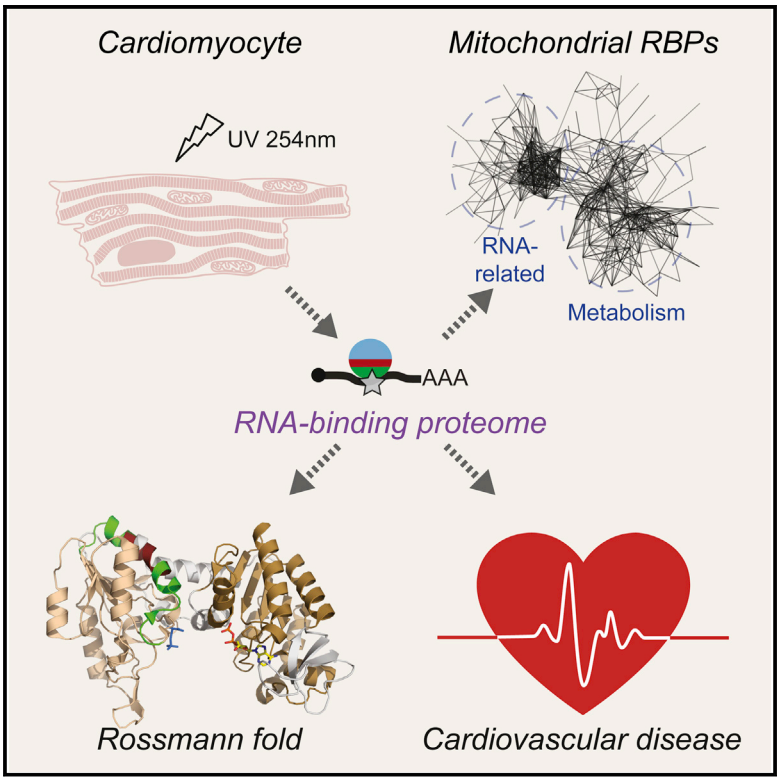
Cell Reports

Resource

# The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease

## Graphical Abstract




## Authors

Yalin Liao, Alfredo Castello, Bernd Fischer, ..., Jeroen Krijgsveld, Matthias W. Hentze, Thomas Preiss

## Correspondence

thomas.preiss@anu.edu.au

## In Brief

RNA functions through dynamic interactions with RNA-binding proteins (RBPs) in all clades of life. In this article, Liao et al. present the RBP repertoire of murine cardiomyocytes. Their findings reflect the unique cardiomyocyte biology and raise the prospect of previously hidden RNA-mediated regulatory interactions between gene expression, physiology, and metabolism.

## Highlights

- mRNA interactome capture and RBDmap reveal the cardiomyocyte RNA-binding proteome
- 1,148 RBPs are identified, 393 of which are thus far unique to cardiomyocytes
- Many cardiac RBPs have links to heart disease and mitochondrial metabolism
- Contacts of metabolic enzymes with RNA frequently involve Rossmann fold domains

## Accession Numbers

PXD002541
PXD002543

Liao et al., 2016, Cell Reports *16*, 1456–1469
August 2, 2016 
http://dx.doi.org/10.1016/j.celrep.2016.06.084

CellPress

OPEN ACCESS
CellPress

Cell Reports
Resource

# The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease

Yalin Liao,[1] Alfredo Castello,[2,4] Bernd Fischer,[2,5] Stefan Leicht,[2] Sophia Föehr,[2] Christian K. Frese,[2] Chikako Ragan,[1] Sebastian Kurscheid,[1] Eloisa Pagler,[1] Hao Yang,[1] Jeroen Krijgsveld,[2,5] Matthias W. Hentze,[2] and Thomas Preiss[1,3,*]
[1]EMBL–Australia Collaborating Group, Department of Genome Sciences, The John Curtin School of Medical Research (JCSMR), The Australian National University, Acton (Canberra) ACT 2601, Australia
[2]European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany
[3]Victor Chang Cardiac Research Institute, Darlinghurst (Sydney), NSW 2010, Australia
[4]Present address: Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK
[5]Present address: German Cancer Research Centre, 69120 Heidelberg, Germany
*Correspondence: thomas.preiss@anu.edu.au
http://dx.doi.org/10.1016/j.celrep.2016.06.084

## SUMMARY

**RNA functions through the dynamic formation of complexes with RNA-binding proteins (RBPs) in all clades of life. We determined the RBP repertoire of beating cardiomyocytic HL-1 cells by jointly employing two in vivo proteomic methods, mRNA interactome capture and RBDmap. Together, these yielded 1,148 RBPs, 391 of which are shared with all other available mammalian RBP repertoires, while 393 are thus far unique to cardiomyocytes. RBDmap further identified 568 regions of RNA contact within 368 RBPs. The cardiomyocyte mRNA interactome composition reflects their unique biology. Proteins with roles in cardiovascular physiology or disease, mitochondrial function, and intermediary metabolism are all highly represented. Notably, we identified 73 metabolic enzymes as RBPs. RNA-enzyme contacts frequently involve Rossmann fold domains with examples in evidence of both, mutual exclusivity of, or compatibility between RNA binding and enzymatic function. Our findings raise the prospect of previously hidden RNA-mediated regulatory interactions among cardiomyocyte gene expression, physiology, and metabolism.**

## INTRODUCTION

RNA-binding proteins (RBPs) are critical interaction partners for all cellular RNAs. RNAs recruit RBPs to form dynamic ribonucleoprotein particles (RNPs), and it is these assemblies that execute RNA function (Castello et al., 2013; Chen and Shyu, 2014; Singh et al., 2015). Multiple sequencing-based RBP footprinting studies (König et al., 2012) have now attested to the long-held view that RBPs interact with RNA in an intricate and highly combinatorial manner (Keene, 2007). The identification of proteins that co-purify with polyadenylated RNA from native cellular contexts by mass spectrometry, dubbed mRNA interactome capture, has yielded the first comprehensive views of active RBPs in eukaryotic cells. Specifically, mRNA interactome capture from human cervical cancer (HeLa) (Castello et al., 2012) and embryonic kidney (HEK293) cells (Baltz et al., 2012), murine embryonic stem cells (ESCs) (Kwon et al., 2013), *S. cerevisiae* (Beckmann et al., 2015; Matia-González et al., 2015; Mitchell et al., 2013), *C. elegans* (Matia-González et al., 2015), as well as human hepatoma cells (HuH-7) (Beckmann et al., 2015) together identified over 1,000 RBPs, many of which had no prior RNA-related annotation. Despite this already considerable expansion of mRNP componentry, more distinct cellular contexts need to be studied to define context-specific RBP repertoires.

The identification of so many RBPs indicates hitherto-unknown connections between seemingly disparate cellular processes and unexpected "moonlighting" activities that proteins carry out in a highly compartmentalized cellular environment (Copley, 2012). One such area deserving of further exploration is suggested by the reported ability of several metabolic enzymes to interact with RNA (Castello et al., 2015; Cieśla, 2006; Hentze, 1994; Hentze and Preiss, 2010). Enzymes could moonlight to regulate mRNA utilization in response to co-factor or metabolite levels, as documented for the cytosolic aconitase (ACO1)/iron regulatory protein (IRP1) paradigm (Muckenthaler et al., 2008). Conversely, RNA could affect enzyme activity, and collectively these interactions could form regulatory RNA-enzyme-metabolite (REM) networks (Hentze and Preiss, 2010).

Here, we report the adaptation and parallel application of mRNA interactome capture (Castello et al., 2012) and RBDmap (Castello et al., 2016) to beating murine HL-1 cardiomyocytes (Claycomb et al., 1998). We chose cardiomyocytes because of their unique metabolic requirements and importance to human disease (Rosca et al., 2013), gaining insight especially into the scope of mitochondrial RBPs.

## RESULTS

### mRNA Interactome Capture and RBDmap Uncover the RNA-Binding Proteome of Cardiomyocytes

Both proteomic methods require large numbers of cells, and thus we employed the murine HL-1 cardiomyocyte cell line as readily renewable source material. HL-1 cells are widely used

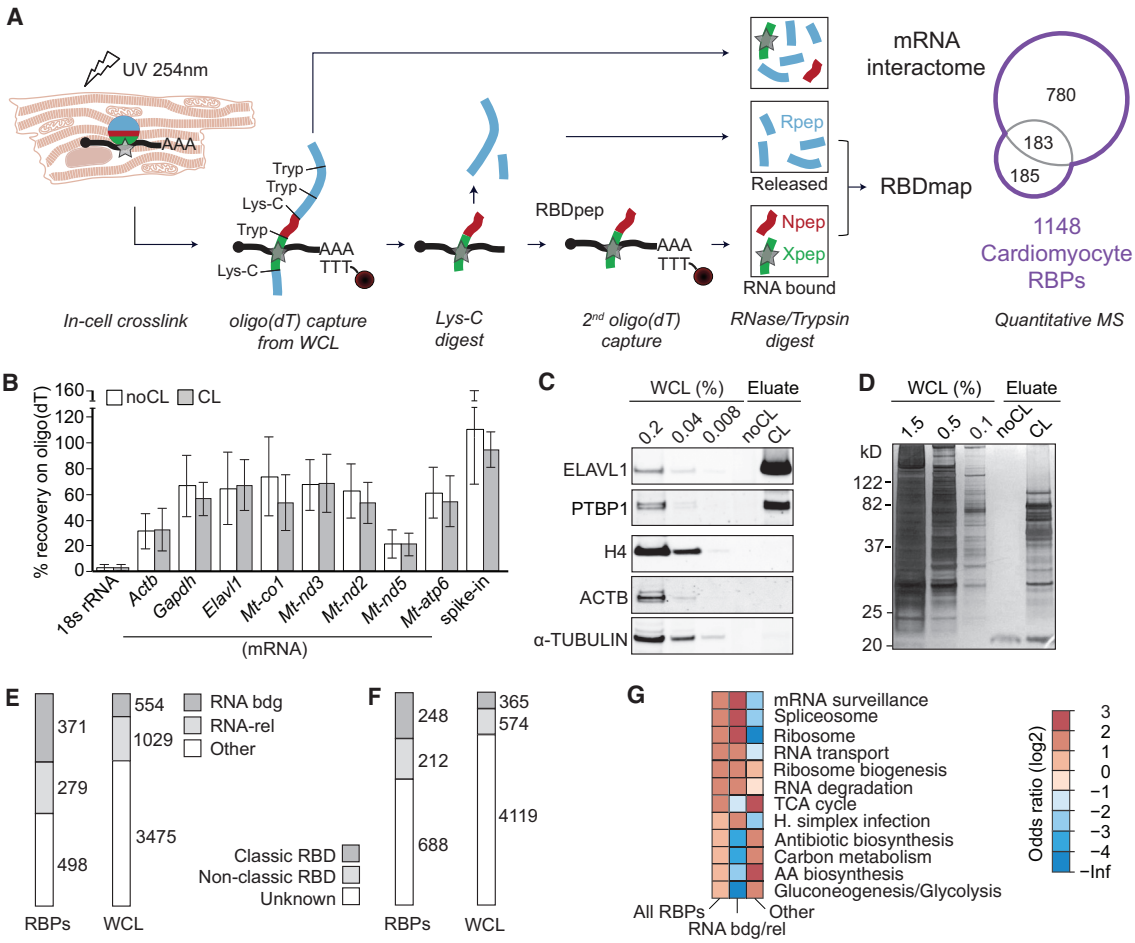



**Figure 1. Identification of Cardiomyocyte RBPs**

(A) Schematic of mRNA interactome capture and RBDmap approaches. Proteins identified by one or both approaches constituted a superset of cardiomyocyte RBPs.

(B) RNA-protein complexes captured on oligo(dT) beads were digested with proteinase K and RNA recovery levels monitored by quantitative PCR. Shown are averages of six biological replicates. Error bars, SD.

(C and D) Complexes captured on oligo(dT) beads were eluted by RNase-digestion, resolved by SDS-PAGE alongside input WCL (percentage equivalent to loaded eluate amount as indicated), and analyzed by western blot (C; see Experimental Procedures for antibody details) or silver stain (D). Results are representative of three independent interactome capture experiments. (See Figures S1A and S1B for equivalent RBDmap controls.)

(E) Proportion of cardiomyocyte RBPs or WCL proteins with GO annotation "RNA binding" or RNA-related annotations (see Supplemental Experimental Procedures for details on annotation sources). Left, cardiomyocyte RBPs; right, WCL.

(F) Analyses as in (E) for RBDs.

(G) Matrix plot of enriched/depleted KEGG pathways among RBP groups as defined in (E), compared to WCL.

See also Figures S1 and S2 and Table S1.

as a model as they retain many characteristics of adult cardiomyocytes (Claycomb et al., 1998), although compared to native myocardium their energy metabolism is less organized (e.g., Eimre et al., 2008). Confluent and spontaneously beating murine HL-1 cells were irradiated with UV light (150 mJ/cm$^2$ at 254 nm) to induce covalent crosslinks (CLs) between proteins and RNA within native complexes. Under these conditions, crosslink formation will typically be sub-stoichiometric and selectively occur at "zero" distance protein-RNA interactions (Castello et al., 2012, 2016). After denaturing cell lysis, RNA-protein complexes were captured on oligo(dT) beads and washed with high-salt/anionic detergent buffer to remove non-crosslinked proteins. Bound material was then processed for "mRNA interactome capture" to determine the scope of cardiomyocyte RBPs or for "RBDmap", which maps protein regions contacting RNA, as schematized in Figure 1A. Three independent biological replicates were processed for each approach, as well as parallel non-crosslinked controls (noCL). Aliquots taken after the first round of capture demonstrated selective purification of polyadenylated RNAs (cytosolic and mitochondrial; Figure 1B), known RBPs such as ELAVL1 and PTBP1 (Figures 1C and S1A), and a distinct subset of cellular proteins as the mRNA interactome (Figures 1D and S1B). For mRNA interactome capture (Castello et al., 2012), RBPs were liberated by RNase treatment, digested

with Trypsin/Lys-C protease mix, and analyzed by quantitative mass spectrometry. This identified 963 high-confidence RBPs (false discovery rate [FDR], 1%; see Figure S1D for reproducibility between replicates). Mass spectrometry was also performed on whole-cell lysate (WCL) samples, which identified 4,749 proteins (Table S1).

The RBDmap approach is described and validated in detail elsewhere (Castello et al., 2016). In applying RBDmap here, complexes eluted after a first round of capture were digested with Lys-C only, before a second round of oligo(dT) purification (Figure S1C). This separated protein fragments into a "released" and an "RNA bound" pool, which were both treated with RNase and Trypsin. All tryptic fragments within the released pool, termed Rpeps, can in principle be detected by mass spectrometry (Figure 1A). By contrast, tryptic digestion of the "RNA-bound" pool will yield two types of peptides. One type (termed Xpep, depicted in green, Figure 1A) will still carry a remnant, or remnants, of crosslinked RNA. Due to this heterogeneous mass shift, Xpeps are difficult to be detected by mass spectrometry (see Castello et al., 2016; Kramer et al., 2014). The other type (termed neighboring peptide or Npep, depicted in red, Figure 1A) is of predictable mass and thus readily detectable. Extension from the Npep boundaries to adjacent Lys-C sites can nevertheless still predict the Xpep coordinates, although sub-stoichiometric crosslinking and protease cleavage will add some complexity to the assessment of individual examples. Note that proteolytic processing efficiency is close to 100% (Figure S1C). Therefore, matching Npep and Xpep(s) together constitute the original RNA-bound Lys-C fragment (termed RBDpep, Figure 1A). Proteins derived from 368 genes exhibited RBDpeps that were enriched in the RNA bound over the released pool at an FDR of 1% (see Figure S1E for reproducibility between replicates).

Proteins that lack a Trypsin cleavage site within their RBDpeps will not be "visible" by RBDmap but may still be detected by mRNA interactome capture. Conversely, the additional proteolytic step and enrichment by a second oligo(dT) capture round in RBDmap will reduce sample complexity and experimental noise, thus improving detection for another subset of proteins (Castello et al., 2016). Consequently, proteins identified by RBDmap substantially (65%), but not completely, overlap with those of the mRNA interactome approach (183 RBPs are shared with the HL-1 mRNA interactome Figure 1A; an additional 51 are shared with the following mRNA interactomes: HeLa, HEK293, HuH-7). Taken together, both approaches define a "superset" of 1,148 cardiomyocyte RBPs (Tables S1, S2, and S3).

## Known and Uncharacterized Cardiomyocyte RBPs Share Similar Functional Features

Compared to the WCL proteome, cardiomyocyte RBPs were enriched for RNA-related functions (56% have RNA-related annotation, Figure 1E; see Figure S1F for the most enriched/depleted Gene Ontology [GO] terms), as expected. A similar GO term enrichment was seen with either the mRNA interactome or RBDmap set individually (data not shown). KEGG pathway enrichment analysis confirms this but interestingly also indicates overrepresentation of several pathways of intermediary metabolism among RBPs without prior RNA-related annotation (Figure 1G; see below). Similarly, proteins with classic or non-classic RNA-binding domains (RBD) are also overrepresented (Figure 1F). Many known RBD types are enriched among the cardiomyocyte RBPs (Figures S1G and S1H). For instance, we captured most of the RRM-containing proteins expressed in HL-1 cells (139 out of 168; Figure S1G). Similar trends were seen for other RBDs, including DEAD box helicase, KH, PWI, and PUF domains as well as many of the expected zinc finger domain subtypes (Figures S1I and S1J). We further identified all eight proteins with the MIF4G fold expressed in HL-1 cells as RBPs (i.e., EIF4G1-3, UPF2, CWC22, CTIF, NOM1, NCBP1). These proteins relate to the nexus of mRNA splicing, translation initiation, and nonsense-mediated decay. Enrichment of the Nol1_Nop2_Fmu domain was driven by capture of four members of the NSUN family of RNA:$m^5C$ (5-methylcytosine) methyltransferases (NSUN1, -2, -4, -5). Nucleoside modifications in polyadenylated RNA have received much attention lately and indeed, 29 proteins with annotations related to "RNA modification" were detected as cardiomyocyte RBPs (Table S3), prominently covering the enzymology of $m^5C$, $m^6A$ ($N^6$-methyladenosine), and pseudouridine modifications, as well as adenosine to inosine editing, all shown to occur in mRNA (Jaffrey, 2014; Sibbritt et al., 2013). Beyond that, we find componentry involved in several additional RNA modifications, including $m^5U$ (5-methyluridine), $m^6{}_6A$ ($N^6$,$N^6$-dimethyladenosine), and D (dihydrouridine), suggesting that the diversity of RNA modifications in polyadenylated RNAs may be richer than currently documented. All seven mammalian pentatricopeptide repeat (PPR)-containing proteins were identified here. PPR proteins function in mitochondrial RNA metabolism (Lightowlers and Chrzanowska-Lightowlers, 2013) and recognize RNA in a modular manner (Filipovska and Rackham, 2013). Several other protein domains lacking reported RNA-binding activity were also well represented in our datasets (Figure S1K), possibly reflecting direct involvement in RNA binding in some cases. Histones were notable (mostly H1 and H2A variants), possibly relating to the known association of RNA with chromatin (Mondal et al., 2010). Consistent with observations in the HeLa cell mRNA interactome (Castello et al., 2012), we detected RNA binding by 11 peptidyl-prolyl isomerases (PPIases, Table S3). Several of these are known for their involvement in nuclear gene expression (i.e., chromatin structure, gene transcription, mRNA splicing, and export) and some have recognized RBDs (Schiene-Fischer, 2015); however, for others these findings suggest hitherto-unknown roles for RNA in PPIase biology and pathology (see below).

Cardiomyocyte RBPs also display the biophysical and sequence features expected of bona fide RBPs. They range from low to high abundance, with some tendency toward higher abundance compared to the WCL proteome (Figure S2A), reflecting similar observations reported elsewhere (Castello et al., 2016; Matia-González et al., 2015). Cardiomyocyte RBPs have no length bias, but they are shifted toward a more alkaline isoelectric point and lower hydrophobicity (Figures S2B–S2D), as reported for other mRNA interactomes (Castello et al., 2012; Kwon et al., 2013). They further display an elevated content of intrinsically disordered regions and are enriched for the amino acids arginine (R), lysine (K), tyrosine (Y), and glycine (G)

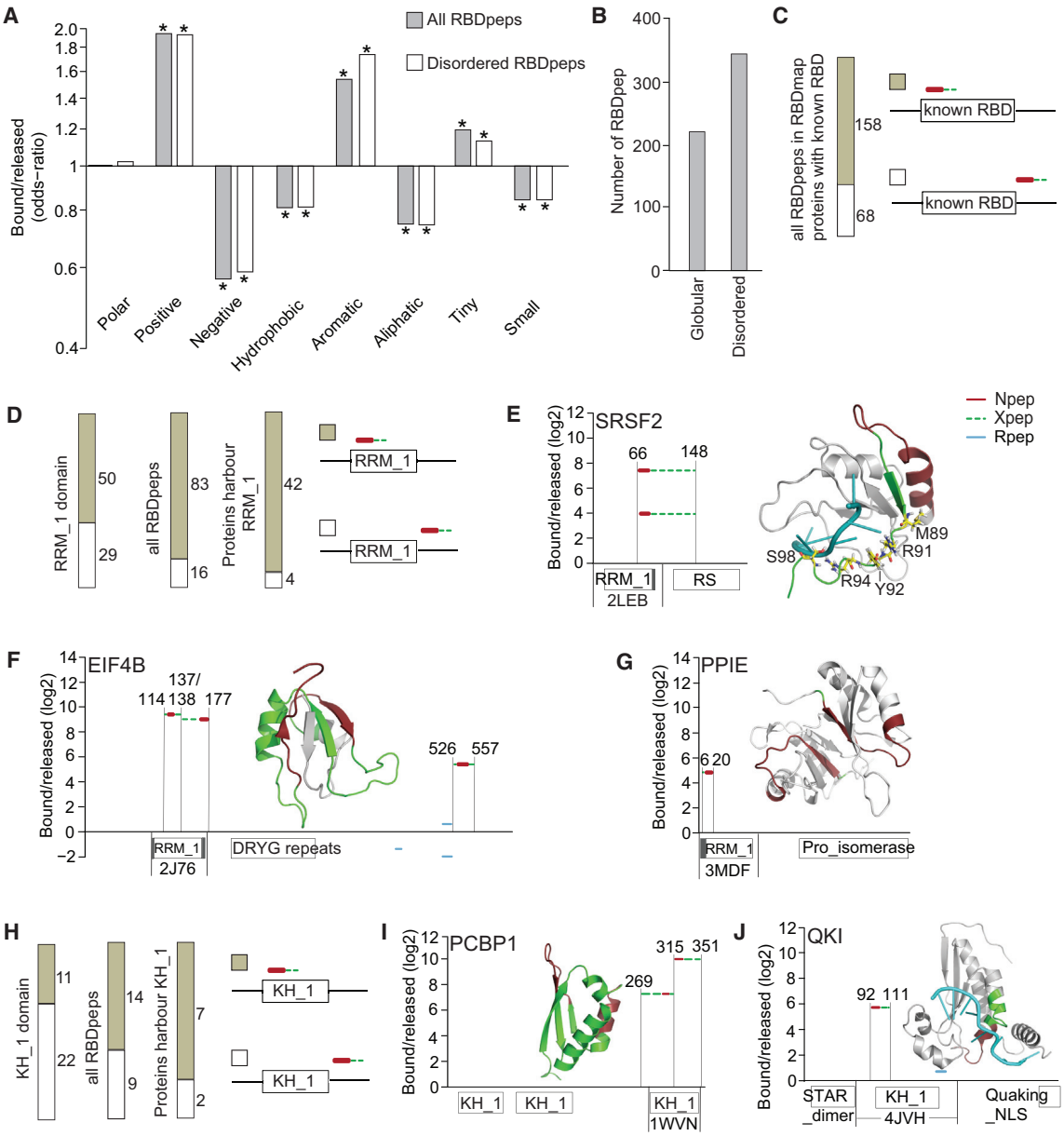



**Figure 2. Performance of the RBDmap Approach**

(A) Analysis of amino acid bias in RBDpeps (cf. Figure 1A) and disordered RBDpeps versus corresponding released fragments. *$p < 0.01$.

(B) RBDpep distribution between globular and disordered protein domains.

(C) Proportion of RBDpeps that overlap with a known RBD.

(D) In RBDmap proteins containing RRM_1, proportion of RRM_1 motifs with RBDpep coverage (left), of RBDpeps that overlap with an RRM_1 (middle) and of RRM_1-containing RBPs with at least one such overlap (right).

(E) Graphic representation of RBDmap data for SRSF2. Here, as in subsequent panels/figures displaying RBDmap data, Npeps (red) and Xpeps (green dashes; the crosslinked amino acid could reside anywhere within the Xpep), jointly termed RBDpep (positions of N- and C- termini are indicated), are mapped onto the linear protein sequence. y axis indicates enrichment (log2) in RNA bound over released fraction. Boxes underneath the x axis indicate Pfam-annotated domains with extensions (shaded gray) based on crystal structures, unless otherwise specified. See Supplemental Experimental Procedures for RS domain annotation. RBDpeps were also highlighted in the co-crystal structure (PDB: 2LEB) of the SRSF2 RRM_1 domain (ribbon diagram) with 5′-UCCAGU-3′ RNA (teal). Amino acids contacting RNA in structure are rendered as stick models.

(F) RBDmap data for EIF4B and crystal structure of the EIF4B RRM_1 (PDB: 2J76). Rpeps are shown in blue. DRYG repeats position was obtained from Méthot et al. (1996).

(G) RBDmap data for PPIE and crystal structure of PPIE RRM_1 dimer (PDB: 3MDF).

*(legend continued on next page)*

(Figures S2E and S2F). Low-complexity and repetitive regions, known to favor formation of RNA-protein granules (Kato et al., 2012) and to frequently occur in RBPs (Castello et al., 2012), are also overrepresented (Figures S2G and S2H). All of these features also apply to the subset of 688 cardiomyocyte RBPs lacking previously known RBDs. Amino acid patterns that center around the enriched residues R, K, Y, S, and G include RS repeats often found in splicing factors and resemble RGG and YGG boxes or poly(K) stretches, each previously implicated in RNA binding (Figure S2I) (Castello et al., 2012).

Overall, the RNA-binding proteome of cardiomyocytes shows substantial overlap with other mRNA interactomes: 429 RBPs are shared with the murine ESC set (Kwon et al., 2013), and 717 are in common with three available human mRNA interactomes from HEK293, HeLa, and HuH-7 cells (Baltz et al., 2012; Beckmann et al., 2015; Castello et al., 2012). 391 "core" RBPs emerge as "in common" between all reported mammalian mRNA interactomes, while 393 RBPs are thus far unique to the cardiomyocyte RBP set (Figure S2J). As expected, the core RBPs are particularly enriched for RNA-related GOMF annotations and recognized RBDs. The "unique" cardiomyocyte RBPs have lower proportions of these attributes and instead feature elevated proportions of proteins with mitochondrial localization, links to cardiovascular disease and development, genetic disease, and metabolic enzyme function (Figure S2K). This likely reflects the unique physiology of cardiomyocytes, e.g., their heavy reliance on mitochondrial metabolism, and their importance to human disease, aspects that we explore further below.

## Benchmarking Cardiomyocytic RBD Assignments by RBDmap

RBDmap identified 568 RBDpeps at 1% FDR, representing RNA-binding regions of 368 cardiomyocyte RBPs. The distances between Lys-C and Trypsin cleavage sites in individual proteins define the resolution achievable by RBDmap by determining the lengths of individual RBDpeps, Npeps and Xpeps (range/median: 7–161/20 amino acids (aa) for RBDpeps, 7–30/11 aa for Npeps, 1–148/11 aa for Xpeps) (Figure S3A). Limitations of RBDmap are that some protein-RNA contacts can be absent from the data due to (1) absence of suitable crosslink geometry; (2) lack of trypsin cleavage site(s) within RBDpeps; and (3) peptide detection biases of mass spectrometry. Moreover, Xpep can only give the boundaries within which the actual RNA-protein crosslink site is situated but cannot locate it further (see Castello et al., 2016 for a comprehensive evaluation). That said, RBDpeps and released peptides exhibit the expected divergent properties. Positively charged (R, K), aromatic (F, W, Y, H), and tiny amino acids (A, G, S, T, C) are enriched among RBDpeps, whereas negatively charged (D, E) and aliphatic amino acids (I, L, V) are more prevalent in released peptides (Figure 2A). A majority (347, 61.1%) of RBDpeps map to disordered regions (Figure 2B), and this subset retains the same amino acid bias (Figure 2A), evoking a broader role of disordered regions in RNA binding. These observations match those with RBDmap of HeLa cells (Castello et al., 2016).

To assess the consistency of RBDmap assignments, we identified proteins with homologs present in both the HL-1 and HeLa datasets (Figure S3A) and tested for concordance of RNA-binding site identifications. Approximately two-thirds of the associated mouse RBDpeps (161 of 237) had an equivalent counterpart in the human dataset (Figure S3B), a high proportion given the use of different cell lines from different organisms as well as the relative inefficiency and spatial limitations of UV crosslinking (Castello et al., 2012). Next, we examined the RBDmap data of 116 proteins with a classic or non-classic RBD. 158 of 226 RBDpeps derived from these proteins mapped to one of the known RBDs (Figure 2C). Forty-six of these proteins contained a total of 79 RRM_1 motifs. Reassuringly, RBDpeps overlapped with 50 of these RRM_1 motifs, and RNA binding mapped to at least one of their RRM_1 motifs for 42 out of the 46 protein examples examined (Figure 2D). Ten of the 12 members of the serine/arginine-rich splicing factor (SRSF) family (Busch and Hertel, 2012) were identified as cardiomyocyte RBPs (Figure S3C); seven of these have coverage in RBDmap. In each case, the RBDpep clearly overlaps with the first RRM_1 motif (Figures 2E and S3D–S3I). For SRSF2, the Xpep interval spans the C-terminal part of RRM_1 and extends into the RS domain. RS domains were proposed to bind RNA as well (Shen and Green, 2006). However, as there is prior evidence that RNA binding to SRSF2 does not require the RS domain (Daubner et al., 2012), the RNA crosslink(s) likely occurred within the C-terminal region of the RRM_1 domain. This would nicely reflect the crystal structure of the RRM_1 domain of SRSF2, which features direct RNA contacts within this region (Figure 2E, right panel). EIF4B and the splicing-associated PPIase PPIE are other proteins confirmed by RBDmap to contact RNA through their RRM_1 domain (Figures 2F and 2G). Nine KH domain proteins, containing a total of 33 KH motifs, were also part of the RBDmap dataset. Here, RBDpeps overlapped with 11 KH motifs, and RNA binding mapped to at least one of their KH motifs for seven out of the nine proteins examined (Figure 2H). Examples are the translational regulator poly(rC) binding protein 1 (PCBP1) (Ostareck et al., 1997) (Figure 2I) and the splicing regulator quaking (QKI) (Hall et al., 2013) (Figure 2J). Of further interest, RBDmap also marked a region adjacent to the NHL repeat domains in the E3 ubiquitin ligase TRIM2, consistent with recent reports that TRIM-NHL proteins are sequence-specific RNA-binding proteins (Kwon et al., 2013; Loedige et al., 2015), and the multiple zinc finger domains of the nuclear pore component and PPIase RANBP2, known to bind mRNAs encoding secretory proteins to promote their translation (Mahadevan et al., 2013) (data not shown). Taken together, these findings support the ability of RBDmap to identify protein regions in contact with RNA.

(H) Analyses as in (D) but for KH_1 motifs.
(I) RBDmap data for PCBP1 and crystal structure of the third KH_1 motif of PCBP1 (PDB: 1WVN).
(J) RBDmap data for QKI and crystal structure of QKI KH_1 in complex with 5′-ACUAACAA-3′ RNA (teal) (PDB: 4JVH).
See also Figure S3 and Table S2.

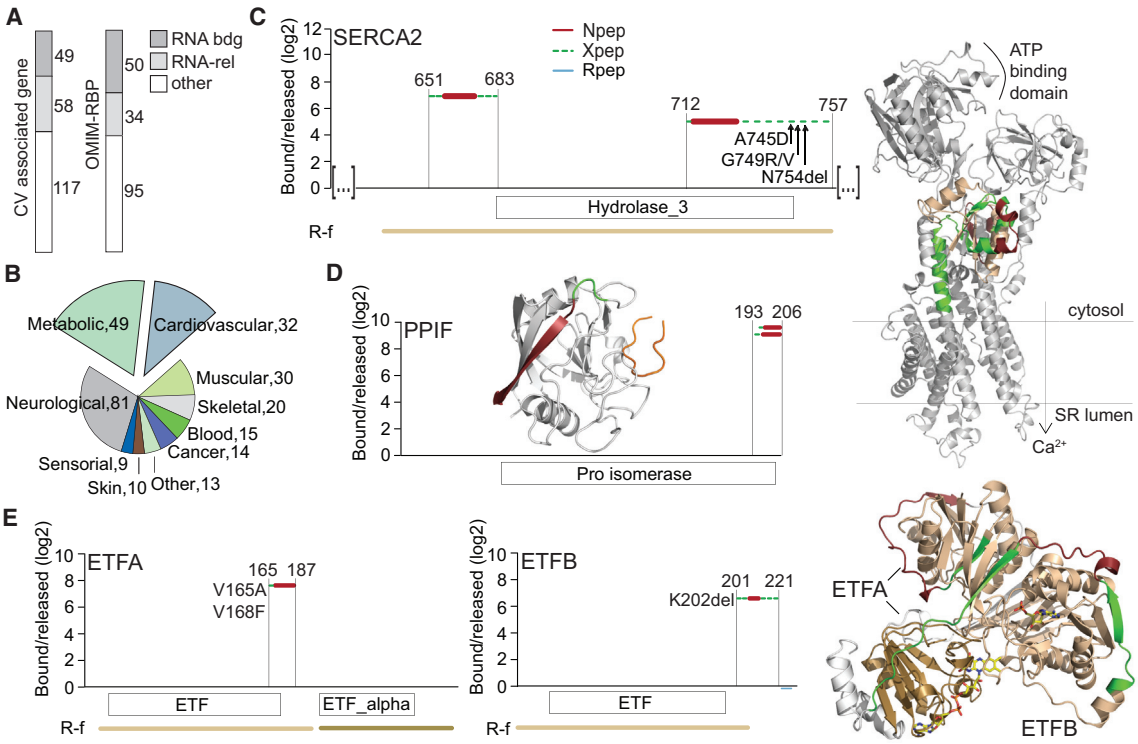



**Figure 3. Association of Cardiomyocyte RBPs with Cardiovascular Disease/Development and Human Mendelian Diseases**

(A) Analyses as in Figure 1E but for RBP association with cardiovascular disease/development (left) and genetic disease (based on OMIM; right).

(B) Spectrum of OMIM-listed genetic diseases caused by mutations in cardiomyocyte RBPs.

(C) RBDmap data for SERCA2 (amino acids 612–758; EC:3.6.3.8) and mapping onto Phyre2-modeled structure. Arrow indicates position of disease-associated missense mutation variants. A R-f domain is also highlighted (wheat).

(D) RBDmap data for PPIF (EC:5.2.1.8) and mapping onto co-crystal structure (PDB: 4TOT) of PPIF with inhibitor NIM258 (orange, PDB: 4TOT).

(E) RBDmap data for ETFA/ETFB and mapping onto crystal structure in complex with FAD and AMP (multicolor; 1EFV). A second R-f domain in ETFA is highlighted with a sand color.

See also Tables S3 and S4.

## Cardiomyocyte RBPs Have Rich Links to Heart Biology and Genetic Disease

Next, we defined human orthologs to murine cardiomyocyte RBPs to investigate links with heart function and disease. 222 cardiomyocyte RBPs relate to cardiovascular disease and development, based on the Cardiovascular Gene Ontology Annotation Initiative, and 117 of these had no prior RNA-related annotation (Figure 3A, left; Table S3). Similarly, 179 cardiomyocyte RBPs, when mutated, are known to cause genetic disease based on the Online Mendelian Inheritance in Man (OMIM) database; 84 of these have prior RNA-related annotation, while 95 lack it (Figure 3A, right; Table S3). A notable example of the former is the RRM_1/SR domain-containing splicing factor RBM20, known to cause dilated cardiomyopathy when mutated (Brauch et al., 2009) RBM20 is primarily expressed in striated muscle, with the highest levels found in the heart (Guo et al., 2012). Consistently, among the available mammalian mRNA interactomes it was uniquely detected in the HL-1 cardiomyocytes. Collectively, the OMIM-RBPs are associated with a spectrum of 273 genetic diseases (Table S4). We find neurological, metabolic, and cardiovascular disorders as the most prominent types of disease (Figure 3B). Together, these account for 59% of the diseases, highlighting the potential relevance of RNA binding to a molecular understanding of multiple genetic disorders and reflecting the common co-occurrence of neurological and cardiac manifestations of disease due to their high energy requirements (see below).

RBDmap data were available for 71 of the OMIM-RBPs (35 without prior RNA-related annotation) and for 12 of these RBDpeps cover known missense mutations or amino acid deletions (Table S3). Examples are HNRNPA2B1 and TARDBP, RRM-containing proteins that also harbor low-complexity regions or prion-like domains (PrLDs) with roles in RNA granule formation. Mutations in the PrLD of RBPs are implicated in neurodegenerative diseases, e.g., in HNRNPA2B1 and TARDBP, they cause multisystem proteinopathy and amyotrophic lateral sclerosis, respectively (Kim et al., 2013; King et al., 2012). Separate RBDpeps cover not only the RRM domain, but also the PrLD, overlapping with known disease mutations for both proteins (data not shown), raising the prospect of defective RNA binding as an additional disease mechanism in affected tissues. Seven of the 12 OMIM-RBPs lack prior RNA-related annotation and are of particular interest. For example, sarcoplasmic/endoplasmic reticulum calcium ATPase 2 (SERCA2)

has cardiovascular relevance as well as links to genetic disease. It regulates heart muscle contractile function by sequestration of calcium to the lumen of the sarcoplasmic reticulum during relaxation. SERCA2 activity is decreased during heart failure, and gene therapy trials to restore its function have shown promise (Gorski et al., 2015). Mutations in SERCA2 further cause the skin disorder Darier's disease (Zheng et al., 2015). RBDmap data locate the RNA-binding activity of this protein to a region adjacent to the ATP binding site in the cytosolic domain (Figure 3C). Interestingly the RBDpep overlaps with four known Darier's disease mutations (G749R, A745D, N754del [annotated in OMIM], G749V [Zheng et al., 2015]), suggesting a possible pathogenic role of altered RNA binding. Ten members of the aforementioned PPIase family have been implicated in cardiovascular disease (Perrucci et al., 2015). We identified four of these as RBPs: PPIC, -D, -F, and FKBP1A, the latter three also have RBDmap data (Table S3). PPIF regulates the mitochondrial permeability transition pore to affect mitochondrial calcium homeostasis and is involved in multiple cardiovascular pathologies (Perrucci et al., 2015). RBDmap locates RNA binding to the pro-isomerase domain of PPIF, in close proximity to a pharmacological inhibitor-binding site (Figure 3D).

Electron transfer flavoprotein (ETF) is a heterodimer located in the mitochondrial matrix and functions in the transfer of electrons from fatty acid oxidation (FAO), amino acid, and choline catabolism to oxidative phosphorylation (OXPHOS). The complex binds FAD in a cleft formed by the C-terminal domain of the α subunit (ETFA), and the β subunit (ETFB); AMP is bound entirely within ETFB. Mutations in ETF lead to the metabolic disorder multiple acyl-CoA dehydrogenase deficiency (MADD or GA-II) (Schiff et al., 2006). RBDmap detects RNA binding to both subunits, with RBDpeps mainly residing on the exterior of the complex, away from the FAD and AMP sites (Figure 3E). The RBDpeps overlap with three known MADD mutations (ETFA: V165A, V168F; ETFB: K202del [annotated in OMIM]). The occurrence of disease-causing mutations within regions of RNA contact thus suggests a role of RNA in a number of disease pathologies.

## Domains of the Rossmann Fold Topology Commonly Bind RNA

The enrichment of annotation relating to intermediary metabolism (Figure 1G) is driven by the presence of 73 metabolic enzymes (for a definition, see Supplemental Experimental Procedures) among the cardiomyocyte RBPs, densely covering cytosolic glycolysis as well as mitochondrial FAO, the TCA cycle, and OXPHOS (see below). Among these RNA-binding enzymes (Table S5; 24 with RBDmap data) are ten examples with prior "classic" literature (Castello et al., 2015; Cieśla, 2006; Copley, 2012; Hentze, 1994; Hentze and Preiss, 2010); another 13 were identified as RBPs by previous mRNA interactome studies, which identified a total of 23 metabolic enzymes (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013). Enzyme Commission (EC) classification EC1 (oxidoreductases; 33 examples) and EC2 (transferases; 25 examples) are particularly prevalent among the set of 73 RNA-binding enzymes (Figure 4A). Notably, most of these interact with either mono or di-nucleotides, e.g., ATP/GTP or $NAD(P)^+$/FAD, a function often carried out by a globular domain referred to as Rossmann fold (R-f). This domain is characterized by a six-stranded parallel β sheet with interlinking α helices on either side of the sheet (Caetano-Anollés et al., 2007; Rossmann et al., 1974). Thus, we used CATH/Gene3D annotation (Sillitoe et al., 2015) to search for R-f topology (CATH id: 3.40.50). 692 proteins in WCL match this topology, and, while there was no significant enrichment of R-f proteins among cardiomyocyte RBPs, it was still striking that 173 of them harbored a R-f domain. Among the R-f RBPs, 87 are with and 86 are without prior RNA-related annotation (Figure 4B; Table S3). 43 of R-f RBPs are covered by RBDmap data, and 33 of 61 RBDpeps derived from these proteins mapped to a R-f domain (Figure 4C). The group of R-f RBPs without prior RNA-related annotation is rich in metabolic enzymes but also includes other functions such as the aforementioned $Ca^{2+}$ ATPase SERCA2 (Figure 3C) and the flavoproteins ETFA&B (Figure 3E). Among the R-f superfamilies 3.40.50.300 ("P loop containing nucleotide triphosphate hydrolases"; 96 examples) is the most prevalent (Figure 4D). It includes 42 RNA helicases (27 DEAD-box, eight DEAH/RHA, three UPF1-, one Ski2-like, and three others) representing one-half of the 87 R-f RBPs with prior RNA-related annotation (Figures 4B and 4E). Notably, other members of the superfamily 3.40.50.300 include several DNA helicases. The large number of cardiomyocyte RBPs classified within the superfamily 3.40.50.720 ("NAD(P)-binding Rossmann-like Domain"; 18 examples) reflects the prevalence of metabolic enzymes among the RBPs lacking prior RNA-related annotation: 13 of the EC1 enzymes among the RBPs display the 3.40.50.720 R-f topology. Other well-represented superfamilies are 3.40.50.150 ("Vaccinia Virus protein VP39"; 18 examples), which references the presence of RNA-modifying enzymes, and 3.40.50.620 (HUPs; nine examples), common among tRNA synthetases. Taken together, the R-f topology emerges prominently among RBPs, both with and without prior RNA-related annotation.

The structural core of superfamily 1 and 2 helicases is composed of two helicase domains, each with R-f topology. ATP binds in a cleft formed by the two domains, while the nucleic acid contacts with the R-f domains are on the opposite surface of this core (Fairman-Williams et al., 2010), as exemplified by the co-crystal structure of EIF4A3 with RNA (Figure S4A). RBDmap data locate RNA-binding sites within nine helicases (DDX5, -17, -21, -27, -46; DHX8, -15, UPF1 and ASCC3). Structural modeling shows that the RBDpep overlaps with the expected RNA contact surface for the pre-mRNA splicing factor DHX15 (Figure 4F) and three additional examples (Figures S4B, S4C, and S4I). Interestingly, ASCC3 was previously described as a DNA helicase (Dango et al., 2011), while it is identified as an RBP here (Figure S4I). For DDX17, DDX27, DDX5, DHX8, and UPF1 the RBDpeps map to N-or C-terminal extensions of the helicase (Figures S4D–S4H), indicating additional RNA contacts beyond the helicase core. In UPF1, for example, the RBDpep maps to the C-terminal SQ domain that physically interacts with the central core and inhibits helicase activity (Figure S4H) (Fiorini et al., 2013). Notwithstanding these latter observations, our RBDmap data broadly concur with the expected arrangement of RNA contact sites within helicase R-f topology domains.

29 of the 73 metabolic enzymes identified here as RBPs have at least one annotated R-f. Inspection of available crystal

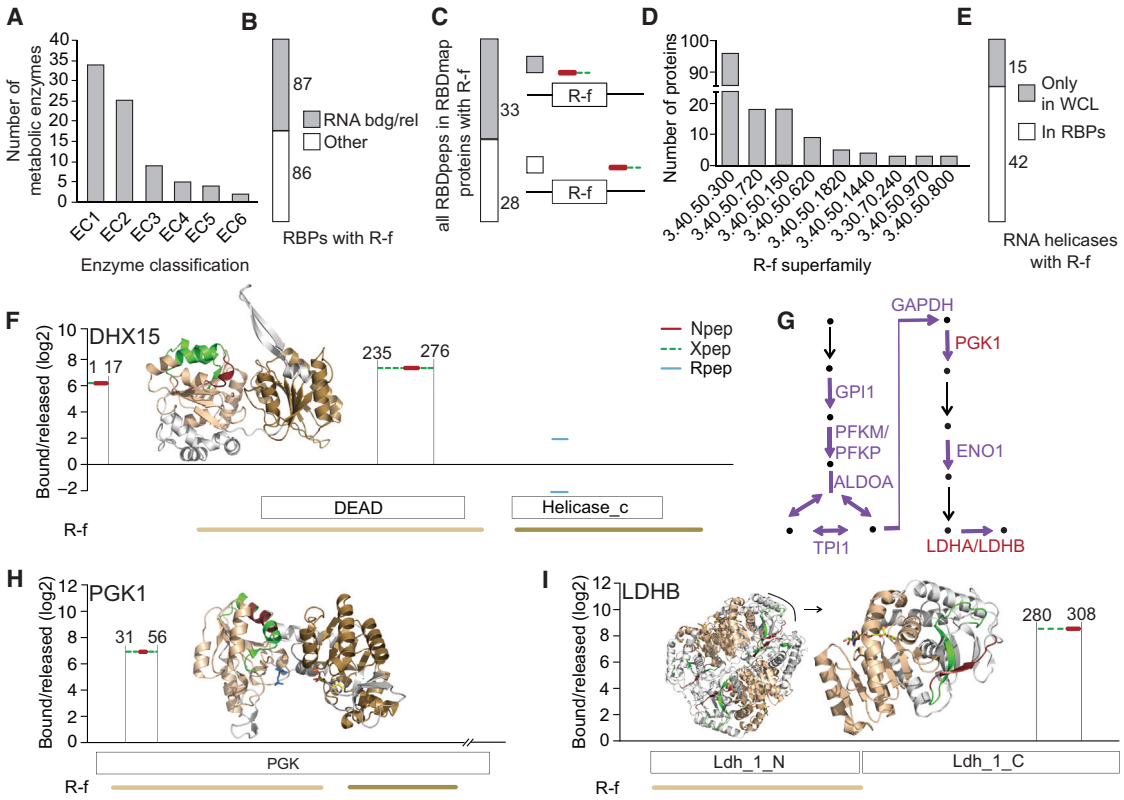



**Figure 4. Rossmann Fold Topology and RNA Binding**

(A) Classification of metabolic enzymes among cardiomyocyte RBPs.

(B) Analysis as in Figure 1E for R-f cardiomyocyte RBPs.

(C) Proportion of RBDpeps that overlap with a R-f domain.

(D) Distribution of cardiomyocyte RBPs across R-f superfamilies (by CATH id; only families with more than three members are shown).

(E) Proportion of WCL RNA helicases with R-f identified as RBPs.

(F) RBDmap data for DHX15 and mapping onto Phyre2-modeled structure. Here, and in the following panels, R-f domains are highlighted as applicable (N-terminal, wheat; C-terminal, sand).

(G) Schematic of glycolysis. Purple and red color indicate enzyme(s) present in cardiomyocyte RBPs and the RBDmap dataset, respectively.

(H) RBDmap data for PGK1 (EC:2.7.2.3) and mapping to tetrameric crystal structure in complex with 3-phosphoglyceric acid (orange) and ADP (multicolor) (PDB: 2XE7).

(I) RBDmap data for LDHB (EC:1.1.1.27) and mapping to tetramer crystal structure in complex with $NAD^+$ (multicolor) (PDB: 1I0Z). Enlarged view shows monomer.

See also Figures S4, S6, and S7 and Tables S3 and S5.

structures identified further enzymes that have domains of similar topology, increasing this list to 41 entries (Table S5). 27 of these bind either mono- (nine) or di-nucleotides (19). Available RBDmap data for 16 examples (25 RBDpeps) shows overlap between RNA binding and the R-f in 12 cases, including PGK1 from the densely covered glycolysis pathway (Figures 4G and 4H; further examples are shown in Figures 7, S6, and S7). In four additional cases the RBDpeps are spatially close to the R-f, including another glycolysis enzyme, LDHB (Figures 4I). This shows that the involvement of R-f topology domains is common among metabolic enzymes that bind RNA, confirming long-held expectations (Hentze, 1994). Unlike the helicases, R-f topology as well as tertiary and quaternary structure is more diverse among the metabolic enzymes, and thus we found the arrangement of RNA binding relative to R-f domain orientation also to be more varied.

## In-Depth Characterization of the Mitochondrial mRNA Interactome

Cardiomyocytes belong to the most highly energy-consuming cell types and are rich in mitochondria. 18.0% of WCL proteins are annotated with the GO term "mitochondrion," and this proportion is similar among cardiomyocyte RBPs (16.3% or 187 proteins; Figure S5A; 49 with RBDmap data). RNA and RBD-related annotations are highly represented among these mitochondrial mtRBPs (Figures S5B and S5C). Protein network analysis reveals two major groupings among mtRBPs: the first represents RBPs with known roles in RNA biology, while the second is centered on intermediary metabolism (Figure 5A). Among the 88 mtRBPs with known RNA roles (17 had RBDmap data), we found that all steps in the mitochondrial RNA life cycle (Rackham et al., 2012) are richly represented (Figure 5B). Cross-referencing the proteome of mitochondrial RNA processing granules

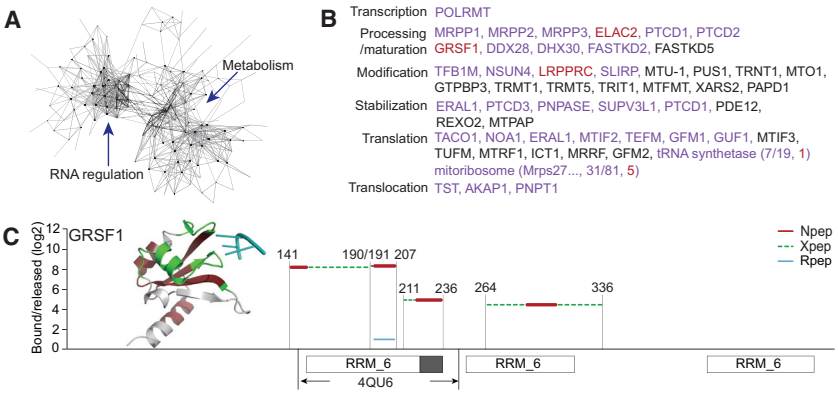



**Figure 5. Analysis of Cardiomyocyte mtRBPs**

(A) Functional protein association networks of mtRBPs based on STRING analysis (see Supplemental Experimental Procedures).

(B) Spectrum of known proteins involved in the mitochondrial RNA life cycle. Color scheme is as in Figure 4F.

(C) RBDmap data for GRSF1 isoform 1 and mapping to the co-crystal structure of N-terminal RRM_6 in complex with 5′-GGG-3′ RNA (teal) (PDB: 4QU6).

See also Figure S5 and Tables S1, S2, S3, and S5.

(Antonicka and Shoubridge, 2015), we further noticed an overlap with 54 of our mtRBPs. These include six PPR proteins, ranging from the mitochondrial RNA polymerase POLRMT, through the polycistronic pre-RNA processing factors LRPPRC, MRPP3 and PTCD1, to the translation regulators PTCD3 and MRPS27. All three subunits of the mitochondrial RNase P complex belong to this cross-section, as well as three recently described RNA granule components, FASTKD2, DDX28, and DHX30 (Antonicka and Shoubridge, 2015). RBDmap demarcates RNA binding to the first two RRM_6 domains of GRSF1 (Figure 5C), a central mitochondrial RNA granule component (Jourdain et al., 2013).

The 99 mtRBPs without prior RNA-related annotation (32 with RBDmap data) are dominated by 49 metabolic enzymes, densely covering FAO, TCA cycle, and OXPHOS (Figure 6A). Interestingly, the mitochondrial RNA granule proteome of 143B cells similarly contains multiple mitochondrial metabolic enzymes (Antonicka and Shoubridge, 2015); nine of these are shared with the list reported here. Is this an indication of a widespread role of RNA binding by mitochondrial metabolic enzymes? Of note, the MRPP3 subunit of RNase P has a PPR RBD, MRPP1 moonlights as an $m^1G_9$ RNA modifying enzyme and HSD17B10/MRPP2, is also an $NAD^+$-dependent short-chain fatty acid dehydrogenase (Beckmann et al., 2015; Rackham et al., 2012). All three RNase P subunits are identified by HL-1 cell interactome capture, indicating they each directly contact RNA. Remarkably, mutations in HSD17B10 cause progressive neurological abnormalities and cardiomyopathy in a manner that is unrelated to the protein's dehydrogenase activity (Deutschmann et al., 2014).

Twelve TCA cycle-related enzymes or enzyme subunits were identified as RBPs, covering most reaction steps (Figure 6A). Seven of these are further validated as RBPs by RBDmap data (depicted in Figure 7 and Figures S6D and S6E). We performed small-scale mRNA interactome capture experiments from HL-1 cells and probed western blots of material eluted from oligo(dT) beads with antibodies against six TCA cycle-related enzymes. This shows selective RNA-mediated purification of four of the six enzymes (Figure 6B; ACO2, CS, IDH2, OGDH). We also constructed a series of plasmids expressing enzymes C-terminally tagged with eGFP, transfected them into HeLa cells alongside positive (Pumilio-eGFP) and negative controls (eGFP alone, eGFP carrying an N-terminal mitochondrial localization signal), and confirmed their expression (Figure S5D) and intracellular localization (Figure S5E). Next, we immunopurified these fusion proteins from UV-crosslinked or control cells and subjected RNase-treated isolates to $^{32}P$ 5′ end labeling with polynucleotide kinase. This confirmed selective direct RNA association of fusion proteins in four cases (Figure 6C, left; ACO2, CS, IDH2, PDHA1). For ACO2, CS, IDH2, and MDH2, we also replicated these findings when performing pull-down and labeling from lysates of purified mitochondria (Figure 6C, right; Figure S5F). In summary, we obtained additional validation of RNA binding for ten TCA cycle-related enzymes by at least one method and two of them, IDH2 and ACO2, score positively in all four validation assays.

## RNA Binding and Metabolic Enzyme Function

We used RBDmap data for 24 metabolic enzymes to explore the relationship between RNA, substrate, and cofactor binding, taking into account available information on tertiary and quaternary structure. Focusing on the TCA cycle, the monomeric enzyme ACO2 is particularly interesting as its cytoplasmic counterpart ACO1/IRP1 is well known for its dual enzyme/RBP functions (Muckenthaler et al., 2008). An RBDpep is situated on one flank of the iron-sulfur cluster-containing cleft of ACO2 (Figures 7A and 7A′). Comparison to the crystal structure of IRP1 bound to the ferritin IRE-RNA stem-loop (Walden et al., 2006) (Figure 7A′, right) suggests a similar mode of RNA binding to the active site cleft of ACO2, although, as the predicted Xpep also extends further, RNA contact with the outer surface of the enzyme is also plausible. The mitochondrial TCA cycle enzyme $NAD^+$-dependent IDH3 is a hetero-tetramer consisting of two $\alpha$, one $\beta$, and one $\gamma$ subunit. No mammalian structures have been reported; all subunits are similar in sequence yet carry out specialized functions. An RBDpep is found at the C terminus of the $\alpha$ subunit (IDH3A; Figure 7B), and structural modeling locates it to the periphery of the monomer, away from regions important for catalysis (Figure 7B′). The two $NADP^+$-utilizing IDH isoforms, cytosolic IDH1 and mitochondrial IDH2, each have R-f topology and form homodimers. Here, the Npeps overlap with the active site cleft, while the extended Xpeps reside just outside of the R-f on the protein surface (Figures S6D, S6D′, S6E, and S6E′).

DLD forms homodimers that serve as the E3 component of two large hetero-oligomeric assemblies, the pyruvate dehydrogenase complex, which links glycolysis to the TCA cycle, and the TCA cycle-integral oxoglutarate dehydrogenase complex. Here, the Xpep regions extend throughout the surface that binds

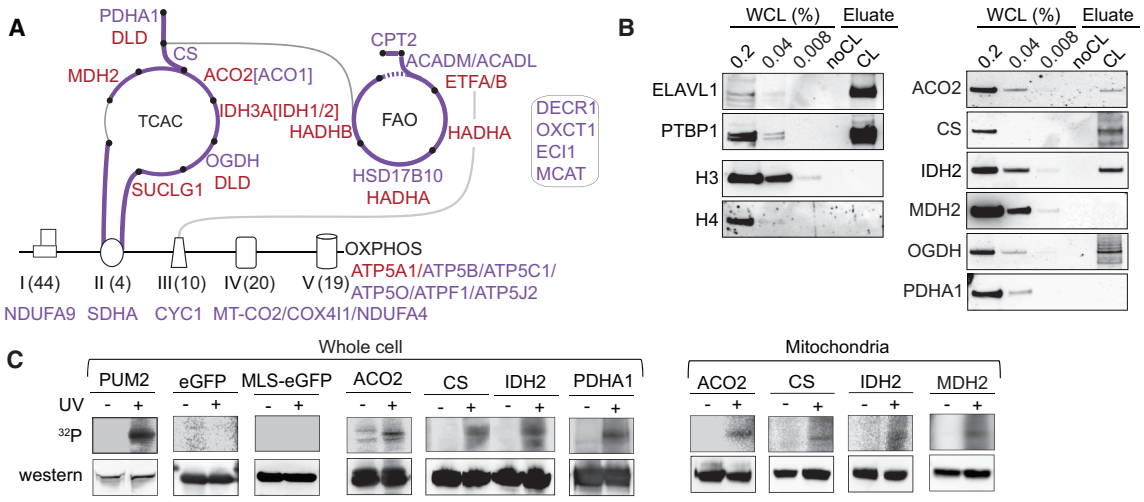



**Figure 6. Validation of RNA Binding by Mitochondrial Enzymes**

(A) Schematic of mitochondrial TCA cycle, FAO, and OXPHOS. Color scheme is as in Figure 4F. Number of known subunit variants for each OXPHOS complex is shown in brackets. NDUFA4 is considered as complex IV subunit (Balsa et al., 2012). Box to the right indicates FAO-associated enzymes.

(B) Small-scale mRNA interactome capture from HL-1 cells and western blots with antibodies against positive and negative controls (left; as in Figure 1C) as well as six endogenous TCA cycle-related enzymes (right; n = 1).

(C) Immunoprecipitation of eGFP-tagged proteins from UV-treated and control HeLa cells and $^{32}$P 5′ end labeling of crosslinked RNA fragments. Samples were processed for storage phosphor imaging (top) and western blot with anti-GFP antibody (bottom). PUM2-eGFP (at lower dose) is shown as a positive control, while eGFP alone, with or without a mitochondrial localization signal, and mock purifications from non-crosslinked cells served as negative controls (n = 1–3). Immunoprecipitation was done from WCL (left) or from purified mitochondria (right). Controls for intracellular localization of fusion proteins and mitochondrial isolation are shown in Figure S5.

See also Figure S5.

FAD but also include sections on the protein surface (Figures 7C and 7C′). Succinyl-CoA ligase is a heterodimer formed by the substrate-binding subunit, SUCLG1, and SUCLG2, which binds GDP. RBDpeps tag both R-f of SUCLG1 with solvent-exposed Npep portions located away from the enzyme's active site (Figures 7D and 7D′). MDH2 functions as a homodimer and the two RBDpeps in the central and C-terminal regions lie outside the R-f on the surface of the dimer (Figures 7E and 7E′).

The two glycolysis enzymes PGK1 and LDHB continue an emerging theme. PGK1 (Figure 4H) possesses two R-f and functions as a monomer. The RBDpep is located on the outer edge of the substrate-binding N-terminal R-f of PGK1 with one extended Xpep adjacent to the active site and the other facing away from it on the outer surface, while ATP binds to the C-terminal R-f. LDHB (Figure 4I) and the aforementioned non-enzymatic ETF complex (Figure 3E) present similar cases. Altogether, we examined 24 metabolic enzymes (additional examples shown in Figures S6 and S7) and noted diverse spatial relationships between RBDmap peptides and enzyme active sites. While mutually exclusive RNA binding and catalytic function must be expected in a few cases, surprisingly, a common theme shared by most examples is the potential for concurrent RNA binding and enzyme activity, suggesting the possibility of allosteric or scaffolding functions for RNA.

## DISCUSSION

We present here a comprehensive analysis of RNA-binding proteins in cardiomyocytes, constituting a rich resource of 1,148 cardiomyocyte RBPs and 568 regions of RNA contact within 368 of these RBPs. The data document and characterize the activity of many previously known RBPs in the cardiac context. For instance, just over half of the RBPs had prior RNA-related annotation and RBDmap identified RNA contact regions within recognized RBDs for many previously uncharacterized examples. Most interestingly, 393 of our RBPs were not seen in published interactomes. Among these will be some that remained undetected in past studies for technical reasons, but many were likely revealed here due to specific aspects of cardiomyocytic gene expression, metabolism, and function.

Over 200 of our RBPs have annotated roles in the cardiovascular system and/or are encoded by OMIM disease genes, and roughly half of these have no prior association with RNA biology. As expected, neurological disorders are prominent among the diseases linked to RBPs (Castello et al., 2013). Uniquely, almost half of the OMIM-RBPs identified here have links to metabolic or cardiovascular disorders. RBDmap identifies RNA contacts within 71 OMIM-RBPs, frequently overlapping with disease-causing mutations, and thus providing entry points into molecular exploration of disease relevance. The calcium channel SERCA2 is a case in point. RNA binding to its cytosolic domain may affect the function of SERCA2 in cardiac contractility and may thus represent a therapeutic target during heart failure. The RNA-binding region of SERCA2 further coincides with the location of mutations causing Darier's disease (Figure 3C).

Another feature of our data is the rich coverage of mitochondrial RNA biology. Around 190 proteins from mitochondria

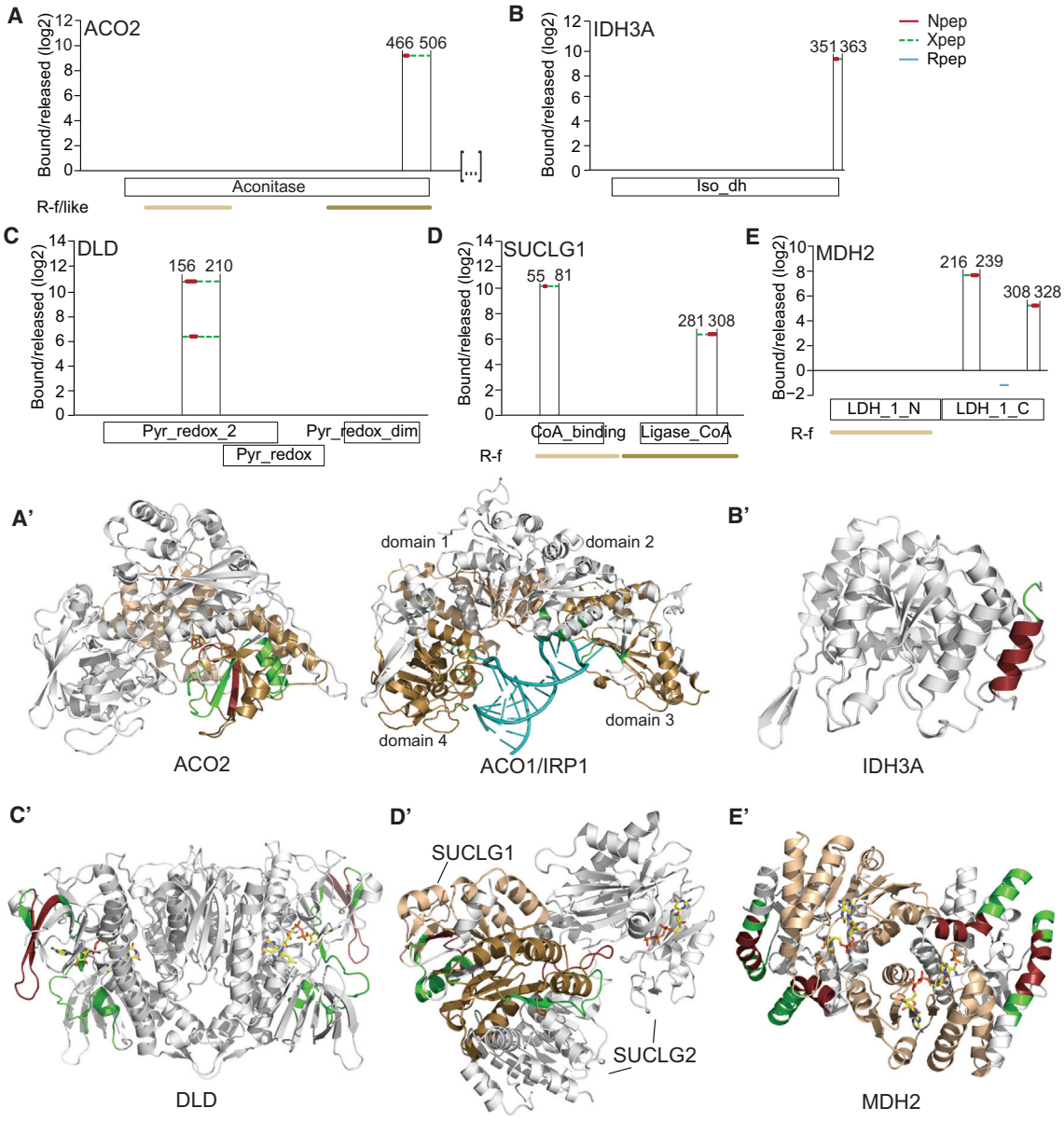



**Figure 7. RNA-Binding Modalities among TCA Cycle Enzymes**

(A–E) RBDmap data for ACO2 (amino acids 1–523; EC:4.2.1.3), IDH3A (EC:1.1.1.41), DLD (EC:1.8.1.4), SUCLG1 (EC:6.2.1.4, EC:6.2.1.5), and MDH2 (EC:1.1.1.37).

(A′–E′) RBDpeps were mapped to enzyme crystal or Phyre2-modeled structures.

(A′) ACO2 crystal structure with 4Fe-4S cluster (multicolor) (PDB: 6ACN) shown on left. IRP1/ACO1 co-crystal structure with IRE-RNA (teal) (PDB: 3SNP) shown for comparison on right; IRE contacts in IRP1 are depicted in green.

(B′) Phyre2-modeled structure of IDH3A.

(C′) Crystal structure of DLD dimer in complex with FAD (multicolor) (PDB: 1ZMC).

(D′) Crystal structure of SUCLG1-SUCLG2 heterodimer in complex with GTP (multicolor) (PDB: 2FP4).

(E′) Crystal structure of MDH2 dimer in complex with D-malate (orange) and $NAD^+$ (multicolor) (PDB: 2DFD).

See also Figures S6 and S7 and Tables S2 and S5.

were found to be active as RBPs in cardiomyocytes, and only roughly half of them had prior RNA-related annotation. This work independently confirms the link to RNA for the latter group, which includes many examples for which a clear role in the mitochondrial RNA life cycle has yet to be defined (Rackham et al., 2012). The other half of the mtRBPs lacks prior RNA-related annotation, opening up rich prospects for discovery. This might particularly apply to the 49 mitochondrial metabolic enzymes we identified as RBPs, densely covering the energy-generating pathways of the TCA cycle, FAO, and OXPHOS. What could

be the function of enzyme-RNA binding? One possibility is that a subset of enzyme molecules, perhaps in response to metabolite levels or flux through the associated pathway, could moonlight as RNA regulators. This is not without precedent in mitochondria: the MRPP2 subunit of mitochondrial RNase P has an alter ego as an $NAD^+$-dependent short-chain fatty acid dehydrogenase (Rackham et al., 2012). It is tempting to speculate that ACO2, in analogy to its cytosolic counterpart ACO1/IRP1 (Muckenthaler et al., 2008), functions in post-transcriptional regulation of mitochondrial mRNA expression. Future work will need to identify RNA targets for ACO2 and the many other interesting leads discovered here to unveil the physiological relevance of their interactions with RNA. In some cases, this might even reveal roles of RBPs outside of their usual mitochondrial localization.

Overall, the 1,148 cardiomyocyte RBPs are less acidic and hydrophobic than the cellular proteome and are enriched in intrinsically disordered, low-complexity, and repetitive regions. They further display enrichment of RNA-related functional and domain annotation. This independently confirms similar observations made with the compositionally quite distinct HeLa mRNA interactome (Castello et al., 2012). The RBDmap approach extends these observations. Relative to the released protein fragments, RBDpeps are enriched for positively charged aromatic and small amino acids, as befits RNA interaction surfaces. The majority of them also lie within disordered protein regions, which retain a similar amino acid bias, indicating that disordered protein regions represent surprisingly common sites of RNA contact. Again, even though made on the basis of a largely distinct peptide set, these findings concur with parallel observations made by applying RBDmap to HeLa cells (Castello et al., 2016). RBDmap marks well-understood RNA-binding domains with high accuracy and implicates several globular protein domains not previously expected to contact RNA, such as the proisomerase domain found in PPIases. Protein domains of the Rossmann fold are common among the cardiomyocyte RBPs, prominently found in RNA helicases and the RNA-binding enzymes. RBDmap affirms the expected and relatively uniform mode of RNA binding to helicases, which mostly share the same R-f topology. By contrast, not all RNA-binding enzymes have R-f domains, and those that do are of more varied topology. Accordingly, RBDmap delineates a diverse range of RNA-binding modalities for the metabolic enzymes.

RNA binding to so many metabolic enzymes and other proteins with functions apparently unrelated to RNA might suggest novel roles of RNA in affecting protein function, rather than the reverse. In the case of enzymes, roles as competitive inhibitor, allosteric regulator, or as an assembly scaffold are conceivable (Castello et al., 2015). We have interpreted RBDmap data for multiple metabolic enzymes in the context of their known tertiary and quaternary structures, and, while this does not resemble formal proof, we have found that the most straightforward explanation for the data as a whole is that RNA binding and catalytic activity need not clash.

RNA binding often applies to multiple enzymes functioning in a given pathway. Assembly of a pathway metabolon with superior metabolic flux properties has been described for the TCA cycle (Vélot et al., 1997; Wu and Minteer, 2015), among other pathways. It is thus an exiting possibility that RNAs may provide assembly scaffolds for metabolons, as suggested by the RNase-sensitivity of a complex of glycolytic enzymes (Mazurek et al., 1996). Such a role of RNA in metabolism would not be too dissimilar to scaffolding functions of noncoding RNAs in RNP aggregation, such as NEAT1 in paraspeckle assembly (Bond and Fox, 2009). Methods to biochemically isolate metabolons should be applied to search for putative RNA scaffolds in future work.

Altogether, we present here a comprehensive compendium of RNA-binding proteins active in cardiomyocytes and a survey of the regions they use to contact RNA. This work provides a fertile resource to study RNA-level regulation and its interfaces to other cellular processes in the context of mitochondrial function, cardiac (patho)physiology, and genetic disease.

## EXPERIMENTAL PROCEDURES

A detailed version of the experimental procedures can be found in the Supplemental Information.

### Cell Culture

HL-1 cardiomyocytes (a gift from W. Claycomb) were maintained as described (Claycomb et al., 1998). HeLa cells were purchased from ATCC and maintained by standard procedures.

### mRNA Interactome Capture

Beating HL-1 cells were irradiated with UV light at 254 nm. Cells were lysed, and RNA-protein complexes were captured on oligo$(dT)_{25}$ magnetic beads, which were washed, eluted, and processed for mass spectrometry essentially as described (Castello et al., 2012).

### RBDmap

Following mRNA interactome capture as detailed above, eluted RNA-protein complexes were digested with Lys-C followed by recapture on oligo$(dT)_{25}$ beads, separating a released fraction from RNA bound peptides, which were again eluted from the beads. Both released and RNA bound fractions were processed for mass spectrometry. RBDmap is described in Castello et al. (2016).

### LC-MS/MS

An LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Scientific) coupled to a nanoAcquity UPLC system (Waters) was used for detection of peptides in WCL, mRNA interactome, and RBDmap fractions.

### Statistical Analysis and Interpretative Bioinformatic Analyses

Approaches to test for significance and for detecting ontology, pathway and disease association/enrichment, as well as protein sequence and other features are detailed in the Supplemental Information.

### RNA-Binding Assays with eGFP-tagged RBPs

eGFP-RBP fusion plasmids were transfected into HeLa cells before irradiation with UV light at 254 nm. Cells or mitochondria were lysed, and RNA-fusion protein complexes were then captured on GFP-Trap beads and digested with RNase T1 before 5′ end labeling with $\gamma$-$^{32}$P-ATP by T4 PNK, anti-GFP western blot, and storage phosphor imaging.

### Data Availability

The R scripts and source code used for data analyses can be found at: http://fischerlab.dkfz.de/cardiomyocyteInteractome/, and https://github.com/PreissLab/cardiomyocyteInteractome.

## ACCESSION NUMBERS

The accession numbers for the mass spectrometry data reported in this paper are PRIDE: PXD002541 and PXD002543.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.06.084.

## AUTHOR CONTRIBUTIONS

Y.L. was responsible for cardiomyocyte culture and mRNA interactome capture experiments, analyzed and interpreted data, and wrote the manuscript. Y.L. and A.C. conducted the RBDmap experiments. A.C. further gave conceptual advice and assisted with data analysis and writing. E.P. and H.Y. performed additional experiments. S.L., S.F., and C.K.F. performed the mass spectrometry experiments and initial data analysis. B.F. performed the bulk of the interpretative data analysis with C.R. and S.K. making additional contributions. J.K. gave conceptual advice and oversaw the mass spectrometry experiments. M.W.H and T.P. conceived and coordinated the study, interpreted the findings, and wrote the manuscript.

## ACKNOWLEDGMENTS

We acknowledge support by the Imaging & Cytometry and Biomolecular Resource Facilities at JCSMR. We thank A. Filipovska, G. Duan, R. Horos, and I. Sillitoe for specialist advice. Y.L. holds a John Stocker Postdoctoral Fellowship (PF14-079) from the Science and Industry Endowment Fund. A.C. holds an MRC Career Development Award (#MR/L019434/1). C.K.F. is supported by a EMBO postdoctoral fellowship (LTF1006-2013). This work was supported by grants from the Gretel and Gordon Bootes Foundation to Y.L. and T.P., the National Health and Medical Research Council of Australia (#1045417) to T.P. and M.W.H, and the European Research Council (Advanced Grant ERC-2011-ADG_20110310) to M.W.H.

Received: October 13, 2015
Revised: April 6, 2016
Accepted: June 7, 2016
Published: July 21, 2016

## REFERENCES

Antonicka, H., and Shoubridge, E.A. (2015). Mitochondrial RNA granules are centers for posttranscriptional RNA processing and ribosome biogenesis. Cell Rep. *10*, 920–932.

Balsa, E., Marco, R., Perales-Clemente, E., Szklarczyk, R., Calvo, E., Landázuri, M.O., and Enríquez, J.A. (2012). NDUFA4 is a subunit of complex IV of the mammalian electron transport chain. Cell Metab. *16*, 378–386.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell *46*, 674–690.

Beckmann, B.M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A.M., Schwarzl, T., Curk, T., Foehr, S., Huber, W., et al. (2015). The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. Nat. Commun. *6*, 10127.

Bond, C.S., and Fox, A.H. (2009). Paraspeckles: nuclear bodies built on long noncoding RNA. J. Cell Biol. *186*, 637–644.

Brauch, K.M., Karst, M.L., Herron, K.J., de Andrade, M., Pellikka, P.A., Rodeheffer, R.J., Michels, V.V., and Olson, T.M. (2009). Mutations in ribonucleic acid binding protein gene cause familial dilated cardiomyopathy. J. Am. Coll. Cardiol. *54*, 930–941.

Busch, A., and Hertel, K.J. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. Wiley Interdiscip. Rev. RNA *3*, 1–12.

Caetano-Anollés, G., Kim, H.S., and Mittenthal, J.E. (2007). The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. Proc. Natl. Acad. Sci. USA *104*, 9358–9363.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393–1406.

Castello, A., Fischer, B., Hentze, M.W., and Preiss, T. (2013). RNA-binding proteins in Mendelian disease. Trends Genet. *29*, 318–327.

Castello, A., Hentze, M.W., and Preiss, T. (2015). Metabolic enzymes enjoying new partnerships as RNA-binding proteins. Trends Endocrinol. Metab. *26*, 746–757.

Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.-M., Föhr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive identification of RNA-binding domains in human cells. Mol Cell *63*. http://dx.doi.org/10.1016/j.molcel.2016.06.029.

Chen, C.Y., and Shyu, A.B. (2014). Emerging mechanisms of mRNP remodeling regulation. Wiley Interdiscip. Rev. RNA *5*, 713–722.

Cieśla, J. (2006). Metabolic enzymes that bind RNA: yet another level of cellular regulatory network? Acta Biochim. Pol. *53*, 11–32.

Claycomb, W.C., Lanson, N.A., Jr., Stallworth, B.S., Egeland, D.B., Delcarpio, J.B., Bahinski, A., and Izzo, N.J., Jr. (1998). HL-1 cells: a cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. Proc. Natl. Acad. Sci. USA *95*, 2979–2984.

Copley, S.D. (2012). Moonlighting is mainstream: paradigm adjustment required. BioEssays *34*, 578–588.

Dango, S., Mosammaparast, N., Sowa, M.E., Xiong, L.J., Wu, F., Park, K., Rubin, M., Gygi, S., Harper, J.W., and Shi, Y. (2011). DNA unwinding by ASCC3 helicase is coupled to ALKBH3-dependent DNA alkylation repair and cancer cell proliferation. Mol. Cell *44*, 373–384.

Daubner, G.M., Cléry, A., Jayne, S., Stevenin, J., and Allain, F.H. (2012). A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. EMBO J. *31*, 162–174.

Deutschmann, A.J., Amberger, A., Zavadil, C., Steinbeisser, H., Mayr, J.A., Feichtinger, R.G., Oerum, S., Yue, W.W., and Zschocke, J. (2014). Mutation or knock-down of 17β-hydroxysteroid dehydrogenase type 10 cause loss of MRPP1 and impaired processing of mitochondrial heavy strand transcripts. Hum. Mol. Genet. *23*, 3618–3628.

Eimre, M., Paju, K., Pelloux, S., Beraud, N., Roosimaa, M., Kadaja, L., Gruno, M., Peet, N., Orlova, E., Remmelkoor, R., et al. (2008). Distinct organization of energy metabolism in HL-1 cardiac cell line and cardiomyocytes. Biochim. Biophys. Acta *1777*, 514–524.

Fairman-Williams, M.E., Guenther, U.P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. Curr. Opin. Struct. Biol. *20*, 313–324.

Filipovska, A., and Rackham, O. (2013). Pentatricopeptide repeats: modular blocks for building RNA-binding proteins. RNA Biol. *10*, 1426–1432.

Fiorini, F., Boudvillain, M., and Le Hir, H. (2013). Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains. Nucleic Acids Res. *41*, 2404–2415.

Gorski, P.A., Ceholski, D.K., and Hajjar, R.J. (2015). Altered myocardial calcium cycling and energetics in heart failure–a rational approach for disease treatment. Cell Metab. *21*, 183–194.

Guo, W., Schafer, S., Greaser, M.L., Radke, M.H., Liss, M., Govindarajan, T., Maatz, H., Schulz, H., Li, S., Parrish, A.M., et al. (2012). RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. Nat. Med. *18*, 766–773.

Hall, M.P., Nagel, R.J., Fagg, W.S., Shiue, L., Cline, M.S., Perriman, R.J., Donohue, J.P., and Ares, M., Jr. (2013). Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. RNA *19*, 627–638.

Hentze, M.W. (1994). Enzymes as RNA-binding proteins: a role for (di)nucleotide-binding domains? Trends Biochem. Sci. *19*, 101–103.

Hentze, M.W., and Preiss, T. (2010). The REM phase of gene regulation. Trends Biochem. Sci. *35*, 423–426.

Jaffrey, S.R. (2014). An expanding universe of mRNA modifications. Nat. Struct. Mol. Biol. *21*, 945–946.

Jourdain, A.A., Koppen, M., Wydro, M., Rodley, C.D., Lightowlers, R.N., Chrzanowska-Lightowlers, Z.M., and Martinou, J.C. (2013). GRSF1 regulates RNA processing in mitochondrial RNA granules. Cell Metab. *17*, 399–410.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., et al. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753–767.

Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. Nat. Rev. Genet. *8*, 533–543.

Kim, H.J., Kim, N.C., Wang, Y.D., Scarborough, E.A., Moore, J., Diaz, Z., MacLea, K.S., Freibaum, B., Li, S., Molliex, A., et al. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature *495*, 467–473.

King, O.D., Gitler, A.D., and Shorter, J. (2012). The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain Res. *1462*, 61–80.

König, J., Zarnack, K., Luscombe, N.M., and Ule, J. (2012). Protein-RNA interactions: new genomic technologies and perspectives. Nat. Rev. Genet. *13*, 77–83.

Kramer, K., Sachsenberg, T., Beckmann, B.M., Qamar, S., Boon, K.L., Hentze, M.W., Kohlbacher, O., and Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nat. Methods *11*, 1064–1070.

Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1122–1130.

Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M. (2013). Human pentatricopeptide proteins: only a few and what do they do? RNA Biol. *10*, 1433–1438.

Loedige, I., Jakob, L., Treiber, T., Ray, D., Stotz, M., Treiber, N., Hennig, J., Cook, K.B., Morris, Q., Hughes, T.R., et al. (2015). The crystal structure of the NHL domain in complex with RNA reveals the molecular basis of Drosophila brain-tumor-mediated gene regulation. Cell Rep. *13*, 1206–1220.

Mahadevan, K., Zhang, H., Akef, A., Cui, X.A., Gueroussov, S., Cenik, C., Roth, F.P., and Palazzo, A.F. (2013). RanBP2/Nup358 potentiates the translation of a subset of mRNAs encoding secretory proteins. PLoS Biol. *11*, e1001545.

Matia-González, A.M., Laing, E.E., and Gerber, A.P. (2015). Conserved mRNA-binding proteomes in eukaryotic organisms. Nat. Struct. Mol. Biol. *22*, 1027–1033.

Mazurek, S., Hugo, F., Failing, K., and Eigenbrodt, E. (1996). Studies on associations of glycolytic and glutaminolytic enzymes in MCF-7 cells: role of P36. J. Cell. Physiol. *167*, 238–250.

Méthot, N., Song, M.S., and Sonenberg, N. (1996). A region rich in aspartic acid, arginine, tyrosine, and glycine (DRYG) mediates eukaryotic initiation factor 4B (eIF4B) self-association and interaction with eIF3. Mol. Cell. Biol. *16*, 5328–5334.

Mitchell, S.F., Jain, S., She, M., and Parker, R. (2013). Global analysis of yeast mRNPs. Nat. Struct. Mol. Biol. *20*, 127–133.

Mondal, T., Rasmussen, M., Pandey, G.K., Isaksson, A., and Kanduri, C. (2010). Characterization of the RNA content of chromatin. Genome Res. *20*, 899–907.

Muckenthaler, M.U., Galy, B., and Hentze, M.W. (2008). Systemic iron homeostasis and the iron-responsive element/iron-regulatory protein (IRE/IRP) regulatory network. Annu. Rev. Nutr. *28*, 197–213.

Ostareck, D.H., Ostareck-Lederer, A., Wilm, M., Thiele, B.J., Mann, M., and Hentze, M.W. (1997). mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3′ end. Cell *89*, 597–606.

Perrucci, G.L., Gowran, A., Zanobini, M., Capogrossi, M.C., Pompilio, G., and Nigro, P. (2015). Peptidyl-prolyl isomerases: a full cast of critical actors in cardiovascular diseases. Cardiovasc. Res. *106*, 353–364.

Rackham, O., Mercer, T.R., and Filipovska, A. (2012). The human mitochondrial transcriptome and the RNA-binding proteins that regulate its expression. Wiley Interdiscip. Rev. RNA *3*, 675–695.

Rosca, M.G., Tandler, B., and Hoppel, C.L. (2013). Mitochondria in cardiac hypertrophy and heart failure. J. Mol. Cell. Cardiol. *55*, 31–41.

Rossmann, M.G., Moras, D., and Olsen, K.W. (1974). Chemical and biological evolution of nucleotide-binding protein. Nature *250*, 194–199.

Schiene-Fischer, C. (2015). Multidomain Peptidyl Prolyl cis/trans Isomerases. Biochim. Biophys. Acta *1850*, 2005–2016.

Schiff, M., Froissart, R., Olsen, R.K., Acquaviva, C., and Vianey-Saban, C. (2006). Electron transfer flavoprotein deficiency: functional and molecular aspects. Mol. Genet. Metab. *88*, 153–158.

Shen, H., and Green, M.R. (2006). RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. Genes Dev. *20*, 1755–1765.

Sibbritt, T., Patel, H.R., and Preiss, T. (2013). Mapping and significance of the mRNA methylome. Wiley Interdiscip. Rev. RNA *4*, 397–422.

Sillitoe, I., Lewis, T.E., Cuff, A., Das, S., Ashford, P., Dawson, N.L., Furnham, N., Laskowski, R.A., Lee, D., Lees, J.G., et al. (2015). CATH: comprehensive structural and functional annotations for genome sequences. Nucleic Acids Res. *43*, D376–D381.

Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The clothes make the mRNA: past and present trends in mRNP fashion. Annu. Rev. Biochem. *84*, 325–354.

Vélot, C., Mixon, M.B., Teige, M., and Srere, P.A. (1997). Model of a quinary structure between Krebs TCA cycle enzymes: a model for the metabolon. Biochemistry *36*, 14271–14276.

Walden, W.E., Selezneva, A.I., Dupuy, J., Volbeda, A., Fontecilla-Camps, J.C., Theil, E.C., and Volz, K. (2006). Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. Science *314*, 1903–1908.

Wu, F., and Minteer, S. (2015). Krebs cycle metabolon: structural evidence of substrate channeling revealed by cross-linking and mass spectrometry. Angew. Chem. Int. Ed. Engl. *54*, 1851–1854.

Zheng, L., Jiang, H., Mei, Q., and Chen, B. (2015). Identification of two novel Darier disease-associated mutations in the ATP2A2 gene. Mol. Med. Rep. *12*, 1845–1849.

**Cell Reports, Volume *16***

## Supplemental Information

# The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease

**Yalin Liao, Alfredo Castello, Bernd Fischer, Stefan Leicht, Sophia Föehr, Christian K. Frese, Chikako Ragan, Sebastian Kurscheid, Eloisa Pagler, Hao Yang, Jeroen Krijgsveld, Matthias W. Hentze, and Thomas Preiss**

# SUPPLEMENTAL FIGURES

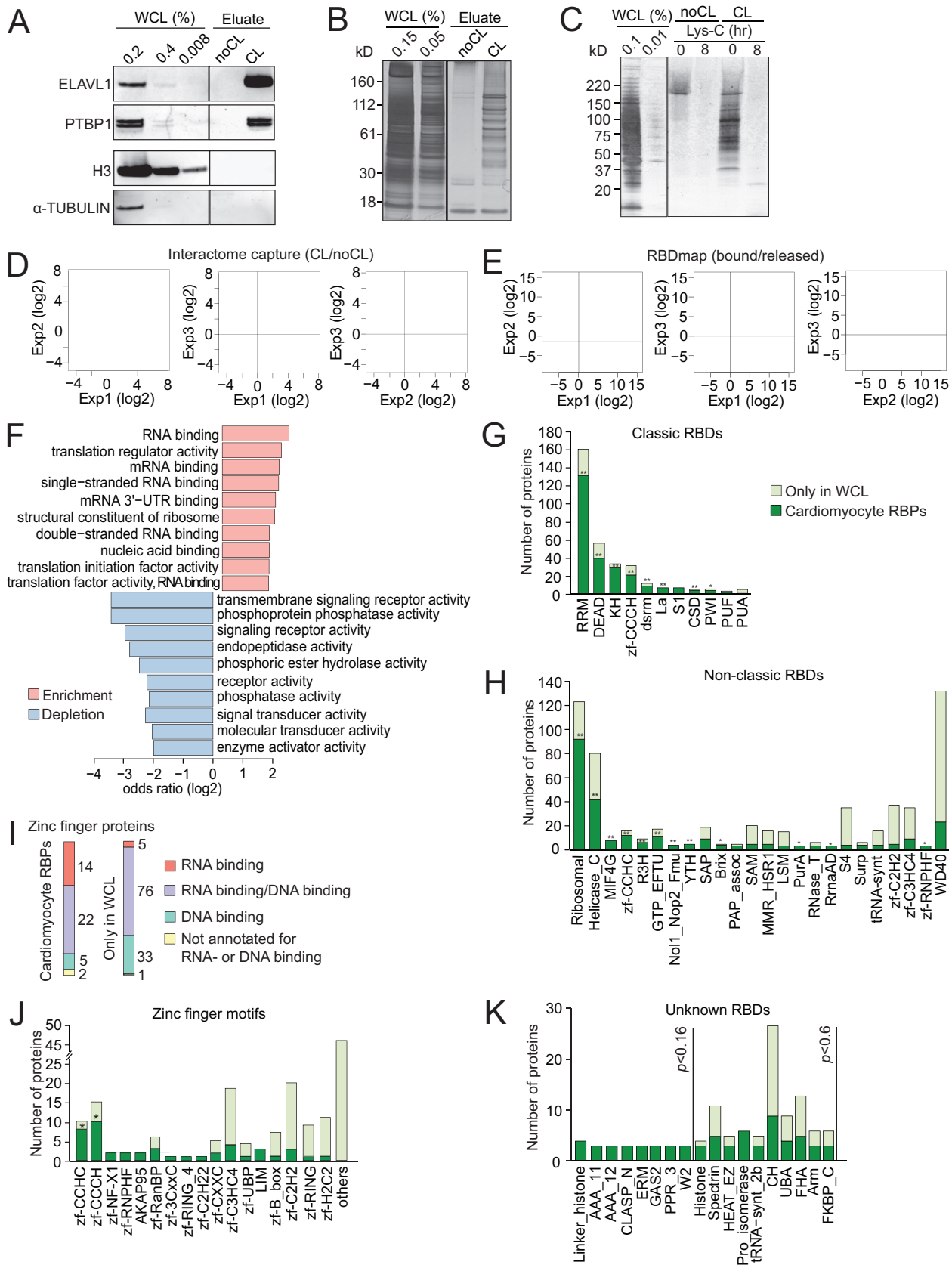

A
WCL (%)
Eluate
0.2
0.4
0.008
noCL
CL
ELAVL1
PTBP1
H3
α-TUBULIN
B
WCL (%)
Eluate
kD
0.15
0.05
noCL
CL
160
112
61
30
18
C
WCL (%)
noCL
CL
Lys-C (hr)
kD
0.1
0.01
0
8
0
8
220
150
100
75
50
37
20
D
Interactome capture (CL/noCL)
Exp2 (log2)
Exp3 (log2)
Exp1 (log2)
Exp2 (log2)
E
RBDmap (bound/released)
F
RNA binding
translation regulator activity
mRNA binding
single−stranded RNA binding
mRNA 3'−UTR binding
structural constituent of ribosome
double−stranded RNA binding
nucleic acid binding
translation initiation factor activity
translation factor activity, RNA binding
transmembrane signaling receptor activity
phosphoprotein phosphatase activity
signaling receptor activity
endopeptidase activity
phosphoric ester hydrolase activity
receptor activity
phosphatase activity
signal transducer activity
molecular transducer activity
enzyme activator activity
Enrichment
Depletion
odds ratio (log2)
G
Classic RBDs
Number of proteins
Only in WCL
Cardiomyocyte RBPs
RRM
DEAD
KH
zf-CCCH
dsrm
La
S1
CSD
PWI
PUF
PUA
H
Non-classic RBDs
Number of proteins
Ribosomal
Helicase_C
MIF4G
zf-CCHC
R3H
GTP_EFTU
Nol1_Nop2_Fmu
YTH
SAP
Brix
PAP_assoc
SAM
MMR_HSR1
LSM
PurA
RNase_T
RrnaAD
S4
Surp
tRNA-synt
zf-C2H2
zf-C3HC4
zf-RNPHF
WD40
I
Zinc finger proteins
Cardiomyocyte RBPs
Only in WCL
14
22
5
2
5
76
33
1
RNA binding
RNA binding/DNA binding
DNA binding
Not annotated for
RNA- or DNA binding
J
Zinc finger motifs
Number of proteins
zf-CCHC
zf-CCCH
zf-NF-X1
zf-RNPHF
AKAP95
zf-RanBP
zf-3CxxC
zf-RING_4
zf-C2H2_2
zf-CXXC
zf-C3HC4
zf-UBP
LIM
zf-B_box
zf-C2H2
zf-RING
zf-H2C2
others
K
Unknown RBDs
Number of proteins
p<0.16
p<0.6
Linker_histone
AAA_11
AAA_12
CLASP_N
ERM
GAS2
PPR_3
W2
Histone
Spectrin
HEAT_EZ
Pro_isomerase
tRNA-synt_2b
CH
UBA
FHA
Arm
FKBP_C

Figure S1

**Figure S1. Efficacy and Reproducibility of mRNA Interactome Capture and RBDmap; GO and RBD Enrichment Analyses of Cardiomyocyte RBPs. Related to Figure 1.**

(A-B) In the RBDmap approach, RNA-protein complexes were digested with RNases after a first round of oligo(dT) capture, resolved by SDS-PAGE alongside aliquot of input WCL (% equivalent to loaded eluate amount as indicated) and analyzed by western blot (A) (see Supplemental Experimental Procedures for antibody details) or silver stain (B). n=3.

(C) RNA-protein complexes from an RBDmap experiment as in (A) were digested with Lys-C protease, and analyzed by silver stain as in (B).

(D) Protein enrichment in CL over noCL eluates, based on differential LC-MS/MS intensities from three biological replicate mRNA interactome capture experiments, were plotted against each other. Each dot represents one protein group. Protein groups enriched in CL or noCL at 1% FDR are depicted by red or black dots, respectively. Grey dots are protein groups that lack enrichment.

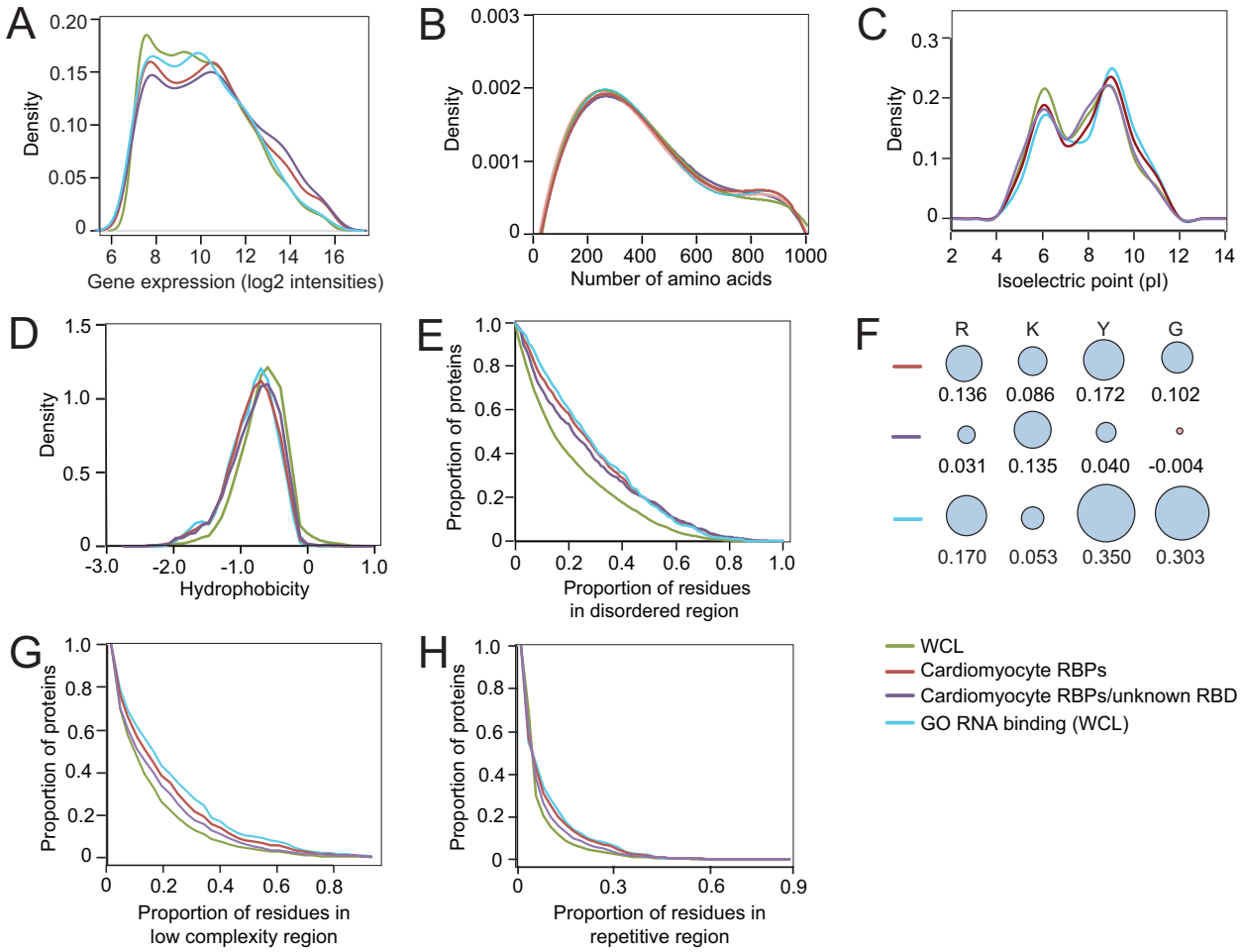
(E) Peptide enrichment in RNA bound fraction over the released fraction, based on differential LC-MS/MS peptide intensity ratio from three biological replicate RBDmap experiments, were plotted against each other. Each dot represents one peptide. Npeps are depicted by red dots. Grey dots indicate peptides that lack enrichment.

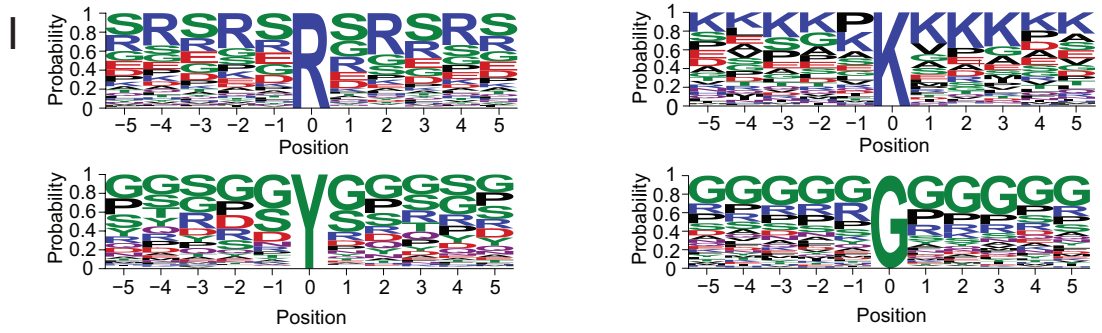
(F) The ten most significantly over- (pink) and under-represented (blue) GOMF terms in cardiomyocyte RBPs, compared to the WCL proteome.

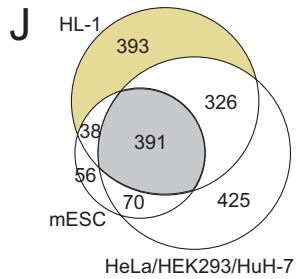
(G) Number of proteins with classic RBDs, in cardiomyocyte RBPs (dark green) or only identified in WCL (light green).

(H) Number of proteins with non-classic RBDs, in cardiomyocyte RBPs (dark green) or only identified in WCL (light green). Only domains represented by three or more hits are shown. Proteins with both classic and non-classic RBDs are considered only in (G).

(I) Distribution of zinc finger-containing proteins with GO ‘RNA binding’, ‘DNA binding’ and their children terms, within cardiomyocyte RBPs (left) or only identified in WCL (right).
(J) Number of proteins with different zinc finger motif subtypes in cardiomyocyte RBPs (dark green) or only identified in WCL (light green).
(K) Number of proteins previously unrelated to RNA biology in cardiomyocyte RBPs (dark green) or only identified in WCL (light green). Only domains represented by three or more hits are shown. Proteins with either classic or non-classic RBDs, are considered only in (G) and (H).


A
Density
Gene expression (log2 intensities)
B
Density
Number of amino acids
C
Density
Isoelectric point (pI)
D
Density
Hydrophobicity
E
Proportion of proteins
Proportion of residues in disordered region
F
R
K
Y
G
0.136
0.086
0.172
0.102
0.031
0.135
0.040
-0.004
0.170
0.053
0.350
0.303
G
Proportion of proteins
Proportion of residues in low complexity region
H
Proportion of proteins
Proportion of residues in repetitive region
WCL
Cardiomyocyte RBPs
Cardiomyocyte RBPs/unknown RBD
GO RNA binding (WCL)


I
Probability
Position


J
HL-1
393
326
38
391
56
70
425
mESC
HeLa/HEK293/HuH-7

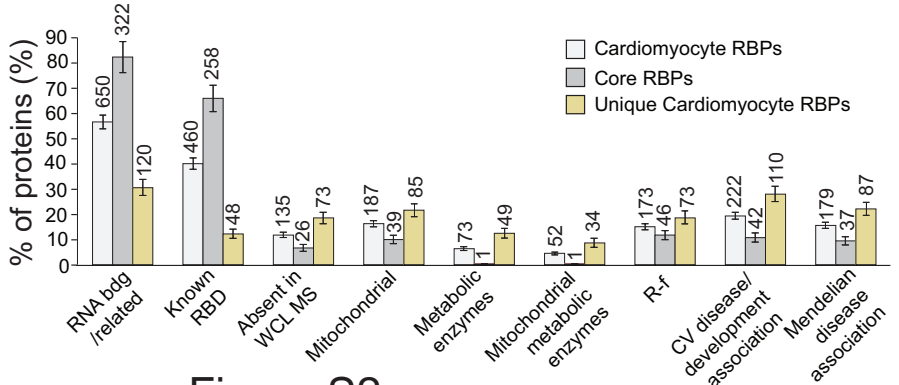

K
% of proteins (%)
Cardiomyocyte RBPs
Core RBPs
Unique Cardiomyocyte RBPs
650
322
120
460
258
48
135
26
73
187
39
85
73
1
49
52
1
34
173
46
73
222
42
110
179
37
87
RNA bdg /related
Known RBD
Absent in WCL MS
Mitochondrial
Metabolic enzymes
Mitochondrial metabolic enzymes
R-f
CV disease/ development association
Mendelian disease association

Figure S2

**Figure S2. Biophysical and Sequence Features of Cardiomyocyte RBPs. Related to Figure 1.**

The following proteins sets were used in panels A-H: WCL proteins (green; used as background in panel F), cardiomyocyte RBPs (red), cardiomyocyte RBPs with unknown RBDs (purple). WCL RBPs with GO 'RNA binding' and its children terms are shown for a reference (blue). For information on bioinformatic methods see Supplemental Experimental Procedures and (Castello et al., 2012; Castello et al., 2013b).

(A) Density of mRNA expression levels in normal HL-1 cells (GEO: GSE56584).

(B) Density of calculated protein length.

(C) Density of calculated isoelectric point (pI).

(D) Density of hydrophobicity.

(E) Proportion of amino acid residues in disordered regions.

(F) Amino acids composition in disordered region of all cardiomyocyte RBPs, cardiomyocyte RBPs with unknown RBDs, or all WCL proteins with GO 'RNA binding' and its children terms. Blue circles indicate enrichment and pink circles indicate depletion relative to the composition of disordered regions in all WCL proteins. Circle area is proportional to the log2 enrichment or depletion ratio.

(G) Distribution of calculated low complexity regions.

(H) Distribution of repetitive dipeptide sequence (dimers of subsequent amino acids, with 0, 1, or 2 amino acids spacing).

The significance of differences between RBP subsets in panels A, B, C, D, E, G, H was tested by two-sample Kolmogorov-Smirnov tests. This showed that distribution of protein size (Number of amino acids) does not differ between all four groups. Compared to WCL, cardiomyocyte RBPs and cardiomyocyte RBPs with unknown RBDs are significantly different in mRNA expression

levels, all three subsets are significantly different in, isoelectric point, hydrophobicity, disordered region, low-complexity region, and repetitive region ($p < 0.01$).

The significance of enrichment/depletion in panel F was tested by two-sample test for population proportions. Comparing all other groups to the WCL proteome, results were significant ($p < 0.01$) for all amino acids, except for glycine (G) in the cardiomyocyte RBP/unknown RBD group.

(I) 11-mers around the 4 repetitive amino acid residues featured in (F) were extracted and aligned by their central residue to show sequence bias. Height of letter indicates the probability of occurrence of a given amino acid at each position.

(J) Venn diagram comparing cardiomyocyte RBPs with four published mRNA interactome datasets: HeLa (Castello et al., 2012), HEK293 (Baltz et al., 2012), mESC (Kwon et al., 2013), and HuH-7 (Beckmann et al., 2015).

(K) Distribution of a range of features among all cardiomyocyte RBPs, core RBPs and unique cardiomyocyte RBPs. Percentage uncertainties were derived using Gaussian error propagation assuming Poisson uncertainties of protein counts. Cardiomyocyte RBPs, 1148; Core RBPs, 391; Unique Cardiomyocyte RBPs, 393.

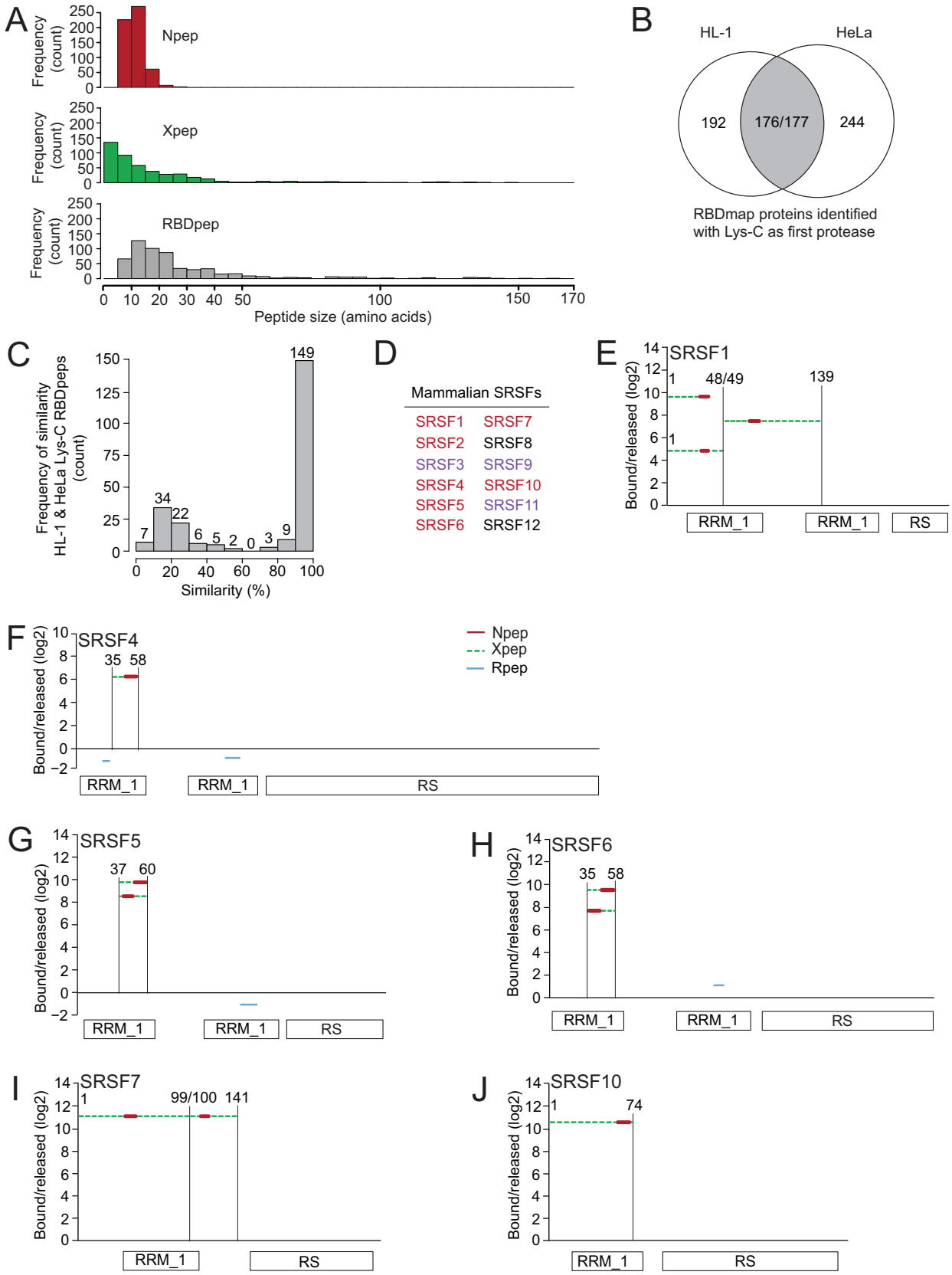

A
Npep
Xpep
RBDpep
Frequency (count)
Peptide size (amino acids)
B
HL-1
HeLa
192
176/177
244
RBDmap proteins identified with Lys-C as first protease
C
Frequency of similarity HL-1 & HeLa Lys-C RBDpeps (count)
Similarity (%)
D
Mammalian SRSFs
SRSF1
SRSF2
SRSF3
SRSF4
SRSF5
SRSF6
SRSF7
SRSF8
SRSF9
SRSF10
SRSF11
SRSF12
E
SRSF1
F
SRSF4
Npep
Xpep
Rpep
G
SRSF5
H
SRSF6
I
SRSF7
J
SRSF10
Bound/released (log2)
RRM_1
RS

Figure S3

**Figure S3. Benchmarking HL-1 RBDmap data. Related to Figure 2.**

(A) Distribution of Npeps, Xpeps (≥ 1 amino acid) and RBDpeps identified in HL-1 RBDmap.

(B) Venn diagram comparing homologous proteins in HL-1 & HeLa RBDmap data (Castello et al., 2016). Both ENSG00000092199 and ENSG00000179172 are human homologs of mouse ENSMUSG00000060373.

(C) Distribution of 237 HL-1 RBDpeps identified from 176 proteins that are also in HeLa RBDmap (Castello et al., 2016), based on sequence similarity (See Supplemental Experimental Procedures).

(D) List of SRSF splicing factor family members found in the cardiomyocyte RBPs (purple) and also in RBDmap (red).

(E-J) RBDmap data for six SRSF family members, SRSF1 (E), SRSF4 (F), SRSF5 (G), SRSF6 (H), SRSF7 (I) and SRSF10 (J), shown as in Figure 2E. See Supplemental Experimental Procedure for RS domain annotation.

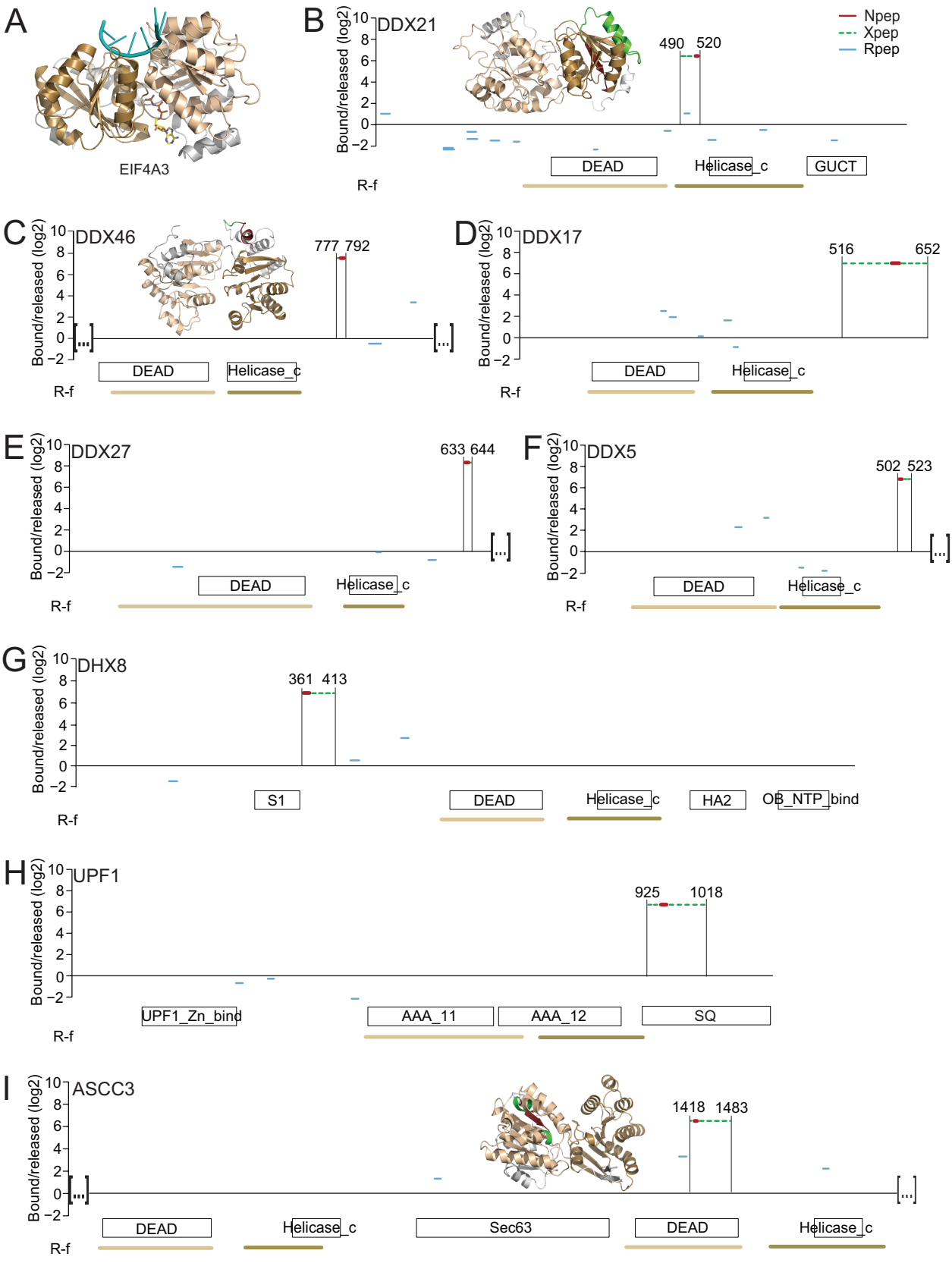

A
EIF4A3
B
DDX21
490 520
Npep
Xpep
Rpep
Bound/released (log2)
DEAD
Helicase_c
GUCT
R-f
C
DDX46
777 792
DEAD
Helicase_c
D
DDX17
516
652
E
DDX27
633 644
F
DDX5
502 523
G
DHX8
361 413
S1
HA2
OB_NTP_bind
H
UPF1
925 1018
UPF1_Zn_bind
AAA_11
AAA_12
SQ
I
ASCC3
1418 1483
Sec63

Figure S4

**Figure S4. RNA-Binding Modalities in Seven RNA Helicases and a DNA Helicase. Related to Figure 4.**

(A) Crystal structure of EIF4A3 (EC:3.6.4.13), shown as a ribbon diagram, in complex with 5'-UUUUUU-3' RNA (teal) and adenosine-5'-triphosphate (ATP, multicolor) (PDB: 2HYI, chain C). N-terminal and C- terminal Rossmann folds (R-f) are depicted in wheat and sand, respectively.

(B-H) RBDmap data for seven RNA helicases, DDX21 (B), DDX46 (C), DDX17 (D), DDX27 (E), DDX5 (F), DHX8 (G) and UPF1 (H) (EC:3.6.4.13), as shown in Figure 2E. RBDpeps were mapped to Phyre2-modeled structures of DDX21, DDX46, DDX5 (not shown) and DDX27 (not shown). Rossmann folds are depicted in wheat or sand.

(I) RBDmap data for ASCC3 (EC:3.6.4.12). RBDpep was mapped to Phyre2-modeled structure.

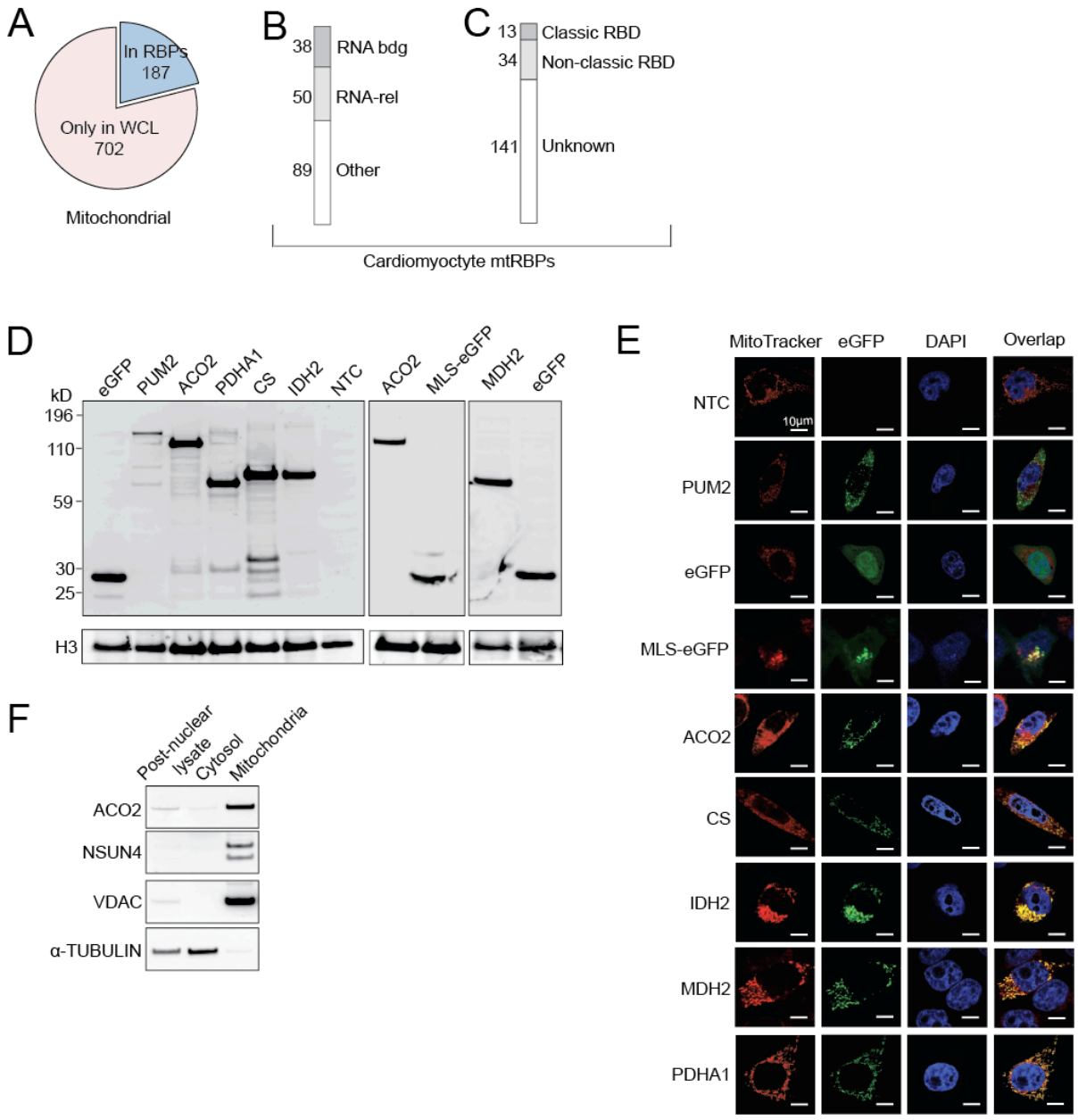

A
In RBPs
187
Only in WCL
702
Mitochondrial
B
38
RNA bdg
50
RNA-rel
89
Other
C
13
Classic RBD
34
Non-classic RBD
141
Unknown
Cardiomyoctyte mtRBPs
D
eGFP
PUM2
ACO2
PDHA1
CS
IDH2
NTC
ACO2
MLS-eGFP
MDH2
eGFP
kD
196
110
59
30
25
H3
E
MitoTracker
eGFP
DAPI
Overlap
NTC
10μm
PUM2
eGFP
MLS-eGFP
ACO2
CS
IDH2
MDH2
PDHA1
F
Post-nuclear lysate
Cytosol
Mitochondria
ACO2
NSUN4
VDAC
α-TUBULIN

Figure S5

**Figure S5. Characterization of Cardiomyocyte mtRBPs. Related to Figure 6.**

(A) Number of WCL mitochondrial proteins present/absent in cardiomyocyte RBPs.

(B) Proportion of cardiomyocyte mtRBPs with GO 'RNA binding' or RNA-related annotation, shown as in Figure 1E.

(C) Proportion of cardiomyocyte mtRBPs harboring classic, non-classic, or unknown RBDs, shown as in Figure 1F.

(D) Analysis of eGFP fusion protein expression by western blot. Probing was done with anti-GFP antibody (top) or anti-histone H3 as a loading control (bottom). Dose of plasmid DNA based on cell culture area is the same as used in RNA $^{32}$P 5' end-labeling experiment. NTC, non-transfected control.

(E) Immunofluorescence analysis of mitochondrial localization for C-terminal tagged eGFP fusion proteins (green) in HeLa cells. Red, MitoTracker; blue, DAPI. Scale bar: 10 μm.

(F) Quality control of HeLa cell mitochondria isolation by western blot, using antibodies against mitochondrial matrix markers (ACO2, NSUN4), mitochondrial outer membrane marker (VDAC) and cytosolic marker (α-TUBULIN).

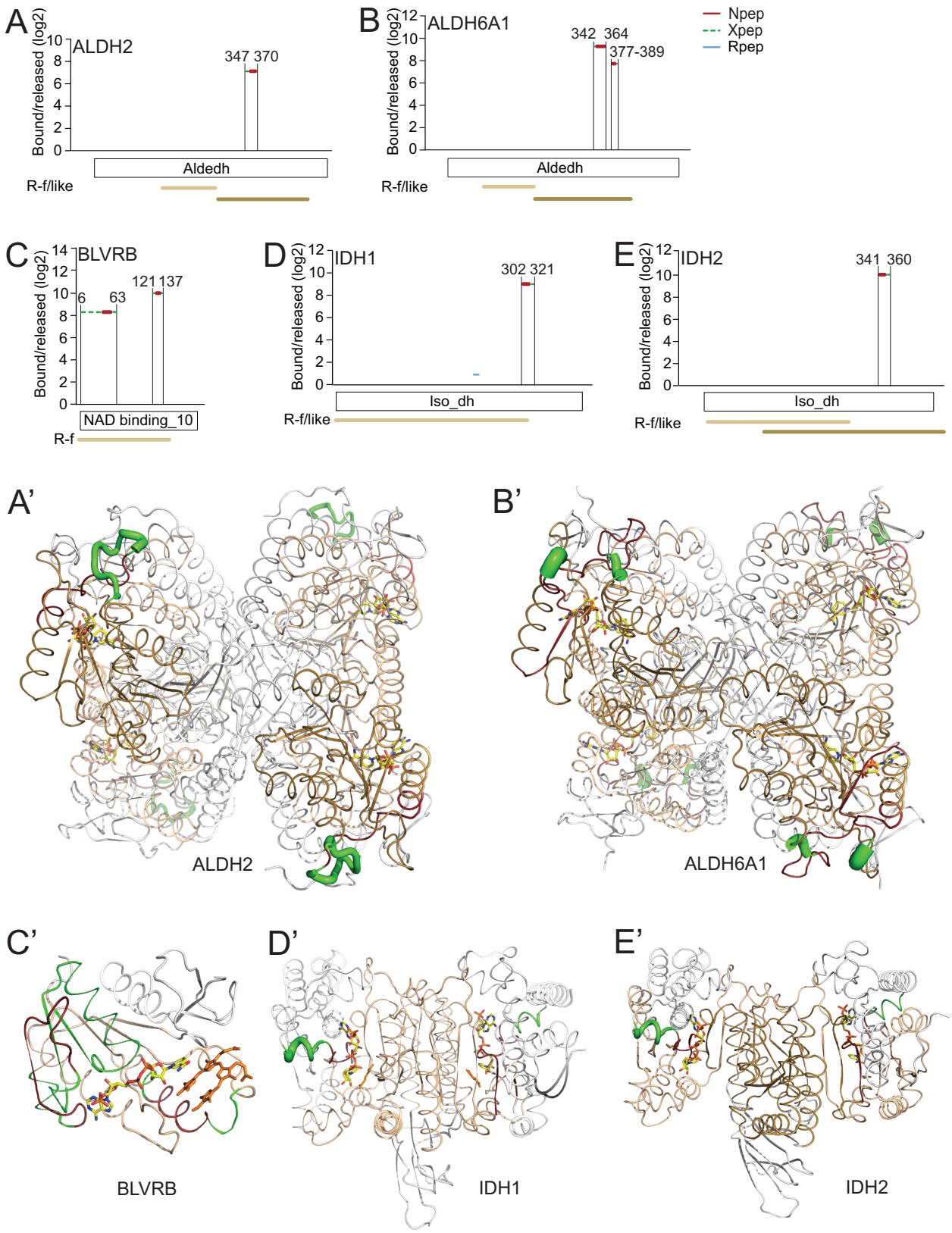

A
ALDH2
Bound/released (log2)
347 370
Aldedh
R-f/like
B
ALDH6A1
342 364
377-389
Aldedh
R-f/like
Npep
Xpep
Rpep
C
BLVRB
6
63
121 137
NAD binding_10
R-f
D
IDH1
302 321
Iso_dh
R-f/like
E
IDH2
341 360
Iso_dh
R-f/like
A'
ALDH2
B'
ALDH6A1
C'
BLVRB
D'
IDH1
E'
IDH2

Figure S6

**Figure S6. RNA-Binding Modalities Among Five Rossmann Fold or Rossmann Fold-Like Enzymes. Related to Figure 4, Figure 7.**

(A-E) RBDmap data for ALDH2 (EC:1.2.1.3), ALDH6A1 (EC:1.2.1.18, EC:1.2.1.27), BLVRB (EC:1.5.1.30), IDH1 (EC:1.1.1.42) and IDH2 (EC:1.1.1.41), shown as in Figure 2E.

(A'-E') RBDpeps were mapped to enzyme crystal structures.

(A') Crystal structure of ALDH2 tetramer, in complex with $NAD^+$ (multicolor) (PDB: 1O01).

(B') Crystal structure of ALDH6A1 tetramer, in complex with $NAD^+$ (multicolor) (PDB: 1T90).

(C') Crystal structure of BLVRB monomer, in complex with biliverdine IX alpha (BLA, orange) and $NADP^+$ Nicotinamide-adenine-dinucleotide phosphate (multicolor) (PDB: 1HE2).

(D') Crystal structure of IDH1 dimer, in complex with 2-oxoglutaric acid (orange) and NADPH (multicolor) (PDB: 3INM).

(E') Crystal structure of IDH2 dimer, in complex with NADPH (multicolor) (PDB: 4JA8).

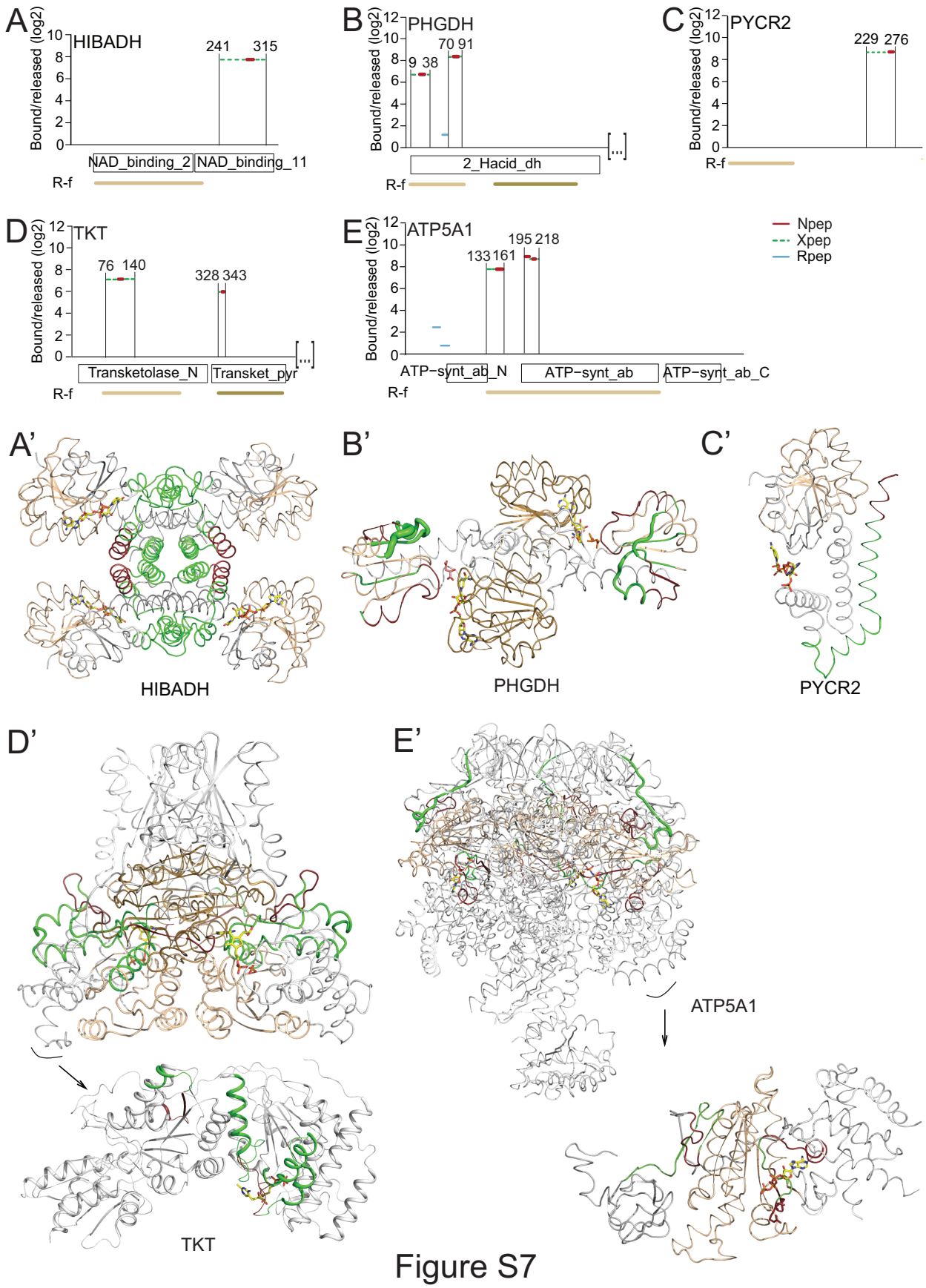

A
HIBADH
241
315
Bound/released (log2)
NAD_binding_2
NAD_binding_11
R-f
B
PHGDH
9
38
70
91
2_Hacid_dh
C
PYCR2
229
276
D
TKT
76
140
328
343
Transketolase_N
Transket_pyr
E
ATP5A1
133
161
195
218
ATP−synt_ab_N
ATP−synt_ab
ATP−synt_ab_C
Npep
Xpep
Rpep
A'
HIBADH
B'
PHGDH
C'
PYCR2
D'
TKT
E'
ATP5A1

Figure S7

**Figure S7. RNA-Bind Modalities Among Five Rossmann Fold or Rossmann Fold Like Enzymes. Related to Figure 4, Figure 7.**

(A-E) RBDmap data for HIBADH (EC:1.1.1.31), PHGDH (EC:1.1.1.95) PYCR2 (EC:1.5.1.2), TKT (EC:2.2.1.1) and ATP5A1, shown as in Figure 2E.

(A'-E') RBDpeps were mapped to enzyme crystal structures.

(A') Crystal structure of HIBADH tetramer, in complex with $NAD^+$ (multicolor) (PDB: 2I9P).

(B') Crystal structure of PHGDH dimer, in complex with D-malate (orange) and $NAD^+$ (multicolor) (PDB: 2G76).

(C') Crystal structure of PYCR2 monomer, in complex with $NADP^+$ (multicolor) (PDB: 2GRA).

(D') Crystal structure of TKT dimer, in complex with D-xylitol-5-phosphate (orange) and thiamin diphosphate (multicolor) (PDB: 4KXV). Arrow indicates enlarged view of monomer crystal structure.

(E') Crystal structure of ATP synthase complex (α3β3γ1δ1ε1) in complex with ATP (multicolor) (PDB: 4YXW). Arrow indicates enlarged view of α subunit ATP5A1.

# SUPPLEMENTAL TABLE LEGENDS

**Table S1. Mass Spectrometry data of cardiomyocyte WCL proteome and mRNA interactome. Related to Figure 1.**

Results are shown as protein groups, their majority Uniprot IDs, and the Ensembl gene IDs. For mRNA interactome capture, the log2 enrichment (CL/noCL) of each protein group in biological replicate samples and statistical significance is also shown. *, detected in whole cell lysate (WCL); #, present in at least two biological replicates.

**Table S2. Peptides identified by cardiomyocyte RBDmap. Related to Figures 2, 3, 4, 6, 7.**

A total of 3342 tryptic peptides were identified. 771 peptides do not have quantitative RNA bound/released enrichment value and therefore were removed. 453 of the remaining 2571 peptides either do not map to unique genes, or do not represent the majority protein coded by a unique gene, or no gene identifier is available; therefore these were also removed (Table S2. "Removed peptides").

The remaining 2118 peptides were classified into 4 groups:

568 peptides were assigned as Npeps (Figure 1A), covering 368 unique genes. They are enriched in the RNA bound compared to the released fraction (log2 enrichment greater than 0) with an FDR of less than 1%. For each gene, the protein with the highest coverage was selected (Table S2. "Npep").

124 peptides were assigned as Candidate Npeps, they are enriched in the RNA bound compared to the released fraction (log2 enrichment greater than 0) with an FDR between 1% and 10% (Table S2. "Candidate Npep").

1287 peptides were assigned as Rpeps (Figure 1), they are peptides with an FDR greater than 15% regardless of enrichment value (Table S2. “Rpep”).

139 peptides remained unassigned as they did not satisfy the selection criteria stated above (Table S2. “Unassigned peptides”).

**Table S3. A collection of cardiomyocyte RBPs features. Related to Figures 2, 3, 4, 6, 7.**

Cardiomyocyte RBPs are listed by Ensembl gene ID and gene name, and the presence of the following features are indicated by a ‘+’ for each entry: identified by mRNA interactome capture; identified by RBDmap; mitochondrial localization; metabolic enzyme; Mendelian disease association; Mendelian RBDpep; categorized as a PPIase; associated with cardiovascular disease & development; unique cardiomyocyte RBP; core RBP. The following features are also listed: RNA-related/unrelated annotation; category of RBD; name of known RBD; top ten enriched and depleted GOMF terms listed against the respective RBP; EC number for metabolic enzyme; type of Rossmann fold homologous superfamily; RNA helicase family; type of RNA modification for RNA modification enzymes. * see Supplemental Experimental Procedure for detail; # RBDpep covers Mendelian disease missense mutation/amino acid deletion; § compared to the following mRNA interactome datasets: HeLa (Castello et al., 2012), HEK293 (Baltz et al., 2012), mESC (Kwon et al., 2013), and HuH-7 (Beckmann et al., 2015).

**Table S4. Spectrum of OMIM diseases associated with cardiomyocyte RBPs. Related to Figure 3.**

Cardiomyocyte OMIM-RBPs are listed by Ensembl gene ID and gene name. For each RBP entry, the associated Mendelian disease is shown by phenotype MIM number, name of disease and type

of disease. For RBPs where the RBDpep covers disease mutation(s), the missense mutation(s) and/or amino acid deletion(s) are also indicated.

**Table S5. Characteristics of metabolic enzymes among cardiomyocyte RBPs. Related to Figures 4, 7.**

Metabolic enzymes among the cardiomyocyte RBPs are listed by Ensembl gene ID and gene name. The presence of the following features are indicated by a '+' for each entry: Rossmann fold; Rossmann-like fold; mitochondrial localization; presence in the mitochondrial RNA processing granule. The following features are also listed: EC number; EC class; type of Rossmann fold homologous superfamily; metabolic pathway; non-substrate ligand. *, see Supplemental Experimental Procedures for details; #, non-substrate ligand annotation was obtained from Uniprot; § as determined in (Antonicka and Shoubridge, 2015).

# SUPPLEMENTAL EXPERIMENTAL PROCEDURES

## Cell Culture

HL-1 cardiomyocytes (a gift from Dr. William Claycomb, Louisiana State University) were maintained as described (Claycomb et al., 1998; Humphreys et al., 2012). HeLa cells were purchased from ATCC and maintained as described (Clancy et al., 2011). All cells were grown in an atmosphere of 5% $CO_2$ at 37 °C and the cell culture media were changed every other day.

## mRNA Interactome Capture

The procedure was performed as previously described (Castello et al., 2012; Castello et al., 2013b) with some modifications to accommodate the properties of cardiomyocyte cultures. Eighteen to forty-eight 145 $cm^2$ culture dishes were seeded with HL-1 cells at $6\times10^6$ cells per dish, and grown for 4 days until confluent to obtain spontaneously beating cells. For in-cell crosslinking (CL), culture dishes were washed twice with ice-cold 1×PBS, placed on ice and irradiated in a Stratalinker (Stratagene) with 150 mJ/$cm^2$ of UV light at 254 nm. Non-crosslinked cells (noCL) were processed in parallel as a specificity control. Unless otherwise stated, the following gives reagent amounts per 145 $cm^2$ culture dish. Cells (CL or noCL) were immediately collected in ice-cold 1×PBS, pelleted and lysed in 10 ml lysis buffer (20 mM Tris•HCl pH 7.5, 500 mM LiCl, 0.5% LiDS, 1 mM EDTA, and 5 mM DTT), then incubated on ice for 10 min. The lysate was then passed three times through a narrow gauge needle (0.4 mm diameter), before addition of 0.67 ng polyadenylated spike-in RNA, which was *in vitro* transcribed from plasmid pT3lucpA (Iizuka et al., 1994; Preiss and Hentze, 1998). RNA-protein complexes were captured on oligo$(dT)_{25}$ magnetic bead (beads from 150 µl original bead suspension, New England Biolabs), by incubating for 1 hr at 4 °C on a rotating wheel. Subsequently, beads were collected

on a magnet and washed twice, with 3 ml of each the following buffers (I: 20 mM Tris•HCl pH 7.5, 500 mM LiCl, 0.1% LiDS, 1 mM EDTA, 5 mM DTT; II: 20 mM Tris•HCl pH 7.5, 500 mM LiCl, 1 mM EDTA, 5 mM DTT; III: 20 mM Tris•HCl pH 7.5, 200 mM LiCl, 1 mM EDTA, 5 mM DTT), for 5 min at 4 °C on a rotating wheel. Finally, RNA-protein complexes were eluted by incubation of the beads with 67 µl 20 mM Tris•HCl pH 7.5, 1 mM EDTA at 50 °C for 3 min. After elution, the oligo$(dT)_{25}$ beads were reactivated in lysis buffer according to the manufacturer recommendations, and two additional bead binding/elution rounds were carried out for each sample.

Finally, all three eluates (from the three capture rounds) were combined and an aliquot was taken for qRT-PCR (equivalent to 50 $cm^2$ culture area), this aliquot was supplemented with ¼ volume of 5×proteinase K buffer (50 mM Tris•HCl pH 7.5, 750 mM NaCl, 1% SDS, 50 mM EDTA, 2.5 mM DTT, 25 mM $CaCl_2$), and treated with proteinase K (Thermo Scientific) at 42 °C for 1 hr, at final concentration of 1 µg/µl, prior to RNA extraction and cDNA synthesis. The remaining eluate was supplemented with ¼ volume of 5×RNase buffer (50 mM Tris•HCl pH 7.5, 750 mM NaCl, 0.25% NP-40, 2.5 mM DTT) and treated with 0.61 U of RNase T1 (Sigma) and 0.2 µg RNase A (Sigma) for 1 hr at 37 °C, further aliquots were taken for protein analysis (equivalent to 100 $cm^2$ or 200 $cm^2$ culture area for western blot and silver stain, respectively). The bulk of the eluate was then processed for mass spectrometry.

**RBDmap**

Initial purification of RBPs was carried out following the RNA interactome capture protocol described above using ten culture dishes per sample. A noCL control was carried out for RNA and protein recovery control purposes. Eluted RNA-protein complexes (1.6 ml/sample) were supplemented with 1 µl (40 U) of RNaseOUT (Life Technologies) and then treated with 1 µg of

endoproteinase Lys-C (Sigma) at 37 °C for 8 hr. An aliquot (75 µl) from each sample was taken prior to and after Lys-C digestion to monitor protein and RNA integrity. The rest of the Lys-C digested sample (1.45 ml) was then diluted with 2.4 ml $H_2O$ and 1 ml 5×dilution buffer (2.5 M LiCl, 100 mM Tris•HCl pH 7.5, 5 mM EDTA, 25 mM DTT), and incubated with oligo$(dT)_{25}$ magnetic beads (beads from 2 ml original bead suspension, New England Biolabs) for 1 hr at 4 °C on a rotating wheel. Beads were collected with a magnet, and the supernatant was kept, representing the released fraction. Beads were washed once with lysis buffer, buffer I, II and III before elution of the RNA bound fraction from the beads following the elution protocol described above. Three bead binding/elution rounds were carried out as described for mRNA interactome capture above. Both, released and RNA bound fractions were treated with RNase T1 and RNase A and processed for mass spectrometry. Details on the RBDmap method development are described elsewhere (Castello et al., 2016).

**HL-1 Whole Cell Lysate**

$2.3\times10^6$ HL-1 cells were seeded on 56.7 $cm^2$ culture dish, and grown for 4 days. Cells were washed twice with ice-cold 1×PBS, and immediately collected in 5 ml ice-cold 1×PBS, pelleted and incubated on ice in 4 ml lysis buffer (as described in Supplemental Experiment Procedure mRNA interactome capture section). The lysate was then passed three times through a narrow gauge needle (0.4 mm diameter), incubated on ice for 10 min, followed by Trichloroacetic acid (TCA) precipitation: 1 ml of ice-cold TCA was added to the lysate, then incubated on ice for 30 min. The lysate was centrifuged at 14,000 × *g* for 20 min. The supernatant was removed, the pellet was then washed with 1 ml of 10% TCA (in $ddH_2O$) and vortexed before centrifugation at 14,000 × *g* for 20 min. The wash step was repeated with 1 ml of cold acetone (chilled at -20 °C).

The pellet was air-dried, then processed for mass spectrometry. All centrifugations were performed at 4 °C.

**Mass Spectrometry**

*Sample Preparation for Mass Spectrometry*

Samples were processed according to the standard filter aided sample preparation protocol with minor modifications (Wisniewski et al., 2009). Cysteines were reduced (10 mM DTT, 55 °C, 30 min) and alkylated (20 mM Iodoacetamide, 30 min in the dark).

HL-1 cell lysate and mRNA interactome capture samples were processed according to the following protocol: Samples were buffer-exchanged into digestion buffer (8 M Urea, 50 mM Triethylammoniumbicarbonate (TEAB, Sigma-Aldrich)) using 10 kD centrifugal devices (Millipore) and incubated with 1 µg sequencing grade Lys-C/Trypsin (Promega) at 37 °C for 4 hr. The buffer was diluted with 50 mM TEAB to a final Urea concentration of <2 M and incubated over night at 37 °C. Resulting peptides were desalted and labeled using stable isotope reductive methylation (Boersema et al., 2009) on OASIS HLB solid phase extraction cartridges (Waters). Labels were swapped between replicates. Peptides were fractionated into 12 fractions on a 3100 OFFGEL Fractionator (Agilent) using Immobiline DryStrips (pH 3-10 NL, 13 cm; GE Healthcare) according to the manufacturer's protocol. Isoelectric focusing was carried out at a constant current of 50 mA allowing a maximum voltage of 8000 V. When 20 kVh were reached the fractionation was stopped, fractions were collected and desalted using StageTips. Samples were dried in a vacuum concentrator and reconstituted in MS loading buffer (5% DMSO, 1% formic acid).

RBDmap samples were processed according to the following protocol: RNA bound and RNA released fractions were buffer-exchanged into 50 mM TEAB using 3 kD centrifugal devices

(Millipore) and incubated with 1 µg sequencing grade Trypsin (Promega) at 37 °C over night. Resulting peptides were desalted and labeled using stable isotope reductive methylation on OASIS HLB solid phase extraction cartridges. Labels were swapped between replicates. Peptides were fractionated by offline high pH reversed phase chromatography (Yang et al., 2012). Briefly, peptides were resuspended in buffer A (20 mM Ammoniumformate, pH 10) and fractionated over a Gemini 3U C18 110A column (100 × 1.00 mm, Phenomenex) using a 60 min linear gradient from 0-35% solvent B (100% Acetonitrile) at a constant flow rate of 0.1 mL/min. Resulting fractions were pooled into 8 samples, desalted, dried in a vacuum concentrator and reconstituted in MS loading buffer.

*LC-MS/MS*

Samples were analyzed on an LTQ Orbitrap Velos Pro mass spectrometer (Thermo Scientific) coupled to a nanoAcquity UPLC system (Waters). Peptides were loaded onto a trapping column (nanoAcquity Symmetry $C_{18}$, 5 µm, 180 µm × 20 mm) at a flow rate of 15 µl/min with solvent A (0.1% formic acid). Peptides were separated over an analytical column (nanoAcquity BEH C18, 1.7 µm, 75 µm × 200 mm) at a constant flow of 0.3 µl/min using the following gradient: 3% solvent B (Acetonitrile, 0.1% formic acid) for 10 min, 7-25% solvent B within 160 min, 25-40% solvent B within 10 min, 85% solvent B for 6 min. Peptides were introduced into the mass spectrometer using a Pico-Tip Emitter (360 µm outer diameter × 20 µm inner diameter, 10 µm tip, New Objective). MS survey scans were acquired from 300-1700 m/z at a nominal resolution of 30,000. The 15 most abundant peptides were isolated within a 2 Da window and subjected to MS/MS sequencing using collision-induced dissociation in the ion trap (activation time 10 msec, normalized collision energy 40%). Only 2+/3+ charged ions were included for analysis. Precursors were dynamically excluded for 30 sec (exclusion list size was set to 500).

*Peptide Identification and Quantification*

Raw data were processed using MaxQuant (version 1.3.0.5) (Cox and Mann, 2008). MS/MS spectra were searched against the UniProt mouse database (WCL and interactome capture: version 12_2013, RBDmap: version 03_2014) concatenated to a database containing protein sequences of common contaminants. Enzyme specificity was set to trypsin/P, allowing a maximum of two missed cleavages. Cysteine carbamidomethylation was set as fixed modification, and methionine oxidation and protein N-terminal acetylation were used as variable modifications. The minimal peptide length was set to six amino acids. The mass tolerances were set to 20 ppm for the first search, 6 ppm for the main search and 0.5 Da for product ion masses. False discovery rates (FDR) for peptide and protein identification were set to 1%. Match between runs (time window 2 min) and re-quantify options were enabled.

*Definition of mRNA interactome proteins*

Statistical analysis for CL/noCL enrichment of protein groups quantified in at least two out of three biological replicates was performed using an empirical Bayes moderated t-test within the R/Bioconductor package limma (Smyth, 2004). p-values were adjusted for multiple testing using the method of Benjamini-Hochberg by controlling for FDR. The UniProt accession numbers of each protein group passing the criteria log2 enrichment >0 at an FDR of 1% were converted into Ensembl gene IDs (Ensembl release 72). Where multiple Ensembl gene IDs applied, the lowest Ensembl gene ID was used. The duplicate Ensembl gene IDs were then removed, and majority proteins corresponding to the remaining genes were defined as the mRNA interactome.

*Definition of RBDpeps*

To identify RNA-binding sites, the log2 intensity ratio of MS-identified peptides in the RNA bound to the released fraction was considered. The distribution of the log2 ratios is bi-modal, representing the released and RNA bound peptides. The log2 ratios are normalized to the location of the left mode using a robust estimate. Log2 ratios of each peptide in three replicated experiments were tested against zero by a moderated t-test from the Limma package in R/Bioconductor (Smyth, 2004), and p-values were corrected for multiple testing by the method of Benjamini-Hochberg. Peptides with a 1% FDR are considered for further analysis and are extended to the closest upstream and downstream Lys-C cleavage sites to recall the original termed 'RBDpep'. Peptides extending this set to a 10% FDR are called 'CandidateRBDpep' (Table S2).

**Bioinformatic Analyses for Cardiomyocyte RBPs**

Links to custom R scripts for bioinformatic analyses can be found in the R scripts availability section of the main paper.

*Gene Annotation Bias*

RNA binding/related terms and protein domain analyses, as well as association with Mendelian disease were essentially performed as previously described (Castello et al., 2012; Castello et al., 2013a), as were analyses of GOMF enrichment, length of protein, pI, hydrophobicity, disordered region, amino acid composition, repetitive region, low complexity region and amino acid sequence logo.

Gene Ontology and Interpro were downloaded from Ensembl release 72, 'RNA binding' annotation indicates GO:0003723 and its children terms, 'RNA-related' annotation indicates a

group of GOBP, GOCC, GOMF terms and their children terms, as well as collected Interpro terms (See R scripts).

Enrichment of GOMF categories was tested for the cardiomyocyte RBPs compared to the background of proteins identified from either the cell lysate or the cardiomyocyte RBPs (designated as WCL collectively throughout all analyses). p-values were computed by Fisher's exact test, and corrected for multiple testing by the method of Benjamini-Hochberg.

*RNA-Binding Domains*

RBDs listed in (Lunde et al., 2007) were considered as classic RBDs, and protein domains with experimental validation in literature were considered non-classic RBDs as described in (Castello et al., 2012).

All Pfam domains information listed in Table S1, Table S3, and RBDmap tracks were obtained through Ensembl release 72. Pfam domains information in Figure 2C/D/H were obtained through Ensembl release 84.

*KEGG Pathway Analysis*

To determine enrichment/depletion of KEGG pathways (Kanehisa and Goto, 2000; Kanehisa et al., 2014) in cardiomyocyte RBPs, Ensembl gene IDs were first mapped to Entrez ID using the Ensembl Biomart and then these identifiers were assigned to their respective KEGG IDs through the KEGG REST interface. WCL genes were mapped to KEGG pathways in order to establish the frequency of KEGG pathway membership in the background proteome. This resulted in a total of 238 pathways with at least one member in the WCL background. Using the same method all cardiomyocyte RBPs were then mapped to KEGG pathways, yielding a total of 200 pathways with at least one member, fully contained within the 238 WCL pathways. Pathway enrichment

was tested using Fisher's exact test, using a one-tailed test (null hypothesis: odds-ratio equals to 1). The resulting 200 p-values were corrected for multiple testing (Benjamini-Hochberg method) using a total of 238 tests, as this was the number of pathways identified in WCL. The list of cardiomyocyte RBPs was further organized into RNA-related and RNA-unrelated proteins according to their annotation as described in section 'Gene annotation bias' section of Supplemental Experimental Procedures. Pathways enriched with an FDR 10% were selected. Odds-ratios were log2 transformed, and binned as shown in Figure 1G.

*STRING Analysis*

Protein-protein interactions of cardiomyocyte mtRBPs were tested by STRING (version 10) (Jensen et al., 2009). The analysis was performed at a confidence level of 0.6 with seven active prediction methods, neighborhood, gene fusion, co-occurrence, co-expression, experiments, databases and textmining.

*Cardiovascular-Associated RBPs*

A list of cardiovascular associated (related to cardiac disease and development) proteins was downloaded from the Cardiovascular Gene Ontology Annotation Initiative website (01_2015) (www.ucl.ac.uk/cardiovasculargeneontology).

*OMIM-RBPs*

RBPs listed in OMIM (Ensembl BioMart, 04_ 2015) were defined as OMIM-RBPs.

*mtRBPs*

RBPs with GO 'mitochondrion' (GO:0005739) were defined as mtRBPs.

*Metabolic Enzymes*

Genes of the cardiomyocyte RBPs were mapped to 'Metabolism' events in Reactome Pathway Database (version 48) (Croft et al., 2014), the results were manually curated to remove genes whose protein products are not classified in the six enzyme commission groups (EC 1-6) from IntEnz database (Fleischmann et al., 2004), unless the gene products are subunits of enzyme complexes. This method was also applied to the HeLa (Castello et al., 2012), the HEK293 (Baltz et al., 2012) and the mESC (Kwon et al., 2013) mRNA interactomes to compare the number of RNA-binding metabolic enzymes. Metabolic pathway assignment in Table S5 is also based on Reactome Pathway Database (version 48).

*Rossmann Fold (R-f) Proteins*

Gene3D domain annotations for all proteins were downloaded from UniProt mouse database (release-2015_05), Figure 4C is based on cathdb release 4.0. CATH id: 3.40.50 indicates the presence of Rossmann fold. The position of Rossmann fold within a protein sequence was obtained by the sequence search tool at CATH v4.0 (Sillitoe et al., 2015).

*RNA Helicases*

RNA helicase families were categorized based on (Fairman-Williams et al., 2010).

*RNA Modifications*

Types of RNA modifications catalyzed by specific enzymes were obtained from the Modomics database (Dunin-Horkawicz et al., 2006; Machnicka et al., 2013).

*Known Mitochondrial RNA-Regulating Proteins*

Proteins with known involvement in the mitochondrial RNA life cycle were curated from the literature (Jourdain et al., 2013; Nicholls et al., 2013; Rackham and Filipovska, 2012; Rackham et al., 2012; Scarpulla et al., 2012; Suzuki et al., 2011).

*OXPHOS Complex Subunits*

Number of subunits from mitochondrial Complex I, Complex II, Complex III, Complex IV and Complex V is retrieved from HUGO Gene Nomenclature Committee (HGNC) Database (04_2015) (Gray et al., 2015).

*Peptidylprolyl Isomerases (PPIases)*

PPIase family members were obtained from the HUGO Gene Nomenclature Committee (HGNC) Database (04_2015) (Gray et al., 2015).

*Human Orthologs*

Human genes orthologous to the murine cardiomyocyte RBP genes were downloaded from Ensembl release 62. The human gene set was used for RBDpep sequence similarity analysis, analyses of cardiovascular associated annotation, Mendelian disease link and comparison to previously published mRNA interactomes. For the latter, a mouse gene was considered as possessing an ortholog in the human mRNA interactome if at least one of its orthologous human genes was contained in the human mRNA interactomes.

*RS Domain Annotation*

RS domain annotation for seven SRSF proteins in RBDmap was based on (Busch and Hertel, 2012; Graveley, 2000; Manley and Krainer, 2010).

**Characterization of RBDpeps**

*Identification of Disordered Fragments*

The intrinsically unstructured or disordered parts of a protein were obtained from the 'Prediction of Intrinsically Unstructured Proteins' (IUPred) database (Dosztanyi et al., 2005). Amino acids with an IUPred score higher than 0.4 were considered as disordered. An RBDpep is regarded as disordered if the average IUPred score is higher than 0.4.

*Amino Acid Composition*

The amino acid composition of all RBDpeps or released fragments is compared to the amino acid composition of interactome proteins. For analysis of disordered or globular RNA-binding sites, RNA bound or released proteolytic fragments overlapping with disordered or globular protein segments were compared to disordered or globular fragments in interactome proteins. Over- and under-representation of a given amino acid was tested by Fisher's exact test, and p-values were corrected for multiple testing by the method of Benjamini-Hochberg.

*RBDpep Sequence Similarity Analysis*

HeLa RBDmap proteins identified with Lys-C as first protease (Castello et al., 2016) were compared to HL-1 RBDmap proteins. Pairwise alignment of HL-1 RBDpeps from the overlapped proteins was performed between mouse and human homologous proteins. Similarity was calculated using the following formula:

100 * (identical positions) / (aligned positions + internal gap positions)

*RBDpep Mapping to Protein Structures*

Protein structures data search and download were performed at Research Collaboratory for Structural Bioinformatics (RCSB) site (www.rcsb.org) of World Wide Protein Databank (wwPDB) (Berman et al., 2003; Berman et al., 2000). The homology modeling of mouse SERCA2, RNA helicases and IDH3A structures were performed with normal modeling mode at Protein Homology/analogY Recognition Engine V 2.0 (PHYRE2) (Kelley and Sternberg, 2009). The presented structures were modeled with 100% confidence. All protein structures were viewed and analyzed with PyMOL v1.3r1.

**General Molecular Biology Methods**

*Protein Gel Electrophoresis*

Protein gel electrophoresis was performed according to standard protocols (Archer et al., 2015). DTT was added to protein samples at a final concentration of 100 mM, the samples were then boiled in 1×NuPAGE LDS sample buffer (Life Technologies) for 3 min. The proteins were loaded onto denaturing polyacrylamide gel (4-12% NuPAGE Bis-Tris gel, Life Technologies), electrophoresed at 180 V for 1 hr.

*Western Blot*

Gels were electroblotted onto nitrocellulose membrane (GE Healthcare Life Sciences). The membrane was blocked in 5% non-fat milk in 1×PBST (1×PBS with 0.2% Tween-20) for 30 min at room temperature, then incubated with primary antibodies on a rotating wheel o/n at 4 °C. The membrane was then washed with 1×PBST for 3 times, 5 min each at room temperature. The

membrane was incubated with secondary antibody in 5% non-fat milk in 1×PBST, for 1 hr at room temperature. Protein signals were detected using Super Signal Femto chemiluminescent reagent (Pierce), visualized on ImageQuant LAS 4000 system (GE Healthcare Life Sciences). Antibodies used were PTBP1 (SAB2101904-50UG, Sigma), ELAVL1 (ab28660, Abcam), H3 (ab1791, Abcam), H4 (ab10158, Abcam), ACTB (ab20272, Abcam), α-TUBULIN (sc-8035, Santa Cruz), CS (ab96600, Abcam), OGDH (ab137773, Abcam), IDH2 (ab131263, Abcam), ACO2 (MS793, MitoSciences), GFP (3H9, Chromotek), NSUN4 (16320-1-AP, Proteintech Group Inc.) and VDAC (4661, Cell Signaling Technology). HRP-conjugated antibodies are anti-mouse (sc-2005, Santa Cruz), anti-rat (sc-2006, Santa Cruz) and anti-rabbit (AP132P, EMD Millipore).

*Silver Stain*

Silver stain was performed according to standard protocol (Chevallet et al., 2006). After electrophoresis, the gel was fixed in 50% methanol (v/v), 5% acetic acid (v/v) for 30 min. The gels were washed sequentially in 50% ethanol (v/v) and 30% ethanol (v/v), 5 min for each, and then in dd$H_2O$ for 10 min. The gel was sensitized by soaking in 0.02% sodium thiosulfate (w/v) for 1 min, followed by three washes with dd$H_2O$, 30 sec each. The gel was then impregnated with silver solution (6 mM silver nitrate, 0.0185% formaldehyde) for 20 min, followed by three washes with water, 30 sec each. After the final wash, developing solution (2% sodium carbonate, 0.0185% formaldehyde, 0.0004% sodium thiosulfate) was applied to the gel within 5 min of preparation. When adequate degree of staining was achieved, the gel was transferred to 10% acetic acid to stop the staining reaction.

When restain was necessary, the gel was incubated in destain solution (prepared by mixing two solutions, 30 mM potassium ferricyanide and 100 mM sodium thiosulfate, at equal volume

immediately prior to use) until gels were clear again, typically this occurred within 10 min. The gel was then washed with water to remove any yellow color. The gel was then restained as described above.

All solutions were prepared freshly, and all procedures were performed on a rocking platform at room temperature. Image of the silver stain was taken by a Nikon D5100 DSLR camera.

*RNA Extraction and cDNA Synthesis*

RNA from lysate or digested RNA-protein complexes was extracted with Trizol prior to random hexamer/$(dT)_{12}$VN-primed cDNA synthesis, using SuperScript III Reverse Transcriptase (Life Technologies).

*Quantitative PCR*

Quantitative PCR was performed using Quantifast qPCR premix (Qiagen) as per manufacturer's instructions. Serial dilutions of cDNA amplified from HL-1 total RNA were used to test the PCR efficiency, which was >90% for all primers. 10 μl reactions were generated in 384-well plates (3 technical replicates per condition) and qPCR was performed on a QuantStudio 12 K Flex (Invitrogen) (Archer et al., 2014). The following table provides PCR primer sequence information, each was used at 0.5 μM final concentration.

| ***Gene name*** | ***Forward primer for qPCR (5'–3')*** | ***Reverse primer for qPCR (5'–3')*** |
|---|---|---|
| *18s rDNA* | gtaacccgttgaaccccatt | ccatccaatcggtagtagcg |
| *Actb* | gatcaagatcattgctcctcctg | agggtgtaaaacgcagctca |
| *Gapdh* | aagggctcatgaccacagtc | cagggatgatgttctgggca |
| *Elavl1* | gaagaccacatggcggaaga | cccaagctgtgtcctgctac |
| *Mt-co1* | gagaggcctttgcttcaaaaa | aggttggttcctcgaatgtg |
| *Mt-nd3* | ctagttgcattctgactcccc | atggtagacgtgcagagctt |
| *Mt-nd2* | gggcatgaggaggacttaaccaaac | tgaggttgagtagagtga |
| *Mt-nd5* | tctcacccaaaacgacatca | ttgaagaatgcgtgggtaca |
| *Mt-atp6* | aggattcccaatcgttgtagcc | ccttttggtgtgtggattagca |
| pT3lucpA spike-in RNA | cgccaaaagcactctgattgac | ccttgtcgtatccctggaagatg |

**RNA Binding Assays with eGFP-tagged Candidate RBPs**

RNA 5' end-labeling assays were performed as previously described with modifications (Baltz et al., 2012; Kwon et al., 2013; Spitzer et al., 2014), using HeLa cells expressing eGFP-tagged candidate RBPs as source material.

*Plasmids Construction*

Plasmids that express C-terminally eGFP-tagged fusion proteins were based on an existing construct, *Nsun2-eGFP-pcDNA5/FRT/TO*, which was generated in the following way. The *Nsun2* coding sequence (CDS) was first cloned into *pEGFP-N1* (Addgene) vector through *Kpn*I/*Xma*I sites. The *Nsun2*-eGFP sequence was subsequently excised by *Kpn*I/*Not*I digest and cloned into the cognate sites of *pcDNA5/FRT/TO* (Life Technologies). To prepare the plasmids used for the

$^{32}$P RNA 5' end-labeling assay, the CDS from other genes were amplified from a HL-1 cDNA library using primers that introduced unique restriction sites and cloned into pGEM-T easy vector (Promega). CDS were liberated from this transit vector and replaced for the *Nsun2* portion of *Nsun2-eGFP-pcDNA5/FRT/TO* cDNA using the restriction enzyme combinations detailed in the table below. The control *eGFP-pcDNA5/FRT/TO* vector was constructed by inserting the *eGFP* CDS from *pEGFP-N1* into *pcDNA5/FRT/TO* (Life Technologies) at *Kpn*I/*Not*I sites. Details on the cloning strategy and linker sequences of the plasmids are given in the following table.

| ***Gene name*** | ***GenBank accession number*** | ***Primers for cloning (5′–3′)*** | ***Cut sites on cDNA ends 5′/3′*** | ***Cut sites on pcDNA5/FRT/TO vector 5′/3′*** | ***Linker amino acid sequence between fusion protein and eGFP*** | ***Dose of DNA ϟ*** |
|---|---|---|---|---|---|---|
| *Aco2* | NM_080633.2 | F*-catggtaccatctttgtcagtgcac | *Kpn*I/ *Xma*I | *Kpn*I/*Xma*I | PRDPPVAT | 2 |
| | | R*-catcccggggctgctgcagctccttc | | | | |
| *Cs* | NM_026444.3 | F-catggtacccggtcccttcccgcca | *Kpn*I/ *Stu*I | *Kpn*I/*Sma*I | QGDPPVAT | 2 |
| | | R-cataggcctgcttagagtccacaaac | | | | |
| *Idh2* | NM_173011.2 | F-catggtaccctcggacctcgcgtgc | *Kpn*I/ *Xma*I | *Kpn*I/*Xma*I | PRDPPVAT | 0.5 |
| | | R-catcccggggctgcttgcccagagctc | | | | |
| *MDH2* | M16229.1 | F-catggtacctcctgccagtagctcc | *Kpn*I/ *Xma*I | *Kpn*I/*Xma*I | PRDPPVAT | 0.5 |
| | | R-catcccggggcttcatgttcttgaca | | | | |
| *Pdha1* | NM_008810.3 | F-catggtacccgccgcgtgagtctgc | *Kpn*I/ *Xma*I | *Kpn*I/*Xma*I | PRDPPVAT | 2 |
| | | R-catcccggggactgactgacttaaac | | | | |
| *Pum2* | NM_001160222.1 | F-cataagcttgcagcgcggccgcag | *Hind*III/ *Xma*I | *Hind*III/ *Xma*I | PRDPPVAT | 1.2 |
| | | R-catcccggggcagcatcccatttgg | | | | |
| eGFP | n/a | n/a | n/a | n/a | n/a | 0.25 |
| MLS-eGFP§ | n/a | n/a | n/a | n/a | n/a | 2 |

* F means forward primer; R means reverse primer.

ϟ means amount of DNA for Lipofectamine 2000 transfection in RNA $^{32}$P 5' end-labeling assay (μg/10cm$^2$ culture area). Related to *Cell Transfection* section below.

§ refers to (Popow et al., 2015).

*Cell Transfection*

$7.2\times10^6$ HeLa cells were seeded on each 145 $cm^2$ culture dish 16 hr prior to transfection. Plasmid transfections were performed with 7 ml transfection mixture containing plasmid DNA and Lipofectamine 2000 (Life Technologies) at 1:2.5 ratio (w/v). For the amount of transfected DNA and other detailed information, see the table above. Complete medium was replaced 6 hr after transfection and cells were harvested at 24 hr after transfection.

*Cell Lysate Preparation*

Cells were washed twice with ice-cold 1×PBS, the culture dish was then placed on ice and irradiated with 0.15 J/$cm^2$ UV at 254 nm. Cells were then harvested by centrifugation at 2,000 × *g* for 5 min. The cell pellets were lysed in three volumes of lysis buffer (CytoBuster protein extraction reagent, EMB Millipore), supplemented with RNaseOUT (Life Technologies) at final concentration of 100 U/ml, and complete EDTA-free protease inhibitor cocktail (Roche) at 1×, then incubated on ice for 10 min, the cell lysate was snap frozen on dry ice, thawed, and cleared by centrifugation at 10,000 × *g* for 10 min, and the supernatant was stored at -80 °C until further analysis. All centrifugations were performed at 4 °C. The typical scale of such an experiment was 3×145 $cm^2$ culture dishes per sample.

*Preparation of Mitochondria*

The procedure was performed as previously described with some modifications to accommodate the properties of cardiomyocytes (Mercer et al., 2011; Rackham et al., 2007). Cells were washed twice with ice-cold 1×PBS, the culture dish was then placed on ice and irradiated with 150 mJ/$cm^2$ UV at 254 nm. Cells were harvested by centrifugation at 1,000 × *g* for 5 min, and resuspended in 8 ml/145 $cm^2$ culture dish of homogenization buffer (250 mM sucrose, 10 mM

NaCl, 1.5 mM $MgCl_2$, 10 mM Tris•HCl, pH 7.5). Cells were then homogenized on ice, 5 times with a loose glass homogenizer, and 10 times with a tight glass homogenizer (Sigma). The suspension was centrifuged at 1,300 × *g* for 5 min and the supernatant was retained. The pellet was resuspended in sucrose buffer (5 ml/145 $cm^2$ culture dish; 250 mM sucrose, 10 mM Tris•HCl, pH 7.5, 1 mM EDTA), and centrifuged at 1,300 × *g* for 5 min. The two supernatants were pooled, centrifuged again at 1,300 × *g* for 5 min. The resulting supernatant was centrifuged at 12,000 × *g* for 15 min to sediment the mitochondria. The mitochondria pellet was resuspended in 1 ml sucrose buffer, and centrifuged again at 10,000 × *g* for 10 min. The supernatant was discarded. The mitochondria pellet was lysed in lysis buffer (same composition and equivalent volume as used in whole cell lysis). The mitochondria lysate was incubated on ice for 10 min, then snap frozen on dry ice, thawed, and cleared by centrifugation at 10,000 × *g* for 10 min, and the supernatant was stored at -80 °C until further analysis. All centrifugations were performed at 4 °C.

*Immunoprecipitation*

The cell lysate/mitochondria lysate was thawed on ice, and RNase T1 was added to the lysate at a final concentration of 1 U/µl. The reaction mixture was incubated in a water bath at 22 °C for 10 min and subsequently cooled on ice for 5 min, before addition of pre-equilibrated GFP-Trap agarose beads (Chromotek). 80 µl beads were added per ml of partial RNase T1 digested lysate. The immunoprecipitation mixture was incubated on a rotating wheel for 2 hr at 4 °C. The beads were collected by centrifugation at 2,500 × *g* for 2 min, subsequently washed once with 250 µl of high salt buffer (500 mM NaCl, 1 mM $MgCl_2$, 0.025% SDS, 0.05% NP-40, 20 mM Tris•HCl, pH 7.5), once with medium salt buffer (250 mM NaCl, 1 mM $MgCl_2$, 0.025% SDS, 0.05% NP-40, 20 mM Tris•HCl, pH 7.5), and once with low salt buffer (150 mM NaCl, 1 mM $MgCl_2$, 0.01% NP-

40, 20 mM Tris•HCl, pH 7.5). After stringent wash, the beads were resuspended in one original bead volume of low salt buffer, and RNase T1 was added to obtain a final concentration of 50 U/µl. The bead suspension was incubated in a water bath at 22 °C for 8 min, and subsequently cooled on ice for 5 min. Beads were collected by centrifugation at 2,500 × *g* for 2 min, and washed twice with wash buffer (50 mM Tris•HCl, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$). Beads were then resuspended in one original bead volume of 1×PNK buffer (New England Biolabs).

*$^{32}$P 5' End-Labeling of RNA Segments Crosslinked to eGFP Fusion Proteins*

To the bead suspension described above, γ-$^{32}$P-ATP (Perkin Elmer) was added to a final concentration of 0.2 µCi/µl, and T4 PNK to a final concentration of 1 U/µl.

The suspension was then incubated at 37 °C for 15 min. Thereafter, the beads were washed five times with wash buffer as above, resuspended in 30 µl elution buffer (1×NuPAGE sample buffer (Life Technologies), supplemented with SDS at final concentration 1%), and incubated at 90 °C for 5 min to elute the RNA-protein complexes. The eluate was collected by centrifugation at 2,500 × *g* for 2 min, separated on a 4-12% NuPAGE Bis-Tris gel (Life Technologies), at 180 V for 90 min, and subsequently electroblotted onto nitrocellulose membrane. The membrane was exposed to a phosphorimager screen at room temperature for 40 hr. After detecting the radioactive signal on Typhoon FLA 9000 laser scanner (GE Healthcare Life Sciences), the membrane was blocked in 5% milk in 1×PBST for 1 hr at room temperature, and eGFP fusion protein was detected by GFP antibody (3H9, Chromotek) (room temperature for 1 hr on a rocking platform). Secondary HRP conjugated anti-rat antibody incubation and protein signal detection were as described in Supplemental Experimental Procedures western blot section.

**Confocal Microscopy**

HeLa cells were seeded on a coverslip in 12-well cell culture dish (3.8 $cm^2$ culture area/well) 16 hr prior to transfection with plasmids expressing eGFP-tagged fusion proteins as described in the Supplemental Experimental Procedures *Cell transfection* section, adjusting proportionally for the smaller culture area. 24 hr after transfection, the culture media were removed, and the cells were incubated in 100 nM MitoTracker Deep Red (Life Technologies) at 37 °C for 30 min. Cells were washed three times with serum free cell culture medium at room temperature on a rocking platform, then fixed with fresh 4% paraformaldehyde at 37 °C for 15 min. The cells were then washed 3 times with 1×PBS at room temperature. Cells were mounted with VectaShield mounting medium with DAPI (VectorLabs), and imaged using Leica TCS SP5 confocal microscope.

# SUPPLEMENTAL REFERENCES


Antonicka, H., and Shoubridge, E.A. (2015). Mitochondrial RNA Granules Are Centers for Posttranscriptional RNA Processing and Ribosome Biogenesis. Cell Rep. *10*, 920-932.

Archer, S.K., Shirokikh, N.E., Hallwirth, C.V., Beilharz, T.H., and Preiss, T. (2015). Probing the closed-loop model of mRNA translation in living cells. RNA Biol. *12*, 248-254.

Archer, S.K., Shirokikh, N.E., and Preiss, T. (2014). Selective and flexible depletion of problematic sequences from RNA-seq libraries at the cDNA stage. BMC Genomics *15*, 401.

Baltz, A.G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell *46*, 674-690.

Beckmann, B.M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A.M., Schwarzl, T., Curk, T., Foehr, S., Huber, W., et al. (2015). The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. Nat. Commun. *6*, 10127.

Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. Nat. Struct. Biol. *10*, 980.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235-242.

Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A.J. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat. Protoc. *4*, 484-494.

Busch, A., and Hertel, K.J. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. WIRES RNA *3*, 1-12.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393-1406.

Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.-M., Föhr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive analysis of RNA-binding domains in human cells. Mol. Cell *In press*.

Castello, A., Fischer, B., Hentze, M.W., and Preiss, T. (2013a). RNA-binding proteins in Mendelian disease. Trends Genet. *29*, 318-327.

Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L.M., Krijgsveld, J., and Hentze, M.W. (2013b). System-wide identification of RNA-binding proteins by interactome capture. Nat. Protoc. *8*, 491-500.

Chevallet, M., Luche, S., and Rabilloud, T. (2006). Silver staining of proteins in polyacrylamide gels. Nat. Protoc. *1*, 1852-1858.

Clancy, J.L., Wei, G.H., Echner, N., Humphreys, D.T., Beilharz, T.H., and Preiss, T. (2011). mRNA isoform diversity can obscure detection of miRNA-mediated control of translation. RNA *17*, 1025-1031.

Claycomb, W.C., Lanson, N.A., Jr., Stallworth, B.S., Egeland, D.B., Delcarpio, J.B., Bahinski, A., and Izzo, N.J., Jr. (1998). HL-1 cells: a cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. Proc. Natl. Acad. Sci. USA *95*, 2979-2984.

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.

Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., et al. (2014). The Reactome pathway knowledgebase. Nucleic Acids Res. *42*, D472-477.

Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics *21*, 3433-3434.

Dunin-Horkawicz, S., Czerwoniec, A., Gajda, M.J., Feder, M., Grosjean, H., and Bujnicki, J.M. (2006). MODOMICS: a database of RNA modification pathways. Nucleic Acids Res. *34*, D145-149.

Fairman-Williams, M.E., Guenther, U.P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. Curr. Opin. Struct. Biol. *20*, 313-324.

Fleischmann, A., Darsow, M., Degtyarenko, K., Fleischmann, W., Boyce, S., Axelsen, K.B., Bairoch, A., Schomburg, D., Tipton, K.F., and Apweiler, R. (2004). IntEnz, the integrated relational enzyme database. Nucleic Acids Res. *32*, D434-437.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1197-1211.

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*, D1079-1085.

Humphreys, D.T., Hynes, C.J., Patel, H.R., Wei, G.H., Cannon, L., Fatkin, D., Suter, C.M., Clancy, J.L., and Preiss, T. (2012). Complexity of murine cardiomyocyte miRNA biogenesis, sequence variant expression and function. PLoS One *7*, e30933.

Iizuka, N., Najita, L., Franzusoff, A., and Sarnow, P. (1994). Cap-dependent and cap-independent translation by internal initiation of mRNAs in cell extracts prepared from Saccharomyces cerevisiae. Mol. Cell. Biol. *14*, 7322-7330.

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., et al. (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. *37*, D412-416.

Jourdain, A.A., Koppen, M., Wydro, M., Rodley, C.D., Lightowlers, R.N., Chrzanowska-Lightowlers, Z.M., and Martinou, J.C. (2013). GRSF1 regulates RNA processing in mitochondrial RNA granules. Cell Metab. *17*, 399-410.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27-30.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. *42*, D199-205.

Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. *4*, 363-371.

Kwon, S.C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1122-1130.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479-490.

Machnicka, M.A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M., et al. (2013). MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res. *41*, D262-267.

Manley, J.L., and Krainer, A.R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). Genes Dev. *24*, 1073-1074.

Mercer, T.R., Neph, S., Dinger, M.E., Crawford, J., Smith, M.A., Shearwood, A.M., Haugen, E., Bracken, C.P., Rackham, O., Stamatoyannopoulos, J.A., et al. (2011). The human mitochondrial transcriptome. Cell *146*, 645-658.

Nicholls, T.J., Rorbach, J., and Minczuk, M. (2013). Mitochondria: mitochondrial RNA metabolism and human disease. Int. J. Biochem. Cell Biol. *45*, 845-849.

Popow, J., Alleaume, A.M., Curk, T., Schwarzl, T., Sauer, S., and Hentze, M.W. (2015). FASTKD2 is an RNA-binding protein required for mitochondrial RNA processing and translation. RNA *21*, 1873-1884.

Preiss, T., and Hentze, M.W. (1998). Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast. Nature *392*, 516-520.

Rackham, O., and Filipovska, A. (2012). The role of mammalian PPR domain proteins in the regulation of mitochondrial gene expression. Biochim. Biophys. Acta *1819*, 1008-1016.

Rackham, O., Mercer, T.R., and Filipovska, A. (2012). The human mitochondrial transcriptome and the RNA-binding proteins that regulate its expression. WIRES RNA *3*, 675-695.

Rackham, O., Nichols, S.J., Leedman, P.J., Berners-Price, S.J., and Filipovska, A. (2007). A gold(I) phosphine complex selectively induces apoptosis in breast cancer cells: implications for anticancer therapeutics targeted to mitochondria. Biochem. Pharmacol. *74*, 992-1002.

Scarpulla, R.C., Vega, R.B., and Kelly, D.P. (2012). Transcriptional integration of mitochondrial biogenesis. Trends Endocrinol. Metab. *23*, 459-466.

Sillitoe, I., Lewis, T.E., Cuff, A., Das, S., Ashford, P., Dawson, N.L., Furnham, N., Laskowski, R.A., Lee, D., Lees, J.G., et al. (2015). CATH: comprehensive structural and functional annotations for genome sequences. Nucleic Acids Res. *43*, D376-381.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. *3*, Article3.

Spitzer, J., Hafner, M., Landthaler, M., Ascano, M., Farazi, T., Wardle, G., Nusbaum, J., Khorshid, M., Burger, L., Zavolan, M., et al. (2014). PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. Methods Enzymol. *539*, 113-161.

Suzuki, T., Nagao, A., and Suzuki, T. (2011). Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. Annu. Rev. Genet. *45*, 299-329.

Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359-362.

Yang, F., Shen, Y., Camp, D.G., 2nd, and Smith, R.D. (2012). High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. Expert Rev. Proteomics *9*, 129-134.