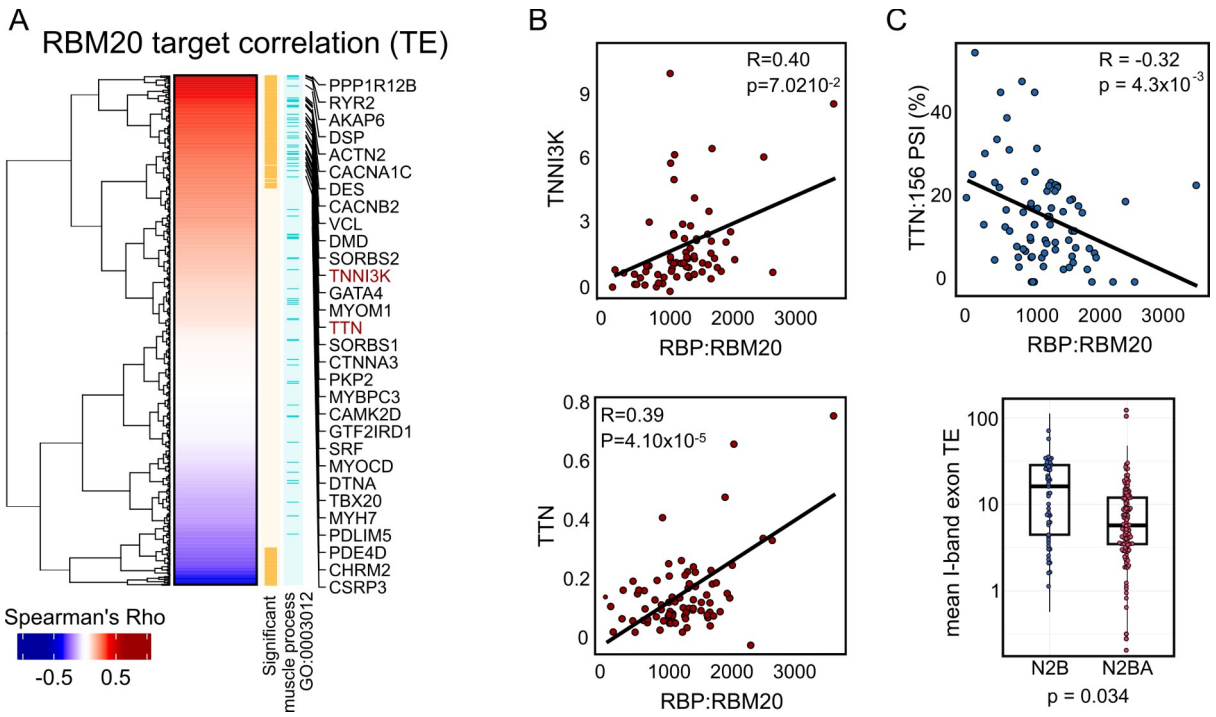


Figure 29: RBM20 as a regulator of translational efficiency.

(A) *Single column heatmap representing hierarchically clustered Spearman's rho correlation values of RBM20 and translational efficiency of its predicted CLIP targets. Significantly correlating targets (n = 163; $p_{adj} \leq 0.05$) and those enriched in the GO term "muscle process" (GO:0003012) are highlighted in orange and blue, respectively.* ***(B)*** *Scatter plots showing the correlation between RBM20 and the translational efficiency levels of two sarcomere genes, TTN and TNNI3K. Correlation coefficients and adjusted p-value are displayed.* ***(C)*** *Scatter plot showing the correlation between RBM20 translation level and TTN:exon156 PSI. Correlation coefficients and the adjusted p-value is displayed (left). Box plot showing the average difference of TTN I-band isoform-specific TEs between TTN isoforms N2B (ENST00000460472) vs. N2BA (ENST00000591111) (p-value = 0.034).*

Our observation showed that RBM20 induced splicing rates have an unexpected impact on translational efficiency potentially generalizable for most muscle-specific RBM20 targets.

The presented results highlight potential capacity splicing factors such as RBM20 to regulate target TE by mutual exclusion of exons with inefficient codon translation rates or exons that negatively impact stability or structure of the transcript, resulting in a different protein composition synthesized (Nott et al., 2004; Zhou et al., 2016).

Dual-function RBPs monitor mRNA abundance and translational efficiency levels by targeting distinct target sets

As mentioned previously, we observed 58 RBPs enriched in target mRNA abundance and 37 with translational efficiency levels, of which 21 could be associated with both traits.
To understand whether this dual association is linked (high mRNA abundance drive increase in TE), we first checked their sets of target genes and found a very limited overlap in target genes between either trait for all 21 RBPs (16.71 ± 8.19%; **Figure 30A**), which decreases for the most strongly correlating targets (1.09 ± 1.49%; r > 0.5; **Figure 30B**). To support our observation, we compared each RBP trait-specific effect size and found no relation between them, such as a carry-over effect from mRNA expression to translational efficiency, supporting our observation that individual RBPs show involvement in the regulation of distinct molecular traits. From now on, these RBPs will be denoted as “dual-function” RBPs, that modulate their function by binding to distinct target genes. One example is multi-functional RBP DDX3X that correlates with mRNA expression of 339 target mRNAs (p_{adj} = 2.83 × 10^{-5}; Glass’ Δ = 6.9) and the TE of 730 target mRNAs (p_{adj} = 5.25 × 10^{-5}; Glass’ Δ = 11.89) but only 43 targets overlap with between sets (**Figure 30A**).

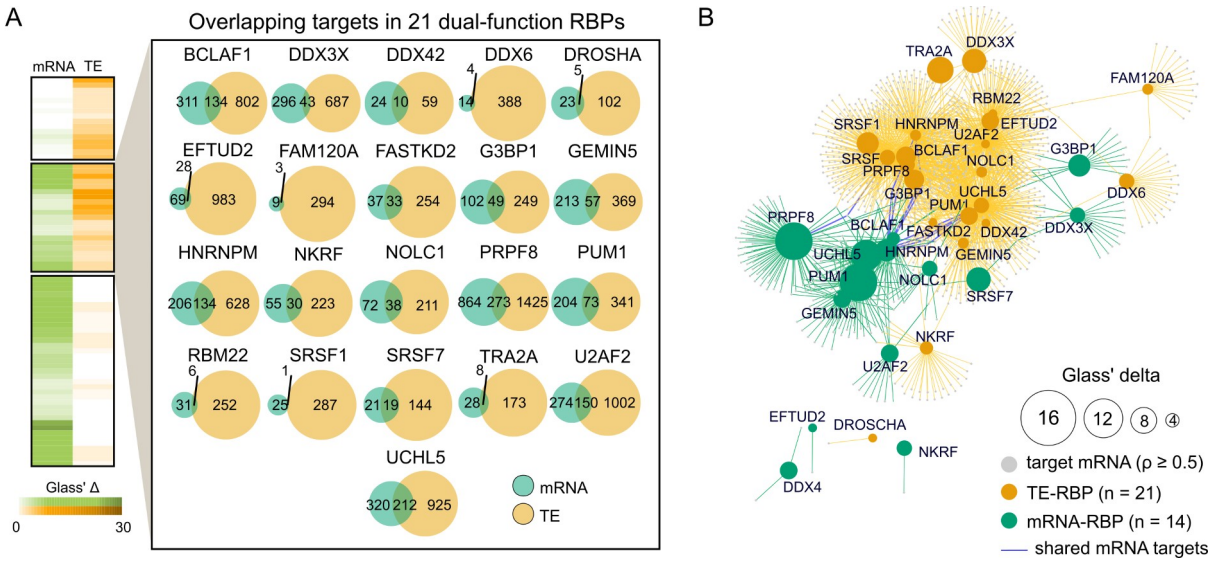


Figure 30: Distinct target sets of dual-function RBPs.

(A) *Two-column heatmap of Glass' delta scores quantifying the effect size of the RBP-target association in the human heart. For the set of 21 RBPs involved in both mRNA abundance and TE regulation, individual overlaps of target genes are shown as a Venn Diagram with an absolute number of significantly correlating target genes.* ***(B)*** *Network displaying dual-function RBPs and their target interaction of both mRNA-RBPs (green) and TE-RBPs (brown) for strong (r ≥ 0.5) correlating targets.*

Interestingly, we observed opposite regulatory behaviour of DDX3X on mRNA targets (positive correlation) and TE (negative correlation). Besides DDX3X, three other RBPs (DDX6, NKRF, GEMIN5) showed an overall negative effect on translation but a positive effect on mRNA abundance and one RBP with the exact opposite behaviour (FAM120A). The remaining RBPs show concordant regulation on their target genes at both traits.

Furthermore, mRNA abundance and TE targets of dual-function RBPs show enrichment in different biological processes, resulting in entirely different biological outcomes. For example, the dual-function RBPs DDX3X and UCHL5 TE targets show enrichment for genes involved in RNA splicing (GO:0008380; $p_{adj} = 7.70 \times 10^{-30}$) but no clear enrichment among the mRNA targets.

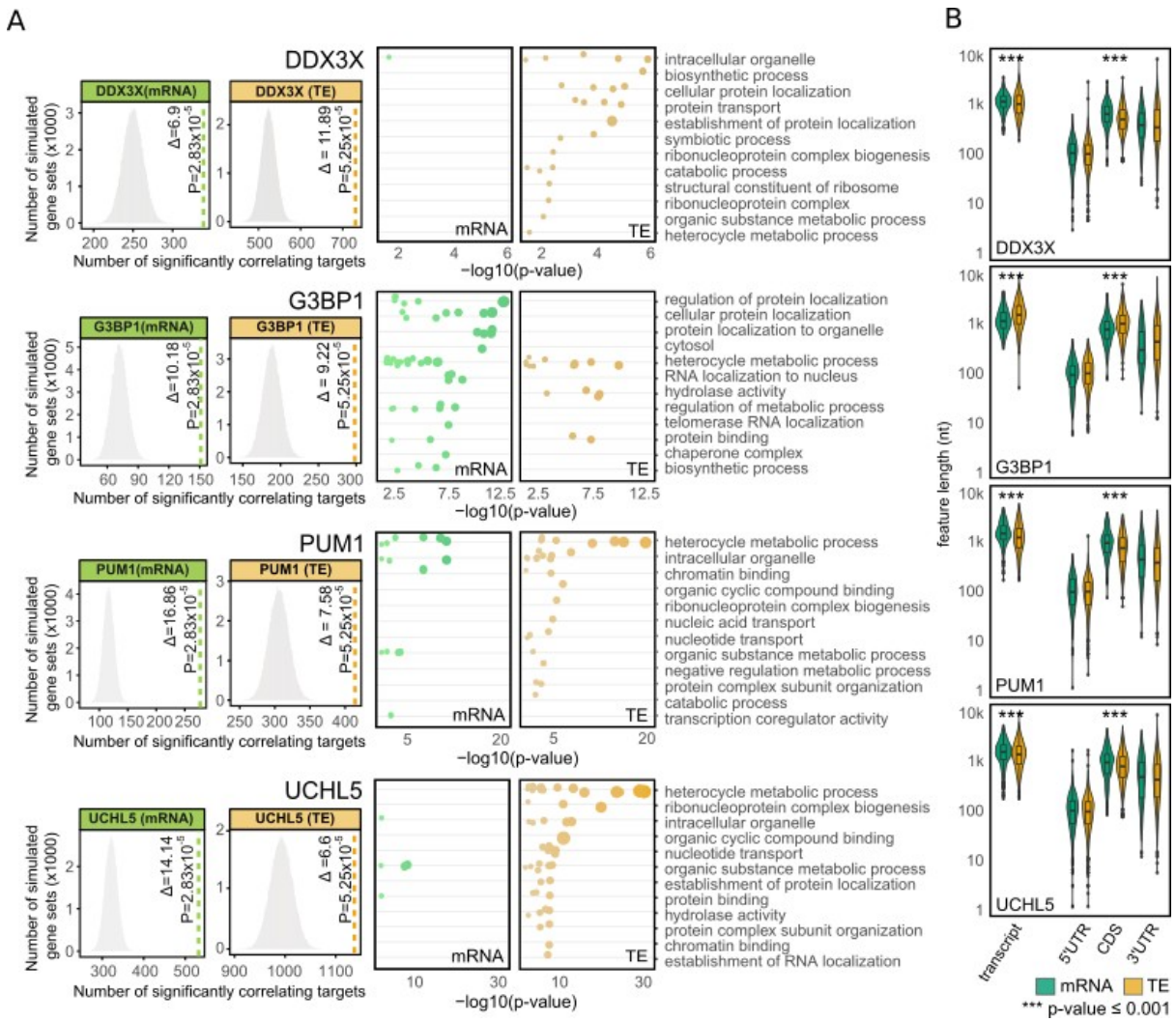


Figure 31: Dual-function RBPs regulate distinct sets of target genes.

(A) *For four selected dual-function RBP examples (DDX3X, G3BP1, PUM1, and UCHL5), histograms showing the significance of RBP-target correlations and dot plot with the 12 most significant parental GO terms are displayed for both traits (mRNA and TE) are displayed.* ***(B)*** *Box plots showing the difference between transcript, 5' UTR, CDS, and 3' UTR sequence length in nucleotides for mRNA vs. TE of corresponding targets are displayed.*

Another dual-function RBP is G3BP1, whose mRNA targets show enrichment for genes involved in localization to the nuclear body (GO:1903405; $p_{adj} = 5.13 \times 10^{-12}$), whereas TE targets encode for proteins involved in RNA splicing (GO:0008380; $p_{adj} = 1.61 \times 10^{-10}$). Among them, we also find examples like PUM1 that show for both molecular traits enrichment in targets coding for proteins involved in mRNA processing (GO:0006397, p_{adj} TE $= 2.64 \times 10^{-22}$ and p_{adj} mRNA $= 6.40 \times 10^{-13}$) (**Figure 31A**).

Dual-functionality achieved through affinity for CDS length and 5' UTR structure

Besides changes in RBPs interactome, dual-functionality can be achieved through context-specific differences in subcellular localization (Buchan, 2014), interaction partners (Cirillo et al., 2020; W. Yang et al., 2019), or the presence of multiple RNA binding domains (Müller-McNicoll & Neugebauer, 2013) by modulating the set of recognized target genes. Based on published immunofluorescence imaging-based evidence of subcellular RBP localization (Van Nostrand et al., 2020), 13 out of 21 dual-function RBPs show localization to both nucleus and cytosol, suggesting localization-dependent function. In the next step, we looked at the relative binding position of CLIP binding sites within a target mRNA (i.e. the position of binding within the mRNA: 5' UTR, coding sequence, 3' UTR or intronic). We could not identify great differences in binding position between both traits. Though for DDX3X and eight other RBPs (DROSHA, FASTKD2, G3BP1, GEMIN5, PRPF8, PUM1, U2AF2, and UCHL5), we observed a significant change in target transcript CDS length, either increasing or decreasing between mRNA and TE targets (**Figure 31B**).

The strongest difference in CDS length we observed for GEMIN5 (decrease for TE targets; 2,226nt vs. 1,519nt; $p_{adj}= 3.66 \times 10^{-9}$) length were seen for GEMIN5 (decrease for TE targets; 2,226nt vs. 1,519nt; $p_{adj}= 3.66 \times 10^{-9}$), PRPF8 (decrease for TE targets; 2,243nt vs. 2,076nt; $p_{adj}= 1.03 \times 10^{-8}$), DDX3X (decrease for TE targets; 1,659nt vs. 1,376nt; $p_{adj}= 2.81 \times 10^{-7}$) and G3BP1 (increase for TE targets; 1,985 vs. 2,798nt; $p_{adj}= 6.11 \times 10^{-8}$).

RBPs like DEAD-box helicase eIF4A (as part of eIF4F complex) (Svitkin et al., 2001) and DDX3X (Calviello et al., 2021; Guenther et al., 2018; Sen et al., 2015; Soto-Rifo et al., 2012) are involved in translational initiation by binding to highly structured transcript leader sequences and subsequently disentangling them. A recently published study showed that DDX3X binds to 5' UTRs and the small ribosomal subunit to initiate translation of the mRNA with a long and structured leader sequence (Calviello et al., 2021). We set out to systematically identify all RBPs involved in translating transcripts with complex 5' UTR sequence by calculating the 5' UTR minimum free energy (MFE) of TE and mRNA targets of all dual-function RBPs. For 17/21 dual-function RBPs, we observed significant differences between positively and negatively correlating targets in structural composition (**Figure 32A**) and weak significance for correlating mRNA targets of the same RBPs.

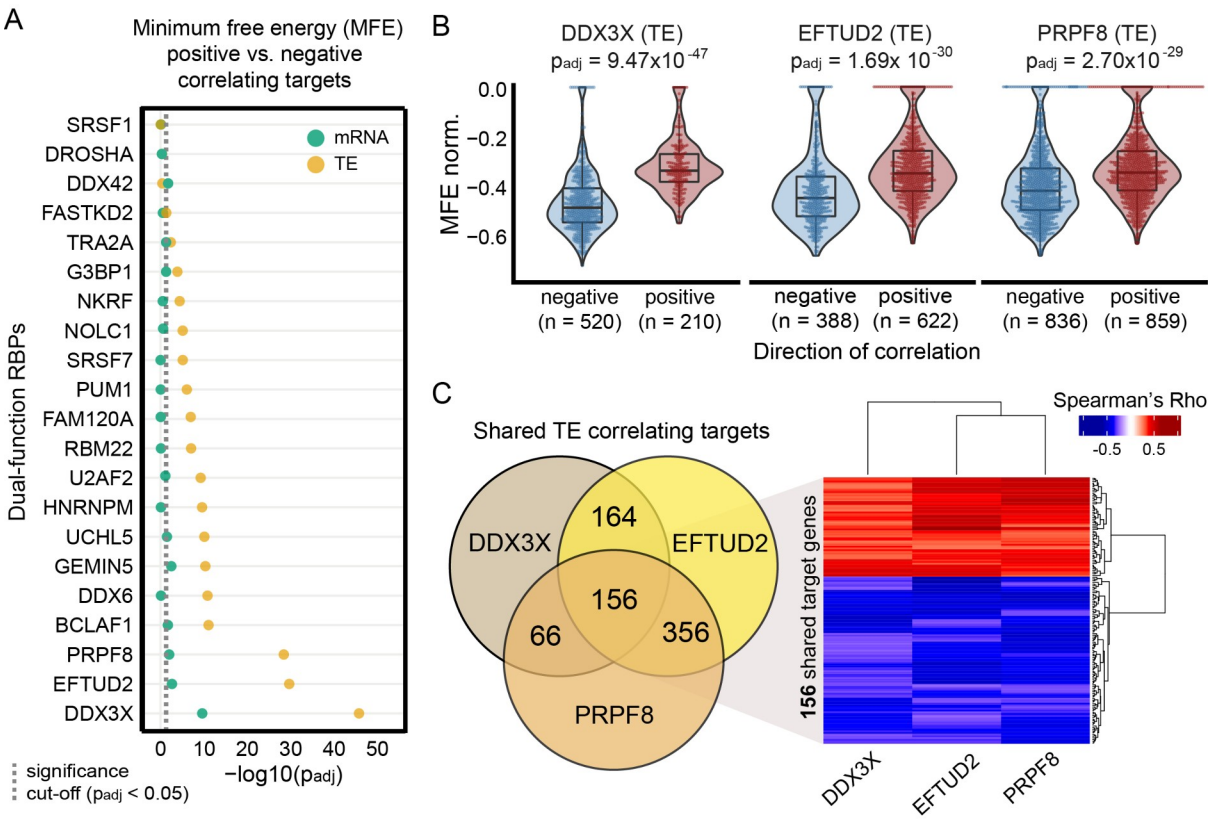


Figure 32: Differential affinity of dual-function RBPs for 5' UTR structures.

(A) *Dot plot displaying the significance of differences in length normalized minimum free energy (MFE) of 5' UTR sequence between positively and negatively correlating target genes for each dual-function RBP. Significance was calculated separately for mRNA (green) and TE (brown) targets. Adjusted p-values are shown on -log$_{10}$ scale, and only 5' UTR sequences longer than 20 nucleotides were evaluated.* ***(B)*** *Box and violin plots for the three most significant dual-function RBPs (DDX3X, EFTUD2, PRPF8) (Figure 23A) displaying the difference between positively and negatively correlating TE targets.* ***(C)*** *Three way Venn Diagram showing the overlap between the selected dual-function RBPs in absolute numbers (left). Heatmap representing TE correlations of 156 shared target genes for the selected dual-function RBPs (right).*

Three TE-RBPs showed by far the strongest difference between positively and negatively correlating targets: as expected DDX3X (p_{adj} = 9.47 × 10^{-47}) and core spliceosome splicing factors PRPF8 (p_{adj} = 2.70 × 10^{-29}) and EFTUD2 (p_{adj} = 1.69 × 10^{-30}) (**Figure 32A** and **32B**).
The 156 targets shared between the three RBPs showed identical correlation directions and encoded for proteins involved in mRNA stabilization (GO:0048255; p_{adj}= 3.09 × 10^{-4}) **(Figure 32C)**. We identified 21 dual-function RBPs involved in target mRNA abundance and translational efficiency regulation through distinct sets of target genes, resulting in different biological consequences but the involvement of similar binding motifs.

Discussion

In this thesis, I presented the first genome-wide investigation of translational information in human cardiac tissue, provided by ribosome profiling. The work focused on the biogenesis of novel microproteins from long noncoding RNAs resulting in the discovery of 339 proteins encoded by 169 lncRNAs of which a large fraction can be replicated in kidney and liver tissues. Surprisingly, dozens of microproteins suggest a mitochondria related function.
Furthermore, I observed extensive translational regulation on cardiac expressed genes involved in biological process that represent hallmarks of cardiac disease. The data set enables a large resource to study translation and translational regulation such as the multifunctional role of RNA-binding proteins, presented in this work.
In the following sections, I will discuss these findings and their implication on our understanding of cardiac biology and translational activity in human heart tissue.

Ribosome profiling: a tool for detection of novel proteins

To identify novel translation events, we, for the first time, applied ribosome profiling (Ribo-seq) in human cardiac tissue. The collected data allowed an unbiased search for novel cardiac expressed proteins without strong nucleotide conservation or minimum ORF length, as has been applied previously to annotate lncRNAs across species (Cabili et al., 2011) or their function in mouse models (Guttman et al., 2009). A potential reason for using conservation approaches to identify novel proteins was simply the absence of a technique reliably providing evidence *in vivo* which is now aided by Ribo-seq and the belief that protein-coding genes must show sequence conservation across species as an indicator of functionality.
The amount of identified translation events largely depends on the gene expression threshold defined within mRNA-seq. In this study, an expressed gene is required to show an average FPKM expression ≥ 1 across all samples of the respective cohort. It has been shown that active translation affects mRNA stability (Herzog et al., 2017; Wu et al., 2019) showing that high expression is not necessarily required for sufficient protein production. Consequently, stable mRNAs with low expression but high ribosome occupancy and high translational efficiency can produce large number of proteins. This likely happens in rare cardiac cell types that can't be captured with bulk ribosome profiling but rather on a single cell level (Litviňuková et al., 2020). Even if ribosome profiling would be applied to single cells, low footprint coverage on individual genes would not allow detection of active translation determined by 3nt periodicity. High coverage in ribosome profiling data across mRNAs is essential to assign 3nt periodicity.

Another circumstance impacting detection of translation events among lncRNAs and protein-coding genes are the used thresholds. Due to the nature of the Ribo-seq protocol, we get a snapshot of active translation within the tissue or cell line and depending on the translational efficiency, we might see a sparse 3nt signal for the ORF of interest. The thresholds applied on the spare translation signals can have multiple distinct outcomes, such as alternative downstream start codon or poorly covered translation termination sites, resulting in insignificant periodicity. Our conservative thresholds aimed to detect robustly translated transcripts and potentially skipped those that slightly miss the requirements due, e.g. too low coverage in a subset of samples.

Nucleotide conservation across species or identification of homolog protein sequences can provide evidence of preserved functionality. We investigated nucleotide conservation across 49 mammals using a PhyloCSF based pipeline and genomic positional and translation initiation conservation between human, mouse, and rat species for all detected lncRNA ORFs. Compared to canonical mRNAs, the majority of the translated lncRNAs reported in this work show limited nucleotide conservation across species. We instead identified conservation limited to primates or no conservation at all, suggesting production of evolutionary young proteins (Ruiz-Orera & Albà, 2019) evolved particularly within the primates or are human-specific. Of course, translation of short open reading frames does not have to result in functional proteins; the translation itself might fulfil a regulatory role such as has been observed with uORFs that frequently repress translation of the downstream main CDS (L. Jia et al., 2020; Johnstone et al., 2016). Upon perturbation of uORF-mediated translation, a wide range of human diseases, covering metabolic or neurologic disorders (Barbosa et al., 2013) have been observed. The act of translation as a regulatory process likely explains the poor sequence conservation of uORFs across species. Identified translated lncRNAs presented here largely lack sequence conservation but are detectable at subcellular level and theier presence can be observed both *in vitro* and *in vivo*, suggesting a stable translational product with potential biological function in the human heart.

Apart from Ribo-seq translation evidence, we were able to validate peptide production for a significant fraction of tested micropeptides *in vitro* using in vitro transcription (IVT) and translation assays. Along with ATG KOs, we got another layer of evidence that lncRNAs form stable proteins whose synthesis can be disturbed by targeting the translation initiation site. To show that these microproteins are stable and detectable *in vivo*, we set out to search them in deep mass spectrometry datasets. Identifying short proteins is problematic in the cardiac tissue due to the high dynamic range of sarcomere proteins in the sample. Also short digested peptides (< 5 amino acids) are not detectable, on the other hand long peptides are hard to detect as well (Bazzini, Johnstone, Christiano, Mackowiak, et al., 2014), although longer

peptides can be additionally digested to overcome this problem. In targeted proteomics high abundant proteins make it nearly impossible to detect lowly expressed ones (Bekker-Jensen et al., 2017), or the high false-discovery rate makes it almost impossible to confirm them (Ezkurdia et al., 2014; Savitski et al., 2015; The et al., 2016). Using the deepest heart shotgun MS dataset to date and a newly generated deep proteome of human iPSC-CM, we were able to find unique evidence for 40.20% of all proteins identified by RiboTaper (**Figure 33**), highlighting the problem of protein detection *in vivo* using e.g. mass spectrometry based methods. Other studies also reported low detection rates using mass spectrometry (MS) methods for potentially translated lncRNAs. Bazzini and colleagues observed, in zebrafish, MS evidence for 98/302 (~32%) for already annotated short ORFs (Bazzini, Johnstone, Christiano, MacKowiak, et al., 2014) and even lower rates were reported by Mackowiak and colleagues (Mackowiak et al., 2015), although this study focused on conserved sORFs across different species making direct comparison difficult.

Among the tested 339 microprotein sequences, we were able to validate 41%; this fraction increased to 55% by using high-throughput selected reaction monitoring (SRM) assay for a randomly selected set of candidates in five independent heart samples. Even though the SRM technique allows to train the mass spectrometer for the spectra of individual micropeptides that were synthetically generated and search only for those within the sample ignoring other proteins, we could identify a bit more than half of the tested micropeptides using this technique.

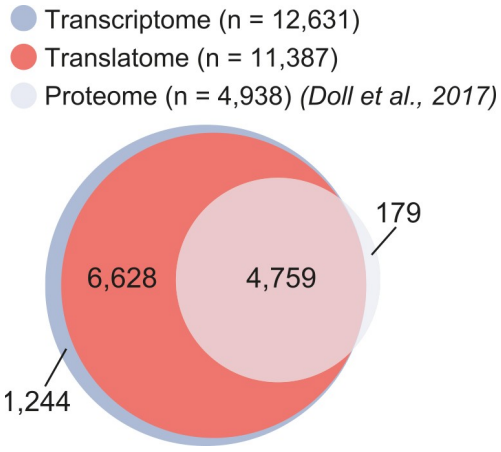


Figure 33: Actively translated genes.

Venn diagram of detected genes in the cardiac transcriptome, translatome and proteome.

Robust mass spectrometry peptide detection is complicated and is constantly improved (Aebersold & Mann, 2016), meaning that micropeptides produced by translated lncRNAs underlie a high false-negative rate due to their short length and low abundance compared to canonical proteins. As we have shown, a combination of multiple validation techniques such as ribosome profiling, IVT assays, high-throughput MS and SRM is needed to get peptide evidence for the majority of tested candidates. Particularly mass spectrometry approaches, which are currently the gold standard to show protein abundance *in vivo*, suffers from high false-positive rates or cannot capture the complete proteome of the sample of interest. Thus a combination of multiple methods seems to be helpful to capture even low abundant proteins. Finally, the development of new high sensitivity mass spectrometry protocols will help us in detection of all cell or tissue residing proteins.

Another source of novel microproteins presented in this work is the class of circular RNAs (circRNAs) that are known for a long time but drew attention over the last few years (Kristensen et al., 2019). Their biogenesis through non-canonical backsplice junctions formation and circular structure lacking polyA site is detectable in, e.g. totRNA (rRNA-) datasets. However, their detection rate relies on reads mapping to the backsplice junctions (Zeng et al., 2017). We used RNase R to show that circRNAs have increased resistance to exonuclease treatment (Jeck et al., 2013) and thus form stable transcripts.

Selected efforts have shown that circRNAs can serve as templates for translation (C. Y. Chen & Sarnow, 1995; Ho-Xuan et al., 2020; Legnini et al., 2017; Pamudurti et al., 2017) and even form an infinite ORF with efficient translation in a rolling cycle (Abe et al., 2015). Using ribosomal footprints that map to the backsplice junctions, we identified a set of 40 potentially translated circRNAs with at least three unique and, in total, five reads mapping to the junction. For 6 out of 40 potentially translated circRNAs, we observed peptide evidence in the same heart shotgun MS dataset used for the micropeptide detection. Additional w*et-lab* work would be required to identify the precise exons composition that forms a mature circRNA (Memczak et al., 2013), which can be used to predict the complete ORF in case of ribosome occupancy. However, a high-throughput method to access circRNA structure is still missing. Thus, our MS search of potentially translated circRNAs is restricted to a relatively short area of the backsplice junction resulting in one or two short tryptic peptides. As discussed before, short tryptic peptides or its amino acid composition might result in high false negative detection rates. Even though we observed ribosome occupancy, it was much lower than for canonical ORFs and suggests low protein abundance, close to noise, during *in vivo* validation. Finally, alternative FDR correction methods are needed to ensure signal conclusions (Hansen, 2021) and avoid current mass-spec estimates which are under debate (Danilova et al., 2019; Gupta et al., 2011).

Alternative library protocols aiming to capture unique circRNA features are required to obtain circRNA transcripts free from linear counterparts of the same host gene. Such information would allow assembly of a circRNA specific transcriptome and ease downstream analysis, such as identification of translation events.

Functional characterization of novel microproteins

A remarkable fraction of translated lncRNAs have well-characterized noncoding functions. A striking example is the well-characterized lncRNA *UPPERHAND* (*HAND2-AS1*). The lncRNA shows positional conservation and hosts several short open reading frames with Ribo-seq evidence. The transcriptional process of *Upperhand* in mouse has been shown to regulated its sense protein-coding gene *HAND2* in *cis*, while the mature transcript with cytosolic localization did not show any biological role (K. M. Anderson et al., 2016). However, others have associated its expression with cancer (Gu et al., 2021). These examples show that lncRNAs have the potential to act as independent non-coding molecules or encode for microproteins, though its proportion among all lncRNAs differs across studies (Guttman et al., 2013; Ingolia et al., 2014). Compared to our study, the difference may be explained by modifications in the library protocol, the definition of lncRNA or treatment with harringtonine as has been done frequently in previous studies (Guttman et al., 2013; Ingolia et al., 2011, 2014; Stern-Ginossar et al., 2012). The hypothesis is supported by several canonical protein-coding genes that produce transcripts with noncoding functions. A prominent example is the TP53 protein, a known tumor suppressor that shows interrelated regulatory roles of both the protein and noncoding RNA (Candeias, 2011) and several other examples (Ivanyi-Nagy et al., 2018; Jenny et al., 2006). Interestingly, the additional noncoding functions of mRNAs have been frequently discovered among 3' UTRs, mostly involved in cancer repression (Fan et al., 1996; Manjeshwar et al., 2003; Rastinejad et al., 1993). Combining the presented results and public investigation, we can conclude that traditional categorization in coding and noncoding genes may come up short and may require classification based on different methods.

Additionally, we predicted lncRNA function by looking at their co-regulated protein-coding genes and their known biological function. We discovered groups of lncRNAs with shared functions such as catabolic processes, nuclear processes, including transcription or DNA binding and/or fatty acid oxidation. The largest group comprising 22 lncRNAs showed coregulation with genes essential for mitochondrial function.

Considering the literature to date, the relatively small mitochondria seem to be an optimal cellular organelle to use small proteins to fulfil its role as an energy factory. Among these

microproteins, we find MOXI, a microprotein involved fatty acid ß-oxidation (Makarewich et al., 2018), MIEF1 involved in mitochondrial translation (Rathore et al., 2018), and mitoregulin, which is involved in protein complex assembly, membrane potential, respiration rates, and CA2+retention (Stein et al., 2018). The microprotein BRAWNIN has been suggested as essential for mitochondrial complex III assembly (S. Zhang et al., 2020), and Mitolamban regulates respiratory complex III function (Makarewich, 2020). Many detected microproteins across species seem to be membrane-bound proteins with specific biological functions (Makarewich, 2020).

Translational regulation of the failing human heart

The large human cohort consists of 80 left ventricular samples that allow us to disentangle transcriptional and translational regulation of the end-stage and healthy hearts for the first time. To overcome differences between RNA-seq and Ribo-seq, the read length of both sets was matched and jointly normalized to achieve equal library sizes. The combined normalization ensures comparison between RNA-seq and Ribo-seq fold changes in downstream analysis. Comparing both diseased and unaffected samples, I was able to identify 4,344 genes differentially expressed at both transcriptional and/or translational levels. By looking only at transcriptional information, we would have missed genes with differential behaviour appearing at the translational level, making this dataset suitable to study active translation without relying on mass spectrometry data that suffers from a poor detection rate of lowly abundant proteins.

Coregulation analysis for all differentially expressed genes resulted in clusters of genes with similar patterns of regulation. We used the fold changes of both quantitative traits (RNA-seq and Ribo-seq) to calculate a delta fold change between those obtaining a measure for the overall direction of regulation. We see clusters with various directions of regulation. For example, five mitochondrial clusters are downregulated at transcription, and for some of them, this effect is enhanced at the translational level. This observation supports our knowledge about mitochondrial dysfunction in DCM. During the onset of DCM, the number of mitochondria in cardiomyocytes increases to compensate for higher energy demand (Rosca & Hoppel, 2010). During the progression of cardiac stress in DCM patients, autophagic flux, responsible for removing dysfunctional mitochondria, causes accumulation of damaged mitochondria leading to cardiomyocyte apoptosis (Campos et al., 2016), an indicator of DCM.

Many of the sarcomere proteins show an upregulated transcriptional basis that is even enhanced at translation. Furthermore, we identified clusters enriched in genes encoding ECM proteins with weak up or downregulation of genes on the transcriptional but strong upregulation

on the translational level. Upon cardiomyocyte death, the heart initiates overproduction of collagen and other ECM proteins to maintain cardiac structure and function. This behaviour is recognized as a hallmark in failing DCM-hearts (Schultheiss et al., 2019; Travers et al., 2016) and serves as an explanation of the observed translational upregulation of genes encoding ECM proteins. The absence of transcriptional upregulation for these clusters suggests that post-transcriptional regulation is likely responsible for higher translational efficiency. Our understanding of this regulation is sparse and warrants further investigation using similar approaches in model organisms such as the rat (Schafer, Adami, et al., 2015) but with sufficient power or large human cohorts as presented in this work.

This work presents a dataset that allows investigation of cardiac biology at the translational level, identifying sets of genes that cooperatively or competitively shape the cardiac landscape. It also allows studying complex mechanisms that are regulated at the translational level, with sufficient power. One example is the mechanistic target of the rapamycin (mTOR) pathway, a master regulator of multiple cellular processes, including translational regulation. Our understanding of mTOR driven translational program is incomplete, and multiple efforts have been looking at uniform cell lines (Hsieh et al., 2012; J.-J. Jia et al., 2021; Thoreen et al., 2012). In the human heart, the mTOR cluster shows both transcriptional and translational upregulation. Among them, we find significantly upregulated translational efficiency levels of 5' terminal oligopyrimidine (TOP) motif-containing mTOR target genes ($p_{adj} = 2.52 \times 10^{-7}$) with a downregulated transcriptional basis. This cohort and particularly the nucleotide resolution provided by ribosome profiling, provides a tool to study mTOR regulation at an unprecedented level.

RNA-binding protein abundance is predictive of target mRNA levels and translational efficiencies

RNA-binding proteins (RBPs) are known to be essential regulators for a wide range of cellular processes such as RNA transcription, splicing, translation, and many more (Gerstberger et al., 2014). However, increasing evidence suggests that RBPs can have multiple functions that regulate gene expression, depending on their localization within different subcellular compartments (Backlund et al., 2020). Our *in silico* pipeline uses genome-wide coregulation analysis to identify RBP-driven mRNA abundance and translational efficiency (TE) regulation of target genes in the human heart. Out of the 143 expressed RBPs, we assigned 74 RBPs to these quantitative traits. We find RBPs with already known function in mRNA abundance of TE level regulation and assign yet uncharacterized function for other RBPs. The mRNA-RBP RBM5

is a known nuclear factor involved in splicing tumor suppression targets including p53 (Jamsai et al., 2017). The TE-RBP EIFG2 is a known protein of the translation initiation complex inducing cap-independent, IRES-driven translation (Liberman et al., 2009). Recently, it has been discovered that EIFG2 can re-initiate the main CDS following uORF translation, though the exact mechanism is unknown (Weber et al., 2021).

The 37 predicted TE-RBPs can be clustered correlation-wise in groups showing either positive or negative direction of regulation to their target mRNAs. It is important to mention that many of them form sets of RBPs sharing at least three mRNA targets. We find examples like the poorly characterized UCHL5, a protease that show opposite effects than the splicing factor U2AF2 on selected targets. This behaviour is reproducible in an independent cohort of primary cardiac fibroblasts suggesting a general complex translational control that likely holds for other RBP groups with competing or cooperative effects.

Future studies need to answer how RBPs are internally regulated and identify master regulator(s), which will be very difficult because of the wide range of RBP classes known to date. With the development of new technologies, a fine-tuned functional roles of RBPs might be elucidated, resulting in a more complex interaction network with multiple regulatory RBP hubs.

Our approach used CLIP-seq information to build a directional RBP-RBP network based on correlation, and a weighted score derived using permutation analysis. We identify five RBPs (HNRNPM, EFTUD2, U2AF2, SF3A3, RPS3) as central regulators of cardiac translation. However, this approach does not consider the amount of non-RBP target mRNAs or the effect size of the individual RBPs. Additional information might help to form a robust network with precise master regulator(s) for the investigated set.

Among the 37 predicted TE-RBPs, we see many splicing factors, which can be explained by the relatively high number of splicing factors within the initial set. Even though RBPs like U2AF2 show exclusive nuclear localization, our pipeline predicted association with TE levels regulation but also some well-characterized splicing factors involved in mRNA translation such as HNRNPM (T.-M. Chen et al., 2019), SRSF1 (Maslon et al., 2014), or SRSF3 (J. Kim et al., 2014). Additional experiments are needed to validate their impact on translation in the human heart, though applying CLIP-seq experiments on post-mortem cardiac tissue is very complex. Furthermore, all RBPs included in this work are ubiquitously expressed across a wide range of tissues and thus likely act as general transcriptional and translational regulators also in the human heart. As discussed previously, CLIP signal is preserved among similarly expressed genes of the same RBP independent of the cell line, and peak differences instead reflect cell type-specific expression than binding affinity (Van Nostrand et al., 2020). Taken together, the predicted transcriptional and translational role likely persist across a wide range of tissues and cell lines with differences specific to the tissue/cell line, as discussed next.

RBM20 mediated isoform production switch defines translational efficiency of target genes

One of the RBPs with novel function is the muscle-specific and disease-relevant TE-RBP RBM20. We have shown its abundance positively correlates with the TE levels but not mRNA abundance of mostly sarcomere genes. Even though the act of splicing occurs in the nucleus, we see an isoform-dependent correlation with translational efficiency, even though translation is a cytosolic process. The most prominent sarcomeric target is *TTN,* which consists of four bands, of which the I-Band is spliced out to switch between N2B and N2BA isoforms (Guo et al., 2012). While N2BA is more elastic due to the inclusion of the I-Band, N2B is shorter and thus stiffer (Freiburg et al., 2000). We observed a negative correlation of RBM20 abundance to the splicing ratios of TTN I-Band specific exons but overall positive TTN TE, indicating that exon inclusion contributes negatively to the translational efficiency. This hypothesis is supported by significantly lower TE levels of N2BA specific exons that are not part of N2B transcript isoform. We have observed this behaviour previously (van Heesch et al., 2019) and can now connect this mechanistically to RBM20 splicing control.

Besides TTN, many other target genes show similar RBM20 driven splicing effects on single or relatively low exon numbers. How exon exclusion affects isoform production and translational efficiency and whether isoform-specific characteristics such as secondary structure (Lim et al., 2018; S. E. Wang et al., 2020) or codon usage determine the differences of TE rates as we have observed for TTN remains elusive and requires additional studies.

The same strategy could be applied for other splicing factors associated with TE levels of target RNAs. There is a gap in our systematic understanding of when and why splicing factors switch isoform production leading to either inclusion or exclusion of selected exons resulting in an isoform switch. Particularly in the heart, whose function is determined by cardiomyocytes, validation experiments are challenging to conduct but are needed to elucidate the precise molecular interactions driving translational regulation by nuclear splicing factors such as RBM20.

Dual-function RBPs regulate mRNA abundance and TE

Along with RBPs showing potential novel functions, we identified 21 RBPs associated with regulation of target mRNA abundance and TE, which we termed “dual-function” RBPs. Only little is known about the multifunctionality of RBPs and how this regulation is activated. Some recent efforts have investigated the molecular mechanisms behind multiple functionalities, which

includes the formation of heterogeneous RBP complexes (Copsey et al., 2017; Damianov et al., 2016), switching from monomers to multimers (Y. Kim & Myong, 2016) and between subcellular localizations (Burgess et al., 2011). As provided by van Nostrand and colleagues (Van Nostrand et al., 2020), subcellular localization by immunofluorescence shows both nuclear and cytosolic localization for most of RBPs, suggesting switching of the cellular compartments to fulfil a distinct function. A prominent example is the splicing factor HNRNPM which our network analysis identified as one of the five central regulators of translation. Within the nucleus HNRNPM is associated with pre-mRNA splicing (S. E. Harvey et al., 2018). Upon viral infection, it switches from the nucleus to the cytosol, where it inhibits RNA virus-triggered innate immunity (Cao et al., 2019) or induces cap-independent IRES-mediated translation upon hypoxia (T.-M. Chen et al., 2019).

Surprisingly, we observed that dual-function RBPs target distinct sets of genes depending on whether mRNA abundance or TE rate is regulated. The limited overlap between groups and independent effect sizes within each RBP indicate distinct modes of regulation. To answer the question of how this regulation is modulated, we investigated different target and RBP features. We set out to identify the usage of different RBP binding motifs and the presence of binding sites located in different transcript regions (i.e. UTRs, CDS, or introns). However, we did not observe any striking differences between different groups of target genes. On the other side, we observed RBPs affinity to protein-coding genes sequence and UTR length, adding up to the independent effects of gene expression. A recent study has investigated the binding kinetics of RBP DAZL and its effect on mRNA abundance and translation. They were able to identify several 3' UTR features (e.g. UTR length, presence of binding clusters and distance to polyadenylation site) correlating with the trait regulation of different groups of target genes.

Among the dual-function RBPs, we identified examples with already known multi-functionality. One of them is G3BP1, a known multi-functional protein that can compartmentalize specific sets of mRNA to arrest translation as a response to arsenite stress, hypoxia and heat shock (Matsuki et al., 2013), UV exposure (Ying & Khaperskyy, 2021) and axonal mRNA translation and nerve regeneration (Sahoo et al., 2018). Besides stress granule formation, G3BP1 binds to specific stem-loop structures to trigger mRNA degradation (Fischer et al., 2020), promoting the formation of large cGAS complexes (Liu et al., 2019) or binds to viral dsRNA and RIG-I to enhance IFN-β response (S. S. Y. Kim et al., 2019). Another example is DEAD-Box Helicase 3 X-Linked (DDX3X) which has been shown to play important roles in transcription, splicing, transport, translation initiation, and cell cycle regulation (Rosner & Rinkevich, 2007) but also information of stress granule and impair mRNA translation (Samir et al., 2019; Valentin-Vega et al., 2016). At the same time, we provide dual-function evidence for the Ubiquitin C-Terminal Hydrolase L5 (UCHL5). Target genes correlating to UCHL5 abundance show mRNA abundance

trait involvement in chromatin organization, while targets with changes in TE rates encode for proteins involved in splicing. For many other dual-function RBPs, we observed different biological functions of trait-specific target genes such as G3BP1 but somewhat similar GO terms for PUM1, and some trait-specific genes showed no enrichment.

Since none of the investigated attributes revealed trait-specific regulation uniformly, a combination of several RBP and target gene features likely achieve modulation of RBP dual-functionality. Thus, combining multiple approaches *in vitro* or *in vivo* might help dissecting modulation of RBP function. One example was proposed by Liao and colleagues (Liao et al., 2016) by simultaneously combining mRNA interactome capture (Castello et al., 2012) and RBDmap (Castello et al., 2016) to identify known and novel RBPs in mouse cardiomyocytes. Furthermore, RBDmap allows the identification of RNA contact regions within RBPs. Future studies need to investigate whether trait-specific targets show similarities in the RBP-RNA interaction regions or if the inclusion of other techniques shed light on the complex interaction of RBP and its target mRNAs.

Conclusion and future outlook

With this work, I presented transcriptional and translational regulation in the human heart, resulting in known processes and unexpected findings. Integration of ribosome profiling as a technique to study the intermediate level of gene expression between transcriptome and proteome allows an unbiased search for novel translation events that would have gone undetected with traditional pipelines. Furthermore, ribosome profiling enables the investigation of global translational differences in gene expression that are not visible at transcription or association of RBP protein abundance with translational efficiency.

Using ribosome profiling, I predicted the translation of hundreds of novel microproteins in the human heart. Many of them are located in the mitochondria, suggesting mitochondrial metabolism, heart homeostasis, and cardiovascular diseases. A significant fraction of these translation events were also found in the kidney and liver suggesting a universal process present across all tissues and their cells. Individual studies are needed to capture the whole translatome, including tissue-specific proteins. Transcriptional and translational profiling on single cells will expand our understanding to unexpected detail, providing information on cell-type specificity for all (short and long) proteins and their role in disease. A combination of different approaches will likely emerge as a gold standard to confirm *in vivo* evidence. False-positive rates and individual tissue complexities make it, to date, impossible to confirm microprotein presence with a single technique. Though the prediction of translation and

confirmation of its abundance may only indicate potential function, future studies need to either apply methods such as CRISPR-CAS9 to screen for potential function (J. Chen et al., 2020) or do it case by case.

Apart from systematic or individual investigations of protein function, micropeptides will likely receive a central role in medicine. Their short protein sequences allows binding to large complexes triggering enhanced function or inhibition, making these interactions or the micropeptides druggable targets in biomedical application. In combination with advanced single cell techniques these proteins might be used for identification of affected cell lines allowing precise diagnosis of affected heart regions and the resulting cardiac disease.

Integration of transcriptional and translational information in diseased and unaffected human hearts revealed gene clusters with determined trait-specific biological roles. These clusters allowed us to gain a deeper insight into processes regulated at transcription, translation, or both levels that transcription alone would miss. Though transcription explains a significant fraction of protein variation, translation that supplements prediction of protein abundance, has been previously suggested (Schafer, Adami, et al., 2015; van Heesch et al., 2019).

Finally, RNA-binding protein abundance can be used as a predictive measure for mRNA abundance and TE rates. Quantitative co-regulation between protein abundance (as measured by Ribo-seq) and targets quantitative traits allows prediction of protein multifunctionality, expanding our understanding of the complex intracellular molecular processes. Even though we extrapolated CLIP-seq peaks from mostly cell lines to cardiac transcriptional and translational information, ubiquitous RBP expression suggest a general function across all types of human tissues. Of course additional CLIP-seq experiments in the heart would fine-tune the results at the same time we expect no large differences.

Bibliography

Abe, N., Matsumoto, K., Nishihara, M., Nakano, Y., Shibata, A., Maruyama, H., Shuto, S., Matsuda, A., Yoshida, M., Ito, Y., & Abe, H. (2015). Rolling Circle Translation of Circular RNA in Living Human Cells. *Scientific Reports*. https://doi.org/10.1038/srep16435

Aebersold, R., & Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature*, *537*(7620), 347–355. https://doi.org/10.1038/nature19949

Ainaoui, N., Hantelys, F., Renaud-Gabardos, E., Bunel, M., Lopez, F., Pujol, F., Planes, R., Bahraoui, E., Pichereaux, C., Burlet-Schiltz, O., Parini, A., Garmy-Susini, B., & Prats, A.-C. (2015). Promoter-Dependent Translation Controlled by p54nrb and hnRNPM during Myoblast Differentiation. *PloS One*, *10*(9), e0136466. https://doi.org/10.1371/journal.pone.0136466

Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H., & Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics*, *33*(21), 3387–3395. https://doi.org/10.1093/bioinformatics/btx431

Amaral, P. P., Clark, M. B., Gascoigne, D. K., Dinger, M. E., & Mattick, J. S. (2011). lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Research*, *39*(Database issue), D146-51. https://doi.org/10.1093/nar/gkq1138

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., Kasaragod, P., Shelton, J. M., Liou, J., Bassel-Duby, R., & Olson, E. N. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*. https://doi.org/10.1016/j.cell.2015.01.009

Anderson, K. M., Anderson, D. M., McAnally, J. R., Shelton, J. M., Bassel-Duby, R., & Olson, E. N. (2016). Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature*, *2*(7629), 1–13. https://doi.org/10.1038/nature20128

Avni, D, Shama, S., Loreni, F., & Meyuhas, O. (1994). Vertebrate mRNAs with a 5'-terminal pyrimidine tract are candidates for translational repression in quiescent cells: characterization of the translational cis-regulatory element. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.14.6.3822

Avni, Dror, Biberman, Y., & Meyuhas, O. (1996). The 5′ Terminal Oligopyrimidine Tract Confers Translational Control on Top Mrnas in a Cell Type-and Sequence Context-Dependent Manner. *Nucleic Acids Research*. https://doi.org/10.1093/nar/25.5.995

Backlund, M., Stein, F., Rettel, M., Schwarzl, T., Perez-Perri, J. I., Brosig, A., Zhou, Y., Neu-Yilik, G., Hentze, M. W., & Kulozik, A. E. (2020). Plasticity of nuclear and cytoplasmic stress responses of RNA-binding proteins. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa256

Bánfai, B., Jia, H., Khatun, J., Wood, E., Risk, B., Gundling, W. E., Kundaje, A., Gunawardena, H. P., Yu, Y., Xie, L., Krajewski, K., Strahl, B. D., Chen, X., Bickel, P., Giddings, M. C., Brown, J. B., & Lipovich, L. (2012). Long noncoding RNAs are rarely translated in two human cell lines. *Genome Research*. https://doi.org/10.1101/gr.134767.111

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395. https://doi.org/10.1038/cr.2011.22

Barbosa, C., Peixeiro, I., & Romão, L. (2013). Gene expression regulation by upstream open reading frames and human disease. *PLoS Genetics*, *9*(8), e1003529. https://doi.org/10.1371/journal.pgen.1003529

Bastide, A., Peretti, D., Knight, J. R. P., Grosso, S., Spriggs, R. V., Pichon, X., Sbarrato, T., Roobol, A., Roobol, J., Vito, D., Bushell, M., von der Haar, T., Smales, C. M., Mallucci, G. R., & Willis, A. E. (2017). RTN3 Is a Novel Cold-Induced Protein and Mediates Neuroprotective Effects of RBM3. *Current Biology*. https://doi.org/10.1016/j.cub.2017.01.047

Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E. S., Vejnar, C. E., Lee, M. T., Rajewsky, N., Walther, T. C., & Giraldez, A. J. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *The EMBO Journal*, *33*(9), 981–993. https://doi.org/10.1002/embj.201488411

Bazzini, A. A., Johnstone, T. G., Christiano, R., MacKowiak, S. D., Obermayer, B., Fleming, E. S., Vejnar, C. E., Lee, M. T., Rajewsky, N., Walther, T. C., & Giraldez, A. J. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO Journal*. https://doi.org/10.1002/embj.201488411

Bekker-Jensen, D. B., Kelstrup, C. D., Batth, T. S., Larsen, S. C., Haldrup, C., Bramsen, J. B., Sørensen, K. D., Høyer, S., Ørntoft, T. F., Andersen, C. L., Nielsen, M. L., & Olsen, J. V. (2017). An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Systems*. https://doi.org/10.1016/j.cels.2017.05.009

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Boisvert, F. M., Van Koningsbruggen, S., Navascués, J., & Lamond, A. I. (2007). The multifunctional nucleolus. In *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm2184

Boo, S. H., & Kim, Y. K. (2020). The emerging role of RNA modifications in the regulation of mRNA stability. In *Experimental and Molecular Medicine*. https://doi.org/10.1038/s12276-020-0407-z

Brauch, K. M., Karst, M. L., Herron, K. J., de Andrade, M., Pellikka, P. A., Rodeheffer, R. J., Michels, V. V., & Olson, T. M. (2009). Mutations in Ribonucleic Acid Binding Protein Gene Cause Familial Dilated Cardiomyopathy. *Journal of the American College of Cardiology*. https://doi.org/10.1016/j.jacc.2009.05.038

Brody, E., & Abelson, J. (1985). The “spliceosome”: Yeast pre-messenger RNA associates with a 40s complex in a splicing-dependent reaction. *Science*. https://doi.org/10.1126/science.3890181

Buchan, J. R. (2014). mRNP granules. Assembly, function, and connections with disease. *RNA Biology*, *11*(8), 1019–1030. https://doi.org/10.4161/15476286.2014.972208

Burgess, H. M., Richardson, W. A., Anderson, R. C., Salaun, C., Graham, S. V., & Gray, N. K. (2011). Nuclear relocalisation of cytoplasmic poly(A)-binding proteins PABP1 and PABP4 in response to UV irradiation reveals mRNA-dependent export of metazoan PABPS. *Journal of Cell Science*. https://doi.org/10.1242/jcs.087692

Bushell, M., Stoneley, M., Kong, Y. W., Hamilton, T. L., Spriggs, K. A., Dobbyn, H. C., Qin, X., Sarnow, P., & Willis, A. E. (2006). Polypyrimidine tract binding protein regulates IRES-mediated gene

expression during apoptosis. *Molecular Cell*, *23*(3), 401–412. https://doi.org/10.1016/j.molcel.2006.06.012

Cabili, M., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes and Development*. https://doi.org/10.1101/gad.17446611

Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., Landthaler, M., Obermayer, B., & Ohler, U. (2016). Detecting actively translated open reading frames in ribosome profiling data. *Nature Methods*. https://doi.org/10.1038/nmeth.3688

Calviello, L., Venkataramanan, S., Rogowski, K. J., Wyler, E., Wilkins, K., Tejura, M., Thai, B., Krol, J., Filipowicz, W., Landthaler, M., & Floor, S. N. (2021). DDX3 depletion represses translation of mRNAs with complex 5'UTRs. *Nucleic Acids Research*, 1–15. https://doi.org/10.1093/nar/gkab287

Camacho, M. P. (2021). Beyond descriptive accuracy: The central dogma of molecular biology in scientific practice. *Studies in History and Philosophy of Science*. https://doi.org/10.1016/j.shpsa.2021.01.002

Campos, J. C., Bozi, L. H. M., Bechara, L. R. G., Lima, V. M., & Ferreira, J. C. B. (2016). Mitochondrial Quality Control in Cardiac Diseases. *Frontiers in Physiology*, *7*. https://doi.org/10.3389/fphys.2016.00479

Candeias, M. M. (2011). The can and can't dos of p53 RNA. *Biochimie*, *93*(11), 1962–1965. https://doi.org/10.1016/j.biochi.2011.06.010

Cao, P., Luo, W.-W., Li, C., Tong, Z., Zheng, Z.-Q., Zhou, L., Xiong, Y., & Li, S. (2019). The heterogeneous nuclear ribonucleoprotein hnRNPM inhibits RNA virus-triggered innate immunity by antagonizing RNA sensing of RIG-I-like receptors. *PLoS Pathogens*, *15*(8), e1007983. https://doi.org/10.1371/journal.ppat.1007983

Capecchi, M. R. (1967). Polypeptide chain termination in vitro: isolation of a release factor. *Proceedings of the National Academy of Sciences of the United States of America*, *58*(3), 1144–1151. https://doi.org/10.1073/pnas.58.3.1144

Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L. A., & Johnson, R. (2016). Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA*. https://doi.org/10.1261/rna.053561.115

Carmody, S. R., & Wente, S. R. (2009). mRNA nuclear export at a glance. *Journal of Cell Science*. https://doi.org/10.1242/jcs.041236

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsveld, J., & Hentze, M. W. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, *149*(6), 1393–1406. https://doi.org/10.1016/j.cell.2012.04.031

Castello, A., Fischer, B., Frese, C. K., Horos, R., Alleaume, A.-M., Foehr, S., Curk, T., Krijgsveld, J., & Hentze, M. W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular Cell*, *63*(4), 696–710. https://doi.org/10.1016/j.molcel.2016.06.029

Chen, C. Y., & Sarnow, P. (1995). Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science*. https://doi.org/10.1126/science.7536344

Chen, F. X., Smith, E. R., & Shilatifard, A. (2018). Born to run: control of transcription elongation by RNA polymerase II. *Nature Reviews. Molecular Cell Biology*, *19*(7), 464–478. https://doi.org/10.1038/s41580-018-0010-5

Chen, J., Brunner, A. D., Cogan, J. Z., Nuñez, J. K., Fields, A. P., Adamson, B., Itzhak, D. N., Li, J. Y., Mann, M., Leonetti, M. D., & Weissman, J. S. (2020). Pervasive functional translation of noncanonical human open reading frames. *Science*. https://doi.org/10.1126/science.aav5912

Chen, J., Tsai, A., O'Leary, S. E., Petrov, A., & Puglisi, J. D. (2012). Unraveling the dynamics of ribosome translocation. In *Current Opinion in Structural Biology*. https://doi.org/10.1016/j.sbi.2012.09.004

Chen, T.-M., Lai, M.-C., Li, Y.-H., Chan, Y.-L., Wu, C.-H., Wang, Y.-M., Chien, C.-W., Huang, S.-Y., Sun, H. S., & Tsai, S.-J. (2019). hnRNPM induces translation switch under hypoxia to promote colon cancer development. *EBioMedicine*, *41*, 299–309. https://doi.org/10.1016/j.ebiom.2019.02.059

Chew, G. L., Pauli, A., Rinn, J. L., Regev, A., Schier, A. F., & Valen, E. (2013). Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. *Development (Cambridge)*. https://doi.org/10.1242/dev.098343

Chew, G. L., Pauli, A., & Schier, A. F. (2016). Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nature Communications*. https://doi.org/10.1038/ncomms11663

Chothani, S., Adami, E., Ouyang, J. F., Viswanathan, S., Hubner, N., Cook, S. A., Schafer, S., & Rackham, O. J. L. (2019). deltaTE: Detection of Translationally Regulated Genes by Integrative Analysis of Ribo-seq and RNA-seq Data. *Current Protocols in Molecular Biology*. https://doi.org/10.1002/cpmb.108

Chothani, S., Schäfer, S., Adami, E., Viswanathan, S., Widjaja, A. A., Langley, S. R., Tan, J., Wang, M., Quaife, N. M., Jian Pua, C., D'Agostino, G., Guna Shekeran, S., George, B. L., Lim, S., Yiqun Cao, E., van Heesch, S., Witte, F., Felkin, L. E., Christodoulou, E. G., … Rackham, O. J. L. (2019). Widespread Translational Control of Fibrosis in the Human Heart by RNA-Binding Proteins. *Circulation*, *140*(11), 937–951. https://doi.org/10.1161/CIRCULATIONAHA.119.039596

Cirillo, L., Cieren, A., Barbieri, S., Khong, A., Schwager, F., Parker, R., & Gotta, M. (2020). UBAP2L Forms Distinct Cores that Act in Nucleating Stress Granules Upstream of G3BP1. *Current Biology*. https://doi.org/10.1016/j.cub.2019.12.020

Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A., & Lawrence, J. B. (2009). An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Molecular Cell*, *33*(6), 717–726. https://doi.org/10.1016/j.molcel.2009.01.026

Cloonan, N., Forrest, A. R. R., Kolle, G., Gardiner, B. B. A., Faulkner, G. J., Brown, M. K., Taylor, D. F., Steptoe, A. L., Wani, S., Bethel, G., Robertson, A. J., Perkins, A. C., Bruce, S. J., Lee, C. C., Ranade, S. S., Peckham, H. E., Manning, J. M., McKernan, K. J., & Grimmond, S. M. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods*, *5*(7), 613–619. https://doi.org/10.1038/nmeth.1223

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2009). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkp1137

Cockman, E., Anderson, P., & Ivanov, P. (2020). Top mrnps: Molecular mechanisms and principles of regulation. In *Biomolecules*. https://doi.org/10.3390/biom10070969

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*, 13. https://doi.org/10.1186/s13059-016-0881-8

Conibear, A. C. (2020). Deciphering protein post-translational modifications using chemical biology tools. In *Nature Reviews Chemistry*. https://doi.org/10.1038/s41570-020-00223-8

Copsey, A. C., Cooper, S., Parker, R., Lineham, E., Lapworth, C., Jallad, D., Sweet, S., & Morley, S. J. (2017). The helicase, DDX3X, interacts with poly(A)-binding protein 1 (PABP1) and caprin-1 at the leading edge of migrating fibroblasts and is required for efficient cell spreading. *The Biochemical Journal*, *474*(18), 3109–3120. https://doi.org/10.1042/BCJ20170354

Coughlin, S. S., Gottdiener, J. S., Baughman, K. L., Wasserman, A., Marx, E. S., Tefft, M. C., & Gersh, B. J. (1994). Black-white differences in mortality in idiopathic dilated cardiomyopathy: the Washington, DC, dilated cardiomyopathy study. *Journal of the National Medical Association*.

Couso, J. P., & Patraquim, P. (2017). Classification and function of small open reading frames. In *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm.2017.58

Crick, F. H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology*.

D'Lima, N. G., Ma, J., Winkler, L., Chu, Q., Loh, K. H., Corpuz, E. O., Budnik, B. A., Lykke-Andersen, J., Saghatelian, A., & Slavoff, S. A. (2017). A human microprotein that interacts with the mRNA decapping complex. *Nature Chemical Biology*. https://doi.org/10.1038/nchembio.2249

Damianov, A., Ying, Y., Lin, C. H., Lee, J. A., Tran, D., Vashisht, A. A., Bahrami-Samani, E., Xing, Y., Martin, K. C., Wohlschlegel, J. A., & Black, D. L. (2016). Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. *Cell*. https://doi.org/10.1016/j.cell.2016.03.040

Danilova, Y., Voronkova, A., Sulimov, P., & Kertész-Farkas, A. (2019). Bias in False Discovery Rate Estimation in Mass-Spectrometry-Based Peptide Identification. *Journal of Proteome Research*. https://doi.org/10.1021/acs.jproteome.8b00991

Darling, A. L., & Uversky, V. N. (2018). Intrinsic disorder and posttranslational modifications: The darker side of the biological dark matter. In *Frontiers in Genetics*. https://doi.org/10.3389/fgene.2018.00158

Davis, C. A., Hitz, B. C., Sloan, C. A., Chan, E. T., Davidson, J. M., Gabdank, I., Hilton, J. A., Jain, K., Baymuradov, U. K., Narayanan, A. K., Onate, K. C., Graham, K., Miyasato, S. R., Dreszer, T. R., Strattan, J. S., Jolanki, O., Tanaka, F. Y., & Cherry, J. M. (2018). The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkx1081

Dever, T. E., Dinman, J. D., & Green, R. (2018). Translation elongation and recoding in eukaryotes. *Cold Spring Harbor Perspectives in Biology*. https://doi.org/10.1101/cshperspect.a032649

Díaz-Muñoz, M. D., & Turner, M. (2018). Uncovering the Role of RNA-Binding Proteins in Gene Expression in the Immune System. *Frontiers in Immunology*, *9*(MAY), 1094. https://doi.org/10.3389/fimmu.2018.01094

Diaz de Arce, A. J., Noderer, W. L., & Wang, C. L. (2018). Complete motif analysis of sequence requirements for translation initiation at non-AUG start codons. *Nucleic Acids Research*, *46*(2), 985–994. https://doi.org/10.1093/nar/gkx1114

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/bts635

Doll, S., Dreßen, M., Geyer, P. E., Itzhak, D. N., Braun, C., Doppler, S. A., Meier, F., Deutsch, M. A., Lahm, H., Lange, R., Krane, M., & Mann, M. (2017). Region and cell-type resolved quantitative proteomic map of the human heart. *Nature Communications*. https://doi.org/10.1038/s41467-017-01747-2

Dong, J., Aitken, C. E., Thakur, A., Shin, B.-S., Lorsch, J. R., & Hinnebusch, A. G. (2017). Rps3/uS3 promotes mRNA binding at the 40S ribosome entry channel and stabilizes preinitiation complexes at start codons. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(11), E2126–E2135. https://doi.org/10.1073/pnas.1620569114

Dries, D. L., Exner, D. V., Gersh, B. J., Cooper, H. A., Carson, P. E., & Domanski, M. J. (1999). Racial Differences in the Outcome of Left Ventricular Dysfunction. *New England Journal of Medicine*. https://doi.org/10.1056/nejm199902253400804

Elliott, P., Andersson, B., Arbustini, E., Bilinska, Z., Cecchi, F., Charron, P., Dubourg, O., Kühl, U., Maisch, B., McKenna, W. J., Monserrat, L., Pankuweit, S., Rapezzi, C., Seferovic, P., Tavazzi, L., & Keren, A. (2008). Classification of the cardiomyopathies: A position statement from the european society of cardiology working group on myocardial and pericardial diseases. *European Heart Journal*. https://doi.org/10.1093/eurheartj/ehm342

Emanuelsson, O., Nielsen, H., Brunak, S., & von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*, *300*(4), 1005–1016. https://doi.org/10.1006/jmbi.2000.3903

Erickson, S. L., Corpuz, E. O., Maloy, J. P., Fillman, C., Webb, K., Bennett, E. J., & Lykke-Andersen, J. (2015). Competition between Decapping Complex Formation and Ubiquitin-Mediated Proteasomal Degradation Controls Human Dcp2 Decapping Activity. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.01517-14

Ezkurdia, I., Vázquez, J., Valencia, A., & Tress, M. (2014). Analyzing the first drafts of the human proteome. *Journal of Proteome Research*, *13*(8), 3854–3855. https://doi.org/10.1021/pr500572z

Fan, H., Villegas, C., Huang, A., & Wright, J. A. (1996). Suppression of malignancy by the 3' untranslated regions of ribonucleotide reductase R1 and R2 messenger RNAs. *Cancer Research*, *56*(19), 4366–4369. http://www.ncbi.nlm.nih.gov/pubmed/8813126

Fatkin, D., MacRae, C., Sasaki, T., Wolff, M. R., Porcu, M., Frenneaux, M., Atherton, J., Vidaillet, H. J., Spudich, S., De Girolami, U., Seidman, J. G., Muntoni, F., Müehle, G., Johnson, W., McDonough, B., & Seidman, C. E. (1999). Missense Mutations in the Rod Domain of the Lamin A/C Gene as Causes of Dilated Cardiomyopathy and Conduction-System Disease. *New England Journal of Medicine*. https://doi.org/10.1056/nejm199912023412302

Fernandopulle, M. S., Lippincott-Schwartz, J., & Ward, M. E. (2021). RNA transport and local translation in neurodevelopmental and neurodegenerative disease. In *Nature Neuroscience*. https://doi.org/10.1038/s41593-020-00785-2

Fields, A. P., Rodriguez, E. H., Jovanovic, M., Stern-Ginossar, N., Haas, B. J., Mertins, P., Raychowdhury, R., Hacohen, N., Carr, S. A., Ingolia, N. T., Regev, A., & Weissman, J. S. (2015). A Regression-Based Analysis of Ribosome-Profiling Data Reveals a Conserved Complexity to Mammalian Translation. *Molecular Cell*, *60*(5), 816–827. https://doi.org/10.1016/j.molcel.2015.11.013

Fischer, J. W., Busa, V. F., Shao, Y., & Leung, A. K. L. (2020). Structure-Mediated RNA Decay by UPF1 and G3BP1. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2020.01.021

Fisher, R. A. (1915). Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. *Biometrika*, *10*(4), 507. https://doi.org/10.2307/2331838

Flockhart, R. J., Webster, D. E., Qu, K., Mascarenhas, N., Kovalski, J., Kretz, M., & Khavari, P. A. (2012). BRAFV600E remodels the melanocyte transcriptome and induces BANCR to regulate melanoma cell migration. *Genome Research*, *22*(6), 1006–1014. https://doi.org/10.1101/gr.140061.112

Fonseca, B. D., Zakaria, C., Jia, J. J., Graber, T. E., Svitkin, Y., Tahmasebi, S., Healy, D., Hoang, H. D., Jensen, J. M., Diao, I. T., Lussier, A., Dajadian, C., Padmanabhan, N., Wang, W., Matta-Camacho, E., Hearnden, J., Smith, E. M., Tsukumo, Y., Yanagiya, A., … Damgaard, C. K. (2015). La-related protein 1 (LARP1) represses terminal oligopyrimidine (TOP) mRNA translation downstream of mTOR complex 1 (mTORC1). *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.M114.621730

Freeman, P. R., Hedges, L. V., & Olkin, I. (1986). Statistical Methods for Meta-Analysis. *Biometrics*, *42*(2), 454. https://doi.org/10.2307/2531069

Freiburg, A., Trombitas, K., Hell, W., Cazorla, O., Fougerousse, F., Centner, T., Kolmerer, B., Witt, C., Beckmann, J. S., Gregorio, C. C., Granzier, H., & Labeit, S. (2000). Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity. *Circulation Research*. https://doi.org/10.1161/01.RES.86.11.1114

Frendewey, D., & Keller, W. (1985). Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell*. https://doi.org/10.1016/S0092-8674(85)80131-8

Gerstberger, S., Hafner, M., & Tuschl, T. (2014). A census of human RNA-binding proteins. *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3813

Gerull, B., Gramlich, M., Atherton, J., McNabb, M., Trombitás, K., Sasse-Klaassen, S., Seidman, J. G., Seidman, C., Granzier, H., Labeit, S., Frenneaux, M., & Thierfelder, L. (2002). Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nature Genetics*. https://doi.org/10.1038/ng815

Gilbert, W. (1978). Why genes in pieces? In *Nature*. https://doi.org/10.1038/271501a0

Gingold, H., Tehler, D., Christoffersen, N. R., Nielsen, M. M., Asmar, F., Kooistra, S. M., Christophersen, N. S., Christensen, L. L., Borre, M., Sørensen, K. D., Andersen, L. D., Andersen, C. L., Hulleman, E., Wurdinger, T., Ralfkiær, E., Helin, K., Grønbæk, K., Ørntoft, T., Waszak, S. M., … Pilpel, Y. (2014). A dual program for translation regulation in cellular proliferation and differentiation. *Cell*, *158*(6), 1281–1292. https://doi.org/10.1016/j.cell.2014.08.011

Glažar, P., Papavasileiou, P., & Rajewsky, N. (2014). circBase: a database for circular RNAs. *RNA*, *20*(11), 1666–1670. https://doi.org/10.1261/rna.043687.113

Grabowski, P. J., Seiler, S. R., & Sharp, P. A. (1985). A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell*. https://doi.org/10.1016/S0092-8674(85)80130-6

Graham, L. D., Pedersen, S. K., Brown, G. S., Ho, T., Kassir, Z., Moynihan, A. T., & Vizgoft, E. K. (2011). Colorectal Neoplasia Differentially Expressed (CRNDE), a Novel Gene with Elevated Expression in Colorectal Adenomas and Adenocarcinomas. *Genes and Cancer*, *2*(8), 829–840. https://doi.org/10.1177/1947601911431081

Granados-Riveron, J. T., & Aquino-Jarquin, G. (2016). The complexity of the translation ability of circRNAs. In *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. https://doi.org/10.1016/j.bbagrm.2016.07.009

Gray, K. A., Daugherty, L. C., Gordon, S. M., Seal, R. L., Wright, M. W., & Bruford, E. A. (2013). Genenames.org: the HGNC resources in 2013. *Nucleic Acids Research*, *41*(Database issue), D545-52. https://doi.org/10.1093/nar/gks1066

GTEx Consortium. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, *45*(6), 580–585. https://doi.org/10.1038/ng.2653

Gu, X., Zheng, Q., Chu, Q., & Zhu, H. (2021). HAND2-AS1: A functional cancer-related long non-coding RNA. *Biomedicine and Pharmacotherapy*, *137*(February), 0–5. https://doi.org/10.1016/j.biopha.2021.111317

Guenther, U.-P., Weinberg, D. E., Zubradt, M. M., Tedeschi, F. A., Stawicki, B. N., Zagore, L. L., Brar, G. A., Licatalosi, D. D., Bartel, D. P., Weissman, J. S., & Jankowsky, E. (2018). The helicase Ded1p controls use of near-cognate translation initiation codons in 5' UTRs. *Nature*, *559*(7712), 130–134. https://doi.org/10.1038/s41586-018-0258-0

Guo, W., Schafer, S., Greaser, M. L., Radke, M. H., Liss, M., Govindarajan, T., Maatz, H., Schulz, H., Li, S., Parrish, A. M., Dauksaite, V., Vakeel, P., Klaassen, S., Gerull, B., Thierfelder, L., Regitz-Zagrosek, V., Hacker, T. A., Saupe, K. W., Dec, G. W., … Gotthardt, M. (2012). RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nature Medicine*, *18*(5), 766–773. https://doi.org/10.1038/nm.2693

Gupta, N., Bandeira, N., Keich, U., & Pevzner, P. A. (2011). Target-decoy approach and false discovery rate: When things may go wrong. *Journal of the American Society for Mass Spectrometry*. https://doi.org/10.1007/s13361-011-0139-3

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., Huarte, M., Zuk, O., Carey, B. W., Cassady, J. P., Cabili, M. N., Jaenisch, R., Mikkelsen, T. S., Jacks, T., Hacohen, N., Bernstein, B. E., Kellis, M., Regev, A., Rinn, J. L., & Lander, E. S. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*. https://doi.org/10.1038/nature07672

Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., & Lander, E. S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell*. https://doi.org/10.1016/j.cell.2013.06.009

Gysi, D. M., Voigt, A., Fragoso, T. de M., Almaas, E., & Nowick, K. (2018). wTO: an R package for computing weighted topological overlap and a consensus network with integrated visualization tool. *BMC Bioinformatics*, *19*(1), 392. https://doi.org/10.1186/s12859-018-2351-7

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., Ulrich, A., Wardle, G. S., Dewell, S., Zavolan, M., & Tuschl, T. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, *141*(1), 129–141. https://doi.org/10.1016/j.cell.2010.03.009

Han, P., Li, W., Lin, C.-H., Yang, J., Shang, C., Nuernberg, S. T., Jin, K. K., Xu, W., Lin, C.-Y., Lin, C.-J., Xiong, Y., Chien, H.-C., Zhou, B., Ashley, E., Bernstein, D., Chen, P.-S., Chen, H.-S. V., Quertermous, T., & Chang, C.-P. (2014). A long noncoding RNA protects the heart from pathological hypertrophy. *Nature*, *514*(7520), 102–106. https://doi.org/10.1038/nature13596

Hanan, M., Soreq, H., & Kadener, S. (2017). CircRNAs in the brain. In *RNA Biology*. https://doi.org/10.1080/15476286.2016.1255398

Hang, J., Wan, R., Yan, C., & Shi, Y. (2015). Structural basis of pre-mRNA splicing. *Science*. https://doi.org/10.1126/science.aac8159

Hansen, T. B. (2021). Signal and noise in circRNA translation. *Methods*. https://doi.org/10.1016/j.ymeth.2021.02.007

Harakalova, M., Kummeling, G., Sammani, A., Linschoten, M., Baas, A. F., Van Der Smagt, J., Doevendans, P. A., Van Tintelen, J. P., Dooijes, D., Mokry, M., & Asselbergs, F. W. (2015). A systematic analysis of genetic dilated cardiomyopathy reveals numerous ubiquitously expressed and muscle-specific genes. *European Journal of Heart Failure*. https://doi.org/10.1002/ejhf.255

Harvey, R. F., Smith, T. S., Mulroney, T., Queiroz, R. M. L., Pizzinga, M., Dezi, V., Villenueva, E., Ramakrishna, M., Lilley, K. S., & Willis, A. E. (2018). Trans-acting translational regulatory RNA binding proteins. *Wiley Interdisciplinary Reviews. RNA*, *9*(3), e1465. https://doi.org/10.1002/wrna.1465

Harvey, S. E., Xu, Y., Lin, X., Gao, X. D., Qiu, Y., Ahn, J., Xiao, X., & Cheng, C. (2018). Coregulation of alternative splicing by hnRNPM and ESRP1 during EMT. *RNA (New York, N.Y.)*, *24*(10), 1326–1338. https://doi.org/10.1261/rna.066712.118

Hellen, C. U. T. (2018). Translation Termination and Ribosome Recycling in Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, *10*(10). https://doi.org/10.1101/cshperspect.a032656

Hellens, R. P., Brown, C. M., Chisnall, M. A. W., Waterhouse, P. M., & Macknight, R. C. (2016). The Emerging World of Small ORFs. In *Trends in Plant Science*. https://doi.org/10.1016/j.tplants.2015.11.005

Herman, D. S., Lam, L., Taylor, M. R. G., Wang, L., Teekakirikul, P., Christodoulou, D., Conner, L., DePalma, S. R., McDonough, B., Sparks, E., Teodorescu, D. L., Cirino, A. L., Banner, N. R., Pennell, D. J., Graw, S., Merlo, M., Di Lenarda, A., Sinagra, G., Bos, J. M., … Seidman, C. E. (2012). Truncations of Titin Causing Dilated Cardiomyopathy. *New England Journal of Medicine*. https://doi.org/10.1056/NEJMoa1110186

Hershberger, R. E., Hedges, D. J., & Morales, A. (2013). Dilated cardiomyopathy: The complexity of a diverse genetic architecture. In *Nature Reviews Cardiology*. https://doi.org/10.1038/nrcardio.2013.105

Hershey, J. W. B., Sonenberg, N., & Mathews, M. B. (2012). Principles of Translational Control: An Overview. *Cold Spring Harbor Perspectives in Biology*, *4*(12), a011528–a011528. https://doi.org/10.1101/cshperspect.a011528

Herzog, V. A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T. R., Wlotzka, W., Von Haeseler, A., Zuber, J., & Ameres, S. L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. *Nature Methods*. https://doi.org/10.1038/nmeth.4435

Hinnebusch, A. G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. In *Annual Review of Microbiology*. https://doi.org/10.1146/annurev.micro.59.031805.133833

Hinnebusch, A. G. (2014). The scanning mechanism of eukaryotic translation initiation. In *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev-biochem-060713-035802

Ho-Xuan, H., GlaÅar, P., Latini, C., Heizler, K., Haase, J., Hett, R., Anders, M., Weichmann, F., Bruckmann, A., Van Den Berg, D., Hüttelmaier, S., Rajewsky, N., Hackl, C., & Meister, G. (2020). Comprehensive analysis of translation from overexpressed circular RNAs reveals pervasive translation from linear transcripts. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa704

Holdt, L. M., Kohlmaier, A., & Teupser, D. (2018). Molecular functions and specific roles of circRNAs in the cardiovascular system. In *Non-coding RNA Research*. https://doi.org/10.1016/j.ncrna.2018.05.002

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Ridwan Amode, M., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Charkhchi, M., Cummins, C., da Rin Fioretto, L., Davidson, C., Dodiya, K., El Houdaigui, B., Fatima, R., … Flicek, P. (2021). Ensembl 2021. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa942

Hsieh, A. C., Liu, Y., Edlind, M. P., Ingolia, N. T., Janes, M. R., Sher, A., Shi, E. Y., Stumpf, C. R., Christensen, C., Bonham, M. J., Wang, S., Ren, P., Martin, M., Jessen, K., Feldman, M. E., Weissman, J. S., Shokat, K. M., Rommel, C., & Ruggero, D. (2012). The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature*. https://doi.org/10.1038/nature10912

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M. J., Kenzelmann-Broz, D., Khalil, A. M., Zuk, O., Amit, I., Rabani, M., Attardi, L. D., Regev, A., Lander, E. S., Jacks, T., & Rinn, J. L. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*, *142*(3), 409–419. https://doi.org/10.1016/j.cell.2010.06.040

Ingolia, N. T. (2016). Ribosome Footprint Profiling of Translation throughout the Genome. In *Cell*. https://doi.org/10.1016/j.cell.2016.02.066

Ingolia, N. T., Brar, G. A., Stern-Ginossar, N., Harris, M. S., Talhouarne, G. J. S., Jackson, S. E., Wills, M. R., & Weissman, J. S. (2014). Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Reports*. https://doi.org/10.1016/j.celrep.2014.07.045

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science (New York, N.Y.)*, *324*(5924), 218–223. https://doi.org/10.1126/science.1168978

Ingolia, N. T., Lareau, L. F., & Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. https://doi.org/10.1016/j.cell.2011.10.002

Ivanyi-Nagy, R., Ahmed, S. M., Peter, S., Ramani, P. D., Ong, P. F., Dreesen, O., & Dröge, P. (2018). The RNA interactome of human telomerase rna reveals a coding-independent role for a histone mRNA in telomere homeostasis. *ELife*. https://doi.org/10.7554/eLife.40037

Jackson, R. J., Hellen, C. U. T., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews. Molecular Cell Biology*, *11*(2), 113–127. https://doi.org/10.1038/nrm2838

Jackson, R. J., Hellen, C. U. T., & Pestova, T. V. (2012). Termination and post-termination events in eukaryotic translation. In *Advances in Protein Chemistry and Structural Biology*. https://doi.org/10.1016/B978-0-12-386497-0.00002-5

Jamsai, D., Watkins, D. N., O'Connor, A. E., Merriner, D. J., Gursoy, S., Bird, A. D., Kumar, B., Miller, A., Cole, T. J., Jenkins, B. J., & O'Bryan, M. K. (2017). In vivo evidence that RBM5 is a tumour suppressor in the lung. *Scientific Reports*. https://doi.org/10.1038/s41598-017-15874-9

Janer, A., van Karnebeek, C. D., Sasarman, F., Antonicka, H., Al Ghamdi, M., Shyr, C., Dunbar, M., Stockler-Ispiroglu, S., Ross, C. J., Vallance, H., Dionne, J., Wasserman, W. W., & Shoubridge, E. A. (2015). RMND1 deficiency associated with neonatal lactic acidosis, infantile onset renal failure, deafness, and multiorgan involvement. *European Journal of Human Genetics : EJHG*, *23*(10), 1301–1307. https://doi.org/10.1038/ejhg.2014.293

Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., Marzluff, W. F., & Sharpless, N. E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA (New York, N.Y.)*, *19*(2), 141–157. https://doi.org/10.1261/rna.035667.112

Jenny, A., Hachet, O., Závorszky, P., Cyrklaff, A., Weston, M. D. J., Johnston, D. S., Erdélyi, M., & Ephrussi, A. (2006). A translation-independent role of oskar RNA in early Drosophila oogenesis. *Development (Cambridge, England)*, *133*(15), 2827–2833. https://doi.org/10.1242/dev.02456

Jia, J.-J., Lahr, R. M., Solgaard, M. T., Moraes, B. J., Pointet, R., Yang, A.-D., Celucci, G., Graber, T. E., Hoang, H.-D., Niklaus, M. R., Pena, I. A., Hollensen, A. K., Smith, E. M., Chaker-Margot, M., Anton, L., Dajadian, C., Livingstone, M., Hearnden, J., Wang, X.-D., … Fonseca, B. D. (2021). mTORC1 promotes TOP mRNA translation through site-specific phosphorylation of LARP1. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa1239

Jia, L., Mao, Y., Ji, Q., Dersh, D., Yewdell, J. W., & Qian, S.-B. (2020). Decoding mRNA translatability and stability from the 5' UTR. *Nature Structural & Molecular Biology*, *27*(9), 814–821. https://doi.org/10.1038/s41594-020-0465-x

Johnstone, T. G., Bazzini, A. A., & Giraldez, A. J. (2016). Upstream ORFs are prevalent translational repressors in vertebrates. *The EMBO Journal*, *35*(7), 706–723. https://doi.org/10.15252/embj.201592759

Kamisago, M., Sharma, S. D., DePalma, S. R., Solomon, S., Sharma, P., McDonough, B., Smoot, L., Mullen, M. P., Woolf, P. K., Wigle, E. D., Seidman, J. G., Jarcho, J., Shapiro, L. R., & Seidman, C. E. (2000). Mutations in Sarcomere Protein Genes as a Cause of Dilated Cardiomyopathy. *New England Journal of Medicine*. https://doi.org/10.1056/nejm200012073432304

Karakas, D., & Ozpolat, B. (2021). The role of lncrnas in translation. *Non-Coding RNA*. https://doi.org/10.3390/ncrna7010016

Khan, M. A. F., Reckman, Y. J., Aufiero, S., Van Den Hoogenhof, M. M. G., Van Der Made, I., Beqqali, A., Koolbergen, D. R., Rasmussen, T. B., Van Der Velden, J., Creemers, E. E., & Pinto, Y. M. (2016). RBM20 Regulates Circular RNA Production from the Titin Gene. *Circulation Research*, *119*(9), 996–1003. https://doi.org/10.1161/CIRCRESAHA.116.309568

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*. https://doi.org/10.1038/nmeth.3317

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*. https://doi.org/10.1186/gb-2013-14-4-r36

Kim, J., Park, R. Y., Chen, J. K., Kim, J., Jeong, S., & Ohn, T. (2014). Splicing factor SRSF3 represses the translation of programmed cell death 4 mRNA by associating with the 5′-UTR region. *Cell Death and Differentiation*. https://doi.org/10.1038/cdd.2013.171

Kim, S. S. Y., Sze, L., & Lam, K. P. (2019). The stress granule protein G3BP1 binds viral dsRNA and RIG-I to enhance interferon-β response. *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.RA118.005868

Kim, Y., & Myong, S. (2016). RNA Remodeling Activity of DEAD Box Proteins Tuned by Protein Concentration, RNA Length, and ATP. *Molecular Cell*, *63*(5), 865–876. https://doi.org/10.1016/j.molcel.2016.07.010

Köhler, A., & Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. *Nature Reviews. Molecular Cell Biology*, *8*(10), 761–773. https://doi.org/10.1038/nrm2255

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D. J., Luscombe, N. M., & Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature Structural & Molecular Biology*, *17*(7), 909–915. https://doi.org/10.1038/nsmb.1838

Koonin, E. V. (2012). Does the central dogma still stand? *Biology Direct*. https://doi.org/10.1186/1745-6150-7-27

Kozak, M. (1986). Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(9), 2850–2854. https://doi.org/10.1073/pnas.83.9.2850

Kozak, Marilyn. (1999). Initiation of translation in prokaryotes and eukaryotes. In *Gene*. https://doi.org/10.1016/S0378-1119(99)00210-3

Kozak, Marilyn. (2001). Constraints on reinitiation of translation in mammals. *Nucleic Acids Research*. https://doi.org/10.1093/nar/29.24.5226

Kretz, M., Webster, D. E., Flockhart, R. J., Lee, C. S., Zehnder, A., Lopez-Pajares, V., Qu, K., Zheng, G. X. Y., Chow, J., Kim, G. E., Rinn, J. L., Chang, H. Y., Siprashvili, Z., & Khavari, P. A. (2012). Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes and Development*, *26*(4), 338–343. https://doi.org/10.1101/gad.182121.111

Kristensen, L. S., Andersen, M. S., Stagsted, L. V. W., Ebbesen, K. K., Hansen, T. B., & Kjems, J. (2019). The biogenesis, biology and characterization of circular RNAs. In *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-019-0158-7

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315

Kuhn, R. M., Haussler, D., & James Kent, W. (2013). The UCSC genome browser and associated tools. *Briefings in Bioinformatics*. https://doi.org/10.1093/bib/bbs038

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Latos, P. A., Pauler, F. M., Koerner, M. V., Şenergin, H. B., Hudson, Q. J., Stocsits, R. R., Allhoff, W., Stricker, S. H., Klement, R. M., Warczok, K. E., Aumayr, K., Pasierbek, P., & Barlow, D. P. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science*. https://doi.org/10.1126/science.1228110

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*(2), R29. https://doi.org/10.1186/gb-2014-15-2-r29

Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., Fatica, A., Santini, T., Andronache, A., Wade, M., Laneve, P., Rajewsky, N., & Bozzoni, I. (2017). Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Molecular Cell*, *66*(1), 22-37.e9. https://doi.org/10.1016/j.molcel.2017.02.017

Leprivier, G., Remke, M., Rotblat, B., Dubuc, A., Mateo, A.-R. F., Kool, M., Agnihotri, S., El-Naggar, A., Yu, B., Prakash Somasekharan, S., Faubert, B., Bridon, G., Tognon, C. E., Mathers, J., Thomas, R., Li, A., Barokas, A., Kwok, B., Bowden, M., … Sorensen, P. H. (2013). The eEF2 Kinase Confers Resistance to Nutrient Deprivation by Blocking Translation Elongation. *Cell*, *153*(5), 1064–1079. https://doi.org/10.1016/j.cell.2013.04.055

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-12-323

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, L., Zhao, Q., & Kong, W. (2018). Extracellular matrix remodeling and cardiac fibrosis. *Matrix Biology : Journal of the International Society for Matrix Biology*, *68–69*, 490–506. https://doi.org/10.1016/j.matbio.2018.01.013

Li, Q., Brown, J. B., Huang, H., & Bickel, P. J. (2011). Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics*, *5*(3), 1752–1779. https://doi.org/10.1214/11-AOAS466

Li, W., Wang, W., Uren, P. J., Penalva, L. O. F., & Smith, A. D. (2017). Riborex: Fast and flexible identification of differential translation from Ribo-seq data. *Bioinformatics*, *33*(11), 1735–1737. https://doi.org/10.1093/bioinformatics/btx047

Li, X., Brock, G. N., Rouchka, E. C., Cooper, N. G. F., Wu, D., O'Toole, T. E., Gill, R. S., Eteleeb, A. M., O'Brien, L., & Rai, S. N. (2017). A comparison of per sample global scaling and per gene normalization methods for differential expression analysis of RNA-seq data. *PloS One*, *12*(5), e0176185. https://doi.org/10.1371/journal.pone.0176185

Liao, Y., Castello, A., Fischer, B., Leicht, S., Föehr, S., Frese, C. K., Ragan, C., Kurscheid, S., Pagler, E., Yang, H., Krijgsveld, J., Hentze, M. W., & Preiss, T. (2016). The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. *Cell Reports*, *16*(5), 1456–1469. https://doi.org/10.1016/j.celrep.2016.06.084

Liberman, N., Marash, L., & Kimchi, A. (2009). The translation initiation factor DAP5 is a regulator of cell survival during mitosis. *Cell Cycle (Georgetown, Tex.)*, *8*(2), 204–209. https://doi.org/10.4161/cc.8.2.7384

Licatalosi, D. D., Mele, A., Fak, J. J., Ule, J., Kayikci, M., Chi, S. W., Clark, T. A., Schweitzer, A. C., Blume, J. E., Wang, X., Darnell, J. C., & Darnell, R. B. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, *456*(7221), 464–469. https://doi.org/10.1038/nature07488

Lim, C. S., Wardell, S. J. T., Kleffmann, T., & Brown, C. M. (2018). The exon–intron gene structure upstream of the initiation codon predicts translation efficiency. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky282

Lin, M. F., Jungreis, I., & Kellis, M. (2011). PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*, *27*(13), i275-82. https://doi.org/10.1093/bioinformatics/btr209

Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C. L., Lindberg, E. L., Kanda, M., Polanski, K., Heinig, M., Lee, M., Nadelmann, E. R., Roberts, K., Tuck, L., Fasouli, E. S., DeLaughter, D. M., McDonough, B., Wakimoto, H., Gorham, J. M., Samari, S., … Teichmann, S. A. (2020). Cells of the adult human heart. *Nature*. https://doi.org/10.1038/s41586-020-2797-4

Liu, Z. S., Cai, H., Xue, W., Wang, M., Xia, T., Li, W. J., Xing, J. Q., Zhao, M., Huang, Y. J., Chen, S., Wu, S. M., Wang, X., Liu, X., Pang, X., Zhang, Z. Y., Li, T., Dai, J., Dong, F., Xia, Q., … Li, T. (2019). G3BP1 promotes DNA binding and activation of cGAS. *Nature Immunology*. https://doi.org/10.1038/s41590-018-0262-4

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*. https://doi.org/10.1186/1748-7188-6-26

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Low, T. Y., vanHeesch, S., vandenToorn, H., Giansanti, P., Cristobal, A., Toonen, P., Schafer, S., Hübner, N., vanBreukelen, B., Mohammed, S., Cuppen, E., Heck, A. J. R., & Guryev, V. (2013). Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis. *Cell Reports*. https://doi.org/10.1016/j.celrep.2013.10.041

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S. Y., AlMazroa, M. A., Alvarado, M., Anderson, H. R., Anderson, L. M., Andrews, K. G., Atkinson, C., Baddour, L. M., Barker-Collo, S., Bartels, D. H., … Murray, C. J. L. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. https://doi.org/10.1016/S0140-6736(12)61728-0

Luk, A., Ahn, E., Soor, G. S., & Butany, J. (2009). Dilated cardiomyopathy: A review. In *Journal of Clinical Pathology*. https://doi.org/10.1136/jcp.2008.060731

Maatz, H., Jens, M., Liss, M., Schafer, S., Heinig, M., Kirchner, M., Adami, E., Rintisch, C., Dauksaite, V., Radke, M. H., Selbach, M., Barton, P. J. R., Cook, S. A., Rajewsky, N., Gotthardt, M., Landthaler, M., & Hubner, N. (2014). RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing. *Journal of Clinical Investigation*, *124*(8), 3419–3430. https://doi.org/10.1172/JCI74523

Mackowiak, S. D., Zauber, H., Bielow, C., Thiel, D., Kutz, K., Calviello, L., Mastrobuoni, G., Rajewsky, N., Kempa, S., Selbach, M., & Obermayer, B. (2015). Extensive identification and analysis of conserved small ORFs in animals. *Genome Biology*. https://doi.org/10.1186/s13059-015-0742-x

Makarewich, C. A. (2020). The hidden world of membrane microproteins. *Experimental Cell Research*, *388*(2), 111853. https://doi.org/10.1016/j.yexcr.2020.111853

Makarewich, C. A., Baskin, K. K., Munir, A. Z., Bezprozvannaya, S., Sharma, G., Khemtong, C., Shah, A. M., McAnally, J. R., Malloy, C. R., Szweda, L. I., Bassel-Duby, R., & Olson, E. N. (2018). MOXI Is a Mitochondrial Micropeptide That Enhances Fatty Acid β-Oxidation. *Cell Reports*, *23*(13), 3701–3709. https://doi.org/10.1016/j.celrep.2018.05.058

Makarewich, C. A., & Olson, E. N. (2017). Mining for Micropeptides. *Trends in Cell Biology*, *27*(9), 685–696. https://doi.org/10.1016/j.tcb.2017.04.006

Manjeshwar, S., Branam, D. E., Lerner, M. R., Brackett, D. J., & Jupe, E. R. (2003). Tumor suppression by the prohibitin gene 3′untranslated region RNA in human breast cancer. *Cancer Research*.

Martin, K. C., & Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. *Cell*, *136*(4), 719–730. https://doi.org/10.1016/j.cell.2009.01.044

Martinez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J., & Embarek, A. M. (2018). Insights into structural and mechanistic features of viral IRES elements. In *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2017.02629

Maslon, M. M., Heras, S. R., Bellora, N., Eyras, E., & Cáceres, J. F. (2014). The translational landscape of the splicing factor SRSF1 and its role in mitosis. *ELife*. https://doi.org/10.7554/eLife.02028

Matsuki, H., Takahashi, M., Higuchi, M., Makokha, G. N., Oie, M., & Fujii, M. (2013). Both G3BP1 and G3BP2 contribute to stress granule formation. *Genes to Cells : Devoted to Molecular & Cellular Mechanisms*, *18*(2), 135–146. https://doi.org/10.1111/gtc.12023

Matsumoto, A., Pasut, A., Matsumoto, M., Yamashita, R., Fung, J., Monteleone, E., Saghatelian, A., Nakayama, K. I., Clohessy, J. G., & Pandolfi, P. P. (2017). MTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature*. https://doi.org/10.1038/nature21034

McKenna, W. J., Maron, B. J., & Thiene, G. (2017). Classification, epidemiology, and global burden of cardiomyopathies. *Circulation Research*. https://doi.org/10.1161/CIRCRESAHA.117.309711

McNally, E. M., & Mestroni, L. (2017). Dilated cardiomyopathy: Genetic determinants and mechanisms. In *Circulation Research*. https://doi.org/10.1161/CIRCRESAHA.116.309396

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S. D., Gregersen, L. H., Munschauer, M., Loewer, A., Ziebold, U., Landthaler, M., Kocks, C., Le Noble, F., & Rajewsky, N. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. https://doi.org/10.1038/nature11928

Menschaert, G., Van Criekinge, W., Notelaers, T., Koch, A., Crappé, J., Gevaert, K., & Van Damme, P. (2013). Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. *Molecular & Cellular Proteomics : MCP*, *12*(7), 1780–1790. https://doi.org/10.1074/mcp.M113.027540

Meyuhas, O., & Kahan, T. (2015). The race to decipher the top secrets of TOP mRNAs. In *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. https://doi.org/10.1016/j.bbagrm.2014.08.015

Miller, M. A., & Olivas, W. M. (2011). Roles of Puf proteins in mRNA degradation and translation. In *Wiley Interdisciplinary Reviews: RNA*. https://doi.org/10.1002/wrna.69

Miloslavski, R., Cohen, E., Avraham, A., Iluz, Y., Hayouka, Z., Kasir, J., Mudhasani, R., Jones, S. N., Cybulski, N., Rüegg, M. A., Larsson, O., Gandin, V., Rajakumar, A., Topisirovic, I., & Meyuhas, O. (2014). Oxygen sufficiency controls TOP mRNA translation via the TSC-Rheb-mTOR pathway in a 4E-BP-independent manner. *Journal of Molecular Cell Biology*. https://doi.org/10.1093/jmcb/mju008

Moore, K. S., & von Lindern, M. (2018). RNA binding proteins and regulation of mRNA translation in erythropoiesis. In *Frontiers in Physiology*. https://doi.org/10.3389/fphys.2018.00910

Morita, M., Gravel, S. P., Chénard, V., Sikström, K., Zheng, L., Alain, T., Gandin, V., Avizonis, D., Arguello, M., Zakaria, C., McLaughlan, S., Nouet, Y., Pause, A., Pollak, M., Gottlieb, E., Larsson, O., St-Pierre, J., Topisirovic, I., & Sonenberg, N. (2013). MTORC1 controls mitochondrial activity and biogenesis through 4E-BP-dependent translational regulation. *Cell Metabolism*. https://doi.org/10.1016/j.cmet.2013.10.001

Mudge, J. M., Ruiz-Orera, J., Prensner, J. R., Brunet, M. A., Gonzalez, J. M., Magrane, M., Martinez, T., Schulz, J. F., Yang, Y. T., Albà, M. M., Baranov, P. V, Bazzini, A., Bruford, E., Martin, M. J., Carvunis, A.-R., Chen, J., Couso, J. P., Flicek, P., Frankish, A., … van Heesch, S. (2021). A community-driven roadmap to advance research on translated open reading frames detected by Ribo-seq. *BioRxiv*. https://doi.org/10.1101/2021.06.10.447896

Müller-McNicoll, M., & Neugebauer, K. M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nature Reviews. Genetics*, *14*(4), 275–287. https://doi.org/10.1038/nrg3434

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science (New York, N.Y.)*, *320*(5881), 1344–1349. https://doi.org/10.1126/science.1158441

Navarro Gonzalez, J., Zweig, A. S., Speir, M. L., Schmelter, D., Rosenbloom, K. R., Raney, B. J., Powell, C. C., Nassar, L. R., Maulding, N. D., Lee, C. M., Lee, B. T., Hinrichs, A. S., Fyfe, A. C., Fernandes, J. D., Diekhans, M., Clawson, H., Casper, J., Benet-Pagès, A., Barber, G. P., … Kent, W. J. (2021). The UCSC genome browser database: 2021 update. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa1070

Nelson, B. R., Makarewich, C. A., Anderson, D. M., Winders, B. R., Troupes, C. D., Wu, F., Reese, A. L., McAnally, J. R., Chen, X., Kavalali, E. T., Cannon, S. C., Houser, S. R., Bassel-Duby, R., & Olson, E. N. (2016). A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science (New York, N.Y.)*, *351*(6270), 271–275. https://doi.org/10.1126/science.aad4076

Nesvizhskii, A. I. (2014). Proteogenomics: Concepts, applications and computational strategies. *Nature Methods*, *11*(11), 1114–1125. https://doi.org/10.1038/NMETH.3144

Nieminen, M. S., Harjola, V. P., Hochadel, M., Drexler, H., Komajda, M., Brutsaert, D., Dickstein, K., Ponikowski, P., Tavazzi, L., Follath, F., & Lopez-Sendon, J. L. (2008). Gender related differences in patients presenting with acute heart failure. Results from EuroHeart Failure Survey II. *European Journal of Heart Failure*. https://doi.org/10.1016/j.ejheart.2007.12.012

Nott, A., Le Hir, H., & Moore, M. J. (2004). Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes & Development*, *18*(2), 210–222. https://doi.org/10.1101/gad.1163204

Nowick, K., Gernat, T., Almaas, E., & Stubbs, L. (2009). Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(52), 22358–22363. https://doi.org/10.1073/pnas.0911376106

Okonko, D. O., & Shah, A. M. (2014). Heart failure: Mitochondrial dysfunction and oxidative stress in CHF. *Nature Reviews Cardiology*, *12*(1), 6–8. https://doi.org/10.1038/nrcardio.2014.189

Olexiouk, V., Crappé, J., Verbruggen, S., Verhegen, K., Martens, L., & Menschaert, G. (2016). sORFs.org: a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Research*, *44*(D1), D324-9. https://doi.org/10.1093/nar/gkv1175

Olexiouk, V., Van Criekinge, W., & Menschaert, G. (2018). An update on sORFs.org: A repository of small ORFs identified by ribosome profiling. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkx1130

Omenn, G. S., Lane, L., Lundberg, E. K., Overall, C. M., & Deutsch, E. W. (2017). Progress on the HUPO Draft Human Proteome: 2017 Metrics of the Human Proteome Project. *Journal of Proteome Research*, *16*(12), 4281–4287. https://doi.org/10.1021/acs.jproteome.7b00375

Orr, M. W., Mao, Y., Storz, G., & Qian, S. B. (2021). Alternative ORFs and small ORFs: Shedding light on the dark proteome. *Nucleic Acids Research*. https://doi.org/10.1093/NAR/GKZ734

Pamudurti, N. R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D., Ramberger, E., Shenzis, S., Samson, M., Dittmar, G., Landthaler, M., Chekulaeva, M., Rajewsky, N., & Kadener, S. (2017). Translation of CircRNAs. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2017.02.021

Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Mancia Leon, W. R., Damianov, A., Chronis, C., Papp, B., Chen, C.-K., McKee, R., Wang, X.-J., Chau, A., Sabri, S., Leonhardt, H., Zheng, S., Guttman, M., Black, D. L., & Plath, K. (2020). A protein assembly mediates Xist localization and gene silencing. *Nature*, *587*(7832), 145–151. https://doi.org/10.1038/s41586-020-2703-0

Papaioannou, D., Volinia, S., Nicolet, D., Świerniak, M., Petri, A., Mrózek, K., Bill, M., Pepe, F., Walker, C. J., Walker, A. E., Carroll, A. J., Kohlschmidt, J., Eisfeld, A.-K., Powell, B. L., Uy, G. L., Kolitz, J. E., Wang, E. S., Kauppinen, S., Dorrance, A., … Garzon, R. (2020). Clinical and functional significance of circular RNAs in cytogenetacically normal AML. *Blood Advances*, *4*(2), 239–251. https://doi.org/10.1182/bloodadvances.2019000568

Patursky-Polischuk, I., Kasir, J., Miloslavski, R., Hayouka, Z., Hausner-Hanochi, M., Stolovich-Rain, M., Tsukerman, P., Biton, M., Mudhasani, R., Jones, S. N., & Meyuhas, O. (2014). Reassessment of the Role of TSC, mTORC1 and MicroRNAs in Amino Acids-Meditated Translational Control of TOP mRNAs. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0109410

Pelletier, J, Kaplan, G., Racaniello, V. R., & Sonenberg, N. (1988). Cap-independent translation of poliovirus mRNA is conferred by sequence elements within the 5’ noncoding region. *Molecular and Cellular Biology*, *8*(3), 1103–1112. https://doi.org/10.1128/mcb.8.3.1103

Pelletier, Jerry, & Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature*. https://doi.org/10.1038/334320a0

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295. https://doi.org/10.1038/nbt.3122

Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. In *Nature Methods* (Vol. 8, Issue 10, pp. 785–786). https://doi.org/10.1038/nmeth.1701

Picotti, P., Rinner, O., Stallmach, R., Dautel, F., Farrah, T., Domon, B., Wenschuh, H., & Aebersold, R. (2010). High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nature Methods*, *7*(1), 43–46. https://doi.org/10.1038/nmeth.1408

Piwecka, M., Glažar, P., Hernandez-Miranda, L. R., Memczak, S., Wolf, S. A., Rybak-Wolf, A., Filipchyk, A., Klironomos, F., Jara, C. A. C., Fenske, P., Trimbuch, T., Zywitza, V., Plass, M., Schreyer, L., Ayoub, S., Kocks, C., Kühn, R., Rosenmund, C., Birchmeier, C., & Rajewsky, N. (2017). Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science*, *357*(6357). https://doi.org/10.1126/science.aam8526

Preiss, T., & Hentze, M. W. (2003). Starting the protein synthesis machine: Eukaryotic translation initiation. In *BioEssays*. https://doi.org/10.1002/bies.10362

Pukkila, P. J. (2001). Molecular Biology: The Central Dogma. In *Encyclopedia of Life Sciences*. https://doi.org/10.1038/npg.els.0000812

Qu, S., Zhong, Y., Shang, R., Zhang, X., Song, W., Kjems, J., & Li, H. (2017). The emerging landscape of circular RNA in life processes. In *RNA Biology*. https://doi.org/10.1080/15476286.2016.1220473

Quek, X. C., Thomson, D. W., Maag, J. L. V, Bartonicek, N., Signal, B., Clark, M. B., Gloss, B. S., & Dinger, M. E. (2015). lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Research*, *43*(Database issue), D168-73. https://doi.org/10.1093/nar/gku988

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btq033

Raj, A., Wang, S. H., Shim, H., Harpak, A., Li, Y. I., Engelmann, B., Stephens, M., Gilad, Y., & Pritchard, J. K. (2016). Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. *ELife*. https://doi.org/10.7554/eLife.13328

Raju, H., Alberg, C., Sagoo, G. S., Burton, H., & Behr, E. R. (2011). Inherited cardiomyopathies. In *BMJ (Online)*. https://doi.org/10.1136/bmj.d6966

Rastinejad, F., Conboy, M. J., Rando, T. A., & Blau, H. M. (1993). Tumor suppression by RNA from the 3' untranslated region of alpha-tropomyosin. *Cell*, *75*(6), 1107–1117. https://doi.org/10.1016/0092-8674(93)90320-p

Rathore, A., Chu, Q., Tan, D., Martinez, T. F., Donaldson, C. J., Diedrich, J. K., Yates, J. R., & Saghatelian, A. (2018). MIEF1 Microprotein Regulates Mitochondrial Translation [Research-article]. *Biochemistry*, *57*(38), 5564–5575. https://doi.org/10.1021/acs.biochem.8b00726

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz369

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., & Vilo, J. (2016). g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Research*, *44*(W1), W83–W89. https://doi.org/10.1093/nar/gkw199

Richardson, P., McKenna, R. W., Bristow, M., Maisch, B., Mautner, B., O'Connell, J., Olsen, E., Thiene, G., Goodwin, J., Gyarfas, I., Martin, I., & Nordet, P. (1996). Report of the 1995 World Health Organization/International Society and Federation of Cardiology Task Force on the definition and classification of cardiomyopathies. In *Circulation*. https://doi.org/10.1161/01.CIR.93.5.841

Richter, J. D., & Coller, J. (2015). Pausing on Polyribosomes: Make Way for Elongation in Translational Control. *Cell*, *163*(2), 292–300. https://doi.org/10.1016/j.cell.2015.09.041

Rigatti, R., Jia, J. H., Samani, N. J., & Eperon, I. C. (2004). Exon repetition: A major pathway for processing mRNA of some genes is allele-specific. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkh197

Risso, D., Ngai, J., Speed, T. P., & Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, *32*(9), 896–902. https://doi.org/10.1038/nbt.2931

Risso, D., Schwartz, K., Sherlock, G., & Dudoit, S. (2011). GC-content normalization for RNA-Seq data. *BMC Bioinformatics*, *12*, 480. https://doi.org/10.1186/1471-2105-12-480

Robert, C., & Watson, M. (2015). Errors in RNA-Seq quantification affect genes of relevance to human disease. *Genome Biology*, *16*, 177. https://doi.org/10.1186/s13059-015-0734-x

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btp616

Rosca, M. G., & Hoppel, C. L. (2010). Mitochondria in heart failure. *Cardiovascular Research*, *88*(1), 40–50. https://doi.org/10.1093/cvr/cvq240

Rosner, A., & Rinkevich, B. (2007). The DDX3 subfamily of the DEAD box helicases: divergent roles as unveiled by studying different organisms and in vitro assays. *Current Medicinal Chemistry*, *14*(23), 2517–2525. https://doi.org/10.2174/092986707782023677

Ruiz-Orera, J., & Albà, M. M. (2019). Translation of Small Open Reading Frames: Roles in Regulation and Evolutionary Innovation. In *Trends in Genetics*. https://doi.org/10.1016/j.tig.2018.12.003

Ruiz-Orera, J., Messeguer, X., Subirana, J. A., & Alba, M. M. (2014). Long non-coding RNAs as a source of new peptides. *ELife*. https://doi.org/10.7554/eLife.03523

Ruiz-Orera, J., Verdaguer-Grau, P., Villanueva-Cañas, J. L., Messeguer, X., & Albà, M. M. (2018). Translation of neutrally evolving peptides provides a basis for de novo gene evolution. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-018-0506-6

Sahoo, P. K., Lee, S. J., Jaiswal, P. B., Alber, S., Kar, A. N., Miller-Randolph, S., Taylor, E. E., Smith, T., Singh, B., Ho, T. S.-Y., Urisman, A., Chand, S., Pena, E. A., Burlingame, A. L., Woolf, C. J., Fainzilber, M., English, A. W., & Twiss, J. L. (2018). Axonal G3BP1 stress granule protein limits axonal mRNA translation and nerve regeneration. *Nature Communications*, *9*(1), 3358. https://doi.org/10.1038/s41467-018-05647-x

Samir, P., Kesavardhana, S., Patmore, D. M., Gingras, S., Malireddi, R. K. S., Karki, R., Guy, C. S., Briard, B., Place, D. E., Bhattacharya, A., Sharma, B. R., Nourse, A., King, S. V, Pitre, A., Burton, A. R., Pelletier, S., Gilbertson, R. J., & Kanneganti, T.-D. (2019). DDX3X acts as a live-or-die checkpoint in stressed cells by regulating NLRP3 inflammasome. *Nature*, *573*(7775), 590–594. https://doi.org/10.1038/s41586-019-1551-2

Savitski, M. M., Wilhelm, M., Hahne, H., Kuster, B., & Bantscheff, M. (2015). A Scalable Approach for Protein False Discovery Rate Estimation in Large Proteomic Data Sets. *Molecular & Cellular Proteomics : MCP*, *14*(9), 2394–2404. https://doi.org/10.1074/mcp.M114.046995

Schafer, S., Adami, E., Heinig, M., Rodrigues, K. E. C., Kreuchwig, F., Silhavy, J., van Heesch, S., Simaite, D., Rajewsky, N., Cuppen, E., Pravenec, M., Vingron, M., Cook, S. A., & Hubner, N. (2015). Translational regulation shapes the molecular landscape of complex disease phenotypes. *Nature Communications*, *6*, 7200. https://doi.org/10.1038/ncomms8200

Schafer, S., Miao, K., Benson, C. C., Heinig, M., Cook, S. A., & Hubner, N. (2015). Alternative Splicing Signatures in RNA-seq Data: Percent Spliced in (PSI). *Current Protocols in Human Genetics*, *87*, 11.16.1-11.16.14. https://doi.org/10.1002/0471142905.hg1116s87

Schoenberg, D. R., & Maquat, L. E. (2012). Regulation of cytoplasmic mRNA decay. In *Nature Reviews Genetics*. https://doi.org/10.1038/nrg3160

Schultheiss, H. P., Fairweather, D. L., Caforio, A. L. P., Escher, F., Hershberger, R. E., Lipshultz, S. E., Liu, P. P., Matsumori, A., Mazzanti, A., McMurray, J., & Priori, S. G. (2019). Dilated cardiomyopathy. *Nature Reviews Disease Primers*. https://doi.org/10.1038/s41572-019-0084-1

Sen, N. D., Zhou, F., Ingolia, N. T., & Hinnebusch, A. G. (2015). Genome-wide analysis of translational efficiency reveals distinct but overlapping functions of yeast DEAD-box RNA helicases Ded1 and eIF4A. *Genome Research*. https://doi.org/10.1101/gr.191601.115

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Shen, L., & Pelletier, J. (2020). General and target-specific DEXD/H RNA helicases in eukaryotic translation initiation. In *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms21124402

Shi, Y. (2017). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. In *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm.2017.86

Slavoff, S. A., Mitchell, A. J., Schwaid, A. G., Cabili, M. N., Ma, J., Levin, J. Z., Karger, A. D., Budnik, B. A., Rinn, J. L., & Saghatelian, A. (2013). Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat Chem Biol*, *9*(1), 59–64. https://doi.org/10.1038/nchembio.1120

Smithers, B., Oates, M., & Gough, J. (2019). "Why genes in pieces?" - Revisited. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz284

Somers, J., Pöyry, T., & Willis, A. E. (2013). A perspective on mammalian upstream open reading frame function. In *International Journal of Biochemistry and Cell Biology*. https://doi.org/10.1016/j.biocel.2013.04.020

Soto-Rifo, R., Rubilar, P. S., Limousin, T., De Breyne, S., Décimo, D., & Ohlmann, T. (2012). DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. *EMBO Journal*. https://doi.org/10.1038/emboj.2012.220

Spriggs, K. A., Bushell, M., & Willis, A. E. (2010). Translational regulation of gene expression during conditions of cell stress. *Molecular Cell*, *40*(2), 228–237. https://doi.org/10.1016/j.molcel.2010.09.028

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. In *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-019-0150-2

Stein, C. S., Jadiya, P., Zhang, X., McLendon, J. M., Abouassaly, G. M., Witmer, N. H., Anderson, E. J., Elrod, J. W., & Boudreau, R. L. (2018). Mitoregulin: A lncRNA-Encoded Microprotein that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. *Cell Reports*, *23*(13), 3710-3720.e8. https://doi.org/10.1016/j.celrep.2018.06.002

Stern-Ginossar, N., Weisburd, B., Michalski, A., Le, V. T. K., Hein, M. Y., Huang, S.-X., Ma, M., Shen, B., Qian, S.-B., Hengel, H., Mann, M., Ingolia, N. T., & Weissman, J. S. (2012). Decoding human cytomegalovirus. *Science (New York, N.Y.)*, *338*(6110), 1088–1093. https://doi.org/10.1126/science.1227919

Sutherland, L. C., Wang, K., & Robinson, A. G. (2010). RBM5 as a putative tumor suppressor gene for lung cancer. In *Journal of Thoracic Oncology*. https://doi.org/10.1097/JTO.0b013e3181c6e330

Svitkin, Y. V., Pause, A., Haghighat, A., Pyronnet, S., Witherell, G., Belsham, G. J., & Sonenberg, N. (2001). The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5′ secondary structure. *RNA*. https://doi.org/10.1017/S135583820100108X

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Von Mering, C. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky1131

Tahmasebi, S., Khoutorsky, A., Mathews, M. B., & Sonenberg, N. (2018). Translation deregulation in human disease. In *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/s41580-018-0034-x

Tan, W. L. W., Lim, B. T. S., Anene-Nzelu, C. G. O., Ackers-Johnson, M., Dashi, A., See, K., Tiang, Z., Lee, D. P., Chua, W. W., Luu, T. D. A., Li, P. Y. Q., Richards, A. M., & Foo, R. S. Y. (2017). A landscape of circular RNA expression in the human heart. *Cardiovascular Research*. https://doi.org/10.1093/cvr/cvw250

Team, R. C. (2016). R: A Language and Environment for Statistical Computing. In *R Foundation for Statistical Computing*.

The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, *47*(D1), D330–D338. https://doi.org/10.1093/nar/gky1055

The, M., MacCoss, M. J., Noble, W. S., & Käll, L. (2016). Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *Journal of the American Society for Mass Spectrometry*. https://doi.org/10.1007/s13361-016-1460-7

Theis, J. L., Zimmermann, M. T., Larsen, B. T., Rybakova, I. N., Long, P. A., Evans, J. M., Middha, S., de Andrade, M., Moss, R. L., Wieben, E. D., Michels, V. V., & Olson, T. M. (2014). TNNI3K mutation in familial syndrome of conduction system disease, atrial tachyarrhythmia and dilated cardiomyopathy. *Human Molecular Genetics*, *23*(21), 5793–5804. https://doi.org/10.1093/hmg/ddu297

Thoreen, C. C., Chantranupong, L., Keys, H. R., Wang, T., Gray, N. S., & Sabatini, D. M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature*, *485*(7396), 109–113. https://doi.org/10.1038/nature11083

Tian, D., Sun, S., & Lee, J. T. (2010). The long noncoding RNA, Jpx, Is a molecular switch for X chromosome inactivation. *Cell*, *143*(3), 390–403. https://doi.org/10.1016/j.cell.2010.09.049

Travers, J. G., Kamal, F. A., Robbins, J., Yutzey, K. E., & Blaxall, B. C. (2016). Cardiac Fibrosis: The Fibroblast Awakens. *Circulation Research*, *118*(6), 1021–1040. https://doi.org/10.1161/CIRCRESAHA.115.306565

Trotta, E. (2014). On the normalization of the minimum free energy of RNAs by sequence length. *PloS One*, *9*(11), e113380. https://doi.org/10.1371/journal.pone.0113380

Turro, E., Su, S. Y., Gonçalves, Â., Coin, L. J. M., Richardson, S., & Lewin, A. (2011). Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome Biology*. https://doi.org/10.1186/gb-2011-12-2-r13

Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A., & Darnell, R. B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science (New York, N.Y.)*, *302*(5648), 1212–1215. https://doi.org/10.1126/science.1090095

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B. J., & Darnell, R. B. (2006). An RNA map predicting Nova-dependent splicing regulation. *Nature*, *444*(7119), 580–586. https://doi.org/10.1038/nature05304

Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet*, *17*(10), 601–614. https://doi.org/10.1038/nrg.2016.85

Ulitsky, I., & Bartel, D. P. (2013). Leading Edge Review lincRNAs: Genomics, Evolution, and Mechanisms. *Cell*.

Valentin-Vega, Y. A., Wang, Y.-D., Parker, M., Patmore, D. M., Kanagaraj, A., Moore, J., Rusch, M., Finkelstein, D., Ellison, D. W., Gilbertson, R. J., Zhang, J., Kim, H. J., & Taylor, J. P. (2016). Cancer-associated DDX3X mutations drive stress granule assembly and impair global translation. *Scientific Reports*, *6*, 25996. https://doi.org/10.1038/srep25996

Van Heesch, S., Van Iterson, M., Jacobi, J., Boymans, S., Essers, P. B., De Bruijn, E., Hao, W., MacInnes, A. W., Cuppen, E., & Simonis, M. (2014). Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biology*. https://doi.org/10.1186/gb-2014-15-1-r6

van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J. F., Adami, E., Faber, A. B., Kirchner, M., Maatz, H., Blachut, S., Sandmann, C.-L., Kanda, M., Worth, C. L., Schafer, S., Calviello, L., Merriott, R., Patone, G., Hummel, O., Wyler, E., Obermayer, B., … Hubner, N. (2019). The Translational Landscape of the Human Heart. *Cell*, *178*(1), 242-260.e29. https://doi.org/10.1016/j.cell.2019.05.010

Van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., Xiao, R., Blue, S. M., Chen, J.-Y., Cody, N. A. L., Dominguez, D., Olson, S., Sundararaman, B., Zhan, L., Bazile, C., Bouvrette, L. P. B., Bergalet, J., Duff, M. O., Garcia, K. E., Gelboin-Burkhart, C., … Yeo, G. W. (2020). A large-scale binding and functional map of human RNA-binding proteins. *Nature*, *583*(7818), 711–719. https://doi.org/10.1038/s41586-020-2077-3

Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., Blue, S. M., Nguyen, T. B., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M., & Yeo, G. W. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature Methods*. https://doi.org/10.1038/nmeth.3810

Vattem, K. M., & Wek, R. C. (2004). Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0400541101

Venø, M. T., Hansen, T. B., Venø, S. T., Clausen, B. H., Grebing, M., Finsen, B., Holm, I. E., & Kjems, J. (2015). Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. *Genome Biology*. https://doi.org/10.1186/s13059-015-0801-3

Wagner, G. P., Kin, K., & Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences = Theorie in Den Biowissenschaften*, *131*(4), 281–285. https://doi.org/10.1007/s12064-012-0162-3

Wang, K. C., & Chang, H. Y. (2011). Molecular Mechanisms of Long Noncoding RNAs. In *Molecular Cell*. https://doi.org/10.1016/j.molcel.2011.08.018

Wang, S. E., Brooks, A. E. S., Poole, A. M., & Simoes-Barbosa, A. (2020). Determinants of translation efficiency in the evolutionarily-divergent protist Trichomonas vaginalis. *BMC Molecular and Cell Biology*. https://doi.org/10.1186/s12860-020-00297-8

Wang, Zhihua, Zhang, X.-J., Ji, Y.-X., Zhang, P., Deng, K.-Q., Gong, J., Ren, S., Wang, X., Chen, I., Wang, H., Gao, C., Yokota, T., Ang, Y. S., Li, S., Cass, A., Vondriska, T. M., Li, G., Deb, A., Srivastava, D., … Wang, Y. (2016). The long noncoding RNA Chaer defines an epigenetic checkpoint in cardiac hypertrophy. *Nature Medicine*, *22*(10), 1131–1139. https://doi.org/10.1038/nm.4179

Wang, Zhong, Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, *10*(1), 57–63. https://doi.org/10.1038/nrg2484

Watson, J. D., & Crick, F. H. C. (1953a). Genetical implications of the structure of deoxyribonucleic acid. *Nature*. https://doi.org/10.1038/171964b0

Watson, J. D., & Crick, F. H. C. (1953b). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*. https://doi.org/10.1038/171737a0

Weber, R., Kleemann, L., Hirschberg, I., Chung, M.-Y., Valkov, E., & Igreja, C. (2021). DAP5 enables translation re-initiation on structured messenger RNAs. *BioRxiv*.

Webster, M. W., Stowell, J. A., & Passmore, L. A. (2019). RNA-binding proteins distinguish between similar sequence motifs to promote targeted deadenylation by Ccr4-Not. *ELife*. https://doi.org/10.7554/eLife.40670

Wilkinson, M. E., Charenton, C., & Nagai, K. (2020). RNA Splicing by the Spliceosome. *Annual Review of Biochemistry*, *89*, 359–388. https://doi.org/10.1146/annurev-biochem-091719-064225

Williams, C. R., Baccarella, A., Parrish, J. Z., & Kim, C. C. (2017). Empirical assessment of analysis workflows for differential expression analysis of human samples using RNA-Seq. *BMC Bioinformatics*, *18*(1), 38. https://doi.org/10.1186/s12859-016-1457-z

Witte, F., Ruiz-Orera, J., Ciolli Mattioli, C., Blachut, S., Adami, E., Schulz, J. F., Schneider-Lunitz, V., Hummel, O., Patone, G., Mücke, M. B., Šilhavý, J., Heinig, M., Bottolo, L., Sanchis, D., Vingron, M., Chekulaeva, M., Pravenec, M., Hubner, N., & van Heesch, S. (2020). *Trans control of cardiac mRNA translation in a protein length-dependent fashion*. 1–25. https://doi.org/10.1101/2020.06.05.133298

Witten, J. T., & Ule, J. (2011). Understanding splicing regulation through RNA splicing maps. In *Trends in Genetics*. https://doi.org/10.1016/j.tig.2010.12.001

Wu, Q., Medina, S. G., Kushawah, G., Devore, M. L., Castellano, L. A., Hand, J. M., Wright, M., & Bazzini, A. A. (2019). Translation affects mRNA stability in a codon-dependent manner in human cells. *ELife*. https://doi.org/10.7554/eLife.45396

Xi, Y., Honeywell, C., Zhang, D., Schwartzentruber, J., Beaulieu, C. L., Tetreault, M., Hartley, T., Marton, J., Vidal, S. M., Majewski, J., Aravind, L., Gollob, M., Boycott, K. M., & Gow, R. M. (2015). Whole exome sequencing identifies the TNNI3K gene as a cause of familial conduction system disease and congenital junctional ectopic tachycardia. *International Journal of Cardiology*. https://doi.org/10.1016/j.ijcard.2015.03.130

Xiao, M. S., & Wilusz, J. E. (2019). An improved method for circular RNA purification using RNase R that efficiently removes linear RNAs containing G-quadruplexes or structured 3' ends. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz576

Xiao, Z., Zou, Q., Liu, Y., & Yang, X. (2016). Genome-wide assessment of differential translations with ribosome profiling data. *Nature Communications*, *7*(1), 11194. https://doi.org/10.1038/ncomms11194

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.-W., Li, Y., Xu, X., Wong, G. K.-S., & Wang, J. (2014). SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics (Oxford, England)*, *30*(12), 1660–1666. https://doi.org/10.1093/bioinformatics/btu077

Yang, W., Ru, Y., Ren, J., Bai, J., Wei, J., Fu, S., Liu, X., Li, D., & Zheng, H. (2019). G3BP1 inhibits RNA virus replication by positively regulating RIG-I-mediated cellular antiviral response. *Cell Death and Disease*. https://doi.org/10.1038/s41419-019-2178-9

Yang, Y., & Wang, Z. (2019). IRES-mediated cap-independent translation, a path leading to hidden proteome. In *Journal of Molecular Cell Biology*. https://doi.org/10.1093/jmcb/mjz091

Yang, Y. Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., Jin, Y., Yang, Y. Y., Chen, L. L., Wang, Y., Wong, C. C. L. C., Xiao, X., & Wang, Z. (2017). Extensive translation of circular RNAs driven by N6-methyladenosine. *Cell Research*, *27*(5), 626–641. https://doi.org/10.1038/cr.2017.31

Ying, S., & Khaperskyy, D. A. (2021). UV damage induces G3BP1-dependent stress granule formation that is not driven by mTOR inhibition-mediated translation arrest. *Journal of Cell Science*. https://doi.org/10.1242/jcs.248310

You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G., Akbalik, G., Wang, M., Glock, C., Quedenau, C., Wang, X., Hou, J., Liu, H., Sun, W., Sambandan, S., Chen, T., Schuman, E. M., & Chen, W. (2015). Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nature Neuroscience*. https://doi.org/10.1038/nn.3975

Zeng, X., Lin, W., Guo, M., & Zou, Q. (2017). A comprehensive overview and evaluation of circular RNA detection tools. In *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1005420

Zhang, C., Zhang, Z., Castle, J., Sun, S., Johnson, J., Krainer, A. R., & Zhang, M. Q. (2008). Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes and Development*. https://doi.org/10.1101/gad.1703108

Zhang, H., Dou, S., He, F., Luo, J., Wei, L., & Lu, J. (2018). Genome-wide maps of ribosomal occupancy provide insights into adaptive evolution and regulatory roles of uORFs during Drosophila development. *PLoS Biology*. https://doi.org/10.1371/journal.pbio.2003903

Zhang, H., Wang, Y., & Lu, J. (2019). Function and Evolution of Upstream ORFs in Eukaryotes. In *Trends in Biochemical Sciences*. https://doi.org/10.1016/j.tibs.2019.03.002

Zhang, L., Salgado-Somoza, A., Vausort, M., Leszek, P., Devaux, Y., & Cardiolinc™ network. Electronic address: https://www.cardiolinc.org. (2018). A heart-enriched antisense long non-coding RNA regulates the balance between cardiac and skeletal muscle triadin. *Biochimica et Biophysica Acta. Molecular Cell Research*, *1865*(2), 247–258. https://doi.org/10.1016/j.bbamcr.2017.11.002

Zhang, S., Reljić, B., Liang, C., Kerouanton, B., Francisco, J. C., Peh, J. H., Mary, C., Jagannathan, N. S., Olexiouk, V., Tang, C., Fidelito, G., Nama, S., Cheng, R. K., Wee, C. L., Wang, L. C., Duek Roggli, P., Sampath, P., Lane, L., Petretto, E., … Ho, L. (2020). Mitochondrial peptide BRAWNIN is essential for vertebrate respiratory complex III assembly. *Nature Communications*. https://doi.org/10.1038/s41467-020-14999-2

Zhang, Y., Jiao, L., Sun, L., Li, Y., Gao, Y., Xu, C., Shao, Y., Li, M., Li, C., Lu, Y., Pan, Z., Xuan, L., Zhang, Y., Li, Q., Yang, R., Zhuang, Y., Zhang, Y., & Yang, B. (2018). LncRNA ZFAS1 as a SERCA2a Inhibitor to Cause Intracellular Ca2+ Overload and Contractile Dysfunction in a Mouse Model of Myocardial Infarction. *Circulation Research*, *122*(10), 1354–1368. https://doi.org/10.1161/CIRCRESAHA.117.312117

Zhong, Y., Karaletsos, T., Drewe, P., Sreedharan, V. T., Kuo, D., Singh, K., Wendel, H. G., & Rätsch, G. (2017). RiboDiff: Detecting changes of mRNA translation efficiency from ribosome footprints. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btw585

Zhou, Z., Dang, Y., Zhou, M., Li, L., Yu, C.-H., Fu, J., Chen, S., & Liu, Y. (2016). Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(41), E6117–E6125. https://doi.org/10.1073/pnas.1606724113

Zhu, P., Wu, J., Wang, Y., Zhu, X., Lu, T., Liu, B., He, L., Ye, B., Wang, S., Meng, S., Fan, D., Wang, J., Yang, L., Qin, X., Du, Y., Li, C., He, L., Ren, W., Wu, X., … Fan, Z. (2018). LncGata6 maintains stemness of intestinal stem cells and promotes intestinal tumorigenesis. *Nature Cell Biology*, *20*(10), 1134–1144. https://doi.org/10.1038/s41556-018-0194-0