

The translational landscape of the human heart

Sebastian van Heesch^{1,35,*}, Franziska Witte^{1,34}, Valentin Schneider-Lunitz^{1,34}, Jana Felicitas Schulz^{1,34}, Eleonora Adami^{1,2}, Allison B. Faber¹, Marieluise Kirchner³, Henrike Maatz¹, Susanne Blachut¹, Clara-Louisa Sandmann¹, Masatoshi Kanda¹, Catherine L. Worth¹, Sebastian Schafer^{2,4}, Lorenzo Calviello^{5,6,7}, Rhys Merriott¹, Giannino Patone¹, Oliver Hummel¹, Emanuel Wyler⁵, Benedikt Obermayer^{5,8}, Michael B. Mücke¹, Eric Lars-Helge Lindberg¹, Franziska Trnka¹, Sebastian Memczak^{5,37}, Marcel Schilling⁵, Leanne E. Felkin^{9,10}, Paul J.R. Barton^{9,10}, Nicholas M. Quafe^{9,10,33}, Konstantinos Vanezis^{10,33}, Sebastian Diecke^{11,12,13}, Masaya Mukai¹⁴, Nancy Mah¹⁵, Su-Jun Oh¹⁵, Andreas Kurtz¹⁵, Christoph Schramm^{16,17}, Dorothee Schwinge¹⁷, Marcial Sebode¹⁷, Magdalena Harakalova^{18,19}, Folkert W. Asselbergs^{18,20,21}, Aryan Vink¹⁹, Roel A. de Weger¹⁹, Sivakumar Viswanathan², Anissa A. Widjaja², Anna Gärtner-Rommel²², Hendrik Milting²², Cris dos Remedios²³, Christoph Knosalla^{12,24,25}, Philipp Mertins³, Markus Landthaler^{5,26}, Martin Vingron²⁷, Wolfgang A. Linke^{28,29}, J. G. Seidman³⁰, Christine E. Seidman^{30,31,32}, Nikolaus Rajewsky⁵, Uwe Ohler^{5,6}, Stuart A. Cook^{2,4,10,33} & Norbert Hubner^{1,12,13,25,35,36,*}

¹ Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany.

² Program in Cardiovascular and Metabolic Disorders, Duke-National University of Singapore, Singapore 169857, Singapore.

³ Proteomics Platform, Berlin Institute of Health (BIH) and Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany.

⁴ National Heart Research Institute Singapore (NHRIS), National Heart Centre Singapore, Singapore 169609, Singapore.

⁵ Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 10115 Berlin, Germany.

⁶ Department of Biology, Humboldt Universität Berlin, 10099, Berlin, Germany.

⁷ Present address: Department of Cell & Tissue Biology, University of California San Francisco, San Francisco 94143, CA, USA.

⁸ Core Unit Bioinformatics, Berlin Institute of Health (BIH), Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC) and Charité-Universitätsmedizin, 10117 Berlin, Germany.

⁹ Cardiovascular Research Centre, Royal Brompton and Harefield NHS Trust, SW3 6NP London, United Kingdom.

¹⁰ National Heart and Lung Institute, Imperial College London, SW3 6LY London, United Kingdom.

¹¹ Pluripotent Stem Cells Platform, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany.

¹² DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, 13347 Berlin, Germany.

¹³ Berlin Institute of Health (BIH), 10178 Berlin, Germany.

¹⁴ Department of Rheumatology and Clinical Immunology, Sapporo City General Hospital, Sapporo, Hokkaido 060-8604, Japan

¹⁵ Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité-Universitätsmedizin Berlin, 13353 Berlin, Germany.

¹⁶ Martin Zeitz Center for Rare Diseases, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany

¹⁷ First Department of Medicine, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany

¹⁸ Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht University, 3584 CX Utrecht, the Netherlands.

¹⁹ Department of Pathology, University Medical Center Utrecht, Utrecht University, 3584 CX Utrecht, the Netherlands.

²⁰ Durrer Center for Cardiovascular Research, Netherlands Heart Institute, 3511 EP Utrecht, NL

²¹ Farr Institute of Health Informatics Research and Institute of Health Informatics, University College London, NW1 2DA London, United Kingdom

²² Herz- und Diabeteszentrum NRW, Universitätsklinikum der Ruhr-Universität, Erich und Hanna Klessmann-Institut für Kardiovaskuläre Forschung und Entwicklung, 32545 Bad Oeynhausen, Germany.

²³ Discipline of Anatomy & Histology, Bosch Institute, Anderson Stuart Building, University of Sydney, Sydney 2006, Australia.

²⁴ Department of Cardiothoracic and Vascular Surgery, German Heart Center Berlin, 13353 Berlin, Germany.

²⁵ Charité-Universitätsmedizin, 10117 Berlin, Germany

²⁶ IRI Life Sciences, Institute für Biologie, Humboldt Universität zu Berlin, 10115 Berlin, Germany

²⁷ Max Planck Institute for Molecular Genetics, Department of Computational Molecular Biology, 14195 Berlin, Germany.

²⁸ Institute of Physiology II, University of Münster, 48149 Münster, Germany.

²⁹ DZHK (German Centre for Cardiovascular Research), Partner Site Göttingen, 37075 Göttingen, Germany

³⁰ Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

³¹ Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.

³² Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA.

³³ MRC-London Institute of Medical Sciences, Hammersmith Hospital Campus, London, W12 0NN, United Kingdom.

³⁷ Present address: Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA.

³⁴ These authors contributed equally

³⁵ Senior author

³⁶ Lead contact

*Correspondence: sebastianvanheesch@gmail.com and nhuebner@mdc-berlin.de

Summary

Gene expression in human tissue has primarily been studied on the transcriptional level, largely neglecting translational regulation. Here, we analyze the translomes of 80 human hearts to identify new translation events and quantify the impact of translational regulation. We show extensive translational control of cardiac gene expression, which is orchestrated in a process-specific manner. Translation downstream of predicted disease-causing protein-truncating variants appears frequent, suggesting inefficient translation termination. We identify hundreds of previously undetected microproteins expressed from lncRNAs and circRNAs, for which we validate the protein products *in vivo*. The translation of microproteins is not restricted to the heart and prominent in the translomes of human kidneys and livers. We associate these microproteins with diverse cellular processes and compartments, and find that many locate to the mitochondria. Importantly, dozens of microproteins are translated from lncRNAs with well-characterized noncoding functions, indicating previously unrecognized biology.

Keywords

Ribosome profiling, human heart, translome, translational regulation, heart failure, dilated cardiomyopathy, ORF detection, short ORFs, microproteins, lncRNAs, circRNAs, titin, protein truncating variants.

Introduction

Translational regulation is a key component of gene expression, yet our understanding of its role in human tissue is sparse. Genome-wide translomes can be characterized using ribosome profiling (or Ribo-seq), which captures mRNA footprints protected by translating ribosomes (Ingolia et al., 2009). From these footprints, the codon-by-codon movement of ribosomes can be inferred and used to identify actively translated open reading frames (ORFs) (Calviello and Ohler, 2017). Newly detected ORFs can include regulatory upstream ORFs (uORFs), which may repress the translational efficiency (TE) of mRNAs (Morris and Geballe, 2000), or short ORFs (sORFs) translated from long noncoding RNAs (lncRNAs), indicating potential microprotein production (Andrews and Rothnagel, 2014). For a handful of microproteins, key physiological roles have been uncovered (Anderson et al., 2015, 2016a; Galindo et al., 2007; Kondo et al., 2010; Nelson et al., 2016; Pauli et al., 2014), though a genome-wide catalog of microproteins in human tissue is lacking. Newly detected microprotein translation events can be used to expand and improve protein databases required for mass spectrometry (MS) searches, as it is not trivial to perform *de novo* microprotein discovery with MS in the absence of *a priori* microprotein sequence information.

Here, we elucidate the translational landscape of 80 human hearts, comprising both dilated cardiomyopathy (DCM) patients and non-DCM controls. DCM has a prevalence of up to 1:250 and is the most common reason for heart transplantation (Hershberger et al., 2013). Combining genotypes, transcriptomes and translomes, we show that protein truncating variants (PTVs) – including titin-truncating variants (TTNtv) that commonly cause DCM (Herman et al., 2012) – often inefficiently terminate translation. Moreover, we identify 169 lncRNAs and 40 circRNAs that encode previously unknown microproteins, which we validate *in vivo* and link to specific cellular processes and organelles, predominantly mitochondria. Dozens of microproteins are expressed from functionally characterized lncRNAs, such as *DANCR* (also known as *ANCR*) (Kretz et al., 2012), *TUG1* (Young et al., 2005), *JPX* (Tian et al., 2010), *myheart* (Han et al., 2014) and *UPPERHAND* (Anderson et al., 2016b), suggesting undiscovered roles of these microproteins or their involvement in biological functions assigned to the lncRNA. The majority of these lncRNAs are ubiquitously expressed in many tissues and we demonstrate their translation in human kidney and liver.

We present a detailed assessment of translation in 80 human hearts that may serve as a blueprint for characterizing the translational landscape of other human tissues. The data and analyses presented in this work can be explored via an interactive web application accessible at <https://shiny.mdc-berlin.de/cardiac-translatome>.

Results

A snapshot of active translation in 80 human hearts

To study cardiac mRNA expression and translation, we applied mRNA-seq and Ribo-seq to human left-ventricular cardiac tissue of 65 end-stage DCM patients and 15 non-DCM controls (Figure 1A, Figure S1A+B and Table S1). Sequenced ribosomal footprints show expected size distributions (Figure S1C), map primarily to coding sequences (CDS) of genes (Figure S1D) and display the 3-nt codon movement characteristic of actively translating ribosomes (Figure 1B and Figure S1E). To catalog translated sequences in the human heart, we created a *de novo* transcriptome assembly and performed an unsupervised search for actively translated ORFs using RiboTaper (Calviello et al., 2016) (Figure 1C, Figure S1F+G). Among the 22,335 identified ORFs are 1,090 uORFs (Figure 1D) and 339 sORFs in non-repetitive sequences of 169 presumed lncRNAs (Figure 1E). Compared to left ventricle protein identifications from the deepest human heart proteome to date (Doll et al., 2017), we infer translation from our Ribo-seq data for twice as many gene products (Figure 1F), possibly because highly expressed cardiac sarcomere proteins hinder the detection of lower expressed proteins by mass spectrometry (MS). In agreement with translation being the intermediate step between transcription and the proteome, Ribo-seq has a higher predictive value of final protein levels than mRNA-seq (Pearson's $r = 0.40$ vs 0.32 ; Figure S1H). All cardiac translation events can be found in Table S1 and have been compiled into an annotated search database for MS-based proteomics that can be downloaded from the shiny webserver.

Dissecting transcriptional and translational control in human tissue

Comparing DCM patients with controls, we detect 2,660 genes with differential mRNA expression levels and 2,648 genes with expression differences in the Ribo-seq data, of which 964 appear to have a transcriptional basis (Figure S2A and Table S2). To identify the fraction of genes that is specifically translationally regulated, we applied an interaction model that accounts for the transcriptional contribution to gene expression regulation (Chothani et al., 2017), yielding 327 translationally downregulated and 474 translationally upregulated genes (Table S2).

We next correlated the translation levels of all differentially expressed genes across the 80 hearts to find process-specific expression coregulation. This identified 30 clusters of jointly regulated genes, of which 22 are enriched for distinct cellular processes (Figure 2A and Figure S2B). To define the contribution of transcription and translation to the expression regulation of each cluster, we performed a principle component analysis (see STAR methods; Figure 2B). This revealed specific translational upregulation of extracellular matrix (ECM) production (Figure 2B+C), likely a manifestation of the hallmark fibrotic response to cardiac damage and failure (Travers et al., 2016). In addition, we find that the downregulation of mitochondrial processes is initiated during transcription and significantly enhanced on the translational level, reflecting the energy deficient state of the failing heart (Okonko and Shah, 2014). Sarcomere components are mostly transcriptionally controlled (Figure 2C), similar to many genes known to cause DCM (of which most encode sarcomere proteins) (Figure S2C).

Components of the mTOR signaling pathway, a known master regulator of cardiac translation (Sciarretta et al., 2018), locate to a gene cluster highly upregulated in DCM hearts (Figure S2B). As a consequence, the translation of 5' terminal oligopyrimidine (TOP) motif-containing mTOR target genes (Thoreen et al., 2012) is significantly upregulated ($p = 2.52 \times 10^{-7}$) (Figure

S2D+E). Most 5' TOP genes are ribosomal proteins and their translational upregulation, in response to decreased mRNA expression (Figure 2C), suggests autonomous control of the translation machinery over ribosome production, and subsequently translational activity, in diseased hearts.

Upstream ORFs influence translational efficiency independent of translation rates

We detect a total of 1,090 actively translated uORFs in 919 genes (8% of all translated genes) (Table S3), which display an expected decrease in translational efficiency (median TE 0.90 vs. 0.65; $p = 6.15 \times 10^{-28}$; Mann-Whitney U test) (Figure 2D and Figure S2F+G). Surprisingly, we mostly find no decreasing linear relationship (i.e. anticorrelation) between the translation rates of uORFs and primary ORFs, but an overall mildly positive correlation (Figure 2E and Table S3). This also holds true for uORFs that show overlap with the primary ORF start, or for uORFs that possess a particularly strong AUG context (e.g. a Kozak sequence (Kozak, 1987) or TISU element (Elfakess and Dikstein, 2008)) (Figure S2H). We then hypothesized that uORFs that are positionally conserved to other species may have a more profound impact on TE, as only 23 out of 1,090 uORFs display amino acid conservation (Lin et al., 2011). In translomes of rat and mouse hearts we find 281 human uORFs with translation initiation site conservation to rodent uORFs (Table S1 and S3), though these do not display a stronger impact on, or anticorrelation with, the primary ORF TE (Figure S2H).

These analyses illustrate that, for most uORFs, there is no detectable quantitative dependency between the frequency of uORF translation and the observed decrease in primary ORF TE. However, a handful of uORFs are differentially translated in DCM hearts and do anticorrelate with the primary ORF TE, including *ZMPSTE24* and *EIF4G2* (Figure 2F). *EIF4G2* contains a 5' UTR IRES and can autoregulate its own translation rates when cap-dependent translation is suppressed (Henis-Korenblit et al., 2000), potentially contributing to the observed anticorrelation. The regulation of *ZMPSTE24* may be of particular importance to cardiac physiology as *ZMPSTE24* specifically processes prelamin A (LMNA). Defective LMNA processing due to mutations in *ZMPSTE24* leads to pathological cardiac dilation, phenotypically identical to DCM caused by *LMNA* mutations (Galant et al., 2016; Pendás et al., 2002).

Naturally occurring genetic variation influences cardiac translation

The influence of natural genetic variation on translational regulation in human tissue has remained unexplored. We therefore identified single-nucleotide variants (SNVs) and small insertions/deletions (indels) from constitutive exons of cardiac expressed genes and tested their local association with mRNA abundance, ribosome occupancy and TE (see STAR methods, Figure S3A-C and Table S4). We detect variants associated with mRNA abundance of 421 genes ($FDR \leq 0.05$) (Figure S3A), with effects similar to GTEx (Aguet et al., 2017) and known left-ventricle eQTLs (Heinig et al., 2017). Most of these variants are not associated with ribosome occupancy (Table S4), consistent with extensive buffering observed previously in a similarly-sized cohort of HapMap lymphoblastoid cell lines (Battle et al., 2015; Cenik et al., 2015). Vice versa, we detect genetic associations with ribosome occupancy for 81 genes, of which variants in 31 genes are not associated with mRNA expression. Both observations indicate translational regulation and indeed, variants in 37 genes are significantly associated with altered gene TEs (Figure S3A; examples in Figure S3D). None of the exonic variants that are associated with TE are located in regulatory features such as uORFs or Kozak sequences, but 8 are predicted to affect RBP binding and RNA secondary structure (Table S4) (Mao et al., 2016).

Protein truncating variants frequently do not truncate proteins

PTVs can have dramatic consequences on protein function, yet medical relevance has only been established for a fraction of PTVs (DeBoever et al., 2017), potentially due to gene haplosufficiency, functional redundancy, or premature stop codon readthrough (Bartha et al., 2015; Huang et al., 2010; Jia et al., 2017). In our cohort we detect 346 potential PTVs: 144 nonsense mutations and 202 frameshift indels (Table S4, Figure S3E-G and STAR methods). For all detected PTVs, we analyzed mRNA allele ratios and Ribo-seq coverage to estimate allele specific expression (ASE) and the ability of the premature stop to terminate translation in the absence of complete nonsense-mediated decay (NMD). Only 32 out of the 346 detected PTVs (9.2%) display an allelic imbalance of heterozygous SNVs indicative of NMD (Figure 3A+B and Figure S3H), suggesting that many alleles with truncating mutations do not undergo extensive NMD. As a measure of premature translation termination, we calculated ribosome drop-off rates by comparing ribosome occupancy before and after the PTV (see STAR methods). For only 59 out of 346 PTVs (17.1%), ribosome occupancy is significantly lower downstream of the introduced stop than upstream (Figure 3C+D and Figure S3I). Thus, for most of the PTVs that can be detected at the RNA level, translation either appears to be terminated inefficiently or reinitiated downstream of the PTV, likely influencing the functional impact of these PTVs.

Truncated *TTN* alleles are translated

Truncating variants in the large sarcomeric protein titin (*TTN*tv) are the most prevalent cause of genetic DCM (Herman et al., 2012), albeit with variable penetrance and expressivity (McNally and Mestroni, 2017). In our cohort, 13 DCM patients have *TTN*tv located in different constitutive exons of *TTN* (Figure 3E and Table S4). In line with previous work in human hearts (Hinson et al., 2015; Roberts et al., 2015), but in contrast to two rat models with *Ttntv* (Schafer et al., 2017a), we find no compelling evidence for NMD in *TTN*tv carriers (Figure 3F). Based on heterozygous SNVs covered by ribosome footprints downstream of the *TTN*tv, premature translation termination is efficient for 4 of the 13 *TTN*tv carriers. For 4 other *TTN*tv carriers, translation appears to persist or reinitiate after the *TTN*tv (Figure 3G), sometimes reaching translated allele ratios close to canonical *TTN* translation (Figure 3H). The remaining 5 *TTN*tv carriers also do not display NMD, suggesting that both alleles are translated, but we lack sufficient Ribo-seq coverage at heterozygous variant positions to distinguish the mutated from the non-mutated allele.

Translation downstream of nonsense *TTN*tv likely results from stop codon suppression, since we observe ribosomes moving past these stop codons and in-frame without being released (Figure S3J). In contrast, translation downstream of frameshifting *TTN*tv likely results from reinitiation at IRESs or timely ribosomal frameshifting back into the primary *TTN* ORF. To test whether translation downstream of a *TTN*tv can lead to the production of stable *TTN*, we profiled the cardiac proteomes of two rat models that carry a genetically engineered, heterozygous frameshift in the *Ttn* Z-disc (*TtntvZ*) or A-band (*TtntvA*) of rat *Ttn* (Schafer et al., 2017a). To obtain allele-specific proteomes, we used F1 hybrids derived from mutant *Ttn* F344 rats and wild-type *Ttn* BN rats. *TtntvZ* alleles, but not *TtntvA* alleles, show translation signals downstream of the *Ttntv* (Schafer et al., 2017a) (Figure 3I) and concordantly, we detect in-frame peptides specific to the *TtntvZ* allele downstream of the *Ttntv* (Figure 3J). In contrast, *TtntvA* animals do not show any N- or C-terminal peptides specific to the mutant allele.

Our data illustrate extensive translational control of TTN production (Figure S3K-N). Not all TTNtv terminate translation efficiently and the rates of these translational patterns differ across mutations and individuals, raising the possibility that the impact of these TTNtv on cardiac function may vary as well.

Extensive translation of lncRNAs in human heart, liver and kidney

Proteins smaller than 100 amino acids (microproteins) have frequently been overlooked and their prevalence, regulation and putative function in human tissue remains largely unknown (Makarewich and Olson, 2017). To discover cardiac microproteins, we searched for actively translated sORFs in cardiac lncRNAs. Out of 783 transcribed lncRNAs, 169 (22%) are translated into potential microproteins with a median length of 49 aa (Figure 4A and Table S5). We independently validate these translation events in the translomes of primary cardiac fibroblasts (Chothani et al., 2018) and iPSC-derived cardiomyocytes (iPSC-CM). Known cardiac microproteins are accurately detected (190/199; 95%), including the recently discovered DWORF (Nelson et al., 2016), SPAR (Matsumoto et al., 2016) and ALN (also known as C4orf3) (Anderson et al., 2016a). Similar to DWORF, 16 out of the 169 translated lncRNAs are specifically expressed in heart or skeletal muscle tissue (Table S5), suggesting a muscle-specific function. To validate the translation potential of the identified sORFs, we performed *in vitro* translation (IVT) assays on complete transcripts of 58 randomly selected translated human lncRNAs, successfully producing microproteins for 44 of these (75%; Figure 4B; Figure S4A+B; Table S5). Subsequent start codon mutation prevented translation and caused a loss of signal in the predicted size range (Figure 4B and Figure S4A).

The expression of most translated lncRNAs (> 90%) is not restricted to the heart. In fact, 122 are expressed in at least 10 other tissues and 44 are expressed in all tissues profiled within the GTEx catalogue (Aguet et al., 2017). To address whether such translation also takes place in other tissues, we generated translomes of 6 human liver and 6 human kidney tissues. Out of the 169 lncRNAs detected as translated in human hearts, 71 (42%) and 116 (69%) are expressed in liver and kidney, respectively. Of these, 56 (liver) and 87 (kidney) are actively translated, with 50 lncRNAs translated in all three tissues (Figure S4C). Importantly, for most of these (85-91%), at least 1 sORF is exactly identical to the sORF detected in the heart. According to the sORFs.org database (Olexiouk et al., 2018), 72 sORFs in 51 translated lncRNAs have previously been detected as translated in human cell lines. Our data confirm the translation of these sORFs in human tissue and furthermore highlight the previously undetected translation of 272 sORFs in 118 lncRNAs.

In agreement with previous observations in human cell lines (Bazzini et al., 2014; Calviello et al., 2016), we only detect a few sORFs with strong amino acid conservation across vertebrates (17 sORFs in 12 lncRNAs) (Lin et al., 2011; Mackowiak et al., 2015). However, many lncRNAs can be aligned to other hominid species (chimpanzee, gorilla and orangutan) ($n = 79$) or to other primates or mammals ($n = 31$ or 43 , respectively), with only 16 being completely specific to humans (Table S5). In rat and mouse hearts, we find comparable lncRNA translation rates (13-22%; Figure 4C and Table S5), with TEs similar to mRNAs (Figure 4D and Figure S4D). Despite limited amino acid conservation, 76 out of 169 human translated lncRNAs are positionally conserved to rodents, i.e. they flank orthologous protein-coding genes with the same relative orientation (Ulitsky, 2016). Of these, 18 are also translated in rodents and 7 share the exact same translation initiation site (Table S5).

Detection of microproteins in human hearts *in vivo*

In vivo microprotein detection is challenging and searches in deep MS datasets using custom search databases can result in false-positive peptide identifications (Bánfai et al., 2012; Bazzini et al., 2014; Low et al., 2013; Mackowiak et al., 2015; Nesvizhskii, 2014; Omenn et al., 2017; Slavoff et al., 2013). Searching extremely deep human heart shotgun MS data (Doll et al., 2017) and a newly generated deep proteome of human iPSC-CM, we detect unique peptide evidence for microproteins translated from 140 out of 339 sORFs, encoded by 93 out of 169 translated lncRNAs (Table S5). For 28 microproteins we detect more than 1 unique peptide and 100 microproteins are detected in more than 1 sample (Table S5). To define the false-positive rate for these searches, we employed a target-decoy strategy followed by a statistical subsampling analysis, executed in addition to the reversed hit target-decoy strategy already implemented in MaxQuant (Cox and Mann, 2008; Elias and Gygi, 2010) (see STAR methods). Although we observe a clear enrichment of true microproteins over artificial ones (empirical p-value < 0.001; effect size = 5.99-7.57; Figure S4E), false-positive peptides could still be detected, reflecting a significant false-discovery rate of \pm 50-60%. For that reason, we next designed a high-throughput selected reaction monitoring (SRM) assay (Picotti et al., 2010). SRM is a highly sensitive targeted mass-spectrometry approach that uses synthetic signature peptides to detect the exact fragmentation patterns of precursor into fragment ions ("transitions"), thereby increasing the sensitivity and specificity of microprotein detection. In 5 human hearts (2 technical replicates each), we positively identify 76 out of 137 randomly selected microproteins (55.4%), translated from 50 out of 83 (60.2%) lncRNAs (Table S5). These results substantiate that many translated sORFs produce microproteins detectable *in vivo*, but illustrate that it is crucial to use a continuum of independent methods (Ribo-seq across samples, IVT assays, shotgun MS and SRM) to provide confidence in microprotein discovery.

Microproteins are produced from "noncoding" RNAs with known functions

Aided by improved transcript annotations (e.g. Figure 4E) we identify translated sORFs in 27 human and 5 mouse lncRNAs with previously assigned noncoding functions, including *LINC-PINT* (also known as *lincRNA-Mkl1*) (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* (Wang et al., 2016), *UPPERHAND* (also known as *UPH* or *HAND2-AS1*) (Anderson et al., 2016b), *ZFAS1* (Zhang et al., 2018b) and *TRDN-AS* (also known as *RP11-532N4.2*) (Zhang et al., 2018a) (Figure 4F, Figure S4F and Table S6). Of the aforementioned lncRNAs, *NEAT1*, *GATA6-AS1* and *UPPERHAND* are positionally conserved and translated in human and rodent hearts, and microproteins expressed from these lncRNAs can all be detected *in vivo* (Table S5). Twenty-two out of the 27 human lncRNAs with characterized noncoding functions are also detected as translated in human kidney and liver and we previously showed that some of these, including *DANCR*, were cytosolically located and associated with ribosomes in human cell lines (van Heesch et al., 2014; Mukherjee et al., 2017).

Driven by these findings, we looked at additional functionally characterized lncRNAs with a known cytosolic localization and ribosome association (Cabili et al., 2015; van Heesch et al., 2014), but no detected canonical AUG ORF. We detect a highly conserved CUG ORF (153 aa; PhyloCSF score = 350) in a previously misannotated 5' leader sequence of the lncRNA *TUG1* (Figure 4G). We validate *TUG1* translation *in vitro* (Figure 4H) and show that TUG1 protein localizes to both the nucleus and mitochondria, or either one of these compartments (Figure 4I). *TUG1* is ubiquitously translated in human and rodent tissues and *TUG1* overexpression drives a gene expression change that we can attribute to the TUG1 protein

(Figure S4G). Interestingly, full-gene ablation of the *Tug1* locus in mice results in male infertility with mid-piece defects, underscoring the importance of the *Tug1* locus independent of the molecular product (DNA, RNA or protein) (*John Rinn lab, personal communication*).

Expression regulation of translated lncRNAs in healthy and diseased hearts

Several of the above-described functionally characterized lncRNAs are genomic antisense lncRNAs reported to be involved in the *cis*-regulation of neighboring protein-coding genes (Anderson et al., 2016b; Han et al., 2014; Zhang et al., 2018a). We find 18 significantly correlating sense-antisense gene pairs (Spearman's rho 0.52 - 0.76; $p = 3.3 \times 10^{-5} - 1.9 \times 10^{-12}$) that involve translated lncRNAs antisense to major cardiac transcription factors (*HAND2*, *TBX5* and *GATA6*) and regulatory or structural cardiac genes (*CORIN*, *TRDN* and *TNNI3*) (Figure S4H). During translation, coregulation of most pairs decreases, with the exception of *TRDN-TRDN-AS1* (Spearman's rho 0.23 vs 0.53; $p = 0.0136$) (Figure S4I). *TRDN-AS1* was recently identified as a *cis* regulator of cardiac and skeletal muscle triadin production (Zhang et al., 2018a) and translational coregulation suggests co-functionality of both proteins. Of all translated lncRNAs, 34 are up- and 7 are downregulated in diseased hearts (Figure S4J+K and Table S2), warranting further investigation into the potential roles of these microproteins.

Microproteins localize to mitochondria and associate with mitochondrial processes

Gene expression correlation across samples can be an indication of functional coregulation (Saha et al., 2017). Clustering genome-wide expression correlations, we find significant enrichment of translated lncRNAs (93 out of 169; $p = 2.17 \times 10^{-15}$; Fisher's Exact Test) in a cluster dominated by nuclear-encoded mitochondrial genes (GO:0005739 mitochondrion; $p = 8.83 \times 10^{-149}$). Genes involved in oxidative phosphorylation (OXPHOS; KEGG:hsa00190; $p = 6.43 \times 10^{-40}$) (Figure 5A) are particularly strongly correlated with select translated lncRNAs (top 3 highlighted in Figure 5B).

For each translated lncRNA, we then compiled all coregulated protein-coding genes (Spearman's rho ≥ 0.5) and searched for functional commonalities. We associate 42 translated lncRNAs with distinct cellular processes (Figure S5A), of which 22 include mitochondrial functions (Figure 5C). We selected 3 out of these 22 microproteins to demonstrate a specific mitochondrial localization (Figure 5D). For 18 additional microproteins, which we predicted to be mitochondrial based on protein sequence features and/or expression coregulation (Table S5), we could also corroborate a mitochondrial localization (Figure S5B). These include microproteins translated from 4 snoRNA host genes (*GAS5*, *SNHG6*, *SNHG8* and *SNHG16*) and from a uORF-derived sORF in the lncRNA *JPX* (Hezroni et al., 2017), further establishing what seems to be a general tendency for many microproteins to be mitochondrially localized.

One mitochondrial microprotein, PDZRN3-AS1, is a 47 aa predicted single-pass transmembrane helix protein, for which we corroborate the helical structure by 3D modeling (Figure 6A). Using (co-) immunoprecipitation, (co-) localization and proteinase K digestion experiments, we show that PDZRN3-AS1 specifically interacts with RMND1 at the mitochondrial inner membrane (Figure 6B-E), where RMND1 is required for the translation of OXPHOS subunits (Janer et al., 2015).

Moreover, and distinct from mitochondrial processes, signal peptide cleavage site predictions suggest that not all microproteins remain in the cell (Figure S6A). We tested 2 potentially secreted microproteins (RP11-432J24.5 and AC093642.6) and indeed find interactions with other secreted proteins and components of the secretory pathway (Figure S6B). Next to

PDZRN3-AS1, multiple other microproteins are predicted to have a transmembrane helix, including SOX9-AS1, BANCR (Flockhart et al., 2012) and UPPERHAND (Anderson et al., 2016b) (Figure S6C and Table S5). Gene expression coregulation implicates UPPERHAND as an integral membrane component of the endoplasmic reticulum (ER) (Figure S6D). Indeed, UPPERHAND localizes to the ER (Figure S6E), where it interacts almost exclusively with membrane proteins (Figure S6F). Strikingly, the *UPPERHAND* lncRNA is strongly downregulated in primary cardiac fibroblasts upon TGF- β 1 stimulation (Chothani et al., 2018), displaying expression regulation opposite to the pro-fibrotic cytokine *IL-11* (Figure S6G) (Schafer et al., 2017b). Both siRNA-mediated knockdown of the *UPPERHAND* lncRNA and mutation of the endogenous UPPERHAND AUG result in increased expression of fibrotic marker genes (Figure S6H+I). The mechanistic basis for a potential antifibrotic role of UPPERHAND would need to be further established, but could possibly be mediated via direct interaction with TGF- β 1 (Miao et al., 2018) or via alleviation of ER stress and the unfolded protein response, known enhancers of fibrosis (Heindryckx et al., 2016; Tanjore et al., 2013).

Translation of human cardiac circRNAs

In addition to lncRNAs, circRNAs are another class of noncoding RNAs with the potential to be translated (Legnini et al., 2017; Pamudurti et al., 2017; Yang et al., 2017). We detect 8,878 human heart circRNAs in 3,181 genes (Table S7; Figure S7A-C), of which 2,070 had previously not been detected (Table S7) (Glažar et al., 2014; Khan et al., 2016). Strikingly, we detect ribosome association, and thus possible protein translation, of 40 circRNAs produced from 39 genes (Figure 7A; Table S7). These circRNAs are largely present in CircBase (85%; Figure 7B) (Glažar et al., 2014) and show increased resistance to RNase R (Figure S7D and Table S7). To ensure specific alignment of Ribo-seq reads at circRNA backsplice junctions, we aligned the Ribo-seq data to matched sets of simulated backsplice junctions to show that circRNA-ribosome association does not arise by chance ($p_{\text{adj}} = 1.5 \times 10^{-4}$) (Figure 7C). Examples of newly detected ribosome-associated circRNAs include the highly occupied *circCFLAR* (Figure 7D) and the cardiac-specific circRNAs *circSLC8A1* (Figure 7E), *circMYBPC3* and *circRYR2*. Moreover, one of the ribosome-associated circRNAs is the microRNA sponge *circCDR1-AS*, and 5 others have been previously reported as translated in human cells (Yang et al., 2017). Importantly, we detect *in vivo* peptide evidence diagnostic for the translated backsplice junctions for 6 out of 40 circRNAs in shotgun MS data (Figure 7F and Table S7). Although this indicates that the identified ribosome-associated cardiac circRNAs can produce detectable peptides, further tailored detection and targeting strategies are required to reassure their identification and establish potential functional roles.

Discussion

Dissecting transcriptional and translational control in human tissue

Translational regulation has a prominent process- and pathway-specific role in shaping gene expression in human hearts. We specifically highlight a role for mTOR signaling in end-stage cardiac dilation, which has previously been implicated as a global regulator of cardiac translation in mouse models with genetic cardiomyopathies (Sciarretta et al., 2018), but only incidentally in humans (Yano et al., 2016). We furthermore find that uORF and primary ORF translation rates are generally not anticorrelated, an observation corroborated by similarly positive correlations in yeast, fruit flies and mammalian cells (Aspden et al., 2014; Brar et al., 2012; Chew et al., 2016). It is possible that uORF peptides directly interfere with the translation machinery to act as a structural roadblock (Lovett and Rogers, 1996), reducing the need for a quantitative dependency. Despite limited sequence conservation we detect unique peptides for 29% of all uORFs (316 out of 1,090; Table S3) (Doll et al., 2017), suggesting a structural or regulatory function less dependent on sequence.

Protein truncating variants frequently do not truncate proteins

We show that genetic variants can influence translation in human tissues and that many mRNAs with PTVs escape NMD and inefficiently terminate translation – two tightly related processes (Keeling et al., 2004). Importantly, actively translating ribosomes show identical in-frame codon movement before and after each PTV, indicating that only the primary reading frame is being translated and downstream ribosome association is not stochastic. Mechanistic differences between PTVs and canonical stop codons could explain why translation termination at PTVs is less efficient (Amrani et al., 2004; Loughran et al., 2014; Peixeiro et al., 2012; Raimondeau et al., 2018), although readthrough also occurs at canonical stop codons (Dunn et al., 2013). Despite considerable attempts we did not identify any motifs or sequences that could facilitate stop codon suppression or IRES-mediated translation reinitiation.

Efficient translation termination or reinitiation could lead to the production of truncated protein, which can have beneficial or damaging physiological consequences. Recently, an IRES in the *DMD* gene, located downstream of *DMD* truncating mutations, was shown to produce highly functional N-terminally truncated dystrophin, thereby attenuating dystrophinopathy (Wein et al., 2014). Similarly, *TTN* isoforms that lack the N-terminal Z-disc and large parts of the *TTN* I-band may be capable of rescuing part of *TTN*'s functionality (Deo, 2016; Zou et al., 2015). In contrast, the production of N-terminally truncated protein can have deleterious (dominant-negative) effects, as previously proposed for C-terminally truncated *TTN* (Herman et al., 2012) and cardiac troponin T (Watkins et al., 1996).

Previously unrecognized transcript isoforms may also lead to apparent translational read-through or reinitiation. The recently discovered *TTN Cronos* isoform (Zou et al., 2015) could theoretically contribute to downstream ribosome occupancy at 2 out of 4 *TTN*tv alleles (Table S4). However, *Cronos* appears lowly expressed in adult hearts and heterozygous positions in the Ribo-seq data can already be detected prior to the *Cronos* start in both human (Figure 3H) and rat hearts (Figure 3I).

It should be noted that all 13 DCM patients with *TTN*tv were end-stage heart failure patients and we could not determine if any translational signals contributed to alterations in disease progression or severity, since they were collected retrospectively. We propose that, amongst

other factors, *TTN* translation dynamics are likely to contribute to the variable expressivity of *TTN* in genetic DCM and the general population. Future studies on phenotypically silent *TTN* carriers, which are frequent in the general population (Schafer et al., 2017a), will be necessary to assess the full scope and consequences of such varying translation activities at mutated *TTN* alleles.

Cardiac lncRNAs produce microproteins detectable *in vivo*

In this study, we detail a discovery, validation and characterization pipeline for previously undetected microproteins in human tissue, which we find to be widespread in human heart, liver and kidney. In contrast to some previous efforts (reviewed in (Makarewich and Olson, 2017)), our *in vivo* microprotein detection pipeline functions independent of sequence conservation, expanding on detection methods previously developed by us and others (Bazzini et al., 2014; Calviello et al., 2016; Mackowiak et al., 2015). This is an important consideration, as limited conservation does not exclude the production of functional microproteins. Rather, lowly conserved proteins may represent evolutionary young genes (Ruiz-Orera et al., 2018) and provide insights into recently evolved human- or primate-specific proteins.

Employing both experimental and computational analyses, we find that many microproteins are conserved amongst primates and can be linked to the mitochondrion – an organelle for which functionally relevant microproteins have been described (Makarewich et al., 2018; Rathore et al., 2018; Stein et al., 2018). We find a particularly strong microprotein expression coregulation with OXPHOS subunits, including multiple small accessory, also known as supernumerary OXPHOS proteins (Zickermann et al., 2010). Interestingly, these accessory proteins are evolutionary dynamic and show variable conservation across eukaryotes (Elurbe and Huynen, 2016), raising the possibility that some of the newly discovered microproteins may have similar functions. Our results position the mitochondrion as a potential evolutionary playground for a subset of recently evolved small proteins, either facilitated by yet unknown localization signals or import systems, or purely driven by microprotein size or (positively charged) amino acid composition (Couso and Patraquim, 2017). For instance, the mitochondrial import and folding protein CHCHD4 (also known as Mia40) may mediate microprotein import as it shows a strong preference for low molecular weight substrates with simple helix-loop-helix structures, connected by parallel disulfide bonds (Backes and Herrmann, 2018; Banci et al., 2009).

Strikingly, multiple microproteins are translated from lncRNAs with well-characterized noncoding roles in health and disease. The coding potential of these lncRNAs was unknown and improved transcript annotations aided in the detection of previously unnoticed sORFs (e.g. *BANCR* and *TUG1*; Figure 4E+G). A prominent example of a positionally conserved, translated and functionally characterized lncRNA is *UPPERHAND*. In mice, *upperhand* transcription rather than the transcriptional product itself was shown to regulate *Hand2* in *cis* (Anderson et al., 2016b) and mature *upperhand* transcripts were found to dominantly localize to the cytosol without clear functionality (Anderson et al., 2016b; Kopp and Mendell, 2018). We identify *UPPERHAND* protein isoforms of up to 134 aa (Figure S4F and Table S5), including a predicted single-pass transmembrane microprotein that interacts with other membrane proteins and localizes to the ER, where it may exert anti-fibrotic properties. Importantly, the mRNA expression and translation of *myheart* (Han et al., 2014) and *chaer* (Wang et al., 2016) could, despite previously claimed conservation to humans, only be detected in rodent hearts (Table S6).

We show that a number of translated lncRNAs are likely to possess both coding and noncoding roles. Such dual roles have been proposed previously (Rinn and Chang, 2012) and are known to exist for several mRNAs (Jenny, 2006; Leygue, 2007) and lncRNAs (Anderson et al., 2015; Yu et al., 2017). Clearly, categorizing genes into coding or noncoding classes based on criteria such as ORF length and sequence conservation would benefit from alternative methods based e.g. on RNA metabolism profiles (Mukherjee et al., 2017). Although dual roles complicate proper dissection of the mechanistic function of any gene, such multifunctionality likely forms a more truthful representation of biological complexity, creating opportunities for exploring the relevance of these microproteins in human health and disease.

Acknowledgments

S.v.H. was supported by an EMBO Long-Term Fellowship (ALTF 186-2015; LTFCOFUND2013, GA-2013-609409). N.H. is recipient of an ERC advanced grant under the European Union's Horizon 2020 research and innovation programme (grant agreement n° AdG788970). C.E.S., J.G.S., S.A.C., and N.H. are supported by a grant from the Leducq Foundation (11 CVD-01). N.H., M.L. and U.O. were supported by the "Bundesministerium für Bildung und Forschung" under grant CaRNAtion. W.A.L. was supported by the German Research Foundation (SFB1002, TPA08). S.A.C., P.J.R.B. and L.E.F. received support from the British Heart Foundation, Medical Research Council (UK), the NIHR Imperial College Biomedical Research Centre, the NIHR Cardiovascular BRU of Royal Brompton & Harefield NHS Foundation Trust UK, the Imperial College Academic Health Science Centre and Heart Research UK. N.M.Q. is funded by the Imperial College Academic Health Science Centre. M.Ka. received funding provided by the Alexander von Humboldt Foundation. LC and UO were supported by the German Federal Ministry of Education and Research (RNA Bioinformatics Center of the German Network for Bioinformatics Infrastructure [de.NBI; BMBF 031 A538C]).

Author Contributions

Conceptualization, S.v.H. and N.H.; Methodology, S.v.H., F.W., V.S.L., J.F.S., E.A., L.C., U.O. and N.H.; Software, F.W., V.S.L., A.B.F., C.L.W., B.O., M.Sc., L.C. and U.O.; Validation, J.F.S., C.L.S., S.B., S.v.H., E.W., R.M., M.Ki., S.M., E.L-H.L., F.T. and P.M.; Formal Analysis, F.W., V.S.L., A.B.F., C.L.W., G.P., M.V., and O.H.; Investigation, S.v.H., J.F.S., E.A., S.B., M.Ki., M.Ka., N.M.Q., K.V., M.B.M., Y.K., M.L., N.R., F.W. and V.S.L.; Resources, M.H., F.W.A., A.V., R.d.W., C.K., W.A.L., C.d.R., H.Mi., A.G.R., P.J.R.B., L.E.F., J.G.S., C.E.S., M.Ka., M.M., N.M., S-J.O., A.K., M.Se., C.S., D.S., S.D., S.S., S.V., A.A.W. and S.A.C.; Data Curation, F.W., V.S.L., C.L.W., A.B.F. and S.v.H.; Writing - Original Draft, S.v.H.; Writing - Review & Editing, S.v.H., H.Ma., C.L.S., J.F.S., F.W., V.S.L., A.B.F., C.L.W., S.A.C. and N.H., with input from all authors; Visualization, S.v.H., F.W., V.S.L., A.B.F., J.F.S., C.L.W. and E.A.; Supervision, S.v.H. and N.H.; Project Administration, S.v.H. and N.H.; Funding Acquisition, S.v.H. and N.H.

Declaration of Interests

The authors declare no competing interests.

Main Figure Titles and Legends

Figure 1) A snapshot of active translation in 80 human hearts. (A) Schematic overview of the experimental approach. (B) Bar plot displaying the P-site positions derived from ribosome footprints across the first 100 nt of annotated ORFs (left) and the percentage of footprints that match these primary reading frames (right). (C) Schematic overview of ORF detection by RiboTaper (Calviello et al., 2016). (D) Summary of detected upstream ORFs (uORFs), primary ORFs and downstream ORFs (dORFs). (E) Stacked bar plot displaying numbers and percentages of translated noncoding RNAs separated by gene biotype. Green biotypes represent the lncRNA fraction. (F) Venn diagram of detected gene products in the cardiac transcriptome, translome and proteome. Left-ventricular protein detections were obtained from (Doll et al., 2017). See also Figure S1 and Table S1.

Figure 2) Dissecting transcriptional and translational control in human tissue. (A) Heatmap with hierarchically clustered expression correlations (Ribo-seq; Spearman's rho) of differentially transcribed and translated genes. Clusters with enrichment for GO or KEGG are highlighted. (B) Principle component analysis separating transcriptionally and translationally controlled gene clusters. Each dot represents a cluster. Numbers of clusters correspond to the position of the cluster in Figure 2A. (C) Examples of functionally coregulated gene clusters. Every bar represents a gene. Genes related to the top GO term are visualized at the top of each panel and the top-5 GO terms and corresponding p-values are given. Genes are sorted based on the extent of translational regulation. (D) Violin box plot with a TE comparison (Mann-Whitney U) between 919 mRNAs with a uORF and 6,769 translated mRNAs without uORF, though with matched 5' UTR lengths (\pm 2nt) and 5' UTR GC content (\pm 5%). (E) Beeswarm dot plot showing the correlation between uORF occupancy and primary ORF TE for individual genes, across all 80 individuals. Genes highlighted in red possess uORFs that are differentially expressed in DCM patients and have opposite fold changes compared to the primary ORF. (F) Expression comparison and correlation scatter for *ZMPSTE24* and *EIF4G2*. Line and whiskers indicate the mean \pm SD. See also Figure S2, Tables S2 and S3.

Figure 3) The impact of protein truncating variants on cardiac translation. (A) Density plots with allele ratios of genes with PTVs that do (left) or do not (right) display allele-specific expression (ASE) as a result of nonsense-mediated decay (NMD). (B) Examples of PTVs that do (*FASTKD1*; left) or do not (*SLC26A10*; right) result in ASE ($*p < 0.05$). (C) Density plots showing the Ribo-seq coverage-based ribosome drop-off score (see STAR methods), grouping PTVs that do (left) or do not (right) show a significant decrease in ribosome coverage after the PTV. (D) Examples of genes with PTVs that do (*TLR5*; left) or do not (*SMPD4*; right) lead to a decrease in ribosome occupancy ($*p < 0.05$). (E) Schematic representation of multiple cardiac *TTN* isoforms and the positions of the 13 TTNtv. (F) Box plots showing *TTN* mRNA-seq RPKMs and mean SNV allele ratios, illustrating that TTNtv do not induce NMD. Whiskers indicate the 10-90 percentile. (G) Measured allele ratios in TTNtv carriers with and without translation downstream of the TTNtv. Allele ratios are shown on a het (0.5) to hom (0 or 1) axis. Each dot represents a genetic variant. (H) Examples of TTNtv carriers with and without translation downstream of the TTNtv. The red vertical line indicates the position of the TTNtv and the pink area contains the variant allele-ratios after the TTNtv. (I) Ribo-seq allele ratios of F1 hybrid F344 x BN rat hearts indicate the translation of Ttn tv alleles (data obtained from (Schafer et al., 2017a)). The trendline uses a moving average of 2. (J) MS-based detection of Ttn protein in the same F1 hybrid F344 x BN rat hearts as used for Ribo-seq. MS intensities

are given for 16 F344-specific peptides that cover missense mutations in *Ttn*, and their corresponding wild type BN peptides. See also Figure S3 and Table S4.

Figure 4) Cardiac lncRNAs produce microproteins detectable *in vivo*. (A) Histogram showing the length distribution (in aa) of newly detected short ORFs in translated lncRNAs. (B) Schematic overview of the *in vitro* translation (IVT) and AUG mutagenesis of lncRNA sORFs, with a representative radiolabeled blot on the right (trimmed at 20 kDa). Predicted microprotein molecular weights are indicated in kDa. Additional assays can be found in Figure S4A. (C) Donut chart with the number of translated lncRNAs in human, mouse and rat hearts. (D) Violin box plots with lncRNA and mRNA translational efficiencies across species. (E) Genomic view of the *BANCR* locus (reverse strand) and the translated sORF. The sORF starts upstream of the previously annotated *BANCR* gene start. (F) IVT assays for functionally characterized lncRNAs. Predicted microprotein molecular weights are indicated in kDa. Full size blots can be found in Figure S4A. (G) Genomic view of the *TUG1* locus, with PhyloCSF and PhyloP tracks displaying amino acid and nucleotide conservation. The identified *TUG1* sORF is located in a *de novo* annotated 5' region of *TUG1*. (H) IVT assay of unmodified and CUG-mutated *TUG1* lncRNA, which includes the endogenous 5' UTR and a fraction of the 3' UTR. The predicted molecular weight is indicated in kDa. (I) Immunofluorescence (IF) of *TUG1* in HeLa cells. The scale bar represents 20 μ M. See also Figure S4, Tables S5 and S6.

Figure 5) Microproteins localize to mitochondria and associate with mitochondrial processes. (A) Heatmap with genome-wide gene-gene expression correlations (Spearman's rho). Co-clustering translated lncRNAs (black) and OXPHOS subunits (green) are highlighted. (B) Circos plot displaying the 3 translated lncRNAs with the strongest expression correlation with OXPHOS proteins. Each connection represents an expression correlation of ≥ 0.5 . (C) GO term clustering of 22 microproteins coregulated with mitochondrial processes (see also Figure S5A; cluster A). (D) Immunofluorescence (IF) depicting the mitochondrial localization of microproteins (selected from Figure 5C; additional examples in Figure S5B). Scale bars represent 20 μ M. Predicted alpha-helical 3D structures are modelled by I-Tasser. See also Figure S5, Tables S5 and S6.

Figure 6) Characterization of microproteins with predicted functional domains. (A) Structure prediction of PDZRN3-AS1 indicating an alpha-helical single-pass transmembrane protein with an outward-facing C-terminus. (B) Volcano plot with IP-MS results for PDZRN3-AS1, identifying RMND1 as the most highly enriched interaction partner ($p = 4.5 \times 10^{-5}$; two-sided *t*-test; FDR ≤ 0.005). Significant interaction partners (FDR 0.005; yellow dots) are enriched for mitochondrial proteins (GO: mitochondrial part: $p = 0.0257$). (C) Reciprocal co-IP of FLAG-tagged PDZRN3-AS1 and RMND1-HA. I: input, U: unbound fraction, E: eluate, IP: immunoprecipitation, WB: western blot, HA: anti-HA antibody, FLAG: anti-FLAG antibody. The presence of HA-tagged RMND1 in the eluate of the PDZRN3-AS1-3xFLAG IP, or *vice versa* PDZRN3-AS1-3xFLAG in the eluate of the RMND1-HA IP, indicate interaction of both overexpressed proteins. (D) IF depicting the colocalization of PDZRN3-AS1-3xFLAG and RMND1-HA. The scale bar represents 20 μ M. (E) Western blot with mitochondrial digestion results upon increasing concentrations of proteinase K (PK). Known outer- (OMM; VDAC1 and TOM20) and inner mitochondrial membrane proteins (IMM; LETM1 and RMND1) are shown; COX4 functions as a protein-length matched IMM control to PDZRN3-AS1. See also Figure S6, Tables S5 and S6.

Figure 7) Translation of human cardiac circRNAs. **(A)** Schematic overview of ribosome association detection at circRNA backsplice junctions. **(B)** Venn diagram showing the overlap in identified human left-ventricular circRNAs with CircBase (Glažar et al., 2014) and previously identified human heart circRNAs (Khan et al., 2016). **(C)** Simulation analysis results highlighting the specificity of Ribo-seq read mapping to true vs simulated backsplice junctions (empirical $p_{\text{adj}} = 1.5 \times 10^{-4}$; see STAR methods). On the x-axis, the minimum required overhang of the Ribo-seq read to map to the backsplice junction is indicated. **(D)** Ribosome occupancy (min. 9 nt overhang) at the backsplice junction of *circCFLAR*. **(E)** mRNA-seq and Ribo-seq read mapping to the *SLC8A1* gene. *SLC8A1* exon 2 forms the ribosome-associated *circSLC8A1*, which is elevated in the totRNA-seq data ($p_{\text{adj}} = 9.4 \times 10^{-28}$) and slightly more occupied by ribosomes ($p_{\text{adj}} = 0.018$; Mann-Whitney *U*) compared to coding exons specific to the linear isoform. **(F)** Table with ribosome-associated circRNAs. CircRNAs with *in vivo* shotgun MS peptide evidence are highlighted in blue. See also Figure S7 and Table S7.

Supplemental Figure Titles and Legends

Figure S1: A snapshot of active translation in 80 human hearts, related to Figure 1. (A)

Dot plot displaying the fraction of raw sequence reads derived from tRNAs, the mitochondrial genome (MT), ribosomal RNA (rRNA) and canonical genes (clean) for human left ventricle Ribo-seq and mRNA-seq data. Only the 'cleaned' reads are used for subsequent data analyses. The mRNA-seq reads are trimmed to footprint size, filtered and further processed identically to the Ribo-seq reads. **(B)** Histograms showing the expression level of genes as measured by mRNA-seq. Expression levels of all genes across all 80 individuals are included. Genes that met our RNA expression cutoff of 1 RPKM are colored red. **(C)** Beeswarm plot visualizing the sequenced ribosome footprint lengths across all 80 samples. **(D)** Bar plot showing the percentage of reads mapping to the coding sequence (CDS) and untranslated regions (5' and 3' UTR) of annotated protein-coding genes. Each line represents a separate sample. **(E)** Bar plot showing the percentage of in-frame reads for all 80 Ribo-seq libraries. This percentage illustrates the codon movement of actively translating ribosomes along the coding sequence of an mRNA, and indicates the sensitivity and efficiency with which these Ribo-seq reads can be used for ORF identification. **(F)** Histograms showing the expression level of genes as measured by Ribo-seq. Expression levels of all genes across all 80 individuals are included. Genes that met our RNA expression cutoff of 1 RPKM and were subsequently identified as actively translated by RiboTaper are colored red. **(G)** Bar plot of the number of ORFs identified by RiboTaper in each of the 80 samples, illustrating the contribution of each of the 80 libraries to ORF annotation in the human heart. **(H)** Gene-based scatterplot showing the correlation between mean mRNA-seq (blue) or Ribo-seq expression levels (red) and protein IBAQ values derived from public left ventricle MS data (Doll et al., 2017). Correlation coefficients are Pearson's r values. Expression measurements of ribosome footprints are more highly correlated with protein levels in comparison to mRNA-seq data. Parts of the variance in protein abundance remain unexplained as protein levels are not only defined by transcription and translation rates, but also by post-translational variables such as protein stability. Additionally, differences in biological samples (different human hearts) and technologies (sequencing vs. mass spectrometry) influence the presented correlations.

Figure S2) Dissecting transcriptional and translational control in human tissue, related to Figure 2. (A)

Fold-change - fold-change (FC/FC) scatter plot depicting all genome-wide significantly differentially expressed genes between diseased (DCM) and unaffected control hearts, as measured by mRNA-seq and/or Ribo-seq. **(B)** Correlating the translation levels of all differentially transcribed and translated genes, we identified 30 coregulated gene clusters of which 22 were significantly enriched for specific GO terms or KEGG pathways. This figure depicts the genes of 4 coregulated clusters with significant GO or KEGG terms. The cluster numbers relate to the position of the cluster in Fig. 2A, from top to bottom. For each cluster, the name is given based on the overarching GO term. The top-5 GO terms are plotted on the side with corresponding p-values. Genes related to the most significant GO term are visualized as dark-grey dots on top of each panel. All genes in a cluster are sorted based on the delta FC (DCM vs non-DCM control) of translation (Ribo) over transcription (RNA). For each gene, the corresponding transcriptional and translational expression fold changes are depicted below the delta FC plot. Color coding indicates the intensity of the expression change, with genes upregulated in DCM in red and genes downregulated in blue. The mean translational (Ribo) FC of the entire cluster illustrates the general trend in expression regulation between controls and disease. **(C)** Transcriptional and translational regulation of genes previously shown to

cause DCM, listing genes collected from published work (Hershberger et al., 2013; McNally and Mestroni, 2017; Tayal et al., 2017) that are highly confidently labeled by the respective papers to contribute to DCM. Genes are separated by transcriptional regulation (green), translational downregulation (blue) and translational upregulation (red). Examples with the expression data of all 65 DCM and 15 control samples are given on the right of each category. The expression of the majority of genes appears to be purely transcriptionally regulated, with phospholamban (*PLN*) and desmoplakin (*DSP*) being interesting exceptions. **(D)** Violin plot showing the increased translational efficiency (TE) of 79 genes harboring a 5' TOP motif in DCM patients in comparison to non-DCM controls. This trend is not visible in 2,354 translated mRNAs with matched 5' UTR length (+/- 2nt) and 5' UTR GC content (+/- 5%). The p-value is calculated using a Mann-Whitney *U* test. **(E)** Exemplary comparison of 5' TOP mRNA translational efficiencies (TE) between DCM sample 1 and non-DCM control 79, illustrating an increase in TE for 5' TOP mRNAs in DCM hearts. **(F)** Scatter plot showing the mRNA expression and translation levels of genes (black) with and without uORFs (grey). Genes with uORFs show reduced translational activity in comparison to genes without uORFs. **(G)** Violin dot plots showing the TE for the three main cardiac troponins, of which only *TNNT2* has a uORF, likely contributing to the decrease in TE. **(H)** Histograms showing the results of a statistical sampling analysis to test the effect of specific uORF features on the TE of the primary ORF, hypothesizing that these features result in a stronger effect on the primary ORF. The features include (i) the presence of a (short) Kozak sequence ([A|G]CCATGG; 20 uORFs), (ii) the presence of a TISU element (14 uORFs), (iii) direct sense overlap of the uORF with the primary ORF (201 uORFs) and (iv) translation initiation site (TIS) conservation with uORFs in rat and/or mouse hearts. For fair comparison, we sampled 10,000 random and size-matched sets of uORFs without these features, and compared their effects on the primary ORF TE (top histograms) and on the correlation between uORF occupancy and primary ORF TE across the 80 heart samples (bottom histograms). The distribution of the resulting p-values (Mann-Whitney *U* test) is plotted, with a red dashed line indicating the median p-value result from these 10,000 tests.

Figure S3) The impact of protein truncating variants on cardiac translation, related to Figure 3. **(A)** Table with total numbers of identified exonic SNVs and Indels separated by mutation effect and location within the gene. Gene-based results of the association testing are shown (see also Table S4). **(B)** Variant calling concordance rates between the RNA-seq variant calls presented in this paper and previous DNA genotyping performed for 31 of the 80 samples. **(C)** Table showing a comparison between characteristics of genetic variants identified in this study and the Exome Aggregation Consortium (ExAC) database (Lek et al., 2016). Variant identifications in this paper are restricted to expressed genes (RPKM \geq 1) and not all annotated exons genome wide. Ti/tv = ratio of transition over transversion events; Het/Hom = ratio of heterozygous over homozygous variants; Ins/Del = ratio of insertions over deletions. **(D)** Association testing results for *ZNF880* and *RMDN1*, showing transcript levels (blue), translation levels (pink) and TEs (green) separated by genotype. Adjusted p-values indicate the significance of the associations. **(E)** Most detected PTVs (236 out of 346) have known variant IDs. This scatter plot shows their minor allele frequencies (MAFs) as calculated across 80 individuals in this study and ExAC. Despite the limited number of genotyped individuals here, the MAFs show concordance with reported frequencies. **(F)** Bar plot showing the distribution of PTVs and other variant types between the first CDS exon, the last exon + 55nt upstream of the last exon-exon junction (following the NMD rule described by (Nagy and Maquat, 1998), and any other exon. PTVs are enriched in the last exon of the canonical CDS (chi-squared test), possibly because these do not induce NMD and may have reduced

functional impact on the protein. **(G)** Density plots with allele ratios for PTVs in the last exon + 55nt upstream of that exon versus PTVs in any other exon (left plot); $p = 0.004$; Kolmogorov-Smirnov test). Correcting for fluctuations in ASE in non-PTV individuals via the ASE score (see STAR methods), this effect reduces, but both distributions remain significantly different (Kolmogorov-Smirnov test; $p = 0.006$). PTVs in this final exon show little NMD, but since most PTVs do not seem to induce NMD efficiently, the difference between both categories is marginal. **(H)** Histogram with ASE scores (see STAR Methods) of all 346 PTVs. PTVs that induce ASE are depicted in blue. The separation of both groups as presented in Fig. 3A is based on the threshold applied in this histogram. **(I)** Histogram with ribosome drop-off scores (see STAR Methods) of all 346 PTVs. PTVs that show decreased translation rates downstream of the premature stop codon are depicted in blue. The separation of both groups as presented in Fig. 3C is based on the threshold applied in this histogram. **(J)** Zoomed view of ribosome movement at a heterozygous *TTN* stop gain in sample #5, which displays continued translation after the TTNtv (Table S4). Ribosome footprints are separated by presence of the TTNtv and based on that colored by allele (WT or TTNtv). The TTNtv itself (CGA > TGA) is marked by a red cross in reads that carry the TTNtv. The P-site position of the ribosome gives an indication of the ribosome position upon stop codon encounter. At least two canonically-sized ribosome footprints (on top) have moved past the exit (E-) site with several others having passed the stop codon-recognizing A-site and the peptidyl (P-) site, all of which suggests inefficient ribosome release and translation termination. **(K)** Percent spliced-in (PSI) plot for *TTN* showing the mean exon inclusion and exclusion across all 80 samples measured by mRNA-seq and Ribo-seq. This figure illustrates that exons that are part of the *TTN* N2BA isoform (mostly located in the I-band) are particularly inefficiently translated. **(L)** Box plots showing the mean PSI values for I-band exons of *TTN* N2B and N2BA, as measured across all 80 samples by mRNA-seq and Ribo-seq. This supplemental figure shows that exons exclusive to *TTN* N2BA are much less efficiently translated than constitutive *TTN* exons also present in *TTN* N2B. P-values are calculated with a Mann-Whitney *U* test. **(M)** PSI plot of *TTN* exon inclusion and exclusion as measured by mRNA-seq and Ribo-seq in mouse hearts, to illustrate that the reduced translational efficiency of *TTN* N2BA is conserved to rodents. **(N)** PSI plot of *TTN* exon inclusion and exclusion as measured by mRNA-seq and Ribo-seq, including PSI values calculated after mapping mRNA-seq reads with a read length (29nt) similar to Ribo-seq reads, in order to exclude a mapping bias potentially caused by repetitive immunoglobulin-like and PEVK domains in the I-band region of *TTN*.

Figure S4) Cardiac lncRNAs produce microproteins detectable *in vivo*, related to Figure 4. **(A)** *In vitro* translation (IVT) assay results, including ATG knockouts, for all candidates subjected to ATG mutagenesis. The first gel contains myoregulin (MRLN) and DWORF as positive controls, and each gel contains a lane with empty vector as a negative control. Predicted microprotein sizes in kilodaltons (kDa) are given below each gel. Gels were trimmed at the 20 kDa size and thus display products in a size range between 2 and 20 kDa. *In vitro* translation of lncRNAs marked with a red asterisk were repeated and displayed again on the last 3 gels at the bottom right of the panel. IVT assays of lncRNAs that did not produce a product in the first round of testing, and were thus not subjected to ATG mutagenesis, are not shown but can be found in Table S5. **(B)** Dot plot showing that the detection of microproteins with IVT assays depends on the produced protein length ($p = 0.0006$; Mann-Whitney *U* test), with smaller proteins being less likely to be detected after radiolabeling, with an apparent detectability threshold of 2.5 kDa. **(C)** Heatmap displaying Ribo-seq expression levels (scaled DEseq2-normalized counts) for 50 lncRNAs with sORFs detected to be translated in human heart, liver and kidney tissue. For selected examples, normalized expression levels for across

tissues and samples are given on the right. **(D)** Scatter plots showing the transcription and translation levels of translated mRNAs (blue) and translated lncRNAs (red) in human, mouse and rat left ventricles. Like other lncRNAs, translated lncRNAs are mostly expressed less. However, their translational efficiencies are generally similar to mRNAs. Pearson's r correlation coefficients are given. **(E)** Subsampling results showing unique peptide identifications for a target-decoy shotgun MS data search for simulated, untranslated lncRNA sORFs (1,000 sets of 339 sampled sORFs; grey density plot; see STAR methods) versus the 339 true translated sORFs (red dashed arrow). A comparison is shown for identification results on the sORF level (left panel) and the gene (lncRNA) level. For both simulations, a strong enrichment is visible resulting in an empirical p-value of 0.001, as none of the simulated sets results in more hits than the actual sORF set. Effect sizes (Cohen's d) are given and show an increase for gene-level identifications. **(F)** P-site plots illustrating the 3-nt codon movement in ribosome profiling data at 3 newly predicted sORFs in lncRNAs *LINC-PINT*, *DANCR* and *UPPERHAND (HAND2-AS1)*. In-frame P-sites are colored blue, out of frame P-sites are colored red and green for +1 and +2 frames, respectively. **(G)** Gene expression analysis results upon overexpression of (i) empty vector, (ii) *TUG1* lncRNA and (iii) *TUG1-3xFLAG* in HEK293T cells. Expression was measured by RNA-seq of triplicate transfections of each construct. The *TUG1* lncRNA construct contains the endogenous 5' UTR, *TUG1* ORF with CUG start codon and 321 nt of the endogenous 3' UTR, which was trimmed due to size restrictions. The *TUG1-3xFLAG* construct contains the codon-optimized *TUG1* ORF and a C-terminal 3xFLAG. The highly concordant expression changes for both the lncRNA sequence and the *TUG1* ORF indicate that the translated *TUG1* protein is sufficient to trigger this response. **(H)** Sense-antisense correlations of translated lncRNAs with neighboring protein-coding genes, as defined by Spearman's rank correlations across the transcriptomes (blue, left) and translomes (red, right) of all 80 samples. Protein-coding genes are shown in bold. The dark-red colored translome correlation of *MBNL1-AS1::MBNL1*, marked by an asterisk, represents an anticorrelation. **(I)** Scatter plots indicating an increase in correlation between triadin (*TRDN*) and *TRDN-AS1* from transcription (blue) to translation (red). Correlation coefficients are Spearman's rho and calculated across all 80 samples. **(J)** Heatmap with Ribo-seq expression levels (scaled RPKMs) for 41 translated lncRNAs that are differentially expressed between controls and DCM patients ($FDR \leq 0.05$; $FC \leq 1/1.2$ or $\geq 1*1.2$). **(K)** Beeswarm dot plots with selected examples of translated lncRNAs that are downregulated (top) or upregulated (bottom) in diseased hearts. DEseq2-normalized counts are plotted on a log₁₀ y-axis. Genome-wide corrected p-values are given; error bars indicate the mean + SD.

Figure S5) Microproteins localize to mitochondria and associate with mitochondrial processes, related to Figure 5. (A) Clustered heatmaps with translated lncRNAs (x-axis) and significantly associated GO terms (y-axis; GO Cellular component (Cc), Biological process (Bp) and Molecular function (Mf)), derived from genome-wide correlations with transcriptionally coregulated genes ($r \geq 0.5$), using all translated cardiac genes as a background set. Color intensity indicates the significance of GO enrichment. Zoomed views show reclustered subgroups of lncRNAs and highlight GO term categories on the right. ECM = extracellular matrix; ER = endoplasmic reticulum; FA = fatty acid; H⁺ = hydrogen ion / proton; MT = mitochondrial; SRP = signal recognition particle. **(B)** Immunofluorescence (IF) staining depicting the colocalization of 18 FLAG-tagged microproteins with mitochondria, upon overexpression in HeLa cells. These microproteins were selected for IF because they showed strong expression coregulation with mitochondrial processes (e.g. Figure 5A or Figure 5C), and/or were predicted to go to mitochondria based on protein localization prediction software (Table S5). Scale bars represent 20 μ M.

Figure S6) Characterization of microproteins with predicted functional domains, related to Figure 6. (A) Visualizations of signal peptide cleavage predictions of 3 microproteins, modified from the output of SignalP 4.1 (Petersen et al., 2011). (B) Volcano plots with the immunoprecipitation and mass-spectrometry results for potentially secreted C-terminally FLAG-tagged microproteins RP11-432J24.5 and AC093642.6. Significantly interacting (yellow dots; FDR 0.005) and secreted interactions partners (blue dots) are highlighted. Among the significant interaction partners, enrichment of secreted proteins was detected for both RP11-432J24.5 ($p = 1.22 \times 10^{-17}$) and AC093642.6 ($p = 0.000202$). Significant GO enrichment is found for terms related to the secretory pathway, for both RP11-432J24.5 (GO: endoplasmic reticulum: $p = 8.14 \times 10^{-22}$; GO: endomembrane system: $p = 1.22 \times 10^{-12}$) and AC093642.6 (GO: endoplasmic reticulum: $p = 0.00000303$; GO: intracellular vesicle: $p = 0.000421$; GO: cytoplasmic vesicle: 0.000479). (C) Visualizations of predicted transmembrane alpha-helices for 7 microproteins, modified from the output of TMHMM v2.0 (Krogh et al., 2001). Additionally, I-Tasser structural model predictions and snake plots of the newly identified microproteins PDZRN3-AS1 and UPPERHAND indicate an alpha-helical transmembrane structure similar to existing microproteins DWORF and SLN, respectively (PDB ID: 4H1W). (D) Heatmap of p-values for significant GO terms associated with genes strongly correlating with UPPERHAND translation. (E) IF staining depicting the colocalization of FLAG-tagged and overexpressed UPPERHAND with the endoplasmic reticulum in HeLa cells. Scale bar represents 20 μM . (F) Volcano plot with immunoprecipitation and mass-spectrometry results depicting interaction partners upon overexpression of the 3xFLAG-tagged microprotein UPPERHAND. Significant interactions are shown as yellow dots (FDR 0.005). Significantly interacting proteins with a membrane localization are highlighted in blue (GO: integral component of membrane: $p = 2.6 \times 10^{-83}$). (G) UPPERHAND is downregulated upon stimulation with TGF- β 1 stimulation in primary human cardiac fibroblasts (Chothani et al., 2018), following opposite RNA expression patterns to the pro-fibrotic cytokine IL-11. (H) UPPERHAND RNA expression levels as measured by RT-PCR with or without TGF- β 1 stimulation. Expression levels are measured after triplicate transfections and normalized to GAPDH (STAR methods). Upon TGF- β 1 stimulation, UPPERHAND expression decreases to \pm 60-70%. Upon siRNA-mediated knockdown, UPPERHAND levels decrease to 15-25%. Control siRNAs represent scrambled versions of the UPPERHAND siRNAs. (I) RNA expression of fibrosis marker genes POSTN, IL-11 and COL1A1 as measured by RT-PCR and normalized against GAPDH, and secreted protein levels of the profibrotic cytokine IL-11 as measured by ELISA. Both UPPERHAND knockdown by siRNAs, as well as endogenous ATG mutation of the UPPERHAND transmembrane ORF result in increased expression of fibrosis markers. As controls, scrambled versions of the UPPERHAND siRNAs ("control siRNA"), knockdown of the TGF- β 1 receptor (TGFR) and the lncRNA LINC-PINT were included.

Figure S7) Translation of human cardiac circRNAs, related to Figure 7. (A) Total mapped sequencing reads and reads matching backsplice junctions as identified by find_circ2 in mRNA-seq (red) and totRNA-seq (blue) data across all 80 samples. The mRNA-seq data is used as a negative control, as circRNAs are not poly-adenylated and should thus not be captured by poly(A)-purification. Reported backsplice junction reads are unfiltered, i.e. they still contain reads falsely aligned between homologous regions of neighboring genes, which are eliminated during downstream output processing (these largely explain the occurrence of reported backsplice junction reads in the mRNA-seq data). (B) Abundance of backsplice junction reads for 324 circRNAs identified both in totRNA-seq data and in mRNA-seq data. For further analyses, only circRNAs with a 100-fold higher abundance in totRNA-seq data than

mRNA-seq data are included (4 are kept; 320 removed). The other detected backsplice events likely result from *trans*-splicing and/or exon shuffling, and are therefore removed. The top diagonal line indicates the 100-fold abundance cutoff. **(C)** Warning and support flags per circRNA, as provided by find_circ2. Please note the difference in order of magnitude between the scale of the x-axis and y-axis. Ribosome-associated circRNAs are indicated as red triangles. **(D)** We validated the circular nature of 18 out of 40 randomly selected ribosome-associated circRNAs via RNase R digestion followed by qPCR. Assays were performed on two independent human heart samples. Plus and minus signs indicate RNase R or mock control treatment. A difference in resistance to RNase R between the linear mRNA and the circRNA confirms the circular nature of the circRNA.

STAR methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Norbert Hubner (nhuebner@mdc-berlin.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human primary material

Human DCM tissue was collected from the left-ventricular assist device (LVAD) core at time of implant or from transmural samples of explanted hearts, whereas heart tissue from non-DCM controls was mostly collected from unused donor hearts. DCM tissues were obtained from the Cardiovascular Research Centre of the Royal Brompton and Harefield NHS Trust (London, UK; n = 41), University Medical Center Utrecht (UMCU; Utrecht, The Netherlands; n = 4) and the Herz- und Diabeteszentrum NRW (HDZ-NRW; Bad Oeynhausen, Germany; n = 20). Cardiac samples from 15 unaffected, non-DCM controls were obtained from UMC Utrecht (Utrecht, The Netherlands; n = 4), the Sydney heart Bank (Sydney, Australia; n = 10) and DHZB (Berlin, Germany; n = 1). Our study population includes 22 female and 58 male subjects with an average age of 43.6 years (SD = 15.48 years). Subject-specific information on sex and age of all subjects is reported in Table S1. Sex and age of the subjects did not show any significant association with condition (disease state; Fisher's exact test; p = 0.28) or main principle components of gene expression (as measured by mRNA-seq or Ribo-seq; Student's *t*-test for association with first 2 principle components; p = 0.13-0.35). Human samples were allocated to experimental groups based on clinically diagnosed disease state (DCM or unaffected controls). No statistical methods were used to predetermine sample size. Targeted proteomics experiments for *in vivo* detection of microproteins were performed on human heart tissue of adult cardiomyopathy patients with HCM (n = 3; mutations in MYH7 (2x) and MYBPC3) and DCM (n = 2; mutations in LMNA (2x)) obtained from Harvard Medical School, Boston USA. None of these overlap with the samples used for the sequencing experiments.

Royal Brompton and Harefield NHS Trust samples (DCM patients): Studies on human DCM tissues from the Royal Brompton and Harefield NHS Trust complied with the UK Human Tissue Act guidelines and were carried out with approval from the Royal Brompton and Harefield local ethical review committee and the National Research Ethics Service Committee South Central, Hampshire B (reference 09/H0504/104). *Herz- und Diabeteszentrum NRW samples (DCM patients):* Myocardial tissue samples from the left ventricle were obtained from the proband's explanted heart or during the implantation of a ventricular assist device after informed consent. The samples were snap-frozen in liquid nitrogen immediately after removal from the patient and stored at -80 °C till usage. The local ethics committee approved the study protocol (Reg.-No. 21/2013). The study conforms to the principles outlined in the Declaration of Helsinki (2013). *University Medical Center Utrecht samples (DCM patients and non-DCM controls):* Studies on human samples from the University Medical Center Utrecht were approved by the Biobank Research Ethics Committee, University Medical Center Utrecht (UMCU), Utrecht, the Netherlands (protocol number WARB 12/387). *Sydney Heart Bank donor heart samples (non-DCM controls):* Nonfailing left ventricular samples were obtained from braindead human donors for whom normal left ventricular function had been confirmed by

echocardiography. Sample collection was done in full accordance with Australian National Health Medical Research guidelines and approved by the Human Research Ethics Committee of the University of Sydney (HREC approval: 2012/2814). *Deutsches Herzzentrum Berlin (DHZB) sample* (non-DCM control): Heart tissue was obtained from a patient with isolated aortic valve disease, excluded from coronary artery disease, with echocardiographic normal dimensions and normal LV function. Studies on this human sample were approved under the local ethical agreement for the DZHK biobank (reference EA4/028/12).

Human kidney tissue samples were obtained after nephrectomy (n = 4) or after autopsy (n = 2) at either the Berlin-Brandenburg Center for Regenerative Therapies (BCRT) of the Charité Universitätsmedizin in Berlin (Germany; ethical approval EA1/134/12) or the Sapporo City General Hospital in Sapporo (Japan; ethical approval H19-057-437) (Table S1). Kidney nephrectomy samples were obtained from unaffected (non-tumor lesion) parts of the kidney in patients diagnosed with renal cell carcinoma (n = 4). Autopsy samples were obtained from pathologically unaffected kidneys in patients diagnosed with systemic sclerosis (n = 1) or plasma cell leukemia (n = 1). Human liver tissue samples were obtained from patients with primary sclerosing cholangitis at the time of liver transplantation (n = 3) and from the resection margin of benign liver tumors (n = 3). All samples were collected at the University Medical Center Hamburg-Eppendorf (UKE), Hamburg, Germany under the ethical approval PV4081.

Written informed consent was obtained from all study subjects prior to surgery or transplantation. For tissue donors, informed consent was obtained from the next of kin.

Human cell culture experiments

Human iPSC-derived cardiomyocytes (female as defined by *XIST* expression levels) used for ribosome profiling and shotgun MS were differentiated as described previously (BurrIDGE et al., 2014) and obtained from the Pluripotent Stem Cell facility at MDC Berlin. In brief, 2x10⁵ cells were seeded on 6 well plates coated with Geltrex (ThermoFisher). After three days culturing in Essential 8 Media, cells become 80-90% confluent and the early mesoderm differentiation was induced by inhibition of GSK3 signaling pathway using CHIR99021 (#72054, 6 μ M, STEMCELL Technologies). Two days later the media was replaced with RPMI with CDM3 supplemented with 5 μ M IWP2 (# 72122, STEMCELL Technologies). Beating cells clusters were observed earliest 10 days after induction of the differentiation. At day 12 of the differentiation experiment the cells were purified using metabolic selection in RPMI media without glucose supplemented with lactate. Human Flp-In™ T-REx™ 293 Cells (derived from HEK293T cells, female, ThermoFisher, Cat#R78007) and HeLa cell lines (female) were cultured in a humidified incubator at 37 °C with 5% CO₂ using Dulbecco's modified eagle medium (DMEM) with high glucose (4.5 g/l), 10% fetal bovine serum (FBS), 2 mM L-glutamine and 1 mM sodium pyruvate. The handling and processing of human primary cardiac fibroblasts (from male donors) is described in (Chothani et al., 2018; Schafer et al., 2017b). In short, human cardiac fibroblasts were prepared from right atrial biopsies minced into 1–2 mm³ pieces and placed in 6-cm dishes. Human cardiac fibroblasts were grown and maintained in DMEM (11995-065, Gibco), supplemented with 20% fetal bovine serum (FBS, 10500, Hyclone) and 1% penicillin–streptomycin (15140-122, Gibco), in a humidified atmosphere at 37 °C and 5% CO₂. The medium was renewed every 2–3 days. At 80–90% confluence, cells were passaged using standard trypsinization techniques. Ribo-seq experiments were carried out at low cell passage (< P4).

Animal Models

Male, 10-week-old C57BL/6 wild type inbred mice ($n = 6$) were housed in animal facilities in Boston and Singapore and fed *ad libitum*. Mice were sacrificed, left-ventricular heart tissue was extracted and immediately snap frozen in liquid nitrogen for storage until further processing. Mouse studies were conducted in accordance with the principles and procedures outlined in the National Institutes of Health Guide for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee at the Duke–National University of Singapore Medical School. Rat studies were conducted on heart tissue samples previously described in and acquired from (Schafer et al., 2015a) and (Schafer et al., 2017a). These include 6-week-old inbred BN-Lx ($n = 5$) and SHR ($n = 5$) males (Schafer et al., 2015a) and heterozygous 8-week-old F1 $Ttn^{WT/WT}$ ($n = 4$), $Ttn^{WT/TntvZ-disk}$ ($n = 4$) or $Ttn^{WT/TntvA-band}$ ($n = 4$) rats, derived from crosses between genetically engineered F344 rats and wild type BN rats (Schafer et al., 2017a). $Ttn^{WT/TntvA-band}$ rats contain a 12-bp deletion and 2-bp insertion (TA), introducing a stop codon in *Ttn* exon 303 (corresponding to human *TTN* exon 327) in the Ttn A-band (TntvA). $Ttn^{WT/TntvZ-disk}$ rats contain a Z-disc truncating mutation (TntvZ), introduced via a frameshift deletion of *Ttn* exons 2–6 (a 5,286-bp deletion). All mouse and rat animals were drug and test naïve, specific pathogen free (SPF) and not involved in previous procedures.

METHOD DETAILS

Tissue processing

Snap frozen human and mouse tissues were powdered using a pre-cooled mortar and pestle under the continuous addition of liquid nitrogen, on days with a humidity below 30%. Per sample, ideally 100 mg (with a minimum of 40 mg) powdered tissue was set aside for ribosome profiling and 5-10 mg was collected for total RNA isolation and RNA-seq. Technical and biological sample information (e.g. amount of tissue, age of sample, center of origin, RNA integrity numbers, etc.) was collected before and during sample processing (Table S1).

Ribosome profiling

Ribosome profiling on human primary left ventricle ($n = 80$), human kidney ($n = 6$), human liver ($n = 6$) and mouse left ventricle ($n = 6$) tissue was performed using a modified TruSeq Ribo Profile protocol optimized for use on tissue samples. This protocol is described in detail in (Schafer et al., 2015a), where it was used to generate Ribo-seq data from BN-Lx and SHR rat left ventricle tissue (available at the European Nucleotide Archive (ENA) under accession number PRJEB7498). The Ribo-seq procedure on tissue samples was carried out as follows: \pm 50-100 mg powdered tissue was lysed for 10 minutes on ice in 1 mL lysis buffer consisting of 1 \times TruSeq Ribo Profile mammalian polysome buffer, 1% Triton X-100, 0.1% NP-40, 1 mM dithiothreitol, 10 U ml⁻¹ DNase I, cycloheximide (0.1 mg ml⁻¹) and nuclease-free H₂O. Using immediate repeated pipetting and multiple passes through a syringe with a 21G needle we dissociated tissue clumps to create a homogenous lysate that facilitates quick and equal lysis of the tissue powder. Samples were next centrifuged at 20,000g for 10 minutes at 4°C to pellet cell and tissue debris. Per sample, 400-800 μ L of lysate was further processed according to the TruSeq Ribo Profile (Mammalian) Reference Guide with the additional modification of 8% PAGE selection directly after PCR amplification of the final library. Ribosome profiling on human cells was performed using the TruSeq Ribo Profile (Mammalian) Library Prep Kit (Illumina, San Diego, CA; USA) exactly according to the TruSeq Ribo Profile protocol. For all samples, ribosome profiling library size distributions were checked on the Bioanalyzer 2100 using a high sensitivity DNA assay (Agilent; 5067-4626), multiplexed and sequenced on an

Illumina HiSeq 2500 producing single end 1x51nt reads. Samples were always processed in large batches of maximum 30 samples to avoid a sample processing bias. Human heart Riboseq libraries were sequenced to an average depth of 115M (min. 56M, max. 232M) raw reads (Table S1).

Stranded total RNA and mRNA sequencing

Total RNA was isolated using TRIzol Reagent (Invitrogen; 15596018) using 5-10mg of the exact same human, rat and mouse tissues and cells processed for ribosome profiling. Rat RNA was isolated from the same rat tissues used for ribosome profiling in (Schafer et al., 2015a), with the goal to generate new, longer read RNA-seq libraries than published previously. Total RNA was DNase treated and purified using the RNA Clean & Concentrator™-25 kit (Zymo Research; R1018). RIN scores were measured on a BioAnalyzer 2100 using the RNA 6000 Nano assay (Agilent; 5067-1511). Ribosomal RNA-depleted (total) totRNA-seq and poly(A)-purified mRNA-seq libraries were generated from high quality RNA (average RNA Integrity Number (RIN) of 8.1 (human), 9.1 (rat) and 7.9 (mouse); Table S1). RNA-seq library preparation was performed in batches of maximum 48 samples according to the TruSeq Stranded Total RNA and mRNA Reference Guides, using 500ng of total RNA as input. Libraries were multiplexed and sequenced on an Illumina HiSeq 4000 producing paired 2x101nt reads. No totRNA-seq data was generated for human kidney and liver tissues. Human heart mRNA-seq libraries were sequenced to an average depth of 83M (min. 59M, max. 118M) raw reads and human totRNA-seq libraries were sequenced to an average depth of 82M reads (min. 52M, max. 205M) (Table S1).

Constructing *de novo* transcriptome assemblies

To capture the complete cardiac transcriptome including unannotated splice isoforms and not yet annotated lncRNAs, we generated species-specific *de novo* transcriptome assemblies to be used as reference annotations for mapping all RNA-seq and ribosome profiling data. To generate these assemblies, we first mapped the 2x101nt mRNA-seq data of human, rat and mouse to respective reference genomes (GRCh38.p10/hg38, Rnor6.0/rn6 and GRCm38.p5/mm38) using STAR v2.5.2b (Dobin et al., 2013) in the 2-pass mapping mode that allows for unbiased exon splice junction detection. We used the standard STAR settings with the following modified parameters: `--outSAMtype BAM SortedByCoordinate`, `--outFilterMismatchNmax 6`, `--outFilterMultimapNmax 20`, `--alignSJDBoverhangMin 3`, `--outFilterType BySJout`, `--alignSJoverhangMin 10` and `--outSAMattributes All`. Next, we added XS tags to each read in every mRNA-seq BAM and used StringTie v1.3.3 (Pertea et al., 2015) to create *de novo* transcriptomes for each sample separately, guided by a reference annotation (Ensembl release 87). Per species, all novel annotations were merged into one consensus annotation GTF file using the StringTie `--merge` option, requiring minimum transcript isoform lengths of 200nt and expression levels equal to or above 1 FPKM. Next, we filtered out the following types of newly identified transcripts: (i) monoexonic genes that lack strand information, (ii) known non-polyadenylated transcripts (e.g. sncRNAs) that received new gene IDs, (iii) newly annotated genes with no uniquely mapping reads as a result of e.g. segmental duplications, repeats, pseudogenes or other reasons for high levels of sequence similarity and (iv) transcript isoforms that merge exons of two or more (neighboring) existing genes due to high sequence similarity. For each entry, StringTie gene IDs assigned to novel transcript isoforms matching existing genes were replaced by corresponding reference gene IDs. We categorically grouped all noncoding antisense transcripts (AS), long intergenic noncoding RNAs (lincRNAs) and genes with a 'processed transcript' biotype as lncRNAs, requiring them to have no sense overlap (i.e. same genomic DNA strand) with any existing coding exon of a

protein-coding gene and to not have alternative transcript isoforms with a protein-coding biotype. While constructing *de novo* transcriptome assemblies we took particular interest in improving the annotation of existing lncRNAs and newly detected genes already annotated as either a lncRNA or a sORF-containing protein-coding gene (≤ 100 aa) in any of the other species analyzed (human, mouse or rat). Due to differences in annotations of each of these three species, some human lncRNAs might already be annotated as a small protein-coding gene in mouse or rat or *vice versa*. Therefore, we cross-referenced potential novel lncRNAs and small protein-coding genes with the annotations of the two other species using the UCSC Batch Coordinate Conversion (LiftOver) utility (Kuhn et al., 2013), keeping potential orthologous lncRNAs in the new StringTie annotations. Genes that are orthologous but newly identified in all three species and thus without an existing biotype were excluded from further analysis. In the human heart transcriptome annotation, these cross-species comparisons resulted in the addition of 117 potential novel lncRNAs, 978 novel isoforms to already annotated lncRNAs and 224 novel isoforms to already annotated small protein-coding genes.

Sequencing data alignment

Prior to mapping, ribosome-profiling reads were clipped for residual adapter sequences and filtered for mitochondrial, ribosomal RNA and tRNA sequences (Figure S1A). Next, all mRNA and ribosome profiling data were mapped to the filtered *de novo* transcriptome assemblies using STAR v2.5.2b. As all possible *de novo* junctions were already incorporated in the *de novo* transcriptomes, we disabled *de novo* exon junction detection to improve the mapping precision of short (28-30nt) ribosome profiling reads. Next, we trimmed the 2x101nt mRNA-seq reads to 29-mers and processed and mapped those mRNA reads with the exact same settings as the ribosome profiling data (Figure S1A), in order to avoid mapping or quantification bias due to read length or filtering. For the mapping of 2x101nt RNA-seq reads we allowed 6 mismatches per read (default is 10) and for the Ribo-seq and trimmed mRNA-seq reads we allowed 2 mismatches. To account for variable ribosome footprint lengths, we defined the search start point of the read using the option `--seedSearchStartLmaxOverLread`, which was set to 0.5 (half the read, independent of ribosome footprint length).

Detecting actively translated ORFs

Canonical ORF detection using ribosome profiling data was performed with RiboTaper v1.3 (Calviello et al., 2016) with standard settings. For each sample we selected only the read lengths for which at least 70% of the reads matched the primary ORF in a meta-gene analysis (Table S1). Following the standard configuration of RiboTaper we required ORFs to have a minimum length of 8 aa, evidence from uniquely mapping reads and at least 21 P-sites. The final list of translation events (including all uORFs, downstream ORFs (dORFs) and lncRNA sORFs) was stringently filtered requiring the translated gene to have an average RNA RPKM ≥ 1 and be detected as translated in at least 10/80 human, 3/10 rat and 2/6 mouse heart samples. Additionally, each specific ORF was required to have an identical translation termination codon in at least 5/81 human, 2/10 rat and 2/6 mouse heart samples. All detected ORFs can be found in Table S1. Upstream ORF identifications exist in two varieties: (i) uORFs that are fully separated from the primary ORF and (ii) uORFs that show genomic, but out-of-frame overlap with the primary ORF (in the same transcript isoform). The latter are not automatically detected as uORFs by RiboTaper and thus need further filtering to be separated from in frame, 5' extensions of the primary ORF. Therefore, we required each overlapping uORF to have a translation start site before the start of the canonical CDS, to end within the canonical CDS (prior to the annotated termination codon) and to be translated in a different frame than the primary ORF i.e. to produce a different peptide. The final set of uORFs consists

of 889 independent uORFs and 201 uORFs that show out-of-frame overlap with the start of a primary ORF. In none of the presented analyses the effects of both types of uORFs are statistically different (not shown).

Gene and ORF quantification

To quantify gene expression we counted sequenced reads mapping to coding sequence (CDS) regions of genes, including those of newly detected translated lncRNAs, using HTseq-count (Anders et al., 2015). For genes without a CDS, such as untranslated lncRNAs, we included gene counts based on reads mapping to all annotated exons. In total, we detect 783 transcribed lncRNAs in the human heart requiring an average RNA RPKM ≥ 1 (RPKM ≥ 0.1 - 3,256 lncRNAs). It should be noted that, likely because of gene annotation issues, 16 out of 11,387 translated genes have zero read counts in the final quantification by HTseq count, because their CDSs show complete sense overlap with a second protein-coding gene with a wrongly assigned different gene ID, making it impossible to discriminate both genes and define the exact source of expression. All mRNA-seq and Ribo-seq counts can be found in Table S2 or downloaded from the Shiny web application at <https://shiny.mdc-berlin.de/cardiac-translatome>.

The quantification of uORF translation rates was performed independently for the two types of detected uORFs: non-overlapping, independently localized uORFs and uORFs that show out-of-frame overlap with the primary ORF. For independently localized uORFs we counted all reads mapping to the uORF and used DESeq2 v1.12.4 for normalization (Love et al., 2014). The translation levels of overlapping uORFs cannot be discriminated from the primary ORF based on genomic coordinates alone. Because overlapping uORFs are translated in a different frame than the primary ORF, we quantified them by taking the sum of (i) all reads mapping to the non-overlapping part of the uORF and (ii) the fraction of reads in the overlapping part that matches the primary reading frame of the uORF (and thus not that of the primary ORF), as derived from the percentage of in-frame Ribo-seq P-sites. The sum of the uORF reads was then normalized by DESeq2.

Differential expression analysis

To allow for proper comparison and integration of mRNA-seq and Ribo-seq data, all mRNA-seq quantifications were derived from CDS-mapped, single-end reads trimmed from 101nt to ribosomal footprint sizes. Expression quantification was followed by read normalization, size factor estimation (on mRNA-seq and & Ribo-seq simultaneously) and differential expression analysis between DCM and unaffected control samples using DESeq2 v1.12.4 (Love et al., 2014). For this analysis we include all genes that we consider to be translated, i.e. matching the criteria described in the 'detecting active translation' section ($n = 11,387$). We considered a gene differentially expressed when it met genome-wide significance thresholds of FDR ≤ 0.05 and a fold change (FC) $\leq 1/1.2$ or $\geq 1*1.2$. We use these cutoffs to capture all possible interactions between transcription and translation, which help us to define the precise contribution of transcriptional regulation to differences observed at the translational level in downstream analyses. In line with this, we introduced an interaction term to model the non-additive effects of disease status and sequencing approach (Ribo-seq or mRNA-seq) to gene expression differences, in order to identify genes that show specific discordant regulation at the translational level between DCM patients and non-DCM controls (Chothani et al., 2017).

Translational efficiency estimation

Translational efficiency (TE) estimations were calculated by taking the ratio of Ribo-seq over mRNA-seq counts. This approach yields independent TEs for each individual sample and gene, in the absence of true technical or biological replicates that are required by most (EdgeR (McCarthy et al., 2012) / DESeq2-based) tools that estimate translational efficiencies and differential translation between sample groups or conditions (e.g. Xtail, RiboDiff, Riborex, anota and Babel (Larsson et al., 2011; Li et al., 2017; Olshen et al., 2013; Xiao et al., 2016; Zhong et al., 2017)). We need these sample- and gene-specific TE estimates for analyses where group-centric comparisons do not suffice, such as measurements of correlation and coregulation across all samples (e.g. with uORF occupancy), association tests between TE and genetic variation, and measurements of variance in TE (related to the presence or absence of a uORF). Another state-of-the-art tool that can produce individual TE estimates in Scikit-ribo (Fang et al., 2018). However, specific Scikit-ribo features that account for secondary RNA structure could unfortunately not be applied to human samples, because of the high number of alternative splice isoforms and multiple CDSs annotated for individual human genes in comparison to yeast or *E. coli*.

Gene-gene correlations and GO enrichment

Spearman correlations were calculated to test for coregulation among (i) all translated genes, (ii) uORFs and primary ORFs and (iii) lncRNA-mRNA sense-antisense pairs, using the DESeq2-normalized counts of pairwise complete observations. Samples with zero counts for a specific gene in the mRNA-seq or Ribo-seq data were excluded from the correlation calculations, with a minimum of 20 samples required for a gene to be included in the clustering. To test for significant differences between correlations, for example between the transcription and translation levels of sense and antisense gene pairs (e.g. *TRDN* and *TRND-AS1*), correlation coefficients were Fisher z transformed (Fisher Z-Transformation or Fisher r to z transformation) to a normal distribution, enabling statistical comparison. To study genome-wide coregulation between translated genes, the correlation matrix was used to calculate the Euclidean distance followed by hierarchical clustering, resulting in 30 clusters. Cluster visualization was done using heatmap.2 from gplots v3.0.1 (<http://CRAN.R-project.org/package=gplots>), or the modified heatmap.3 (<https://github.com/obigriffith/biostar-tutorials/tree/master/Heatmaps>). GO enrichment (The Gene Ontology Consortium, 2017) and KEGG pathway (Kanehisa et al., 2017) analyses to assign functional annotation to selected (sub)-clusters or pairs of correlating genes were performed with gProfiler v0.6.4 (Reimand et al., 2016) using g:Profiler archive revision 1741 (Ensembl 90, Ensembl Genomes 38). Principle component analysis (PCA) was performed in order to define the main contributing layer of gene expression regulation (transcription, translation or both) for each cluster of coregulated, differentially expressed genes. As input for the PCA, for each gene cluster we provided the fractions of genes that were translationally downregulated ($\Delta \text{Log}_2\text{FC} < 1/1.2$), translationally upregulated ($\Delta \text{Log}_2\text{FC} > 1.2$) and transcriptionally regulated ($\Delta \text{Log}_2\text{FC} > 1/1.2$ and < 1.2). In addition, we provided the mean transcription and translation $\log_2\text{FC}$ values between DCM patients and controls. The placement of each cluster in the PCA plot is thus not only based on the balance between transcriptionally and translationally controlled genes, but also on the directionality of both layers of regulation (up or down) in diseased hearts.

Variant detection and effect prediction

The identification of single nucleotide variants (SNVs) and insertions/deletions (indels) was performed using GATK v3.6 (McKenna et al., 2010) on paired end, 101nt mRNA-seq reads. According to GATK Best Practices (Van der Auwera et al., 2013) duplicate reads that originate from the same RNA fragment were marked with Picard v1.136, so that they are not counted

as additional evidence for the identification of a variant. The reads were then processed using GATK SplitNCigarReads, a critical step for mRNA-seq data that decreases false positives by removing segments of reads extending into intronic regions. Because a MAPQ of 255 as assigned by STAR is “unknown” to GATK, we reassigned mapping quality scores and subsequently performed GATK Base Recalibration to correct possible systematic sequencing errors. BaseRecalibrator uses machine learning to model sources of technical error leading to over or under estimation of base quality scores and adjusts the scores accordingly. dbSNP v138 was used as a source of known variants (Sherry, 2001). We identified variants in the processed BAM files using GATK HaplotypeCaller and the genotypes were called for all samples at all variant positions with GenotypeGVCFs.

SNVs and indels were filtered separately using GATK VariantFiltration. We required both SNVs and indels to have a Quality by Depth > 2.0 (this corrects the quality score for high depth regions to prevent an inflated score due to deep coverage) and have a minimum coverage of 10 non-duplicate reads. The variant confidence score was normalized by available coverage to avoid inflated quality scores in high coverage regions. Then filtering by FisherStrand avoided false positive calls due to sequencing bias of one strand over the other. We filtered variants out with a FisherStrand Score > 30.0 for SNVs and > 200.0 for indels, keeping variants below this threshold. Furthermore, clusters of 3 or more SNVs in a 35bp window were excluded, as well as variants in exons with a "Percent Spliced In" (PSI) score below 0.80 (80%). A PSI cutoff is applied as variants in exons that are not frequently used will have a lower phenotypic impact than variants in constitutive exons. We have previously shown that this is important for truncating mutations in *TTN* (Roberts et al., 2015). Potentially damaging variants in lowly expressed exons of *TTN* are frequent in healthy individuals, whereas mutations in highly expressed exons have a 93% probability of pathogenicity. As a final filtering step, RepeatMasker was used to remove all variants within repetitive sequences to avoid potential SNVs and indels due to misalignment. Removing SNVs in repeats also eliminated the vast majority of RNA editing events that mostly occur in SINE elements, as only 128 known A-to-I editing events from the REDportal database remain (Picardi et al., 2017). In total, we included 101,813 exonic SNVs and 7,077 indels for further analyses.

A subset of 31 samples were previously genotyped on the Affymetrix GW6 array followed by imputation (Heinig et al., 2017). At the 13,420 variant positions for which both genotype and mRNA-seq variant calls passed filtering, we found high concordance rates (Figure S3B), reassuring the quality of our mRNA-seq variant calls. Additionally, the identified variants have characteristics and allele frequencies similar to larger cohorts (Figure S3C and Table S4). To determine the effect of the SNVs and indels on genes and transcripts, we used Ensembl Variant Effect Predictor (VEP) v87 (McLaren et al., 2016). We included information for APPRIS gene labels and transcript support levels to use for further filtering criteria. We included only variants in genes with an APPRIS label and with merged or manually curated transcript support level. We excluded variants in NMD, HLA and incomplete CDS transcripts. Because the variants are called in mRNA-seq data, we only included variants in genes with an RPKM ≥ 1.0 .

Linking exonic variants to gene expression

To identify possible linkage between genomic variants and gene-specific transcription, translation or translational efficiency levels, we performed association testing on normalized mRNA-seq, Ribo-seq and TE values using the R package MatrixEQTL v2.1.1 (Shabalina, 2012). For this analysis we used all identified exonic variants, requiring at least 5 samples per

genotype group (reference or alternative), resulting in 42,988 testable variants. To define significant linkage, we apply the Benjamini-Hochberg correction (significance cutoff $FDR \leq 0.05$). Of the 963 SNVs associating with mRNA expression or translation levels, 97% has a known variant ID. This confirms that the vast majority of common (≥ 5 samples) variants used for the association analysis are genetic by origin.

Characteristics and QC of protein truncating variants

Protein truncating variants include all nonsense SNVs and frameshift indels that cause a premature stop codon. To validate the quality of our RNA-seq based PTV variant calls, we applied multiple QC steps. All DCM patients were genetically screened by clinical diagnostic centers for stop codons in genes frequently associated with DCM. This resulted in PTV identifications in *TTN*, *NKX2-5* and *TPM1*, for which supporting sequencing reads were detectable in the RNA-seq data. Additionally, 42 PTVs in 31 DCM patients were previously identified by Complete Genomics whole-genome sequencing (*not published*), and all of these are concordant with our RNA-seq calls (Table S4). Also, 236 out of 346 detected PTVs have known variant IDs (e.g. a dbSNP ID) and their effect predictions and minor allele frequencies are highly concordant with public variant repositories (Table S4 and Figure S3E).

Importantly, PTV identification rates are slightly lower than genome-wide numbers of large human cohorts not restricted to expressed genes (Lek et al., 2016). This is because our search space is limited to cardiac expressed genes and because NMD would hinder the detection of PTVs in efficiently decayed mRNAs in our RNA expression data. For example, a genetically-confirmed disease-associated PTV in tropomyosin 1 (*TPM1*) did not meet our initial variant identification filtering criteria, as the RNA of the mutated allele had a much lower relative abundance than that of the unmutated allele (PTV allele ratio = 0.02). However, this is clearly not the case for most PTVs: we do identify all DCM-associated heterozygous PTVs in *TTN* (mean allele ratio = 0.38; Figure 3F) and *NKX2-5* (allele ratio = 0.38), where the predicted PTV allele is not decayed.

PTVs are enriched in the final exon of the coding sequence and the 50-55nt upstream of the final exon-exon junction (97 out of 346; $p = 1.093 \times 10^{-12}$; chi-squared test) (Figure S3F). Following the exon-junction complex-based nonsense-mediated decay rule or "NMD rule" (Nagy and Maquat, 1998), mRNAs carrying PTVs in this region of the mRNA should not induce NMD. In comparison to mRNAs with PTVs in different parts of the transcript, these variants indeed display slightly higher allele ratios - i.e. less decay ($p = 0.004$; Kolmogorov-Smirnov test). Separating both groups of PTVs based on the 'ASE score' (see STAR methods section "The consequences of PTVs on gene expression"), which corrects for (variance in) allelic imbalances that occur naturally (i.e. in human samples that do not carry a PTV in this gene), this effect is reduced, though remains significant (Figure S3G; $p = 0.006$; Kolmogorov-Smirnov test). In general, the differences in allele ratios between both groups of PTVs separated by location is only marginal, as most PTVs detected here do not seem to induce NMD (Figure 3A).

The consequences of PTVs on gene expression

To quantify allele ratios and ribosome occupancy upstream and downstream of PTVs, we used GATK's ASEReadCounter on the list of filtered SNV positions generated from all 80 samples to calculate the number of reference and alternative reads at each biallelic position, in each sample. The location of the premature stop codon introduced by a frameshift indel was determined with a custom script based on the Ensembl transcript sequence, which was

selected based on transcript-filtering criteria mentioned above (effect prediction). To define allele-specific expression, allele counts were generated from the mRNA-seq .bam files. We defined potential NMD via an ASE score, which we calculated by measuring the mRNA-seq allele ratio at variant positions. To define a PTV as NMD-causing, we required the mean allele ratio of samples containing a PTV to be at least 1.2 standard deviations lower than that of the samples without the PTV. It is important to take natural allele ratios into account as ASE can be abundant for specific gene loci. Such ASE is not necessarily caused by the PTV, leading to false estimates of the extent of ASE that can be attributed to the PTV. These considerations translated into the following formula: ASE score = (meanAlleleRatio_WT – meanAlleleRatio_PTV) / SD(AlleleRatio_WT). Thus, if the ASE score ≤ -1.2 , we assume the PTV has the potential to cause NMD.

To define the efficiency by which PTVs drive premature translation termination, we calculated a ribosome drop-off score that compares ribosome coverage before and after the PTV in samples with and without the PTV. For this, we calculated length- and library-size normalized Ribo-seq expression levels before and after the PTV, using the primary transcript isoforms predicted to be truncated by each respective PTV. We then calculated the fold change in Ribo-seq coverage before and after the PTV, for samples with and without PTV separately, and calculated the ribosome drop-off score by subtracting the log₂ transformed fold changes of both groups. PTVs resulting in a drop-off score ≤ -1 are labeled as potentially truncating translation.

Detecting differential exon splicing

Exon inclusion analysis was performed to filter out genetic variants (including PTVs) in infrequently used exons. We calculated the percent spliced in (PSI) for every known exon in both mRNA-seq and Ribo-seq data, by determining the ratio of reads included to excluded in an exon as reported previously (Schafer et al., 2015b). To determine isoform-specific regulation of *TTN*, we tested for differential splicing using the Mann-Whitney *U* test and applied a Benjamini-Hochberg correction (significance cutoff FDR ≤ 0.05). To compare gene-specific differential isoform production, we calculated Δ PSI values by taking the PSI(group1) - mean(PSI(wildtype)). For all splicing analyses, 2x10¹nt PE mRNA-seq data were used for the accurate identification and quantification of exon in- or exclusion.

Ttn detection in rats with Ttn truncating variants

Rat heart tissue was obtained from F1 hybrids crossed between Brown Norway (BN; *Ttn* WT) and Fischer 344 rats (F344; heterozygous *Ttn* truncating variant (Ttntv) in the Z-disk or A-band), as described in (Schafer et al., 2017a). The resulting F1 hybrids contained either of the following *Ttn* allele combinations: *Ttn*^{WT/WT}, *Ttn*^{WT/TtntvZ-disk} or *Ttn*^{WT/TtntvA-band}. Ribo-seq data was obtained from (Schafer et al., 2017a) and is accessible at European Nucleotide Archive (ENA) under accession ERP015402. As described previously (Schafer et al., 2017a), reads assigned to either the BN or F344 allele were normalized to library size and to the expression levels of both alleles in the wild type rats. To define F344 allele ratios in the Ribo-seq data, we additionally required at least 10 uniquely mapping reads covering the genetic variant position.

We next performed a shotgun mass spectrometry analysis on 4 biological replicates of each rat strain to find peptide evidence for truncated Ttn protein. Therefore, pulverized heart tissue from the exact same animals used for Ribo-seq was resuspended in lysis buffer (6 M Guanidium HCl in 10 mM Hepes pH 8), boiled for 10 min at 95 °C and further proceed for in-solution digest, first reduced in 12 mM DTT (45 min at RT) and then alkylated using 40 mM

chloroacetamide (30 min at RT). Proteins were digested using endopeptidase LysC (Wako, Osaka, Japan; enzyme:protein ratio of 1:100) for 4 h, followed by dilution in 4 volumes of 50 mM ammonium bicarbonate (pH 8.5) and further digestion with sequence grade trypsin (Promega; enzyme:protein ratio of 1:100) for 16 h. The digestion was stopped by acidifying each sample to pH < 2.5 by adding 10% trifluoroacetic acid solution. After centrifugation to pellet insoluble material (14,000 rpm, 10 min) the peptides were extracted and desalted using stage tip protocol (Rappsilber et al., 2003). In short, pipet tips (200 μ L, Gilson) were packed with C18 chromatographic beads (3M, Minneapolis, MN) to generate stage tips. Beads were activated with methanol and washed with wash buffer (2% acetonitrile and 1% trifluoroacetic acid) before sample loading. Stage tips were washed, samples were eluted with Buffer B (80% acetonitrile, 0.1% formic acid) and organic solvent was evaporated using a speedvac (Eppendorf). Samples were diluted in Buffer A (3% acetonitrile and 0.1% formic acid) and peptides were separated on a reversed-phase column (20 cm fritless silica microcolumns with an inner diameter of 75 μ m, packed with ReproSil-Pur C18-AQ 1.9 μ m resin (Dr. Maisch GmbH)) using a 200 min gradient with a 250 nl/min flow rate of increasing Buffer B concentration (from 2% to 60%) on a High-Performance Liquid Chromatography (HPLC) system (ThermoScientific). Peptides were ionized using an electrospray ionization (ESI) source (ThermoScientific) and analyzed on a Thermo Orbitrap Fusion (Q-OT-qIT, Thermo). Survey scans of peptide precursors from 300 to 1500 m/z were performed at 120K resolution with a 2×10^5 ion count target. Tandem MS was performed by isolation at 1.6 m/z with the quadrupole, HCD fragmentation with normalized collision energy of 30, and rapid scan MS analysis in the ion trap. The MS² ion count target was set to 2×10^3 and the max injection time was 300 ms. Only precursors with charge state 2–7 were sampled for MS². The dynamic exclusion duration was set to 60 s with a 10-ppm tolerance around the selected precursor and its isotopes. The instrument was run in top speed mode with 3 s cycles, meaning the instrument would continuously perform MS² events until the list of non-excluded precursors diminishes to zero or 3 s. Each sample was measured in two technical replicates. Data were analyzed using MaxQuant software package (v1.5.2.8). The internal Andromeda search engine was used to search MS² spectra against a decoy rat UniProt database (RAT.2017-01) plus an in-house database for missense variants in the F344 rat *Ttn* locus, containing forward and reversed sequences. The search included variable modifications of methionine oxidation and N-terminal acetylation, and fixed modification of carbamidomethyl cysteine. Minimal peptide length was set to seven amino acids and a maximum of 4 missed cleavages was allowed. The FDR (false discovery rate) was set to 1% for peptide and protein identifications. Unique and razor peptides were considered for quantification.

LncRNA expression and translation across tissues and cell types

Cardiac and/or skeletal muscle specific expression of translated lncRNAs was calculated using GTEx v6 data (Aguet et al., 2017) by requiring the mean expression of a gene in left ventricle or atrial appendage to be 12-fold higher than in all other tissues (cardiac specific) or by requiring the mean expression of a gene in left ventricle, atrial appendage and skeletal muscle to be 10-fold higher than the mean expression in all other tissues (muscle specific). Further tissue- and cell-type resolution of lncRNA translation was obtained via ribosome profiling of human kidney tissues, human liver tissues, primary cardiac fibroblasts (Chothani et al., 2018) and hiPSC-derived cardiomyocytes (21 days old). To define the presence of a translated lncRNA in either of these two cell or tissue types, a gene was required to be detected as translated by RiboTaper (Calviello et al., 2016).

Conservation of uORFs and translated sORFs in lncRNAs

To determine amino acid conservation of detected sORFs and uORFs we implemented a method and scoring pipeline based on PhyloCSF (Lin et al., 2011) as presented in (Mackowiak et al., 2015). LncRNA sORFs with a PhyloCSF score above 10 were considered to display signs of conservation. Positional conservation was determined using reciprocal LiftOver of the genomic coordinates of translated lncRNAs between human (hg38), rat (rn6) and mouse (mm10) genomes, requiring the overlapping genes to have a noncoding annotation and similar relative orientation to neighboring protein-coding genes. Translation initiation site (TIS) conservation was also defined reciprocally between human, rat and mouse as described in (Fields et al., 2015). In short, we extracted the exact TIS of any translated lncRNA sORF as identified by RiboTaper and performed LiftOver of this coordinate to the other 2 species, requiring a translation initiation site to be present in either of these two species within a window of 9nt up and downstream of the converted TIS coordinate.

***In vitro* translation of cardiac lncRNAs**

Synthetic gene constructs containing 84 complete transcript isoforms (including the predicted 5' and 3' UTR) of 60 translated lncRNA genes were produced by Genewiz Europe (Leipzig, Germany; constructs available upon request). The 84 transcript isoforms correspond to unique ORFs that are specific to different splice isoforms of 60 translated genes. To disrupt the predicted open reading frame, we introduced single nucleotide mutations, deletions or insertions using the QuickChange II site-directed mutagenesis kit (Agilent Technologies, CA, USA) to mutate the AUG directly or to generate a frameshift following the AUG. Mutagenic primers were designed with the QuikChange Primer Design Program (www.agilent.com/genomics/qcpd) and were synthesized and HPLC purified by BioTeZ (Berlin, Germany). Mutagenic PCR reactions, DpnI digests and bacterial transformation were performed according to QuickChange II site-directed mutagenesis kit instructions. DNA was extracted from transformed bacterial cells using Qiagen miniprep kits (Qiagen, Germany) and sequences were verified by Sanger sequencing.

Microproteins from wildtype and mutated constructs were translated *in vitro* from 0.5 µg linearized plasmid DNA using the TnT® Coupled Wheat Germ Extract system (Promega, Mannheim, Germany) in the presence of 10 mCi/mL [³⁵S]-methionine (Hartmann Analytic, Braunschweig, Germany) according to manufacturer's instructions. Five µL lysate was denatured for 2 min at 85 °C in 9.6 µL Novex Tricine SDS Sample Buffer (2X) (Thermo Fisher Scientific) and 1.4 µL DTT (500 mM). Proteins were separated on 16% Tricine gels (Invitrogen) for 1 h at 50 V followed by 3.5 h at 100 V and blotted on PVDF-membranes (Immobilon-PSQ Membrane, Merck Millipore). Incorporation of [³⁵S]-methionine into newly synthesized proteins enabled the detection of translation products by phosphor imaging (exposure time of 1 day). On the transcript-isoform level, we detect peptide products for a total of 58 out of 81 tested human lncRNA constructs (72.0%), which corresponds to the successful *in vitro* translation of 44 out of 58 (75.9%) translated human lncRNA genes. These numbers exclude positive controls (MRLN and DWORF) and the two mouse-specific translated lncRNAs *chaer* and *myheart* that were not expressed in human hearts, of which *myheart* could be successfully translated *in vitro*. Of note, most microproteins that could not be detected, including *chaer*, had predicted product sizes smaller than 3 kDa, suggesting technical detection limitations of products in that lower size range.

Searching public mass spectrometry data

In order to detect protein products of translated cardiac genes we obtained the raw data of the deepest human heart proteome available at the moment (Doll et al., 2017), which we

downloaded from the EMBL-EBI PRIDE archive under accession number PXD006675. For data analyses we used the MaxQuant software package (v1.6.0.1) (Cox and Mann, 2008). We performed two independent MaxQuant searches, one to identify and quantify protein products of protein-coding genes for the analyses presented in Figure 1F and Figure S1H, and one to specifically detect previously not annotated microproteins translated from lncRNAs, circRNAs or uORFs. For both runs, standard search parameters were used and included variable modifications of methionine oxidation, deamidation of asparagine and glutamine and N-terminal acetylation, and fixed modification of carbamidomethyl cysteine. Minimal peptide length was set to seven amino acids and a maximum of 4 missed cleavages was allowed.

For the first search we specifically explored the left-ventricular protein measurements of the Doll et al dataset (so not including other human heart regions). The goal of this search was to define for how many gene products that we detected as translated with Ribo-seq, the protein products could also be identified with deep mass spectrometry. We searched against a library containing protein sequences of genes that have been predicted to be translated according to our data (Cardiac Translatome FASTA search database; available for download at <https://shiny.mdc-berlin.de/cardiac-translatome/>) and merged this with all human UniProt entries (HUMAN.2017-01; with decoy format including reversed sequences). For these searches, we kept the FDR for both peptides and proteins to the standard settings (1%). The match between runs function was activated to allow identification of peptides without MS/MS information. We used the IBAQ intensity values for all correlation analyses between sequencing-based and MS-based quantifications, for which only unique and razor peptides were considered (Figure S1H).

For the second search, with the goal to identify peptides that match newly detected microproteins expressed from lncRNAs, circRNAs or uORFs, we included the protein measurements of all heart regions from the Doll et al dataset (not just the left ventricle). We then searched against the same merged library as before, i.e. the newly annotated translation events merged with the human UniProt database (HUMAN.2017-01; with decoy format including reversed sequences). A FASTA file for predicted circRNA peptides (of which each was required to span the backsplice junction and follow the predicted reading frame) was added separately. We maintained statistical filtering on the peptide level (5% FDR), but similar to searches for small proteins performed previously, we excluded the protein FDR (Calviello et al., 2016; Fesenko et al., 2018; Ma et al., 2018; Mackowiak et al., 2015). For small proteins, for which only single or limited numbers of unique peptides can theoretically be identified, the protein FDR is often overestimated. This is especially true for large proteomics datasets of increasing size (as discussed in e.g. (Kim et al., 2014; Savitski et al., 2015)), such as the deep MS dataset being searched here (Doll et al., 2017). How small proteins are best discovered in large proteomics datasets has been the subject of much debate and may lead to increased numbers of false-positive identifications (Gupta and Pevzner, 2009; Omenn et al., 2017). Complementary detection methods, for example via targeted proteomics as performed in this study, are recommended to increase the confidence of the identifications.

To limit the number of peak matches that occur by chance in large database searches, MaxQuant uses a target-decoy strategy that is (amongst other criteria) required for peptide FDR estimation. This decoy search includes reversed peptides of all (micro)proteins present in the search database, providing an artificial search space for proteins that match in length and amino acid distribution, but should not be detectable. Here, we employed an additional target-decoy strategy, which is complementary to a reversed peptide search (Elias and Gygi,

2010), by screening the shotgun MS data for thousands of in-silico predicted, untranslated sORFs, derived from a 3-frame translation of the 169 actual translated lncRNAs. We created a collection of such 'artificial' sORFs requiring that these (i) did not show any sign of active translation as predicted by RiboTaper and (ii) showed no sense overlap with the actual translated sORF in the same reading frame. All cardiac-expressed transcript isoforms of the here detected 169 translated lncRNAs were included, and artificial sORFs sense-overlapping actual sORFs in alternative reading frames (+1 or +2) were kept. When within a single transcript isoform (i.e. without downstream alternative splicing possibilities), multiple artificial sORFs showed sense-overlap with each other, only the longest sORF was kept to be as inclusive as possible. These filtering steps resulted in a total of 3,623 artificial sORFs, from which we subsampled 1,000 sets of 339 ORFs matching the size distribution of the 339 true (Ribo-seq supported) sORFs. For each set, we summed up the number of unique shotgun MS hits and compared this to the number of identifications for the true set of translated sORFs. The same analysis was repeated, this time quantifying identifications on the gene (i.e. lncRNA) level. In both analyses, none of the 1,000 sets yielded more positive identifications than the true (Ribo-seq supported) sORF set (empirical p-value < 0.001), showing a clear enrichment for true over artificial microproteins. To more accurately quantify the significance of the observed effect, we calculated Cohen's effect sizes (d) for both the sORF-level and gene-level comparisons (Cohen's $d = 5.99$ and 7.57 for translated sORFs and lncRNAs, respectively). Despite this enrichment, many false-positive unique peptide hits could be detected, indicating that microprotein evidence solely detected by deep shotgun MS searches needs to be further substantiated with independent (targeted MS or antibody-based) methods.

Deep proteomic analysis of human iPSC-derived cardiomyocytes

Protein extracts from total human iPSC-derived cardiomyocytes (29 days old) as well as nuclear and cytoplasmic fractions were solubilized in denaturation buffer (6 M Urea/2M Thiourea, 10 mM Hepes, pH 8) and in-solution digest was performed by first reducing proteins in 12 mM DTT (45 min at RT) and then alkylating them by using 40 mM chloroacetamide (30 min at RT). The samples were digested using endopeptidase LysC (Wako, Osaka, Japan, enzyme:protein ratio of 1:100) for 4 h, followed by dilution in 4 volumes of 50 mM ammonium bicarbonate (pH 8.5) and further digestion with sequence-grade trypsin (Promega, enzyme:protein ratio of 1:100) for 16 h. The digestion was stopped by acidifying each sample to pH < 2.5 by adding 10% trifluoroacetic acid solution. After centrifugation to pellet insoluble material (14,000 rpm, 10 min) the peptides were extracted and desalted using stage tip protocol (Rappsilber et al., 2003), as described above. Peptides from total extracts were further fractionated by strong cation exchange (SCX) chromatography, using 6 (50, 75, 125, 200, 300 and 500mM ammonium acetate) fractionation steps. 1 μ g of unfractionated and fractionated peptide samples were analyzed by mass spectrometry using a Thermo Fusion instrument (Thermo). Before ionization on an electrospray ionization (ESI) source (ThermoScientific) peptides were separated on a reversed-phase column (20 cm fritless silica microcolumns with an inner diameter of 75 μ m, packed with ReproSil-Pur C18-AQ 3 μ m resin (Dr. Maisch GmbH)) using a 90, 200 or gradient with a 250 nl/min flow rate of increasing Buffer B concentration (from 2% to 60%) on a High-Performance Liquid Chromatography (HPLC) system (ThermoScientific). Survey scans of peptide precursors from 300 to 1500 m/z were performed at 120K resolution with a 2×10^5 ion count target. Tandem MS was performed by isolation at 1.6 m/z with the quadrupole, HCD fragmentation with normalized collision energy of 30, and rapid scan MS analysis in the ion trap. The MS² ion count target was set to 2×10^3 and the max injection time was 300 ms. Only precursors with charge state 2–7 were sampled for MS². The dynamic exclusion duration was set to 60 s with a 10-ppm tolerance around the selected

precursor and its isotopes. The instrument was run in top speed mode with 3 s cycles, meaning the instrument would continuously perform MS² events until the list of non-excluded precursors diminishes to zero or 3 s. Data were analyzed using MaxQuant software package (v1.6.0.1) (Cox and Mann, 2008). The internal Andromeda search engine was used to search MS² spectra against the in-house library of predicted microproteins and the human UniProt database (HUMAN.2017-01) containing forward and target-decoy reverse sequences. The search included variable modifications of methionine oxidation, deamidation of asparagine and glutamine and N-terminal acetylation, and fixed modification of carbamidomethyl cysteine. Minimal peptide length was set to seven amino acids and a maximum of 4 missed cleavages was allowed. Illustrating the depth of the data, we could identify 5,699 proteins based on 56,766 peptides, using standard MaxQuant settings and excluding contaminants. To search for microproteins, we set the peptide FDR to 5% and eliminated the protein FDR, as discussed in "Searching public mass spectrometry data".

Selected reaction monitoring (SRM) proteomics

Synthetic peptides of crude quality were ordered (JPT Inc., Berlin, Germany) and resuspended in 20% acetonitrile (100 mM ammonium bicarbonate). Spectra were recorded by measuring synthetic peptides (1 pmol per peptide) on a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific) with higher energy collision dissociation method (HCD) with a mass resolution of 70,000 for the MS and 17,500 for the MS/MS scans. The recorded spectra were analyzed using the MaxQuant software package (v1.6.0.1) applying a custom-made database containing the predicted sequences, with carbamidomethylation of cysteines as a fixed and oxidation of methionines as a variable modification. For peptides and proteins an FDR of 1% was applied. Based on the fragmentation pattern of the peptides in the Q-Exactive plus mass spectrometer, an SRM method for a TSQ Quantiva instrument (Thermo Fisher Scientific) was developed monitoring up to 6 of the most intense fragment ions with good library matching values using the Skyline Software package v3.6 (MacLean et al., 2010). Best collision energy for each peptide was calculated and predicted using Skyline. A total of 223 peptides and 908 transitions were selected to identify (i) 5 control proteins (GAPDH, ACTA, TUBA1B, HIST1H2, LMNA;LMNB), 20 annotated small proteins (UniProt; < 100 aa; as matching controls for the microproteins in terms of size), and (iii) 137 microprotein candidates derived from different sORFs of 83 lncRNA genes (see Table S5). Based on the retention time profile (137 min gradient with increasing acetonitrile concentration: 5 to 27% for 117 min, 27 to 54% for 20 min) peptides were grouped into five scheduled SRM-methods. Dwell time was set to 200 ms and scheduled retention windows of 20 min were chosen. TSQ Quantiva parameters for measurements were set to Q1 and Q3 resolution of 0.7 (FWHM).

Pulverized human heart tissue was resuspended in lysis buffer (6 M Guanidium HCl in 10 mM Hepes pH 8) and boiled for 10 min at 95 °C. 100 µg of protein extract was used for in-solution digest, first reduced in 12 mM DTT (45 min at RT) and then alkylated using 40 mM chloroacetamide (30 min at RT). The samples were digested using 2 µg endopeptidase LysC (Wako, Osaka, Japan) for 4 h, followed by dilution in 4 volumes of 50 mM ammonium bicarbonate (pH 8.5) and further digestion with 2 µg trypsin (Promega) for 16 h. The digestion was stopped by acidifying each sample to pH < 2.5 by adding 10% trifluoroacetic acid solution. After centrifugation to pellet insoluble material (14,000 rpm, 10 min) the peptides were extracted and desalted using stage tip protocol (Rappsilber et al., 2003), as described above. Peptide samples were eluted from stage tips (80% acetonitrile, 0.1% formic acid), and were resolved in sample buffer (5% acetonitrile and 0.1% formic acid) after evaporating organic solvent. Two microgram of peptide solution was injected and separated on a reversed-phase

column (20 cm fritless silica microcolumns with an inner diameter of 75 μm , packed with ReproSil-Pur C18-AQ 1.9 μm resin (Dr. Maisch GmbH, Ammerbuch, Germany)) using a 137 min gradient of increasing acetonitrile concentration (5 to 27% for 117 min, 27 to 54% for 20 min) with a 250 nl/min flow rate on a High-Performance Liquid Chromatography (HPLC) system (Thermo Fisher Scientific). Peptides were ionized using an electrospray ionization (ESI) source (Thermo Fisher Scientific) and analyzed on a Thermo TSQ Quantiva instrument (Thermo Fisher Scientific). Dwell time was set to 200 ms and scheduled retention windows of 10 min were chosen. TSQ Quantiva settings for Q1 and Q3 were set as described above (Establishment of SRM-assay). Peak annotation was carried out using Skyline software package with the following settings: Precursor Charges 2 to 4; ion charges 1 and 2; Ion types y, p, b, a, z; up to 6 productions picked; auto-selection of matching transitions enabled; ion match tolerance = 0.05 m/z; method match tolerance = 0.055 m/z; Resolving power of MS2 filtering was set to 17,500 at 200 m/z). A dot product filter of ≥ 0.7 was applied and for significant peaks the dot product, retention time and total peak area was individually reported for each biological (5 hearts) and technical (2 runs per heart) replicate (see Table S5). All five control proteins were robustly identified in all replicates of the five heart samples. Importantly, we detect only 10 out of 20 in UniProt annotated matched control proteins (similarly sized and expressed at similar levels in the human heart), suggesting an estimated detection rate of 50% for our microprotein candidate set. This indicates that the SRM search performed for newly discovered microproteins yields a success rate that is very similar to that of the matched control group of previously annotated small proteins (55.4% vs. 50% of the tested microproteins).

Translation of functional lncRNAs

To identify translation of lncRNAs with previously described noncoding functions, we constructed a comprehensive database of 324 functionally characterized lncRNAs based on manually curated databases (Amaral et al., 2011; Gray et al., 2015; Quek et al., 2015) and extensive recent literature search (Table S6). Sixty-one out of 324 lncRNAs are expressed in human, rat or mouse hearts and 32 of those (52.5%) appear actively translated in the heart (27 in human, 1 in rat, 7 in mouse). Additionally, 28 functional lncRNAs are detected as translated in cardiac fibroblasts and 17 in iPSC-derived cardiomyocytes, of which respectively 10 and 2 had not been detected as translated in the human heart left ventricle data. Combining all sampled cell types and species, this brings the total number of translated, functionally characterized lncRNAs to 42 (Table S6).

Microprotein feature searches and modeling

Predictions of protein localization were performed with TargetP 1.1 (Emanuelsson et al., 2000) and DeepLoc 1.0 (Almagro Armenteros et al., 2017), omitting plant-specific chloroplasts as a possible localization. Prediction of signal peptides and transmembrane helices were carried out using SignalP 4.1 and TMHMM 2.0c respectively (Krogh et al., 2001; Petersen et al., 2011). All 339 microproteins were modeled using a locally installed version of I-TASSER 5.1 using standard settings (Yang et al., 2014). Microprotein features and structure predictions are accessible through the interactive Cardiac Translatome web application at <https://shiny.mdc-berlin.de/cardiac-translatome>.

Identifying microprotein interaction partners by IP-MS

For immunoprecipitation experiments, HEK293T cells (4×10^6) were seeded in triplicates on poly-D-Lysine (Sigma, Germany) coated 10 cm dishes and transfected with 28 μg plasmid-DNA of FLAG-tagged microproteins or empty 3xFLAG-vector (negative control) using TransFectin (BioRad, California) following manufacturer's instructions. Two days post

transfection cells were washed twice with ice-cold phosphate-buffered saline (PBS), scraped in 1.5 mL ice-cold PBS and transferred into Eppendorf tubes. After centrifugation at 950 g for five min at 4 °C, cell pellets were lysed in 200 µL lysis buffer (150 mM NaCl, 50 mM Tris pH 7.5, 1% IGPAL-CA-630, 2x Complete protease inhibitor without EDTA) for 30 min on ice. Lysates were centrifuged at 20,800 g for 15 min at 4 °C and supernatants were added to 30 µL 50% antibody-coupled magnetic bead solution (M2-magnetic beads, Sigma, Germany) and 300 µL wash buffer 1 (150 mM NaCl, 50 mM Tris pH 7.5). Beads were washed 3x in 150 µL wash buffer 1 before usage. After incubating the samples for 2 h at 4 °C in an over-head shaker, samples were washed once with 1 mL wash buffer 2 (150 mM NaCl, 50 mM Tris pH 7.5, 0.05% IGPAL-CA-630) and three times with wash buffer 1. Supernatants were removed and magnetic beads were frozen at -80 °C until analysis by mass spectrometry. Beads from the triplicate immunoprecipitation experiments were resuspended in 20 µL urea buffer (6 M urea, 2 M thiourea, 10 mM HEPES, pH 8.0), reduced for 30 min at 25 °C in 12 mM dithiothreitol solution, followed by alkylation in 40 mM chloroacetamide for 20 min in the dark at 25 °C. Samples were first digested with 0.5 µg endopeptidase LysC (Wako, Osaka, Japan) for 4 h. After adding 80 µL 50 mM ammonium bicarbonate (pH 8.5) samples were digested with 1 µg sequence-grade trypsin (Promega) overnight at 25 °C. The peptide-containing supernatant was removed and collected into a fresh tube. Beads were washed twice with 50 µL 50 mM ammonium bicarbonate (pH 8.5) and the supernatants were pooled. Samples were acidified by adding 1 µL formic acid to stop the digestion. Peptides were extracted and desalted using StageTip protocol (Rappsilber et al., 2003). Peptides were eluted using Buffer B (80% Acetonitrile and 0.1% formic acid) and organic solvent was evaporated using a speedvac (Eppendorf). Samples were diluted in Buffer A (3% acetonitrile and 0.1% formic acid) and separated on a reversed-phase column (20 cm fritless silica microcolumns with an inner diameter of 75 µm, packed with ReproSil-Pur C18-AQ 1.9 µm resin (Dr. Maisch GmbH)) using a 90 min gradient with a 250 nL/min flow rate of increasing Buffer B concentration (from 2% to 60%) on a High-Performance Liquid Chromatography (HPLC) system (Thermo Fisher Scientific). Peptides were ionized using an electrospray ionization (ESI) source (Thermo Fisher Scientific) and analyzed on a Thermo Orbitrap Fusion (Q-OT-qIT) or Thermo Q Exactive Plus instrument. The Orbitrap Fusion was run in data dependent mode selecting the top 20 most intense ions in the MS full scans, selecting ions from 350 to 2000 *m/z*, using 60 K resolution with a 4×10^5 ion count target and 50 ms injection time. Tandem MS was performed by isolation at 0.7 *m/z* with the quadrupole, HCD fragmentation with normalized collision energy of 32 and resolution of 15 K. The MS² ion count target was set to 5×10^4 with a maximum injection time of 250 ms. Only precursors with charge state 2–7 were sampled for MS². The dynamic exclusion duration was set to 30 s with a 10-ppm tolerance around the selected precursor and its isotopes. The Q Exactive Plus instrument was run in data dependent mode selecting the top 10 most intense ions in the MS full scans, selecting ions from 350 to 2000 *m/z*, using 70 K resolution with a 3×10^6 ion count target and 50 ms injection time. Tandem MS was performed at a resolution of 17.5 K. The MS² ion count target was set to 5×10^4 with a maximum injection time of 250 ms. Only precursors with charge state 2–6 were sampled for MS². The dynamic exclusion duration was set to 30 s with a 10-ppm tolerance around the selected precursor and its isotopes. Data were analyzed using MaxQuant software package (v1.5.2.8). The internal Andromeda search engine was used to search MS² spectra against a human UniProt database (HUMAN.2017-01) and an in-house bait protein sequence database containing forward and reverse sequences. The search included variable modifications of methionine oxidation and N-terminal acetylation and fixed modification of carbamidomethyl cysteine. Minimal peptide length was set to seven amino acids and a maximum of 3 missed cleavages was allowed. The FDR was set to 1% for peptide and protein identifications. Unique

and razor peptides were considered for quantification. Retention times were recalibrated based on the built-in nonlinear time-rescaling algorithm. MS² identifications were transferred between runs with the “Match between runs” option for biological replicates, in which the maximal retention time window was set to 0.7 min. IBAQ and LFQ intensities were calculated using the in-built algorithm. The resulting text files were filtered to exclude reverse database hits, potential contaminants, and proteins only identified by site. Statistical data analysis was performed using Perseus software (v1.6.1.3) (Tyanova et al., 2016). Biological replicates for each condition were defined as groups and intensity values were filtered for “minimum value of 3” per group. After log₂ transformation, missing values were imputed with random noise simulating the detection limit of the mass spectrometer. Imputed values are taken from a log normal distribution with 0.25x the standard deviation of the measured, log-transformed values, down-shifted by 1.8 standard deviations. Differences in protein abundance between FLAG tagged bait samples and FLAG control samples were calculated using two-sample *t*-test, and a permutation-based FDR significance cut-off was used to define specific interaction partners.

Co-immunoprecipitation (Co-IP) assay

Reciprocal co-immunoprecipitation was performed to validate the interaction of RMND1 and the microprotein PDZRN3-AS1. HEK293T cells (1.3×10^6) were seeded on poly-D-Lysine (Sigma, Germany) coated 6 cm dishes and transfected with 7.3 µg plasmid-DNA of i) empty vector, ii) PDZRN3-AS1-3xFLAG, iii) RMND1-HA for negative controls and iv) 3.7 µg plasmid-DNA of PDZRN3-AS1-3xFLAG and RMND1-HA using TransFectin (BioRad, California) following manufacturer’s instructions. Two days post transfection cells were harvested as described above (Identifying microprotein interaction partners), but using only 100 µL lysis buffer. For precipitation of PDZRN3-AS1-3xFLAG, supernatants (Input) were added to 25 µL 50% antibody-coupled magnetic bead solution (M2-magnetic beads, Sigma, Germany) and filled up to 500 µL with wash buffer 1 (150 mM NaCl, 50 mM Tris pH 7.5). Beads were washed 3x in 150 µL wash buffer 1 before usage. After incubating the samples for 2 h at 4 °C in an over-head shaker, samples were washed once with 1 mL wash buffer 2 (150 mM NaCl, 50 mM Tris pH 7.5, 0.05% IGPAL-CA-630), two times with wash buffer 1 and one more time with milliQ water. After removal of supernatants 100 µL 0.1M Glycine-HCl pH 2.7 was added and samples were incubated for 10 min in an overhead shaker at RT. The eluate was collected and neutralized with 15 µL 1M Tris pH 10.6. For western blot analysis 10.6 µL of input, unbound fraction and eluate were denatured for 10 min at 70 °C in 4 µL NuPAGE™ LDS Sample Buffer (4X) (Invitrogen; NP0007) and 1.6 µL of NuPAGE™ Sample Reducing Agent (10X) (Invitrogen; NP0009). Proteins were separated on NuPAGE™ 12% Bis-Tris Protein Gels (Invitrogen; NP0343BOX) for 30 min in MES buffer (Invitrogen; NP0002) at 200 V and blotted on PVDF-membranes (Immobilon-PSQ Membrane, Merck Millipore; ISEQ00010). Membranes were stained against PDZRN3-AS1-3xFLAG and RMND1-HA using mouse monoclonal anti-FLAG (M2) antibody (Sigma; M8823-1ML) and rabbit monoclonal anti-HA-Tag antibody (Cell Signaling Technology; 3724), respectively and developed using the ECL procedure according to manufacturer’s instructions (GE Healthcare; RPN2109). The reciprocal immunoprecipitation of RMND1-HA from cell lysates was performed with 25 µL of Pierce™ Anti-HA Magnetic Beads (Thermo Fisher Scientific; 88836) following manufacturer’s instructions. Proteins were separated and detected as described immediately above for immunoprecipitation of PDZRN3-AS1-3xFLAG.

Microprotein localization by immunofluorescence

Human HeLa cells were grown on glass slides for 24 h and transfected with FLAG-tagged plasmids using Lipofectamine 2000 reagent for 24 h. Cells were fixed with 4%

paraformaldehyde for 10 minutes at room temperature and washed three times with ice-cold phosphate-buffered saline (PBS). The cells were permeabilized and blocked for 1 h at room temperature using 2.5% bovine albumin serum, 10% anti-goat serum and 0.1% Triton X and washed again. Expressed microproteins were stained for 1 h at room temperature using an anti-FLAG mouse monoclonal antibody (1:500, F1804, Sigma Aldrich) and co-stained with organelle markers for ER or mitochondria, respectively (1:500, mouse anti-PDI #3501; 1:1000, rabbit ATPIF1 #13268, both Cell Signaling Technology; Danvers, MA, USA). Afterwards, the slides were washed and incubated with fluorescently-labeled secondary antibodies (1:500, Alexa Fluor 488 anti-rabbit & Alexa Fluor 594 anti-mouse (Invitrogen, Carlsbad, CA, USA) for 30 minutes at room temperature. Cells were washed again, stained with 4-6-diamidino-2-phenylindole (NucBlue™ Fixed Cell ReadyProbes™ Reagent, R37606, Thermo Fisher) for 5 minutes at room temperature and mounted onto glass slides using ProLong™ Gold antifade reagent (Molecular Probes; Invitrogen™). Images were visualized using a LEICA SP8 confocal microscope using a 63x objective. Image analysis was performed using Leica confocal software Las X (v3.5.2) and ImageJ (v1.52a) (Schneider et al., 2012).

Mitochondrial isolation and proteinase K digestion

HEK293T cells of three 10-cm dishes were washed twice with cold PBS, transferred to microfuge tubes and centrifuged at 10,000 g for 5 min at 4°C. The cell pellet was resuspended in 8 mL 2 mg/mL BSA and left on ice for at least 15 min to facilitate cell swelling. The cell suspension was transferred to a 5 mL glass homogenizer and homogenized with 50-70 strokes using a drill-fitted pestle. The homogenate was centrifuged at 800 g for 5 min at 4°C. The supernatant was collected to obtain the mitochondrial fraction and centrifuged at 10,000 g for 10 min at 4°C. The pellet, containing crude mitochondria, was resuspended in 500 µL buffer (1 mM EDTA, 20 mM HEPES, 220 mM mannitol, 70 mM sucrose and 0.5 mM PMSF). An additional centrifugation step (800 g for 5 min at 4°C) was added to remove remaining non-lysed cells. Next, mitochondria were centrifuged at 10,000 g for 10 min at 4°C and resuspended in 200 µL of Sucrose buffer (10 mM HEPES and 0.5 M sucrose). Fifty micrograms of solubilized mitochondria were pelleted by centrifugation (10,000 g for 10 min at 4°C) and resuspended in a buffer consisting of 125 mM KCl, 20 mM HEPES, 2 mM MgCl₂, 2 mM KH₂PO₄, 0.04 mM EGTA at a pH of 7.2. Next, 3.5 µg of mitochondria were pelleted and resuspended in 25 µL RIPA buffer containing 2x Protease Inhibitor and analyzed as full mitochondrial lysate control. Remaining mitochondria were equally distributed across 12 samples and incubated at 37°C for 1 hour with proteinase K concentrations ranging between 0.01 µg/mL to 100 µg/mL. One sample incubated without Proteinase K served as a negative control. To terminate the proteinase K digestion, PMSF (1 mM final concentration) and proteinase inhibitors (2x final concentration) were added. Samples were mixed with LDS loading buffer and Reducing Agent and analyzed by SDS-PAGE and Western blotting. The membrane was stained for overexpressed proteins (PDZRN3-AS1-3xFLAG and RMND1-HA) and different mitochondrial proteins, including known outer (TOM20, VDAC1) and inner (LETM1, COX4) mitochondrial membrane proteins.

The role of UPPERHAND in cardiac fibrosis

Knock out the AUG of the transmembrane ORF of *UPPERHAND* (*HAND2-AS1*) was performed in immortalized fibroblasts (BJ-5ta (ATCC® CRL-4001) according to Liang et al. (Liang et al., 2017). Guide RNAs targeting the ATG in the endogenous *UPPERHAND* locus were designed using the CRISPOR design software available at <http://crispor.tefor.net/> (Haeussler et al., 2016) and ssODNs were designed according to Richardson et al., 2016 (sequences are available on request) (Richardson et al., 2016). For the nucleofection reaction

mix, a ribonucleoprotein (RNP) complex was prepared by mixing 5 µg Cas9 protein (0,5 µL) (Integrated DNA Technologies) with 2 µL freshly annealed guide RNA (5 µL of 100 M µtracr RNA and 5 µL of 100 µM target specific crRNA will be incubated for 5 min at 95°C and thereafter cooled to RT to form the active guide RNA), followed by incubation at room temperature. Thereafter, cells were dissociated into single cells with TrypLE express (Thermo Scientific) and resuspended in 30 µL nucleofection buffer R. To this mixture, previously prepared RNP complexes supplemented with additional 7 µL Buffer R and 3 µL target specific ssODN (10 µM) were added. 3 x 10⁶ µL of the suspension was then nucleofected using the 10 µL Neon nucleofection system. Post-nucleofection, the cells were plated with fibroblast media. Four days after transfection, the cell pool was analyzed using amplicon sequencing to validate the targeting efficacy. A single cell-derived cell population was Sanger sequenced for validation of the mutation.

To test the effect of *UPPERHAND* knockdown and ATG knockout on fibrosis, human cardiac fibroblasts were transfected with siRNAs or scrambled controls, and simulated with TGF-β1 or left unstimulated. For TGF-β1 stimulation, human fibroblasts were starved in serum-free DMEM for 16 hours prior to TGF-β1 stimulation. Fibroblasts were then stimulated for 12 hours and compared to unstimulated fibroblasts grown for the same duration and under the same conditions, as described in (Schafer et al., 2017b). For siRNA transfection, human cardiac fibroblasts were seeded in 96-well black CellCarrier (PerkinElmer) plates and transfected with 12.5 nM On-Targetplus siRNAs (Dharmacon) in serum-free Opti-MEM medium and DMEM containing 10% FBS (ratio 1:9) using Lipofectamine RNAiMax (13778-150, Life Technologies). The cells were transfected for 24 h and subsequently cultured in DMEM containing 1% FBS overnight before being subjected to further analyses. siRNA-mediated knockdown of the translated lncRNA *LINC-PINT* and TGF-β1 receptor, as well as a scrambled version of the *UPPERHAND* siRNAs, were included as controls.

For RT-PCRs to measure expression levels of the siRNA-targeted genes (*UPPERHAND*, *TGFR* and *LINC-PINT*) and selected fibrosis markers (*POSTN*, *IL-11* and *COL1A1*), total RNA was extracted from cell lysate using TRIzol reagent (Invitrogen) followed by RNeasy column (Qiagen) purification. The cDNA was prepared using an iScript cDNA synthesis kit, in which each reaction contained 1 µg of total RNA, as per the manufacturer's instructions. Quantitative RT-PCR gene expression analysis was performed on triplicate samples with either TaqMan (Applied Biosystems) or fast SYBR green (Qiagen) technology using a StepOnePlus (Applied Biosystem) over 40 cycles. Expression data were normalized to GAPDH mRNA expression levels and the 2^{-ΔΔCt} method was used to calculate the fold change. Specific TaqMan probes were obtained from Applied Biosystems. siRNA-mediated knockdown of control genes *TGFR* and *LINC-PINT* resulted in a reduction to 40-60% of endogenous expression levels (data not shown). Changes in secreted IL-11 protein levels were measured as described previously (Schafer et al., 2017b) using the following kit: Human IL-11 Quantikine ELISA kit (D1100, R&D Systems).

Circular RNA detection

CircRNA backsplice junction detection was performed simultaneously on ribosomal RNA-depleted totRNA-seq and mRNA-seq data, starting with mapping of the reads to the human genome (hg38) using BWA-MEM (Li and Durbin, 2010) using standard settings except for the following parameters: -t 4 -L 3,3 -E 3,3 -k 14 -T1 (Figure S7A). CircRNA detection on mapped data was performed using find_circ2 (v1.2, https://github.com/rajewsky-lab/find_circ2). Stringent filtering criteria were applied, requiring a circRNA (i) to be detected with at least 2

unique backsplice junction spanning reads in a minimum of 10 out of 80 samples, with a total of at least 50 junction spanning reads across all 80 samples, (ii) to be within an exonic sequence size range of 50bp-10kb, (iii) to originate from within a single gene (to avoid false-positive splice junctions from nearby genes with highly homologous sequences), (iv) to be derived from autosomes or sex chromosomes (not on unplaced chromosome contigs or scaffolds), and (v) to not be detected in mRNA-seq data. For the latter we apply a normalized ratio cutoff of 100:1 for presence in totRNA-seq versus mRNA-seq data, which results in the exclusion of 320 out of 324 circRNAs detected also in mRNA-seq data (Figure S7B). False positive non-circRNA backsplice junctions additionally detected in mRNA-seq data likely arise from *trans*-splicing and / or exon shuffling in polyadenylated transcripts (Guo et al., 2014), but this is not frequently observed. To assure high quality of the identified circRNAs, we compared warning and support flags reported by find_circ2 and observe a general trend in which 87% of all 8,878 circRNAs have > 15x more support than warning flags (Figure S7C).

Ribosome-associated circRNAs

To test for ribosome association of cardiac circRNAs we extracted the exonic sequence surrounding the backsplice junction (40bp on either side) and mapped all unmapped ribosome profiling reads (i.e. Ribo-seq reads that cannot be aligned to the linear transcriptome or genome) to the circRNA backsplice junctions using Bowtie2 (v2.0.6) (Langmead and Salzberg, 2012). We did not allow any mismatches and required a minimum read-junction overlap of 9 nt on either side of the junction. A total of 1,298 Ribo-seq reads (776 unique reads) map to the backsplice junctions of 508 out of 8,878 different circRNAs. Forty circRNAs in 39 genes meet our further filtering criteria for robust ribosome association, requiring the backsplice junction to be covered by at least 3 unique and at least 5 total junction-spanning Ribo-seq reads. To assess the false-positive detection rate of circRNA ribosome association, we constructed a pool of ~3.8 million simulated, intragenic backsplice junctions derived from random combinations of translated cardiac exons. From this set we excluded all backsplice junctions belonging to actual circRNAs or *trans*-splicing events as initially detected by find_circ2 in the mRNA-seq or totRNA-seq data (prior to any filtering). From this collection we subsampled 10,000 datasets with each 8,878 junctions that match the circRNA size distributions of the actual cardiac circRNA dataset. We mapped the unmapped proportion of the Ribo-seq data to each of these sets requiring various minimum junction overlaps (between 1 and 15 nt), showing that none of the 10,000 simulated sets equal the ribosome association observed for the true circRNA set. For a selection of 18 out of 40 ribosome-associated circRNAs we performed RNase R digestions followed by qPCR and Sanger sequencing, as described previously (Memczak et al., 2015). In short, total RNA was isolated from 2 human heart samples using TRIzol reagent (Thermo Scientific, Waltham, Massachusetts), DNase I treated and purified with the RNA Clean & Concentrator™-25 kit (Zymo Research; R1018). For qPCR analysis, total RNA was treated with RNase R (Epicentre, San Diego, California) for 15 min at 37°C, at a concentration of 3 U/μg RNA. After treatment, 5% *C. elegans* total RNA was spiked-in, followed by phenol-chloroform extraction of the RNA mixture. For controls, the RNA was mock treated without the enzyme. Primers used for qPCR are listed in Table S7.

Further information on the experimental design

There was no explicit randomization or blinding procedure for human and animal data comparisons, unless stated otherwise (for instance see STAR methods sections on shotgun MS microprotein randomization searches, the impact of PTVs, association of uORF types with TE, and circRNA detection). For animal and human cell culture experiments, the number of biological and technical replicates is mentioned in the results and/or relevant STAR methods

sections. No human or animal samples were excluded from the analyses presented in this work. No statistical methods were used to predetermine sample size of human DCM and control experimental groups. No human replication cohorts were included in this study.

QUANTIFICATION AND STATISTICAL ANALYSES

Statistical analyses were performed using custom bash scripts and the programming language R (v.3.4.4). Crucial software used for data quantification and statistical analyses is stated in the respective STAR methods sections and the Key Resource Table. Statistical parameters such as the value of *n*, mean/median, standard deviation (SD) and significance level (**p* < 0.05, ***p* < 0.01, ****p* < 0.001 and *****p* < 0.0001) are reported in the figures and/or in the figure legends. The “*n*” represents animal, human or cell culture sample numbers (STAR methods: "EXPERIMENTAL MODEL AND SUBJECT DETAILS"; Figures 1A, 3E+G+J, 7E; Figures S1C, S3C+E), the number of genes, exons, circRNAs, translation events, peptides or proteins analyzed or detected (STAR methods: "Differential expression analysis"; all results text; Figures 1E+F, 3A+C, 5A, 6B, 7B; Figures S2A+H, S3G+L, S4C+E+J, S6B+F), or the number of technical or biological replicates (Figure S4G, S6G-I). The type of statistical test (e.g. Mann-Whitney *U* test or *t*-test) is annotated in the figure legend and STAR Methods segment specific to each analysis. Unless stated otherwise, statistical analyses are two-sided tests performed using R. The type of statistical test (e.g. Mann-Whitney *U* test or Student's *t*-test) is annotated in the figure legend and STAR Methods segment specific to each analysis. For FDR estimation the Benjamini-Hochberg procedure was used and Bonferroni correction was applied to correct for multiple testing.

DATA AND SOFTWARE AVAILABILITY

All identifiable human sequencing data have been deposited in the European Genome-phenome Archive (EGA) under accession number EGAS00001003263, which can be directly accessed at <https://ega-archive.org/studies/EGAS00001003263>. All non-identifiable human sequencing data and rodent left ventricle sequencing data have been deposited in the European Nucleotide Archive (ENA) under accession number PRJEB29208, which can be accessed at <https://www.ebi.ac.uk/ena/data/view/PRJEB29208>. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol et al., 2018) with the dataset identifier PXD012593, which can be accessed at <https://www.ebi.ac.uk/pride/archive/projects/XD012593>. All code used for the analyses in this paper is available upon request.

ADDITIONAL RESOURCES

To make our data easily accessible, we built an interactive app that allows users to query all sequencing data (<https://shiny.mdc-berlin.de/cardiac-translatome>). It enables the user to browse and visualize differential expression results, human cardiac microproteins and genetic associations presented in the paper. Additionally, fully prepared sessions for exploring the sequencing data and detected ORFs with the Integrated Genomics Viewer (IGV) are provided, as well as a custom FASTA database for proteomics searches.

Supplemental Excel Table Titles and Legends (Provided as separate spreadsheets)

Table S1: Human, mouse and rat sample information and sequencing stats, related to Figure 1. Table with sample information, sequencing statistics and all identified actively translated ORFs of human, rat and mouse left ventricle heart data.

Table S2: Differentially transcribed and translated genes in diseased hearts, related to Figure 2. Table with DEseq2 results for differential transcription and translation. Includes all normalized read counts, information on translational correlation, coregulated clusters and associated GO terms.

Table S3: Human heart upstream ORFs, related to Figure 2. Table with RiboTaper identifications of independent and primary ORF-overlapping uORFs. Includes information on uORF translation, host gene translational efficiencies and correlation, conservation, peptide detection and domain predictions.

Table S4: Genetic variation across 80 hearts, related to Figure 3. Table with genotype calls for all 80 samples based on mRNA-seq data. Includes genetic association test results, variant effect predictions, protein truncating variants and TTNtv.

Table S5: Translated lncRNAs in human, rat and mouse hearts, related to Figure 4 - 6. Table with all lncRNA short ORFs identified in human, mouse and rat hearts. The table includes statistics obtained from RiboTaper, gene- and peptide level conservation, cell-type and tissue specificity, expression levels, domain predictions and *in vitro* and *in vivo* validations (IVT assays, shotgun MS evidence and SRM).

Table S6: Translation of functionally characterized lncRNAs, related to Figure 4 - 6. Table with a manually curated list of functional lncRNAs collected from lncRNADB.com, HUGO and recent literature searches. Includes information on lncRNA expression and translation across species and cell types.

Table S7: Human cardiac circRNAs and association with ribosomes, related to Figure 7. Table with all circRNAs identified in human hearts, with information about their abundance, ribosome occupancy and resistance to RNase R.

References

- Aguet, F., Ardlie, K.G., Cummings, B.B., Gelfand, E.T., Getz, G., Hadley, K., Handsaker, R.E., Huang, K.H., Kashin, S., Karczewski, K.J., et al. (2017). Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213.
- Almagro Armenteros, J.J., Sønderby, C.K., Sønderby, S.K., Nielsen, H., and Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**, 3387–3395.
- Amaral, P.P., Clark, M.B., Gascoigne, D.K., Dinger, M.E., and Mattick, J.S. (2011). lncRNADB: a reference database for long noncoding RNAs. *Nucleic Acids Res.* **39**, D146–51.
- Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S., and Jacobson, A. (2004). A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* **432**, 112–118.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169.
- 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., et al. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**, 595–606.
- Anderson, D.M., Makarewich, C.A., Anderson, K.M., Shelton, J.M., Bezprozvannaya, S., Bassel-Duby, R., and Olson, E.N. (2016a). Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides. *Sci. Signal.* **9**, ra119–ra119.
- Anderson, K.M., Anderson, D.M., McAnally, J.R., Shelton, J.M., Bassel-Duby, R., and Olson, E.N. (2016b). Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature* **2**, 1–13.
- Andrews, S.J., and Rothnagel, J.A. (2014). Emerging evidence for functional peptides encoded by short open reading frames. *Nat Rev Genet* **15**, 193–204.
- Aspden, J.L., Eyre-Walker, Y.C., Phillips, R.J., Amin, U., Mumtaz, M.A.S., Brocard, M., and Couso, J.-P. (2014). Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *Elife* **3**, e03528.
- Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1–11.10.33.
- Backes, S., and Herrmann, J.M. (2018). Protein Translocation into the Intermembrane Space and Matrix of Mitochondria: Mechanisms and Driving Forces. *Front. Mol. Biosci.* **4**, 83.
- Banci, L., Bertini, I., Cefaro, C., Ciolfi-Baffoni, S., Gallo, A., Martinelli, M., Sideris, D.P., Katrakili, N., and Tokatlidis, K. (2009). MIA40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. *Nat. Struct. Mol. Biol.* **16**, 198–206.
- Bánfai, B., Jia, H., Khatun, J., Wood, E., Risk, B., Gundling, W.E., Kundaje, A., Gunawardena, H.P., Yu, Y., Xie, L., et al. (2012). Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* **22**, 1646–1657.
- Bartha, I., Rausell, A., McLaren, P.J., Mohammadi, P., Tardaguila, M., Chaturvedi, N., Fellay, J., and Telenti, A. (2015). The Characteristics of Heterozygous Protein Truncating Variants in the Human Genome. *PLoS Comput. Biol.* **11**, e1004647.
- Battle, A., Khan, Z., Wang, S.H., Mitrano, A., Ford, M.J., Pritchard, J.K., and Gilad, Y. (2015). Impact of regulatory variation from RNA to protein. *Science* **347**, 664–667.
- 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., et al. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* **33**, 981–993.
- Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335**, 552–557.
- BurrIDGE, P.W., Matsa, E., Shukla, P., Lin, Z.C., Churko, J.M., Ebert, A.D., Lan, F., Diecke, S., Huber, B., Mordwinkin, N.M., et al. (2014). Chemically defined generation of human cardiomyocytes. *Nat. Methods* **11**, 855–860.
- Cabili, M.N., Dunagin, M.C., McClanahan, P.D., Biaesch, A., Padovan-Merhar, O., Regev, A., Rinn, J.L., and Raj, A. (2015). Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **16**, 20.
- Calviello, L., and Ohler, U. (2017). Beyond Read-Counts: Ribo-seq Data Analysis to Understand the Functions of the Transcriptome. *Trends Genet.* **33**, 728–744.
- Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., Landthaler, M., Obermayer, B., and Ohler, U. (2016). Detecting actively translated open reading frames in ribosome profiling data. *Nat. Methods* **13**, 165–170.
- Cenik, C., Cenik, E.S., Byeon, G.W., Grubert, F., Candille, S.I., Spacek, D., Alsallakh, B., Tilgner, H., Araya, C.L., Tang, H., et al. (2015). Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res.* **25**, 1610–1621.
- Chew, G.-L., Pauli, A., and Schier, A.F. (2016). Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nat. Commun.* **7**, 11663.
- Chothani, S., Adami, E., Viswanathan, S., Hubner, N., Cook, S., Schafer, S., and Rackham, O. (2017). Reliable detection of translational regulation with Ribo-seq. *BioRxiv* 234344.
- Chothani, S., Schafer, S., Adami, E., Viswanathan, S., Widjaja, A.A., Langley, S.R., Tan, J., Chee, J.P., D'Agostino, G., van Heesch, S., et al. (2018). Translational control of cardiac fibrosis. *BioRxiv* 451666.

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol. Cell* 33, 717–726.

Couso, J.P., and Patraquim, P. (2017). Classification and function of small open reading frames. *Nat. Rev. Mol. Cell Biol.* 18, 575–589.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372.

DeBoever, C., Tanigawa, Y., McInnes, G., Lavertu, A., Chang, C., Bustamante, C.D., Daly, M.J., and Rivas, M.A. (2017). Medical relevance of protein-truncating variants across 337,208 individuals in the UK Biobank study. *BioRxiv* 179762.

Deo, R.C. (2016). Alternative Splicing, Internal Promoter, Nonsense-Mediated Decay, or All Three. *Circ. Cardiovasc. Genet.* 9, 419–425.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.

Doll, S., Dreßen, M., Geyer, P.E., Itzhak, D.N., Braun, C., Doppler, S.A., Meier, F., Deutsch, M.-A., Lahm, H., Lange, R., et al. (2017). Region and cell-type resolved quantitative proteomic map of the human heart. *Nat. Commun.* 8, 1469.

Dunn, J.G., Foo, C.K., Belletier, N.G., Gavis, E.R., and Weissman, J.S. (2013). Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *Elife* 2, e01179.

Elfakess, R., and Dikstein, R. (2008). A translation initiation element specific to mRNAs with very short 5'UTR that also regulates transcription. *PLoS One* 28, e3094.

Elias, J.E., and Gygi, S.P. (2010). Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics. In *Proteome Bioinformatics. Methods in Molecular Biology™ (Methods and Protocols)*, H. S., and J. A, eds. (Humana Press), pp. 55–71.

Elurbe, D.M., and Huynen, M.A. (2016). The origin of the supernumerary subunits and assembly factors of complex I: A treasure trove of pathway evolution. *Biochim. Biophys. Acta - Bioenerg.* 1857, 971–979.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016.

Fang, H., Huang, Y.F., Radhakrishnan, A., Siepel, A., Lyon, G.J., and Schatz, M.C. (2018). Scikit-ribo Enables Accurate Estimation and Robust Modeling of Translation Dynamics at Codon Resolution. *Cell Syst.* 6, 180–191.e4.

Fesenko, I., Kirov, I., Kniazhev, A., Khazigaleeva, R., Lazarev, V., Kharlampieva, D., Grafksaia, E., Zgoda, V., Butenko, I., Arapidi, G., et al. (2018). Distinct types of short open reading frames are translated in plant cells. *BioRxiv* 213736.

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., et al. (2015). A Regression-Based Analysis of Ribosome-Profiling Data Reveals a Conserved Complexity to Mammalian Translation. *Mol. Cell* 60, 816–827.

Flockhart, R.J., Webster, D.E., Qu, K., Mascarenhas, N., Kovalski, J., Kretz, M., and Khavari, P.A. (2012). BRAFV600E remodels the melanocyte transcriptome and induces BANCR to regulate melanoma cell migration. *Genome Res.* 22, 1006–1014.

Galant, D., Gaborit, B., Desgrouas, C., Abdesselam, I., Bernard, M., Levy, N., Merono, F., Coirault, C., Roll, P., Lagarde, A., et al. (2016). A Heterozygous ZMPSTE24 Mutation Associated with Severe Metabolic Syndrome, Ectopic Fat Accumulation, and Dilated Cardiomyopathy. *Cells* 5, e21.

Galindo, M.I., Pueyo, J.I., Fouix, S., Bishop, S.A., and Couso, J.P. (2007). Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol.* 5, 1052–1062.

Glažar, P., Papavasileiou, P., and Rajewsky, N. (2014). circBase: a database for circular RNAs. *RNA* 20, 1666–1670.

Graham, L.D., Pedersen, S.K., Brown, G.S., Ho, T., Kassir, Z., Moynihan, A.T., and Vizgoff, E.K. (2011). Colorectal Neoplasia Differentially Expressed (CRNDE), a Novel Gene with Elevated Expression in Colorectal Adenomas and Adenocarcinomas. *Genes and Cancer* 2, 829–840.

Gray, K.A., Yates, B., Seal, R.L., Wright, M.W., and Bruford, E.A. (2015). Genenames.org: The HGNC resources in 2015. *Nucleic Acids Res.* 43, D1079–D1085.

Guo, J.U., Agarwal, V., Guo, H., and Bartel, D.P. (2014). Expanded identification and characterization of mammalian circular RNAs. *Genome Biol* 15, 409.

Gupta, N., and Pevzner, P.A. (2009). False discovery rates of protein identifications: A strike against the two-peptide rule. *J. Proteome Res.* 8, 4173–4181.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* 17, 148.

Han, P., Li, W., Lin, C.-H., Yang, J., Shang, C., Nurnberg, S.T., Jin, K.K., Xu, W., Lin, C.-Y., Lin, C.-J., et al. (2014). A long noncoding RNA protects the heart from pathological hypertrophy. *Nature* 514, 102–106.

van Heesch, S., van Iterson, M., Jacobi, J., Boymans, S., Essers, P.B., de Bruijn, E., Hao, W., MacInnes, A.W., Cuppen, E., and Simonis, M. (2014). Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol* 15, R6.

Heindryckx, F., Binet, F., Ponticos, M., Rombouts, K., Lau, J., Kreuger, J., and Gerwins, P. (2016). Endoplasmic reticulum stress enhances fibrosis through IRE1 α -mediated degradation of miR-150 and XBP-1 splicing. *EMBO Mol. Med.* 8, 729–744.

Heinig, M., Adriaens, M.E., Schafer, S., van Deutekom, H.W.M., Lodder, E.M., Ware, J.S., Schneider, V., Felkin, L.E., Creemers, E.E., Meder, B., et al. (2017). Natural genetic variation of the cardiac transcriptome in non-

diseased donors and patients with dilated cardiomyopathy. *Genome Biol.* 18, 170.

Henis-Korenblit, S., Strumpf, N.L., Goldstaub, D., and Kimchi, A. (2000). A Novel Form of DAP5 Protein Accumulates in Apoptotic Cells as a Result of Caspase Cleavage and Internal Ribosome Entry Site-Mediated Translation. *Mol. Cell. Biol.* 20, 496–506.

Herman, D.S., Lam, L., Taylor, M.R.G., Wang, L., Teekakirikul, P., Christodoulou, D., Conner, L., DePalma, S.R., McDonough, B., Sparks, E., et al. (2012). Truncations of Titin Causing Dilated Cardiomyopathy. *N. Engl. J. Med.* 366, 619–628.

Hershberger, R.E., Hedges, D.J., and Morales, A. (2013). Dilated cardiomyopathy: The complexity of a diverse genetic architecture. *Nat. Rev. Cardiol.* 10, 531–547.

Hezroni, H., Ben-Tov Perry, R., Meir, Z., Housman, G., Lubelsky, Y., and Ulitsky, I. (2017). A subset of conserved mammalian long non-coding RNAs are fossils of ancestral protein-coding genes. *Genome Biol.* 18, 162.

Hinson, J.T., Chopra, A., Nafissi, N., Polacheck, W.J., Benson, C.C., Swist, S., Gorham, J., Yang, L., Schafer, S., Sheng, C.C., et al. (2015). Titin mutations in iPS cells define sarcomere insufficiency as a cause of dilated cardiomyopathy. *Science* 349, 982–986.

Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.* 6, 1–11.

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., et al. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142, 409–419.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-Wide Analysis in Vivo of Resolution Using Ribosome Profiling. *Science* 324, 218–223.

Janer, A., Van Karnebeek, C.D.M., Sasarman, F., Antonicka, H., Al Ghamdi, M., Shyr, C., Dunbar, M., Stockler-Ispiroglu, S., Ross, C.J., Vallance, H., et al. (2015). RMND1 deficiency associated with neonatal lactic acidosis, infantile onset renal failure, deafness, and multiorgan involvement. *Eur. J. Hum. Genet.* 23, 1301–1307.

Jenny, A. (2006). A translation-independent role of oskar RNA in early Drosophila oogenesis. *Development* 133, 2827–2833.

Jia, J., Werkmeister, E., Gonzalez-Hilarion, S., Leroy, C., Gruenert, D.C., Lafont, F., Tulasne, D., and Lejeune, F. (2017). Premature termination codon readthrough in human cells occurs in novel cytoplasmic foci and requires UPF proteins. *J. Cell Sci.* 130, 3009–3022.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361.

Keeling, K.M., Lanier, J., Du, M., Salas-Marco, J., Gao, L., Kaenjak-Angeletti, A., and Bedwell, D.M. (2004). Leaky termination at premature stop codons antagonizes nonsense-mediated mRNA decay in *S. cerevisiae*. *RNA* 10, 691–703.

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., et al. (2016). RBM20 Regulates Circular RNA Production from the Titin Gene. *Circ. Res.* 119, 996–1003.

Kim, M.-S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., et al. (2014). A draft map of the human proteome. *Nature* 509, 575–581.

Kondo, T., Plaza, S., Zanet, J., Benrabah, E., Valenti, P., Hashimoto, Y., Kobayashi, S., Payre, F., and Kageyama, Y. (2010). Small Peptides Switch the Transcriptional Activity of Shavenbaby During Drosophila Embryogenesis. *Science* 329, 336–339.

Kopp, F., and Mendell, J.T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* 172, 393–407.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15, 8125–8148.

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., et al. (2012). Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes Dev.* 26, 338–343.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580.

Kuhn, R.M., Haussler, D., and James Kent, W. (2013). The UCSC genome browser and associated tools. *Brief. Bioinform.* 14, 144–161.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.

Larsson, O., Sonenberg, N., and Nadon, R. (2011). anota: Analysis of differential translation in genome-wide studies. *Bioinformatics* 27, 1440–1441.

Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., Fatica, A., Santini, T., Andronache, A., Wade, M., et al. (2017). Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol. Cell* 66, 22–37.e9.

Lek, M., Karczewski, K.J., Minikel, E. V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291.

Leygue (2007). Steroid receptor RNA activator (SRA1): unusual bifaceted gene products with suspected relevance to breast cancer. *Nucl. Recept. Signal.* 4.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595.

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

Liang, X., Potter, J., Kumar, S., Ravinder, N., and Chesnut, J.D. (2017). Enhanced CRISPR/Cas9-mediated

precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J. Biotechnol.* **241**, 136–146.

Lin, M.F., Jungreis, I., and Kellis, M. (2011). PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**, i275–82.

Loughran, G., Chou, M.Y., Ivanov, I.P., Jungreis, I., Kellis, M., Kiran, A.M., Baranov, P. V., and Atkins, J.F. (2014). Evidence of efficient stop codon readthrough in four mammalian genes. *Nucleic Acids Res.* **42**, 8928–8938.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.

Lovett, P.S., and Rogers, E.J. (1996). Ribosome regulation by the nascent peptide. *Microbiol. Rev.* **60**, 366–385.

Low, T.Y., van Heesch, S., van den Toorn, H., Giansanti, P., Cristobal, A., Toonen, P., Schafer, S., Hubner, N., van Breukelen, B., Mohammed, S., et al. (2013). Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis. *Cell Rep.* **5**, 1469–1478.

Ma, J., Saghatelian, A., and Shokhirev, M.N. (2018). The influence of transcript assembly on the proteogenomics discovery of microproteins. *PLoS One* **13**, e0194518.

Mackowiak, S.D., Zauber, H., Bielow, C., Thiel, D., Kutz, K., Calviello, L., Mastrobuoni, G., Rajewsky, N., Kempa, S., Selbach, M., et al. (2015). Extensive identification and analysis of conserved small ORFs in animals. *Genome Biol* **16**, 179.

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968.

Makarewich, C.A., and Olson, E.N. (2017). Mining for Micropeptides. *Trends Cell Biol.* **27**, 685–696.

Makarewich, C.A., Baskin, K.K., Munir, A.Z., Bezprozvannaya, S., Sharma, G., Khemtong, C., Shah, A.M., McAnally, J.R., Malloy, C.R., Szveda, L.I., et al. (2018). MOXI Is a Mitochondrial Micropeptide That Enhances Fatty Acid β -Oxidation. *Cell Rep.* **23**, 3701–3709.

Mao, F., Xiao, L., Li, X., Liang, J., Teng, H., Cai, W., and Sun, Z.S. (2016). RBP-var: A database of functional variants involved in regulation mediated by RNA-binding proteins. *Nucleic Acids Res.* **44**, D154–D163.

Matsumoto, A., Pasut, A., Matsumoto, M., Yamashita, R., Fung, J., Monteleone, E., Saghatelian, A., Nakayama, K.I., Clohessy, J.G., and Pandolfi, P.P. (2016). mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature* **541**, 228–232.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A.Y., Cibulskis, K., Kernysky, A.M., Garimella, K. V, Altshuler, D., Gabriel, S., Daly, M.J., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.* **17**.

McNally, E.M., and Mestroni, L. (2017). Dilated cardiomyopathy: Genetic determinants and mechanisms. *Circ. Res.* **121**, 731–748.

Memczak, S., Papavasileiou, P., Peters, O., and Rajewsky, N. (2015). Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood. *PLoS One* **10**, e0141214.

Miao, F., Chen, J., Chen, Z., Song, Y., Pang, L., and Shi, M. (2018). LncRNA HAND2-AS1 inhibits non-small cell lung cancer migration, invasion and maintains cell stemness through the interactions with TGF- β 1. *Biosci. Rep.* **39**, BSR20181525.

Morris, D.R., and Geballe, A.P. (2000). Upstream Open Reading Frames as Regulators of mRNA Translation. *Mol. Cell. Biol.* **20**, 8635–8642.

Mukherjee, N., Calviello, L., Hirsekorn, A., De Pretis, S., Pelizzola, M., and Ohler, U. (2017). Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat. Struct. Mol. Biol.* **24**, 86–96.

Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. *Trends Biochem. Sci.* **23**, 198–199.

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., et al. (2016). A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* **351**, 271–275.

Nesvizhskii, A.I. (2014). Proteogenomics: Concepts, applications and computational strategies. *Nat. Methods* **11**, 1114–1125.

Okonko, D.O., and Shah, A.M. (2014). Heart failure: Mitochondrial dysfunction and oxidative stress in CHF. *Nat. Rev. Cardiol.* **12**, 6–8.

Olexiuk, V., Van Crielinge, W., and Menschaert, G. (2018). An update on sORFs.org: A repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res.* **4**, D497–D502.

Olshen, A.B., Hsieh, A.C., Stumpf, C.R., Olshen, R.A., Ruggero, D., and Taylor, B.S. (2013). Assessing gene-level translational control from ribosome profiling. *Bioinformatics* **29**, 2995–3002.

Omenn, G.S., Lane, L., Lundberg, E.K., Overall, C.M., and Deutsch, E.W. (2017). Progress on the HUPO Draft Human Proteome: 2017 Metrics of the Human Proteome Project. *J. Proteome Res.* **16**, 4281–4287.

Pamudurti, N.R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D., Ramberger, E., et al. (2017). Translation of CircRNAs. *Mol. Cell* **66**, 9-21.e7.

Pauli, A., Norris, M.L., Valen, E., Chew, G.-L.L., Gagnon, J.A., Zimmerman, S., Mitchell, A., Ma, J., Dubrulle, J., Reyon, D., et al. (2014). Toddler: An Embryonic Signal That Promotes Cell Movement via Apelin Receptors. *Science* **343**, 1248636–1248636.

Peixeiro, I., Inácio, Â., Barbosa, C., Silva, A.L., Liebhaber, S.A., and Romão, L. (2012). Interaction of PABPC1 with the translation initiation complex is critical to the NMD resistance of AUG-proximal nonsense mutations. *Nucleic Acids Res.* *40*, 1160–1173.

Pendás, A.M., Zhou, Z., Cadiñanos, J., Freije, J.M.P., Wang, J., Hultenby, K., Astudillo, A., Wernerson, A., Rodríguez, F., Tryggvason, K., et al. (2002). Defective prelamin A processing and muscular and adipocyte alterations in Zmpste24 metalloproteinase-deficient mice. *Nat. Genet.* *31*, 94–99.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al. (2018). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* *47*, 442–450.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* *33*, 290–295.

Petersen, T.N., Brunak, S., Von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* *8*, 785–786.

Picardi, E., D'Erchia, A.M., Giudice, C. Lo, and Pesole, G. (2017). REDlportal: A comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.* *45*, D750–D757.

Picotti, P., Rinner, O., Stallmach, R., Dautel, F., Farrah, T., Domon, B., Wenschuh, H., and Aebersold, R. (2010). High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat. Methods* *7*, 43–46.

Quek, X.C., Thomson, D.W., Maag, J.L. V, Bartonicek, N., Signal, B., Clark, M.B., Gloss, B.S., and Dinger, M.E. (2015). IncRNAdb v2.0: Expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.* *43*, D168–D173.

Raimondeau, E., Bufton, J.C., and Schaffitzel, C. (2018). New insights into the interplay between the translation machinery and nonsense-mediated mRNA decay factors. *Biochem. Soc. Trans.* *46*, 503–512.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop And Go Extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* *75*, 663–670.

Rathore, A., Chu, Q., Tan, D., Martinez, T.F., Donaldson, C.J., Diedrich, J.K., Yates, J.R., and Saghatelian, A. (2018). MIEF1 Microprotein Regulates Mitochondrial Translation. *Biochemistry* *57*, 5564–5575.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* *44*, W83–W89.

Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L., and Corn, J.E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* *34*, 339–344.

Rinn, J.L., and Chang, H.Y. (2012). Genome Regulation by Long Noncoding RNAs. *Annu. Rev. Biochem.* *81*, 145–166.

Roberts, A.M., Ware, J.S., Herman, D.S., Schafer, S., Baksi, J., Bick, A.G., Buchan, R.J., Walsh, R., John, S., Wilkinson, S., et al. (2015). Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci. Transl. Med.* *7*, 270ra6.

Ruiz-Orera, J., Verdagué-Grau, P., Villanueva-Cañas, J.L., Messeguer, X., and Albà, M.M. (2018). Translation of neutrally evolving peptides provides a basis for de novo gene evolution. *Nat. Ecol. Evol.* *2*, 890–896.

Saha, A., Kim, Y., Gewirtz, A.D.H., Jo, B., Gao, C., McDowell, I.C., Engelhardt, B.E., and Battle, A. (2017). Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res.* *27*, 1843–1858.

Savitski, M.M., Wilhelm, M., Hahne, H., Kuster, B., and Bantscheff, M. (2015). A Scalable Approach for Protein False Discovery Rate Estimation in Large Proteomic Data Sets. *Mol. Cell. Proteomics* *14*, 2394–2404.

Schafer, S., Adami, E., Heinig, M., Rodrigues, K.E.C., Kreuchwig, F., Silhavy, J., van Heesch, S., Simate, D., Rajewsky, N., Cuppen, E., et al. (2015a). Translational regulation shapes the molecular landscape of complex disease phenotypes. *Nat. Commun.* *6*, 7200.

Schafer, S., Miao, K., Benson, C.C., Heinig, M., Cook, S.A., and Hubner, N. (2015b). Alternative Splicing Signatures in RNA-seq Data: Percent Spliced in (PSI). *Curr. Protoc. Hum. Genet.* *87*, 11.16.1–11.16.14.

Schafer, S., de Marvao, A., Adami, E., Fiedler, L.R., Ng, B., Khin, E., Rackham, O.J.L., van Heesch, S., Pua, C.J., Kui, M., et al. (2017a). Titin-truncating variants affect heart function in disease cohorts and the general population. *Nat. Genet.* *49*, 46–53.

Schafer, S., Viswanathan, S., Widjaja, A.A., Lim, W.-W., Moreno-Moral, A., DeLaughter, D.M., Ng, B., Patone, G., Chow, K., Khin, E., et al. (2017b). IL11 is a crucial determinant of cardiovascular fibrosis. *Nature* *552*, 110–115.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* *9*, 671–675.

Sciarretta, S., Forte, M., Frati, G., and Sadoshima, J. (2018). New Insights Into the Role of mTOR Signaling in the Cardiovascular System. *Circ. Res.* *122*, 489–505.

Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* *28*, 1353–1358.

Sherry, S.T. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* *29*, 308–311.

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., and Saghatelian, A. (2013). Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat Chem Biol* *9*, 59–64.

Stein, C.S., Jادیya, P., Zhang, X., McLendon, J.M., Abouassaly, G.M., Witmer, N.H., Anderson, E.J., Elrod, J.W., and Boudreau, R.L. (2018). Mitoregulin: A lncRNA-Encoded Microprotein that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. *Cell Rep.* *23*, 3710–3720.e8.

Tanjore, H., Lawson, W.E., and Blackwell, T.S. (2013). Endoplasmic reticulum stress as a pro-fibrotic stimulus.

Biochim. Biophys. Acta - Mol. Basis Dis. 1832, 940–947.

Tayal, U., Prasad, S., and Cook, S.A. (2017). Genetics and genomics of dilated cardiomyopathy and systolic heart failure. *Genome Med.* 9, 20.

The Gene Ontology Consortium (2017). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* 45, D331–D338.

Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., and Sabatini, D.M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* 485, 109–113.

Tian, D., Sun, S., and Lee, J.T. (2010). The long noncoding RNA, Jpx, Is a molecular switch for X chromosome inactivation. *Cell* 143, 390–403.

Travers, J.G., Kamal, F.A., Robbins, J., Yutzey, K.E., and Blaxall, B.C. (2016). Cardiac fibrosis: The fibroblast awakens. *Circ. Res.* 118, 1021–1040.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 13, 731–740.

Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet* 17, 601–614.

Wang, Z., Zhang, X.-J., Ji, Y.-X., Zhang, P., Deng, K.-Q., Gong, J., Ren, S., Wang, X., Chen, I., Wang, H., et al. (2016). The long noncoding RNA Chaer defines an epigenetic checkpoint in cardiac hypertrophy. *Nat. Med.* 22, 1131–1139.

Watkins, H., Seidman, C.E., Seidman, J.G., Feng, H.S., and Sweeney, H.L. (1996). Expression and functional assessment of a truncated cardiac troponin T that causes hypertrophic cardiomyopathy: Evidence for a dominant negative action. *J. Clin. Invest.* 98, 2456–2461.

Wein, N., Vulin, A., Falzarano, M.S., Szegedy, C.A.K., Maiti, B., Findlay, A., Heller, K.N., Uhlén, M., Bakthavachalu, B., Messina, S., et al. (2014). Translation from a DMD exon 5 IRES results in a functional dystrophin isoform that attenuates dystrophinopathy in humans and mice. *Nat. Med.* 20, 992–1000.

Xiao, Z., Zou, Q., Liu, Y., and Yang, X. (2016). Genome-wide assessment of differential translations with ribosome profiling data. *Nat. Commun.* 7, 11194.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2014). The I-TASSER suite: Protein structure and function prediction. *Nat. Methods* 12, 7–8.

Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., Jin, Y., Yang, Y., Chen, L., Wang, Y., et al. (2017). Extensive translation of circular RNAs driven by N6-methyladenosine. *Cell Res.* 27, 626–641.

Yano, T., Shimoshige, S., Miki, T., Tanno, M., Mochizuki, A., Fujito, T., Yuda, S., Muranaka, A., Ogasawara, M., Hashimoto, A., et al. (2016). Clinical impact of myocardial mTORC1 activation in nonischemic dilated cardiomyopathy. *J. Mol. Cell. Cardiol.* 91, 6–9.

Young, T.L., Matsuda, T., and Cepko, C.L. (2005). The noncoding RNA Taurine Upregulated Gene 1 is required for differentiation of the murine retina. *Curr. Biol.* 15, 501–512.

Yu, X., Zhang, Y., Li, T., Ma, Z., Jia, H., Chen, Q., Zhao, Y., Zhai, L., Zhong, R., Li, C., et al. (2017). Long non-coding RNA Linc-RAM enhances myogenic differentiation by interacting with MyoD. *Nat. Commun.* 8, 14016.

Zhang, L., Salgado-Somoza, A., Vausort, M., Leszek, P., and Devaux, Y. (2018a). A heart-enriched antisense long non-coding RNA regulates the balance between cardiac and skeletal muscle triadin. *Biochim. Biophys. Acta - Mol. Cell Res.* 1865, 247–258.

Zhang, Y., Jiao, L., Sun, L., Li, Y., Gao, Y., Xu, C., Shao, Y., Li, M., Li, C., Lu, Y., et al. (2018b). LncRNA ZFAS1 as a SERCA2a Inhibitor to Cause Intracellular Ca²⁺ Overload and Contractile Dysfunction in a Mouse Model of Myocardial Infarction Novelty and Significance. *Circ. Res.* 122, 1354–1368.

Zhong, Y., Karaletsos, T., Drewe, P., Sreedharan, V.T., Kuo, D., Singh, K., Wendel, H.G., and Räscher, G. (2017). RiboDiff: Detecting changes of mRNA translation efficiency from ribosome footprints. *Bioinformatics* 33, 139–141.

Zhu, P., Wu, J., Wang, Y., Zhu, X., Lu, T., Liu, B., He, L., Ye, B., Wang, S., Meng, S., et al. (2018). LncGata6 maintains stemness of intestinal stem cells and promotes intestinal tumorigenesis. *Nat. Cell Biol.* 20, 1134–1144.

Zickermann, V., Angerer, H., Ding, M.G., Nübel, E., and Brandt, U. (2010). Small single transmembrane domain (STMD) proteins organize the hydrophobic subunits of large membrane protein complexes. *FEBS Lett.* 584, 2516–2525.

Zou, J., Tran, D., Baalbaki, M., Tang, L.F., Poon, A., Pelonero, A., Titus, E.W., Yuan, C., Shi, C., Patchava, S., et al. (2015). An internal promoter underlies the difference in disease severity between N- and C-terminal truncation mutations of titin in zebrafish. *Elife* 4, e09406.