Title
Control of Gene Expression by RNA Binding Proteins and RNA Secondary Structure Elements

Author
Khoroshkin, Matvei

Publication Date
2022

Peer reviewed|Thesis/dissertation

Control of Gene Expression by RNA Binding Proteins and RNA Secondary Structure Elements

by
Matvei Khoroshkin

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biological and Medical Informatics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Hani Goodarzi
FDD44359FCC6487...

Hani Goodarzi

Chair

DocuSigned by:
Hiten Madhani

Hiten Madhani

DocuSigned by:
Stephen Floor
3709C3C6AF93461...

Stephen Floor

Committee Members

DEDICATION AND ACKNOWLEDGEMENTS

I dedicate this work to my parents and to my sister, who has also been a parent for me for the last 10 years. You created a home I love, and you taught me to appreciate the world around us.

I would like to thank my advisor, Hani Goodarzi, for being very patient with me and shielding me from the stress of navigating the big and scary academic world. I would like to thank my labmates, especially Albertas, Johnny, and Lisa, who went through a lot of good and bad days together with me. I would like to thank my collaborators, who shared their skills with me and helped me with experiments, especially Andrey, Vanya, Faraz, Aidan, Luke, Ryan, Simon, Mark, Christina, and Yifan. I would like to thank my mentees, who taught me many valuable lessons in turn: Yikai, Nadya, Rithvik, Anthony, Fathima, and especially Artemii – I appreciate the time we spent together. I would like to thank my roommates, who created an environment that was fun to be in: Andrew, Sophia, Matan, Keely, Daniel, and Gabe. I would like to thank all my friends at UCSF – I value the time we spent together and hope that our best times are still in the future: Chris, Snow, James, Greyson, Jason, Wes, Garrett, Cole, Nat, Christina, Luca, Ben, Leo, Ilan, Alina, Scott, Amy, Beatrice, Amanda, Yuwei, Athena, Tamas, Candace, and all the rest. I would like to thank Zhenya, Petya, Arina, and Sasha for supporting me throughout these years, even though we didn't see each other very often. I would like to thank my aunt Anya and uncle Lenya for helping me out whenever I needed something. I would like to thank my parents again for being a huge support. I would like to thank my brother Anton, my grandmother Natasha, and once more, my sister Anya: our connection kept me going in the hard moments. And finally, a huge thank you to my wife Yessica, for being with me.

Control of Gene Expression by RNA Binding Proteins and RNA Secondary Structure Elements

Matvei Khoroshkin

ABSTRACT

RNA-binding proteins (RBPs) are multifunctional regulators of gene expression with complex and context-dependent mechanisms of action. The underlying regulatory grammar that underlies RBP-mediated functions remains largely unexplored. Here, we developed and applied a multi-omic data integration platform to systematically decipher the context-specific functions of RBPs. We used *in vivo* proximity-dependent biotinylation (BioID) analysis of 50 human RBPs to generate a comprehensive map of major RBP neighborhoods. In parallel, we took advantage of CRISPR-interference with single-cell RNA-seq read-out (CROP-seq) to effectively capture the overall transcriptomic response downstream of each RBP knockdown. By integrating these physical and functional interaction data with the ENCODE transcriptome-wide atlas of RBP binding from eCLIP assays, we generated a map of RBP functional interaction. The resulting map captures well-studied post-transcriptional pathways and reveals novel RPB-mediated regulatory functions. Here, we report the validation of context-specific functions for several RBPs using biochemical and genetic approaches. We showed that TAF15 controls splicing, translation, and stability for distinct and largely independent RNA regulons. Similarly, we demonstrated that ZNF800 and QKI play non-canonical roles in both transcriptional and post-transcriptional regulation. Taken together, our integrative map reveals that each RBP may participate in multiple post-transcriptional regulatory modules, each with their own target regulon and regulatory function. Deciphering these complex modes of function and interaction are the first step towards a better understanding of post-transcriptional control in health and disease.

RNA structural switches are key regulators of gene expression in bacteria, yet very few have been described and characterized in Metazoa. Here, we present an integrative computational and experimental approach to systematically annotate functional RNA structural switches genome-wide. By applying this approach to the

human transcriptome, we discovered 245 putative RNA switches. Among these, we further characterized and functionally dissected a previously unknown RNA switch in the 3'UTR of the *RORC* transcript. *In vivo* DMS-MaPseq complemented by cryogenic electron microscopy confirmed the existence of this element as an ensemble of two alternative structural conformations. We then used a genome-wide CRISRPi screen to identify the *trans* factors that regulate gene expression through this RNA structural switch. We discovered that the nonsense-mediated mRNA decay pathway acts on this element in a conformation-specific manner. Our findings suggest that RNA structural switches may play an important, yet underappreciated role, in shaping the gene expression landscape of the cell in metazoans.

Table of Contents

List of Figures

Chapter 1: Unified Framework for Systematic Identification of Post-Transcriptional Regulatory Modules

INTRODUCTION

RNA-binding proteins (RBPs) play a fundamental role in post-transcriptional gene expression control. Regulatory factors do not function in isolation; nearly all RBPs are multifunctional and act in a highly context-dependent manner (e.g. Wang et al. 2013). It is common for a single RBP to influence multiple aspects of RNA metabolism, including splicing, export, and stability. This diversity commonly observed in RBP function results from their ability to form multitudes of independent and context-driven complexes. However, we currently lack a comprehensive map of these RBP-RNA complexes that form the building blocks of post-transcriptional regulation. Revealing the identity of these post-transcriptional regulatory modules is a key step towards a better understanding of regulatory networks in health and disease.

Individual modalities, such as knockdown/knockout/overexpression of an RBP of interest followed by transcriptomic readout, often fail to capture the combinatorial complexity of post-transcriptional control of gene expression in isolation (Sternburg and Karginov 2020). The key reasons for this are (1) intrinsic limitations of each individual experimental technique, and (2) biological noise originating from mutual entanglement of regulatory networks. These problems are common for studying any biological system, however, they are especially prominent for the studies of post-transcriptional regulation since RBPs are remarkably multi-functional and redundant. For example, the RBP muscleblind-like 1 (MBNL1) regulates splicing (Lin et al. 2006; Mankodi et al. 2002) when binding pre-mRNAs in the nucleus; whereas, it modulates RNA

stability (Fish et al. 2016) when binding 3'UTRs of mRNAs in the cytoplasm(Therefore, traditional approaches focused on one RBP-one function had limited success in capturing the roles of individual RBPs. This is because any individual assay describes a single specific aspect of a relationship between two RBPs: e.g., physical interaction, or overlapping target RNAs. However, in order to reveal the regulatory grammar underlying RNA metabolism, a generalized understanding of functional interaction between RBPs is required that incorporates multiple data modalities rather than capturing one specific type of protein interaction. To accomplish this, we developed an integrative framework that takes into account three scenarios of RBP-RBP entanglement that are resolved by different assays: (i) two RBPs are co-localized in close physical proximity to each other, (ii) two RBPs share significant overlapping RNA targets, and (iii) modulation of two RBPs results in similar downstream transcriptomic shifts. Combining data from these three independent modalities, within a unified statistical and analytical framework, allows us to capture RBP regulatory modules in their most generalizable form.

To capture the spatial relationship between RBPs, we carried out a large-scale proximity-labeling effort to reliably capture and identify the proteins that reside in proximity of each RBP. In order to reveal functional interactions between RBPs that do not rely on direct physical interactions, we took advantage of loss of function screens with rich single-cell read-outs (Norman et al. 2019). We used the CRISPR-interference technology (Gilbert et al. 2014) to target RBPs and used single-cell RNA sequencing to capture changes in the transcriptomic state of the cell in response to RBP knockdowns. This rich dataset allowed us to explore functional interactions between RBPs with few *a priori* assumptions. Finally, we supplemented the resulting data with RBP-RNA interactions mapped as part of the RNA ENCODE effort using cross-linking immunoprecipitation followed by

sequencing (eCLIP) (Van Nostrand et al. 2016). Incorporation of this data allows for identification of the RNA regulons that are downstream of RBP regulatory modules.

Taken together, in this study, we use a multi-omic data integration approach to reveal combinatorics of RBPs in post-transcriptional gene regulation, taking advantage of both reliability of low-throughput proteomics (BioID labeling) and high-throughput sequencing (Perturb-seq, ENCODE eCLIP). By integrating several experimental modalities together, we generated a general map of RBP regulatory interactions. Here, we show that this map recapitulates many known and canonical functions of individual RBPs, however, we also showcase multiple instances of novel and non-canonical functions. Our discovery of a large repertoire of functional interactions between RBPs, many of which were previously unknown, provides an opportunity for the community to better capture the role of post-transcriptional regulatory programs in health and disease.

RESULTS

Integrating data from complementary approaches captures known RBP complexes and unravels novel connections

In order to broadly and systematically annotate regulatory interactions between RBPs, we combined data from multiple independent modalities, namely (i) BioID2-mediaed proximity protein labeling, (ii) RBP Perturb-seq, and (iii) eCLIP RNA ENCODE dataset (Fig. 1.1). 50 of 120 RBPs studied by the ENCODE consortium (Van Nostrand et al. 2016) are previously cloned as part of the human ORFeome cDNA library (Yang et al. 2011). We used this collection to fuse BioID2 to each of these RBPs, and generated stable clones in the K562 background. We then

performed individual streptavidin-precipitations followed by mass spectrometry. This allowed us to systematically identify co-localized RBPs. In parallel, we performed a Perturb-seq experiment and obtained transcriptome-wide gene expression measurements for 68 RBPs knockdowns using single-cell RNA sequencing. We used the resulting high-dimensional data to systematically delineate functional interactions between RBPs based on their shared transcriptomic fingerprints. And finally, we re-analyzed the ENCODE eCLIP data to capture overlapping regulon memberships between RBPs (Van Nostrand et al. 2016).

To assess the RBPs interactions, we used BioID labeling-based pulldown followed by mass spectrometry, the gold standard for studying protein-protein spatial interactions. Traditional antibody-based pulldowns are limited by the strength of the bait-antibody interaction, which does not allow for harsh washes and leads to higher noise and antibody-dependent bias in the data. They are also prone to loss of lower-affinity interactions, such as the interactions mediated via intrinsically disordered protein regions, which are recognized to be functionally crucial in post-transcriptional regulation. This problem can be solved with proximity-based covalent biotin labeling techniques as biotin-streptavidin interaction is very strong, allows for harsh washes, and therefore tends to have low levels of nonspecific binding (Li et al. 2017) (Roux et al. 2018). The main disadvantage of biotin-based systems (Kim et al. 2016; Cho et al. 2020; Hung et al. 2014; Hung et al. 2016) is limited scalability: a cell line expressing a fusion of bait protein with the labeling enzyme has to be generated and functionally validated for each protein of interest. Because of this, only a few large-scale datasets of dozens or hundreds of bait proteins have been generated to date (Antonicka et al. 2020; Go et al. 2021; Youn et al. 2018). We applied proximity-dependent biotinylation (BioID) analysis to 50 human RBPs.

In order to define the RBP-RNA interactions, we re-analyzed the ENCODE eCLIP (enhanced cross-linking immunoprecipitation) data (Van Nostrand et al. 2016). In this method, proteins and RNA get cross-linked by UV light, the bait RBP with the RNAs attached is pulled down by an antibody, and the respective RNA fragments are sequenced. To reliably identify the RNA targets for the RBPs of interest, we reprocessed all the publicly available eCLIP datasets as in (Agrawal et al. 2021). The knowledge of RBP-RNA interactions is important for two reasons. First, the eCLIP data defines the regulons for each RBP; moreover, if an RBP is multifunctional, the eCLIP data facilitates the annotation of the individual regulons for each of the RBP's functions. Second, this data helps annotate the pathway-like RBP interactions: two RBPs never directly interact with each, yet they act on the same set of RNA targets consecutively.

Specific combinations of RBPs exhibit non-additive effects on gene expression, enabling a variety of phenotypic changes in the cell. Such combinations can be detected by modeling the genetic interactions manifold, that describes the transcriptional states a cell can occupy upon perturbation (Norman et al. 2019). Here, we use Perturb-seq (Adamson et al. 2016) to quantitatively measure the effects of RBP knockdowns on the cellular transcriptional state. We then use this data to model the functional interactions between RBPs.

Together, these three modalities capture complementary aspects of regulatory interactions between RBPs. Therefore, integrating these sources of information is a critical step towards generating a comprehensive and generalizable map of RBP-RBP interaction (Fig. 1.1). To accomplish this, we first generated RBP-RBP maps for individual modalities. We then computed *z*-scores across

features (i.e., protein preys, genes, and target RNAs) for each modality respectively. Next, we estimated pairwise cosine distances between RBPs, and transformed them into empirical p-values to achieve a uniform RBP similarity scale. Finally, the p-values of independent measurements of RBP similarity (i.e. from BioID, Perturb-seq, and eCLIP respectively) were combined into a single probability score expressing the overall level of interaction between a pair of RBPs.

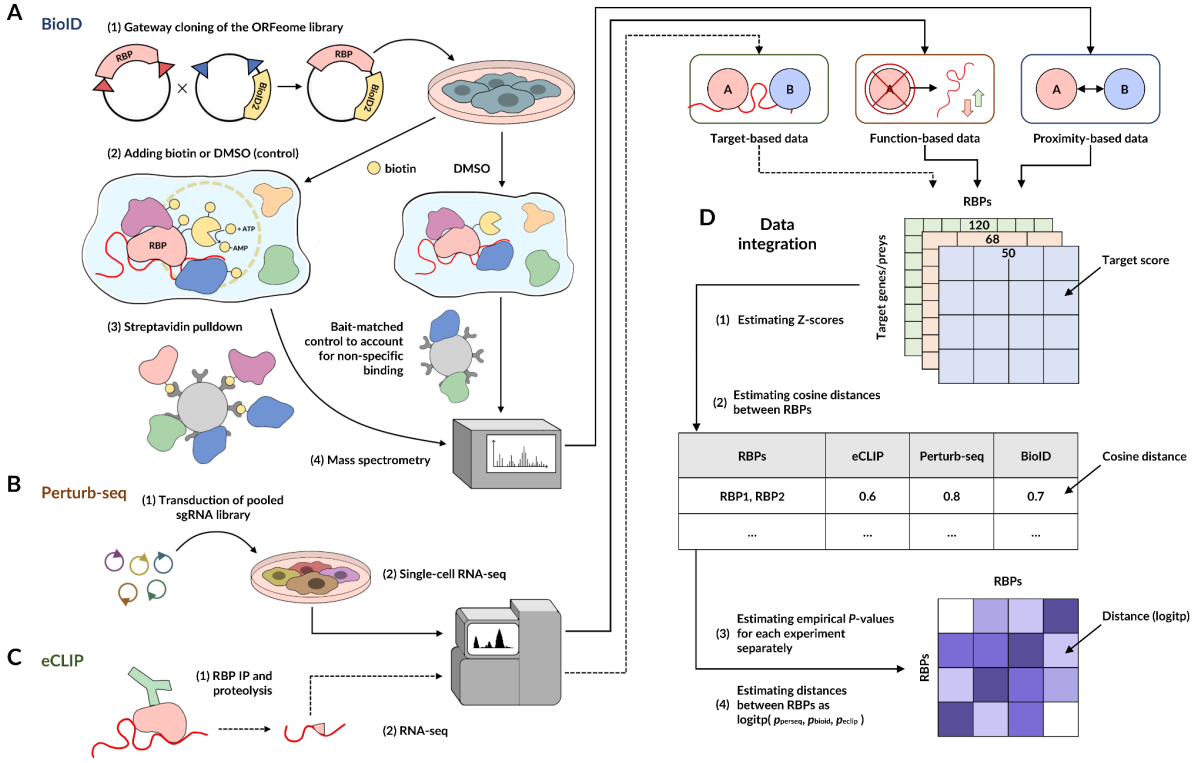

Figure 1.1. Generating an integrated regulatory interaction map of RNA-binding proteins: an overview of our workflow, the experimental data used in the study and the datasets integration procedure.

Results of BioID2 **(A)**, Perturb-seq **(B)** and publicly available ENCODE eCLIP **(C)** assays were independently processed and normalized across RBPs (*z*-score transformation).**(D)**. The resulting Z-scores were used to estimate the cosine distance between all pairs of the tested RBPs. The distances were then replaced by empirical *P*-values reflecting the distance between two particular RBPs according to a particular type of assay. Finally, to generate an integrated regulatory interaction map, for each pair of RBPs, the *p*-values from three assays were then logit-aggregated to obtain a single resulting measure of the distance between RBPs in the pair.

The resulting Integrated RBP Regulatory Map, combining physical and genetic interactions, elucidates the combinatorial regulatory logic underlying RBP-mediated post-transcriptional control of gene expression (Fig. 1.2). Interaction maps are often interpreted by identifying proteins that cluster together into functional complexes. And as shown in Fig. 1.2, our map similarly captures a number of canonical RBP modules involved in key post-transcriptional regulatory programs such as mRNA stability and RNA processing. However, as stated above, RBPs carry out multiple independent functions, which is not captured by this "one RBP-one cluster" scheme. This is clearly highlighted by the many off-diagonal interactions that are revealed in our RBP map.

Analysis of the on-diagonal clusters in the Integrated RBP Regulatory Map clearly captured many canonical RBP complexes involved in translation, splicing, ribosomal RNA processing, and other RNA metabolic pathways (Fig. 1.2). As expected, our integrative approach brings together RBPs that form key regulatory modules and broadly recapitulates what is known about the functions of these RBPs; however, tracking the source of the signal to each of the input modalities is also often illuminating. One example is the group of splicing-related RBPs: eCLIP data suggests that these RBPs bind common targets (mostly snRNAs and mRNAs), whereas Perturb-seq data highlights a sub-group of RBPs presented by HNRNPK and DDX42 that are both involved in the regulation of p53-mediated apoptosis and are upregulated under p53 mutations (Escobar-Hoyos et al. 2020). Another example is the group of RBPs involved in mitochondrial and cytoplasmic translation. While eCLIP data brings together the rRNA-binding RBPs, BioID data clearly distinguishes mitochondrial RBPs (TBRG4, FASTKD2, and SUPV3L1) from the others.

The Integrated RBP Regulatory Map reveals numerous cases of combinatorial interactions and multi-functional roles for RBPs, a number of which have been studied previously. Visually, such combinatorial interactions appear as off-diagonal clusters (Fig. 1.2A); two examples are zoomed in on Fig. 1.2B. FUS, along with several other proteins, is associated with both mRNA splicing and ribosome biogenesis regulatory modules. Both regulatory functions have been shown for FUS experimentally (Fiesel and Kahle 2011; Moore 2016; Rogelj et al. 2012). FASTKD2, depending on the context, can play roles in regulation of mitochondrial mRNA translation and mitochondrial ribosomal biogenesis, respectively (Antonicka and Shoubridge 2015; Popow et al. 2015). Indeed, FASTKD2, along with several other proteins, is associated with both regulatory modules on the integrated map. These and many other examples highlight the combinatorial nature of RBP functioning.

In order to gain insight into the functions of individual RBPs, we analyzed the functions of their closest neighbors on the Integrated RBP Regulatory Map. We observed that the closest neighbors point to the known functions of RBPs (Fig. 1.2C). Thus, strong interactions of XRCC6 (Ku70) with DDX51, UTP18, and WDR43 suggest its role in ribosome biogenesis; this function has been recently proved experimentally (Shao et al. 2020). Interactions of PPIL4 with NPM1 and SAFB suggest that it might take part in transcriptional regulation. Consistent with this finding, PPIL4 was recently shown to interact with JMJD6, an annotated histone demethylase (Barak et al. 2021). The interactions of LIN28B with RPS3, METAP2 and DDX6 recapitulate LIN28B's known role in mRNA translation (Basak et al. 2020).

The use of complementary data types ensures that the Integrated RBP Regulatory Map captures RBP multifunctionality. To demonstrate it, we first analyzed the functions of RBP's closest neighbors within individual data types (Fig. 1.2C). eCLIP data suggests that among the closest neighbors of ZNF800 are chromatin remodeling proteins UCHL5 and GRWD1. On the other hand, the proximity labeling data suggests that ZNF800 interacts primarily with ribosome biogenesis factors, such as DDX21 and RPS11. The Integrated RBP Regulatory Map close neighborhood of ZNF800 consists of both chromatin remodeling factors and ribosome biogenesis factors (Fig. 1.2C), suggesting that the use of complementary data types allows for simultaneous capturing of multiple functions. TAF15 is another example of multifunctional protein. eCLIP data suggests that TAF15 is closely related to transcriptional regulators, such as FUS and EWSR1. Proximity labeling data suggests interactions of TAF15 and splicing machinery, such as SMNDC1 and GPKOW. Perturb-seq data suggests that among TAF15's close neighbors are translational regulators PABPC4 and EIF4G2. The TAF15's neighborhood of the Integrated RBP Regulatory Map contains all of the above proteins (Fig. 1.2C). Thus, the Integrated RBP Regulatory Map covers multiple independent functions of RBPs captured by complementary data sources.

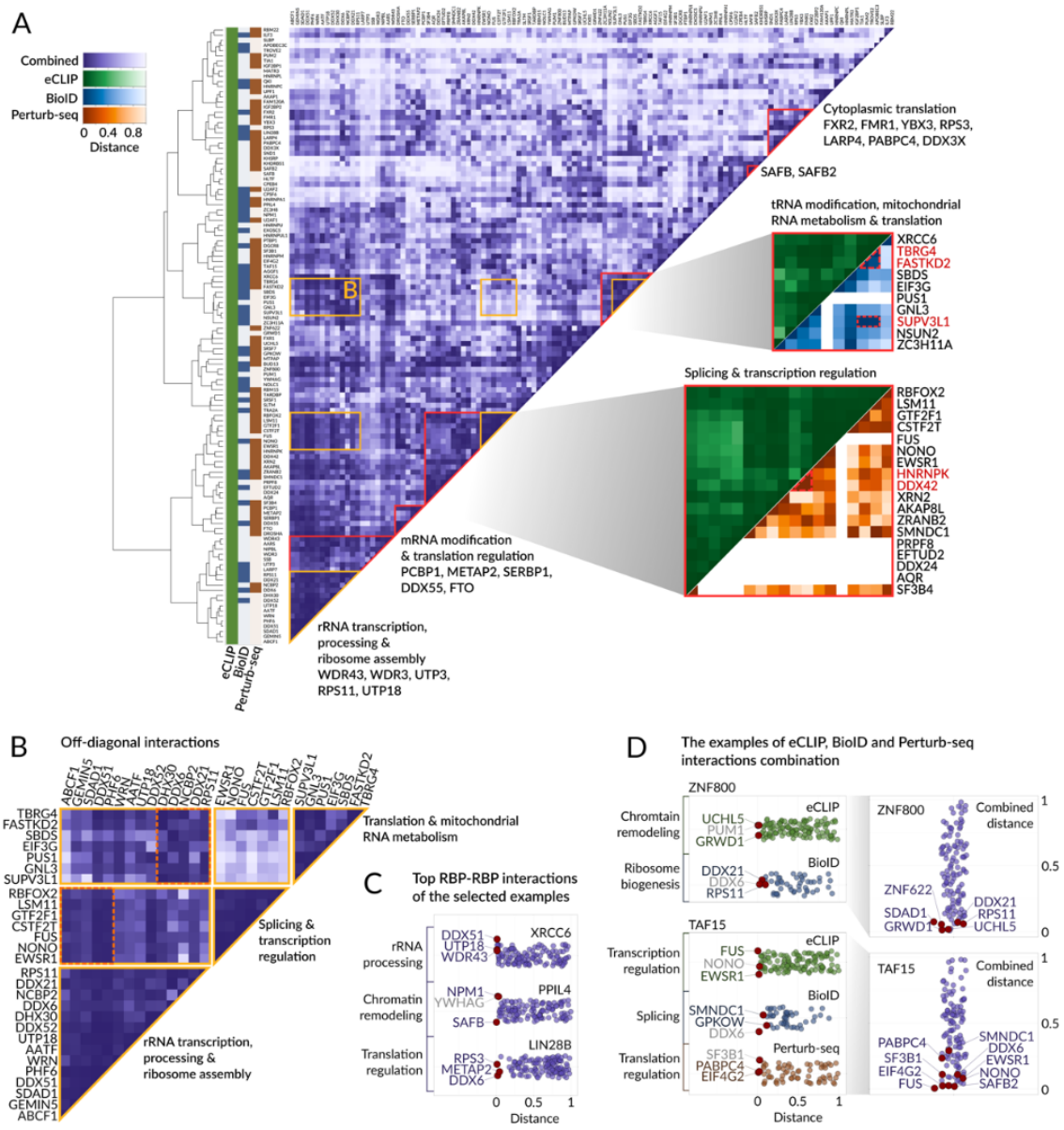

Figure 1.2. Post-transcriptional regulatory modules revealed by integrative analysis of RBP-RBP interactions.

(A) The heatmap of integrated distances between RBPs combining multiple sources of information for each RBP. Each row and column represents an RBP and the colored map on the left denotes the source datasets available and used for each RBPs. RBPs are clustered based on their combined distance using hierarchical clustering (dendrogram included). Known regulatory modules located on the diagonal that consist of previously annotated functional interactions are marked with red borders and labeled. In contrast, the examples of clusters located off the diagonal are highlighted in yellow and additionally illustrated in (B).

(B) The section of the integration heatmap containing off-diagonal clusters. Association of the rRNA maturation factors ABCF1, GEMIN5, SDAD1, DDX51 and PHF6 with splicing machinery

is highlighted in orange (GEMIN5 involved in splicing, other RBPs are involved in transcription regulation). DHX30, DDX6, NCBP2, DDX21 and RPS11 interaction with the regulators of translation and mitochondrial RNA metabolism is highlighted in orange as well (DHX30 is a mitochondrial ribosome assembly factor, other are also cytoplasmic and involved in translation).
(C) RBPs combined distances to the selected examples XRCC6, PPIL4 and LIN28B. Top 3 RBPs with the lowest distance are labeled for each example and their most studied functions are highlighted (RBPs not related to these functions are colored gray).
(D) RBPs distances to ZNF800 (top) and TAF15 (bottom) calculated based on eCLIP, BioID and Perturb-seq separately are illustrated on the left in the same way as in (C). Combined distances are illustrated on the right.

Defining functional RBP neighborhoods using BioID-mediated proximity labeling

Having defined the protein modules that each RBP falls into, we next sought to assign regulatory functions to each RBP module. The proximity labeling data allows us to go beyond the RBP-RBP interactions, and study the functions of both individual RBPs and RBP modules by analyzing their proximity to the other proteins in the cell. Here, we provide improved annotations for all studied RBPs. To accomplish this, for each RBP, we ranked the proximity-labeled "prey" proteins by their enrichment in the biotinylated fraction and used gene-set enrichment analysis (GSEA) to identify the top pathways and protein complexes(Fig. 1.3A). This procedure allows us to systematically estimate the likelihood of an RBP functioning in a given pathway across all "RBP-pathway" pairs of interest. Conceptually, these likelihood scores reflect the uncertainty of a given annotation, where higher scores denote higher confidence in the proposed association. We have visualized the high-confidence annotations in a heatmap along with the major RNA classes that, based on our eCLIP analysis, are the targets of each RBP in Fig. 1.3B (Fig. 1.3B).

Importantly, in many cases, the previously known functions of RBPs are clearly captured by our approach (Fig. 1.3). For example, we have correctly annotated SRSF7, NONO, and HNRNPA1 as splicing-related RBPs that bind predominantly mRNAs. Similarly, we identified RPS11, NPM1,

and DDX52 as RBPs that are involved in ribosome biogenesis and directly bind rRNA. Our BioID-based annotations clearly identifies RBPs that regulate transcription (HNRNPC, NPM1, QKI), initiate and regulate mRNA translation (LARP4, EIF3G, RPS3, LIN28B, YWHAG), participate in snRNA processing (TAF15, NPM1) and mitochondrial metabolism (SUPV3L1, FASTKD2, TBRG4), and modulate centrosome amplification (YWHAG) (Mukhopadhyay et al. 2016).

Our findings, however, go substantially beyond what is known and reveal novel and unexplored "non-canonical" functions. This further highlights the gaps in our existing knowledge of RBP annotations that can be systematically filled using the approach we have provided here. For example, SRSF7 is primarily known as a splicing factor; however, we observed an equally strong enrichment of mRNA 3' end processing and polyadenylation pathways, which are not yet annotated in GO but are alluded to in recent publications (Müller-McNicoll et al. 2016), (Schwich et al. 2021).

Overall, as shown in Fig. 1.3, we have revealed new functional associations for multitudes of RBPs across many regulatory pathways.

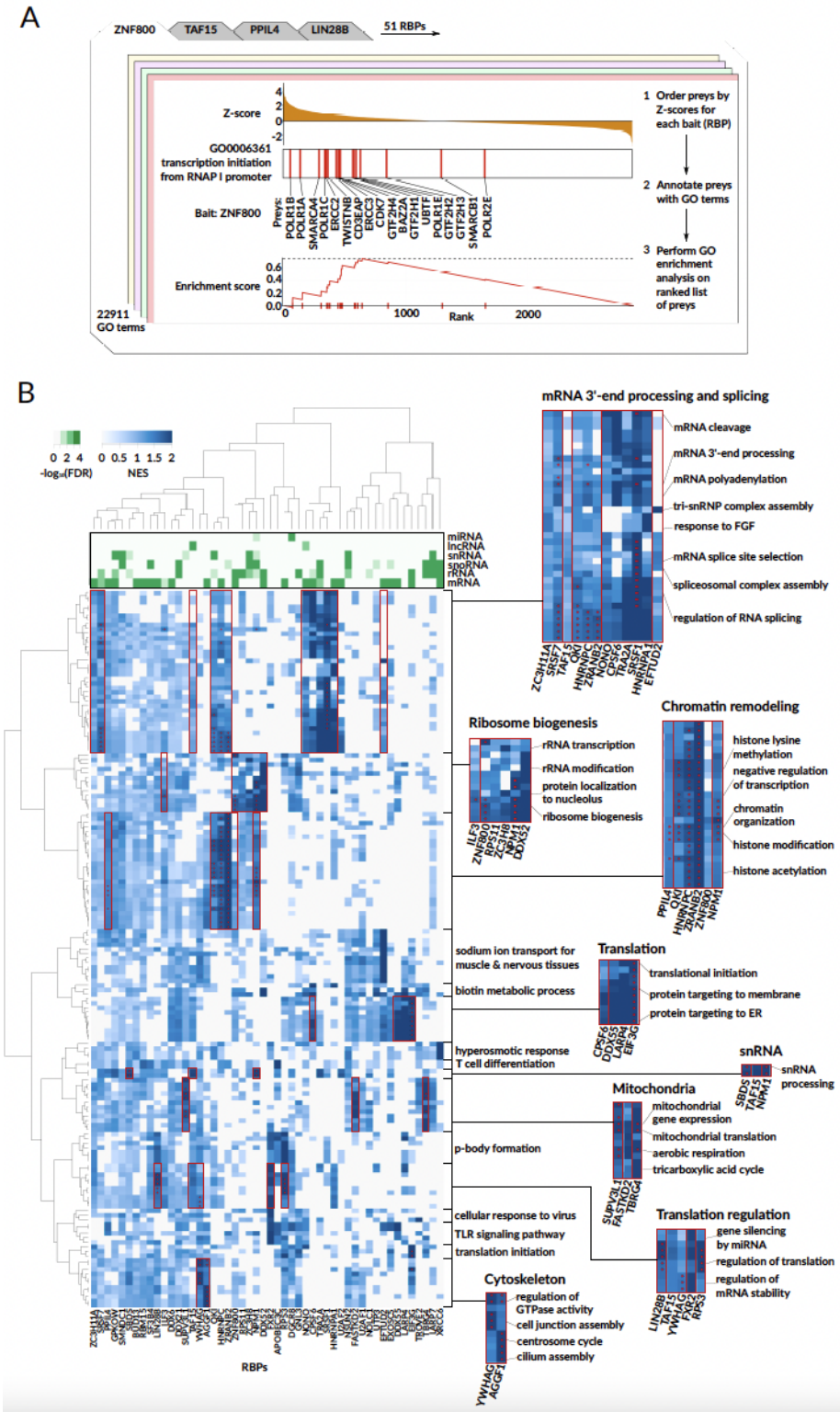
A
ZNF800
TAF15
PPIL4
LIN28B
51 RBPs
Z-score
GO0006361
transcription initiation
from RNAP I promoter
Bait: ZNF800
Preys:
POLR1B
POLR1A
SMARCA4
POLR1C
ERCC2
TWISTNB
CD3EAP
ERCC3
CDK7
GTF2H4
BAZ2A
GTF2H1
UBTF
POLR1E
GTF2H2
GTF2H3
SMARCB1
POLR2E
Enrichment score
Rank
22911
GO terms
1 Order preys by Z-scores for each bait (RBP)
2 Annotate preys with GO terms
3 Perform GO enrichment analysis on ranked list of preys
B
-log10(FDR)
NES
miRNA
lncRNA
snRNA
snoRNA
rRNA
mRNA
mRNA 3'-end processing and splicing
mRNA cleavage
mRNA 3'-end processing
mRNA polyadenylation
tri-snRNP complex assembly
response to FGF
mRNA splice site selection
spliceosomal complex assembly
regulation of RNA splicing
Ribosome biogenesis
rRNA transcription
rRNA modification
protein localization to nucleolus
ribosome biogenesis
Chromatin remodeling
histone lysine methylation
negative regulation of transcription
chromatin organization
histone modification
histone acetylation
sodium ion transport for muscle & nervous tissues
biotin metabolic process
Translation
translational initiation
protein targeting to membrane
protein targeting to ER
hyperosmotic response
T cell differentiation
snRNA
snRNA processing
Mitochondria
mitochondrial gene expression
mitochondrial translation
aerobic respiration
tricarboxylic acid cycle
p-body formation
cellular response to virus
TLR signaling pathway
translation initiation
Translation regulation
gene silencing by miRNA
regulation of translation
regulation of mRNA stability
Cytoskeleton
regulation of GTPase activity
cell junction assembly
centrosome cycle
cilium assembly
RBPs

Figure 1.3. BioID-mediated proximity labeling defines RBP neighborhoods and enables functional annotation of RBPs.

(A) Schematic of our RBP pathway annotations using ZNF800 and GO:0006361 (transcription initiation from RNA polymerase I promoter) as an example. Co-precipitated proteins are ranked by their BioID protein neighborhood *z*-scores and gene-set enrichment analysis is performed on the resulting ranked list.
(B) RBP preference to bind particular RNA types (based upon eCLIP RNA targets) is shown along with GO term enrichment in RBP neighborhoods. Columns correspond to the RBPs we have assayed, rows correspond to individual gene ontology terms (BP), and the color denotes the GSEA normalized enrichment score (NES). GO terms with NES more than 2 for at least one RBP are included in the plot, a red asterisk indicates FDR < 0.05. Some known RBPs functions are highlighted by red boxes and additionally placed to subpanels.

ZC3H11A and TAF15 are involved in multiple post-transcriptional processes including splicing, translation, and RNA stability regulation

Our data highlights ZC3H11A and TAF15 as multifunctional RBPs involved in multiple post-transcriptional processes for distinct RNA regulons. ZC3H11A is known to be a part of the TREX complex responsible for mRNA export and was shown to co-localize with SRSF2 in nuclear splicing speckles (Folco et al. 2012; Younis et al. 2018). TAF15 is predominantly studied as a part of TFIID and RNAPII complexes participating in the regulation of gene expression (Jobert et al. 2009). Yet, TAF15's role in post-transcriptional regulation is still poorly understood. Several studies have shown TAF15 involvement in stability and processing of lncRNA *LINC00665*, *FGFR4*, *GRIN1* mRNAs, and a small subset of other mRNAs in neurons (Ruan et al. 2020; DeJong et al. 2021; Ibrahim et al. 2013; Kapeli et al. 2016). Additionally, it has been shown that TAF15 participates in miRNA-mediated regulation of cell cycle gene expression and its direct role in mRNA transport and translation has been suggested according to its CLIP signal enrichment in 3' UTRs (Ballarino et al. 2013; Kapeli et al. 2016).

Our integrated RBP interaction map suggests that both ZC3H11A and TAF15 are involved in a much wider set of post-transcriptional regulatory processes (Fig. 1.2). In particular, TAF15 has the highest interaction scores with FUS, SAFB2, EIF4G2, NONO, and SAFB, which in addition to transcription are also associated with translation (EIF4G2), and splicing (NONO). On the other hand, ZC3H11A's top interacting partners include GPKOW and DHX30, suggesting splicing-related functions. Gene set enrichment analysis of the BioID data resulted in the highest-scoring GO terms related to mRNA export (for ZC3H11A) and transcription (for TAF15), as expected. However, we noted additional high-scoring terms, including "spliceosomal snRNP assembly" for both RBPs (GO:0000387, NES = 1.5) as well as "mRNA stabilization" (GO:0048255, NES = 1.5) and "positive regulation of translation" (GO:0045727, NES = 1.7) for TAF15 (Fig. 1.4A).

To verify the predicted ZC3H11A and TAF15 roles in splicing regulation, we used CRISPR-interference technology to precisely knockdown these RBPs in K562 cells and performed paired-end total RNA-seq in respective knockdown cells (96% and 98% knockdown, respectively). Upon silencing either of these genes, we observed a large number of significant alternative splicing events (ASEs; Fig. 1.4B, 4C). These include 171 and 94 ASEs passing log10(Bayes factor) > 2 and absolute effect size > 0.2 for ZC3H11A and TAF15 knockdown, respectively. We independently validated multiple of these significant ASEs using quantitative RT-PCR (Fig. 1.4 B,C); therefore confirming the involvement of ZC3H11A and TAF15 involvement in the regulation of alternative splicing.

As mentioned above, in addition to RNA processing and splicing, we also observed a significant association between TAF15 and translational control. To confirm this association, we performed ribosome footprinting (Ribo-seq) and matched RNA sequencing in control and TAF15 knockdown cells. Ribosome footprinting, which captures active translation, allowed us to measure changes in

translational efficiency (TE), defined as ratio of ribosome protected fragments to total RNA. We then asked whether TAF15-bound mRNAs are differentially translated when this RBP is silenced. Consistent with a direct role for TAF15 in translational control, TAF15-bound RNA targets, defined based on eCLIP data of this RBP (Van Nostrand et al. 2016), were strongly enriched among mRNAs that translationally replaced in TAF15 knockdown cells (Fig. 1.4D). To independently confirm this observation, we used protein abundance data collected from TAF15 KD and control cell lines using quantitative mass-spectrometry. As expected, TAF15 targets showed a significant change in their protein abundance without a concomitant change in their mRNA levels (highlighted in Fig. 1.4D). Taken together, our findings demonstrate that TAF15 acts as a direct enhancer of mRNA translation for its target regulon.

In addition to translational control, we had observed a strong association between TAF15 and regulators of RNA stability. To further explore this association, we performed RNA metabolic labeling (SLAM-seq) to deconvolve changes in transcription rate and decay observed in TAF15 knockdown cells. We hypothesized that TAF15 performs its stability-related function through direct RBP-RNA interaction compared to the transcription regulation mediated through the interactions with RNA PolII (Bertolotti et al. 1996). Therefore, we analyzed the changes in transcription rate and total mRNA amount among the TAF15 target mRNAs with iPAGE (Goodarzi et al. 2009) and GSEA (Korotkevich et al. 2021). Both tools demonstrated the enrichment of TAF15-bound RNA targets among downregulated mRNAs (based on the total RNA abundance). However, consistent with our hypothesis that the TAF15 regulon is being destabilized when this RBP is silenced, we did not see a change in the transcription rates of TAF15 targets based on SLAM-seq data (Fig. 1.4E). We independently verified these observations for multiple

TAF15 targets using qRT-PCR, namely HNRNPH2, GADD45A, and SNHG7 (Fig. 1.4E). These observations confirm the involvement of TAF15 in regulating mRNA stability.

Taken together, our data supports the hypothesis of TAF15 being a multifunctional regulator. We demonstrate that it plays multiple roles, each of them affecting its own regulon (set of target RNAs). The regulons don't show significant overlaps, suggesting that TAF15's multiple functions are independent of each other. As such, TAF15 controls splicing of 1894 mRNAs, translation of 1138 mRNAs, and stability of 140 mRNAs; only 96 of these mRNAs are affected by TAF15 in 2 different ways (Fig. 1.4F). The case of TAF15 illustrates the nature of multi-functionality of RBPs.

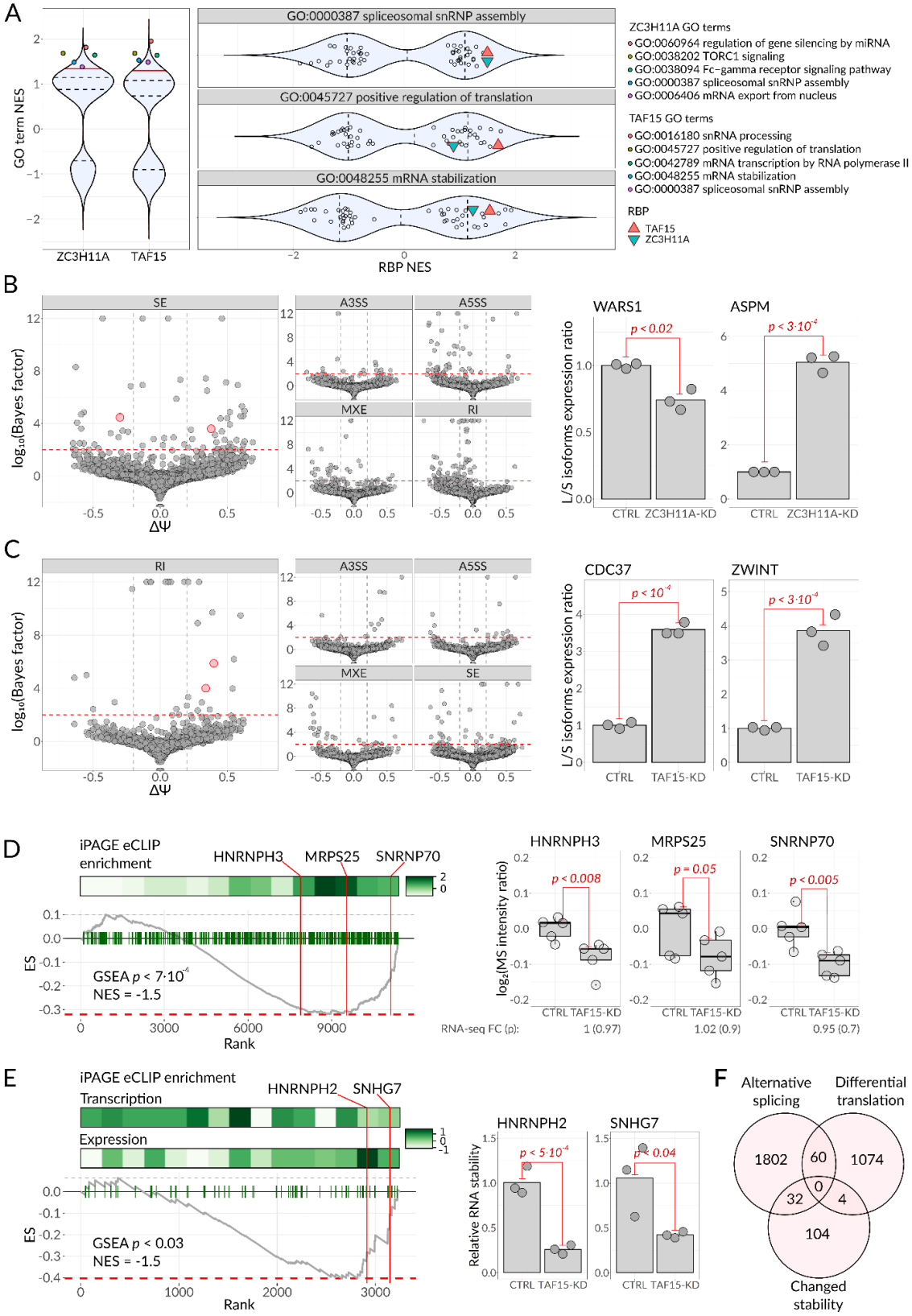
A
GO term NES
ZC3H11A
TAF15
GO:0000387 spliceosomal snRNP assembly
GO:0045727 positive regulation of translation
GO:0048255 mRNA stabilization
RBP NES
ZC3H11A GO terms
GO:0060964 regulation of gene silencing by miRNA
GO:0038202 TORC1 signaling
GO:0038094 Fc−gamma receptor signaling pathway
GO:0000387 spliceosomal snRNP assembly
GO:0006406 mRNA export from nucleus
TAF15 GO terms
GO:0016180 snRNA processing
GO:0045727 positive regulation of translation
GO:0042789 mRNA transcription by RNA polymerase II
GO:0048255 mRNA stabilization
GO:0000387 spliceosomal snRNP assembly
RBP
TAF15
ZC3H11A
B
SE
A3SS
A5SS
MXE
RI
log10(Bayes factor)
ΔΨ
WARS1
ASPM
p < 0.02
p < 3·10-4
L/S isoforms expression ratio
CTRL ZC3H11A-KD
C
RI
A3SS
A5SS
MXE
SE
CDC37
ZWINT
p < 10-4
p < 3·10-4
CTRL TAF15-KD
D
iPAGE eCLIP enrichment
HNRNPH3
MRPS25
SNRNP70
GSEA p < 7·10-4
NES = -1.5
ES
Rank
log2(MS intensity ratio)
p < 0.008
p = 0.05
p < 0.005
RNA-seq FC (p):
1 (0.97)
1.02 (0.9)
0.95 (0.7)
E
iPAGE eCLIP enrichment
Transcription
Expression
HNRNPH2
SNHG7
GSEA p < 0.03
NES = -1.5
Relative RNA stability
p < 5·10-4
p < 0.04
F
Alternative splicing
Differential translation
Changed stability
1802
60
1074
0
32
4
104

Figure 1.4. Non-canonical functions of ZC3H11A and TAF15 predicted from our integrated analyses.

(A) Violin plots illustrating normalized enrichment scores (NES) resulting from gene set enrichment analysis. Left subpanel: all GO-BP terms for ZC3H11A and TAF15; right subpanel: all tested RBPs for GO:0000387, GO:0045727 and GO:0048255 GO terms. Dashed lines: quartiles; solid red line: 0.9 quantile.
(B) Results of ZC3H11A-KD RNA-seq analysis with MISO (left) and qPCR verification (right). Two skipped exon (SE) cases for WARS1 and ASPM transcripts were confirmed by qPCR and labeled in the scatter plot. Dashed lines indicate the following filters: log10(Bayes factor) > 2 and |isoforms levels difference| > 0.2.
(C) Results of TAF15-KD RNA-seq analysis with MISO (left) and qPCR verification (right), as in (B).
(D) Ribo-seq (on the left) and mass spectrometry (on the right) of TAF15-KD and control cells confirm TAF15 involvement in translation control. We assessed the enrichment and/or depletion patterns of RNAs with exonic TAF15 binding across all genes ranked by their changes in translation efficiency upon TAF15 knockdown. Both iPAGE (top) and GSEA (bottom) analytical pipelines were used for this analysis. For three of TAF15 targets (see Figure S4), we have included protein levels from quantitative mass-spectrometry, as well as changes in their mRNA abundance based on RNA-seq. RNA-seq comparisons were performed using edgeR.
(E) SLAM-seq (on the left) and qPCR (on the right) of TAF15-KD confirm its involvement in mRNA stabilization. Enrichments of the TAF15 eCLIP targets among RNAs with the total expression (EXPR) and transcription rate (TR) upon TAF15 knockdown were assessed by iPAGE (top) and GSEA (bottom, only for EXPR) using SLAM-seq logFC for ranking. Only EXPR is significantly enriched by eCLIP suggesting TAF15 RNA stability regulation via direct binding. Three exonic TAF15 eCLIP targets (see Figure S4) were assessed further by qPCR which revealed a significant reduction of their mature RNA levels compared to pre-mRNA upon knockdown.

RNA-binding proteins QKI and ZNF800 are involved in the regulation of transcription

Our integrated RBP interaction map revealed several RNA-binding proteins that are strongly associated with transcriptional control, chief among them we noted ZNF800 and QKI. ZNF800 is a zinc finger protein whose molecular functions are poorly studied, yet it's implicated in diseases such as lung cancer (Zhuo et al. 2020). In contrast, QKI is a well-studied RBP involved in many RNA-related processes including regulation of mRNA expression, splicing, and stability, nuclear export, and translation. This protein plays a major part in neuron myelination and its functions are mostly described for neural cells (Chen et al. 2021; Chen et al. 2021; Zhou et al. 2021; Shin et al. 2021; Åberg et al. 2006).

We hypothesized that ZNF800 and QKI act as transcriptional regulators and we validated this hypothesis with ATACseq. Based on our proximity labeling results, ZNF800's protein neighborhood functions in DNA methylation, transcription by RNA polymerase I, rRNA processing, and chromatin remodeling (Fig. 1.5A). On the other hand, QKI's neighborhood is associated with histone methylation, RNA splicing, transcription by RNA polymerase II and chromatin organization. To validate previously unknown role for ZNF800 in chromatin remodeling and confirm recently discovered chromatin-associated QKI functions, we performed ATAC-seq using control K562 and CRISPRi-generated knockdown cells (87% and 76% knockdown, respectively) (Ren et al. 2021). We observed a significant and global increase in chromatin accessibility across thousands of regions when these RBPs were silenced (2660/2724 significantly differential regions were upregulated for QKI knockdown, and 1399/1417 regions were up-regulated in ZNF800 knockdown). Therefore, both ZNF800 and QKI are involved in transcriptional repression.

To demonstrate that ZNF800 and QKI regulate transcription through direct binding of DNA promoters, we performed ChIP-qPCR on several gene targets. Namely, we tested binding of ZNF800 to the DNA sequences of *RPS15* and *RPL10A* genes, and binding of QKI to the DNA sequences of *PRC1* and *LTBR* genes. The gene target sequences were significantly enriched in ChIP samples compared to control samples (Fig. 1.5E). These results suggest ZNF800 and QKI direct binding of the promoter regions of their target RNA genes.

To study what processes ZNF800 and QKI affect post-transcriptionally, we compared the representation of different RNA types among their targets using Fisher's exact test (see Methods). We observed that ZNF800 binds primarily to mRNA and rRNA targets (Fig. 1.5C). The rRNA binding suggests that ZNF800 might participate in ribosome biogenesis; this hypothesis is

consistent with the ZNF800 functions predicted from the proximity labeling data (Fig. 1.5A). As for QKI, we observe significant enrichment of snRNAs among its RNA targets (Fig. 1.5C); this is consistent with QKI's involvement in splicing, which has been studied experimentally in the past (Zong et al. 2014). We also observe enrichment of splicing factors among the proteins that interact with it spatially, according to the proximity labeling data (Fig. 1.5A). Thus, our data suggests that ZNF800 regulates the ribosome biogenesis pathway post-transcriptionally, while QKI is involved in alternative splicing regulation.

Next, we aimed to study the relationship between the transcriptional and post-transcriptional regulons of both ZNF800 and QKI. We defined the transcriptional regulon as the set of genes whose promoter regions become significantly more accessible upon the RBP's knockdown (based on ATACseq data), and we defined the post-transcriptional regulon as the set of genes whose mRNAs bind the RBP directly (based on eCLIP data). First, we used the hypergeometric test to inquire if the transcriptional and post-transcriptional regulons overlap for either ZNF800 or QKI at the level of individual genes. We did not observe significant overlap for QKI. For ZNF800, the overlap was statistically significant ($p=0.01$), yet very small (6.4% of the DNA targets and 9.7% of the RNA targets; odds ratio 1.4) (Fig. 1.5C). We then performed gene set enrichment analysis on the transcriptional and post-transcriptional regulons; we observed almost no pathways that would be significantly enriched among the two regulons neither for ZNF800, nor for QKI. Thus, we hypothesize independent involvement of the same RBPs in transcriptional and post-transcriptional regulation.

Taken together, we demonstrate the direct involvement of QKI and ZNF800 in multiple regulatory processes. Each function affects a specific group of genes, and these groups don't show major

overlaps between different functions of the same RBP. Thus, we can consider these RBPs multifunctional entities that control several regulons in an independent way.

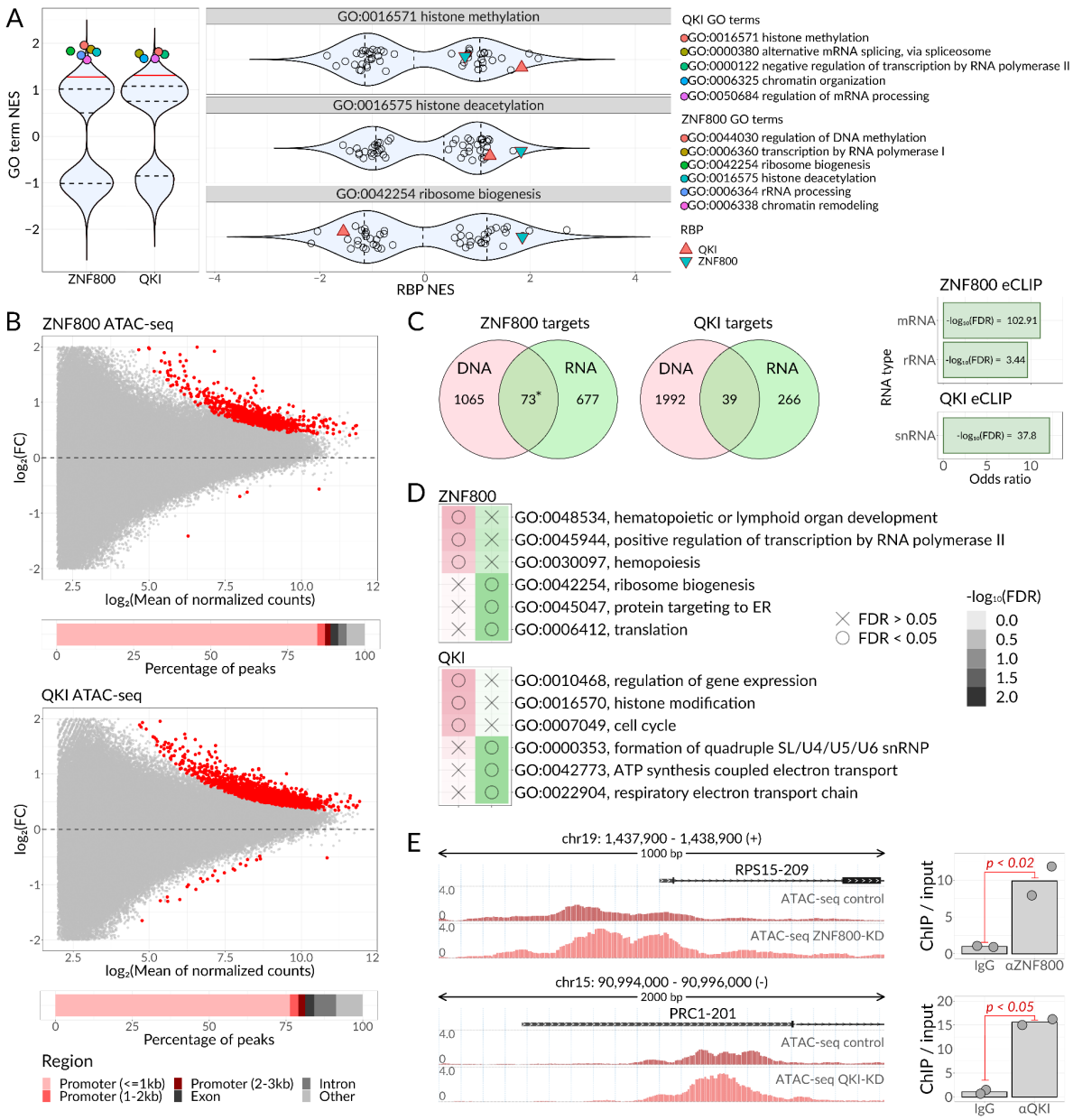

Figure 1.5. ATAC-seq in knockdown cells confirms ZNF800 and QKI involvement in the regulation of transcription.

(A) Violin plots illustrating normalized enrichment scores (NES) resulting from gene set enrichment analysis. Left subpanel: all GO-BP terms for ZNF800 and QKI; right subpanel: all tested RBPs for GO:0016571, GO:0016575 and GO:0042254 GO terms. Dashed lines: quartiles; solid red line: 0.9 quantile.

(B) The results of differential ATAC-seq analysis in control and ZNF800-KD/QKI-KD cells. Each point denotes an ATAC-seq peak; peaks passing 0.05 FDR are colored red.
(C) ZNF800 and QKI DNA and RNA targets based upon eCLIP and differential ATAC-seq data. Preferably bound RNA types (FDR < 0.01) of ZNF800 and QKI are shown on the left. Genes detectable in eCLIP and unambiguously mapped to Entrez IDs were used to plot Venn diagrams (right subpanel).
(D) Extract from GO-BP enrichment analysis of ZNF800 and QKI targets bound at chromatin (DNA, ATAC-seq) and transcript (RNA, eCLIP) levels.
(E) Illustrative examples of genes with promoter regions bound by RBP (ZNF800 or QKI). Genomic loci with ATAC-seq tracks are shown in the left subpanel (Y-axis: counts per million). Results of the ChIP-qPCR verification of the corresponding ATAC-seq peaks are shown in the right subpanel.

DISCUSSION

In the post-genomic era, we can reliably quantify modulations in the RNA dynamics, from processing to translation to decay, across the transcriptome. However, systematic identification of the underlying molecular mechanisms that give rise to these modulations remains a major challenge in systems biology. The regulon hypothesis, first formulated for transcriptional control, has been extended to the post-transcriptional realm to reveal shared regulatory pathways that target functionally related RNA molecules in a coordinated manner. Despite many advances in the field, our knowledge of RNA regulons remains limited. We have long argued that a comprehensive map of regulator-regulon is a crucial step towards creating mechanistic and predictive models of gene expression control in health and disease (Fish et al. 2019; Fish et al. 2021). In this study, we proposed an integrative framework for decoding context-dependent and combinatorial regulatory interactions between RNA-binding proteins and their target regulons.

Many research initiatives are directed towards mapping the regulons as units of gene expression control. The ENCODE consortium undertook a vast effort to map protein-DNA and protein-RNA interactions through large-scale application of ChIP-seq and eCLIP, respectively. Similar data modalities have been generated by other groups in the field as well, which together provide a rich resource for generating biological networks focused on nucleic acid interactions. However, not every interaction is functional and not every function can be generalized to all interactions. This is especially true of RNA-binding proteins as they often bind thousands of RNAs, and their regulatory functions across these targets is often multifaceted and context-dependent. Over the past few years, this understanding has led us to revisit the notion of RNA-binding proteins as units of translational control. Instead, we view “regulatory modules” as collections of RNA-binding

proteins that come together for a specific function with its own target regulon. In this retelling of post-transcriptional control, the same RNA-binding protein can participate in multiple modules and carry out separate and even divergent functions for independent regulon. This many-to-many relationship between RNA-binding proteins and regulatory functions necessitates additional layers of functional annotations that we currently lack. This is the core of the problem we have sought to tackle in this study.

Our framework provides a generalizable strategy for transitioning from descriptive maps of RNA-protein interactions to mechanistic models of regulatory modules. Focus on the interactions between regulons is well suited for studying highly context-dependent regulatory networks, such as post-transcriptional gene expression regulation. This provides a complementary approach to efforts based on explanation of an observed phenotype, e.g CRISPRi screens and gene set enrichment analysis methods. By careful and robust interpretation of regulon relationships and functions, one can start to create a comprehensive model of gene expression regulation in the cell. Such approaches should eventually enable engineering of the cellular response to external factors.

MATERIALS AND METHODS

Cell culture

All cells were cultured in a 37°C 5% CO2 humidified incubator. The 293T cells (ATCC CRL-3216) were cultured in DMEM high-glucose medium supplemented with 10% FBS, glucose (4.5 g/L), L-glutamine (4 mM), sodium pyruvate (1 mM), penicillin (100 units/mL), streptomycin (100 μg/mL) and amphotericin B (1 μg/mL) (Gibco). The K562 cell line was cultured in RPMI-1640 medium supplemented with 10% FBS, glucose (2 g/L), L-glutamine (2 mM), 25 mM HEPES,

penicillin (100 units/mL), streptomycin (100 μg/mL) and amphotericin B (1 μg/mL) (Gibco). All cell lines were routinely screened for mycoplasma with a PCR-based assay.

Western blotting

Cell lysates were prepared by lysing cells in ice-cold RIPA buffer (25 mM Tris-HCl pH 7.6, 0.15 M NaCl, 1% IGEPAL CA-630, 1% sodium deoxycholate, 0.1% SDS) containing 1X protease inhibitors (Thermo Scientific). Lysate was cleared by centrifugation at 20,000 x g for 10 min at 4°C. Samples were denatured for 10 min at 70°C in 1X LDS loading buffer (Invitrogen) and 50 mM DTT. Proteins were separated by SDS-PAGE using 4-12% Bis-Tris NuPAGE gels, transferred to nitrocellulose (Millipore), blocked using 5% BSA, and probed using target-specific antibodies. Bound antibodies were detected using dye-conjugated secondary antibodies (Licor) according to the manufacturer's instructions. Antibodies: HA (BioLegend 901533), eIF3I (BioLegend 646701), beta-tubulin (Proteintech 10094-1-AP), GAPDH (Proteintech 10494-1-AP).

RNA isolation

Total RNA for RNA-seq and RT-qPCR was isolated using the Zymo QuickRNA isolation kit with in-column DNase treatment per the manufacturer's protocol.

RNA-seq

RNA-seq libraries were prepared using SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian (Takara).

CRISPRi-mediated gene knockdown

K562 cells expressing dCas9-KRAB fusion protein were constructed by lentiviral delivery of pMH0006 (Addgene #135448) and FACS isolation of BFP-positive cells.

The lentiviral constructs were co-transfected with pCMV-dR8.91 and pMD2.D plasmids using TransIT-Lenti (Mirus) into 293T cells, following the manufacturer's protocol. The virus was harvested 48 hours post-transfection and passed through a 0.45 µm filter. Target cells were then transduced overnight with the filtered virus in the presence of 8 µg/mL polybrene (Millipore).

Guide RNA sequences for CRISPRi-mediated gene knockdown were cloned into pCRISPRia-v2 (Addgene #84832) via BstXI-BlpI sites. After transduction with sgRNA lentivirus, K562 cells were selected with 20 µg/mL blasticidin (Gibco). Knockdown of target genes was assessed by RT-qPCR as described below.

BioID2-RBP fusion cell line generation

In order to construct the cell lines stably expressing BioID2-RBP fusion proteins, we first cloned in an open reading frame of BioID2 enzyme (Kim et al. 2016), followed by a linker and Gateway LR Clonase cloning site, into the pWPI backbone (Addgene #12254). We then used Gateway LR Clonase II Enzyme mix (Thermo Fisher) to clone the open reading frames of the RBPs of interest (from ORFeome cDNA library) into the destination vector. The lentiviral constructs were co-transfected with pCMV-dR8.91 and pMD2.D plasmids using NanoFect (ALSTEM) into 293T cells, following manufacturer's protocol. The virus was harvested 48 hours post-transfection and passed through a 0.45 µm filter. K562 cells were then transduced for 2 hours while centrifuging (800 RPM) with the filtered virus in the presence of 8 µg/mL polybrene (Millipore). Cells were

selected with 2 µg/mL puromycin for 3 days (Gibco). The expression of the fusion protein was validated by western blotting.

Biotin treatment and pulldown

The pulldown was performed as described in (Kim et al. 2016). Cells were incubated with biotin-depleted media (biotin-free RPMI-1640 medium, supplemented with 10% dialyzed FBS, glucose (2 g/L), L-glutamine (2 mM), 25 mM HEPES, penicillin (100 units/mL), streptomycin (100 µg/mL) and amphotericin B (1 µg/mL) for 72 h before analysis. For BioID2 pulldown, 12×10^6 cells per replicate were incubated with 50 µM biotin for 24 h. For the negative control samples, 12×10^6 cells per replicate were incubated with DMSO. After three times of PBS wash, the cells were lysed in 1 ml of lysis buffer containing 50 mM Tris, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 1% Sodium deoxycholate, 0.1% SDS, 1 × Complete protease inhibitor (Halt Phosphatase Inhibitor Cocktail; Life Technologies), and 250 units benzonase (EMD millipore). The lysates were passed through a 25G needle 10 times and cleared 10 min at 14,000 g at +4°C. The protein concentration was measured with BCA Protein Assay Kit (Thermo Scientific); the lysate was diluted to a concentration of 2 µg/mL. 500 µl of lysate was incubated with 125 µl of Dynabeads (MyOne Streptavidin C1; Life Technologies) overnight with rotation at +4°C. Beads were collected using a magnetic stand and washed with twice with 2% (wt/vol) SDS, twice with wash buffer containing 50 mM Tris, pH 7.5, 500 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% SDS, twice with wash buffer containing 50 mM Tris, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% SDS, then boiled for 5 min in 50 µl of elution buffer containing 2% SDS, 100mM DTT, Tris-HCl pH 7.5. The supernatant was collected and saved for mass spectrometry analysis.

Mass spectrometry analysis

Eluted BioID samples were reduced by the addition of 100 mM DTT and boiling at 95˚C for 10 min, before being subjected to Filter Aided Sample Preparation (FASP) (Wisniewski et al., Anal. Biochem., 2011) to generate tryptic peptides, as described previously (Dermit et al, Dev Cell, 2020). Briefly, samples were diluted 7 fold in UA buffer (8M urea, 100 mM Tris HCl pH 8.5), transferred to to Vivacon 500 Hydrosart centrifugal filters with a molecular cut-off of 30kDa (Sartorius), and concentrated by centrifugation at 14,000 g for 15 min. Filters were then washed twice by addition of 0.2 mL of UA buffer to the filter tops and re-concentrating. Reduced cysteine residues were then alkylated by addition of 100µL of 50 mM iodoacetamide dissolved in UA buffer, and incubation at room temperature in the dark for 30 min. The iodoacetamide solution was then removed by centrifugation at 14,000 g for 10 min, and samples were washed twice with 0.2 mL of UA buffer as before. Urea was then removed from samples by performing three washes with 0.2 mL of ABC buffer (0.04 M ammonium bicarbonate). Filters were then transferred to fresh collection tubes, and proteins were digested by addition of 0.3 µg of MS grade Trypsin (Sigma-Aldrich) dissolved in 50 µL of ABC buffer, and overnight incubation in a thermo-mixer at 37˚C with gentle shaking (600 rpm). The resulting peptides were eluted from the filters by centrifugation at 14,000 g for 10 min. Residual remaining peptides were further eluted by addition of 100 µL ABC to the filter tops and centrifugation. This was repeated once and the combined eluates were then dried by vacuum centrifugation (no heating) and reconstitution in 2% Acetonitrile (ACN), 0.2% Trifluoroacetic acid (TFA), followed by desalting using C18 StageTips (Rappsilber, et al., Nat Protoc. 2007). The desalted peptides were dried again by vacuum centrifugation (45˚C) and re-suspended in A* buffer (2%ACN, 0.5% Acetic

acid, 0.1% TFA in water) before LC-MS/MS analysis. 1/3rd of each sample was analyzed on a Q-Exactive plus Orbitrap mass spectrometer coupled with a nanoflow ultimate 3000 RSL nano HPLC platform (Thermo Fisher). Samples were resolved at a flow rate of 250 nL/min on an Easy-Spray 50 cm x 75 µm RSLC C18 column with 2 µm particle size (Thermo Fisher), using a 123 minutes gradient of 3% to 35% of buffer-B (0.1% formic acid in ACN) against buffer-A (0.1% formic acid in water), and the separated peptides were infused into the mass spectrometer by electrospray (1.95 kV spray voltage, 255˚C capillary temperature). The mass spectrometer was operated in data dependent positive mode, with 1 MS scan followed by 15 MS/MS scans (top 15 method). The scans were acquired in the mass analyser at 375-1500 m/z range, with a resolution of 70,000 for the MS and 17,500 for the MS/MS scans. A 30 second dynamic exclusion of fragmented peaks was applied to limit repeated fragmentation of the same ions.

Perturb-seq

Perturb-seq experiment was performed as previously described (Datlinger et al. 2017). Briefly, a library of 205 sgRNAs (5 non-targeting sgRNAs and 200 sgRNAs targeting 100 genes, 2 sgRNAs per gene) was ordered as a pooled oligonucleotide library from Twist Bioscience with the following design:

[ATCTTGTGGAAAGGACGAAACACCG]-[Protospacer Sequence]-[GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC]

The library was PCR-amplified using Q5 Hot Start High-Fidelity 2X Master Mix (NEB) with the primers with the following sequences: 5'-ATCTTGTGGAAAGGAC-3' and 5'-GCCTTATTTTAACTTGCTA-3'. To clone libraries for this vector or CROPseq-Guide-Puro vector (Addgene 86708), the starting vector was digested with BsmBI following the protocol

outlined in (Sanjana et al. 2014). The library was cloned into the digested backbone using Gibson Assembly method (Thomas et al. 2015). The reaction product was transformed into Takara Stellar competent cells according to manufacturer recommendations, grown overnight in 100 mL LB and purified using ZymoPURE II Plasmid Midiprep Kit. K562 cells were infected with the plasmid library at a low multiplicity of infection to minimize double infection. The infected cells were selected with 2 µg/mL puromycin for 3 days. Live cells were isolated on flow cytometer (FACSAria II) by propidium iodide staining. Approximately 5000 live cells were captured by 10X Chromium Controller using Chromium Single Cell 3' Reagent Kits v2. Sample preparation was performed according to the manufacturer's protocol. Samples were sequenced on a NovaSeq 6000 using the following configuration: Read 1: 28, i7 index: 8, i5 index: 0, Read 2: 98.

To facilitate sgRNA assignment, sgRNA-containing transcripts were additionally amplified by PCR reactions by modifying a previously published approach (Hill et al. 2018). The following primers were used for amplification: 5'-AATGATACGGCGACCACCGAGATCTACAC-3' and 5'-CAAGCAGAAGACGGCATACGAGATTACGACAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTggactatcatatgcttaccgtaacttgaaag-3'. PCR product was cleaned up by 1.0x SPRI beads (SPRIselect; BECKMAN COULTER; Cat. No. B23317). Samples were sequenced using paired-end 150 bp sequencing on an Illumina MiSeq sequencer.

Ribosome profiling

Ribosome profiling was performed as previously described (McGlincy and Ingolia 2017). Briefly, approximately $10x10^6$ cells were lysed in ice-cold polysome buffer (20 mM Tris pH 7.6, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 100 µg/mL cycloheximide) supplemented with 1% v/v Triton

X-100 and 25 U/mL Turbo DNase (Invitrogen). The lysates were triturated through a 27G needle and cleared for 10 min at 21,000 x g at 4°C. The RNA concentration in the lysates was determined with the Qubit RNA HS kit (Thermo). Lysate corresponding to 30 µg RNA was diluted to 200 µl in polysome buffer and digested with 1.5 µl RNaseI (Epicentre) for 45 min at room temperature. The RNaseI was then quenched by 10 µl SUPERaseIN (Thermo).

Monosomes were isolated using MicroSpin S-400 HR (Cytiva) columns, pre-equilibrated with 3 mL polysome buffer per column. 100 µl digested lysate was loaded per column (two columns were used per 200 µl sample) and centrifuged 2 min at 600 x g. The RNA from the flow-through was isolated using the Zymo RNA Clean and Concentrator-25 kit. In parallel, total RNA from undigested lysates was isolated using the same kit.

Ribosome-protected footprints (RPFs) were gel-purified from 15% TBE-Urea gels as 17-35 nt fragments. RPFs were then end-repaired using T4 PNK (NEB) and pre-adenylated barcoded linkers were ligated to the RPFs using T4 Rnl2(tr) K227Q (NEB). Unligated linkers were removed from the reaction by yeast 5'-deadenylase (NEB) and RecJ nuclease (NEB) treatment. RPFs ligated to barcoded linkers were pooled, and rRNA-depletion was performed using riboPOOLs (siTOOLs) as per the manufacturer's recommendations. Linker-ligated RPFs were reverse transcribed with ProtoScript II RT (NEB) and gel-purified from 15% TBE-Urea gels. cDNA was then circularized with CircLigase II (Epicentre) and used for library PCR. First, a small-scale library PCR was run supplemented with 1X SYBR Green and 1X ROX (Thermo) in a qPCR instrument. Then, a larger scale library PCR was run in a conventional PCR instrument, performing a number of cycles that resulted in ½ maximum signal intensity during qPCR. Library PCR was gel-purified from 8% TBE gels and sequenced on a SE50 run on Illumina HiSeq4000 instrument at UCSF Center for Advanced Technologies.

SLAM-seq

K562 TAF15 knockdown and control cells were placed into suspension at 1x10^6 cells per 1mL per well of a 24-well plate in triplicate the day before the SLAM-seq assay. Approximately 3 hours before each assay, growth media was replaced with fresh growth media. Newly synthesized RNA was labeled for 1hr and 3hrs with 4-thiouridine (Sigma-Aldrich T4509-25MG) (Muhar et al. 2018; Fuchs et al. 2015). Cells were then spun down and washed twice with 1X PBS before being resuspended in Trizol. RNA was extracted and alkylated using the LEXOGEN SLAMseq Kinetics Kit - Anabolic Kinetics Module according to the manufacturer's instructions. Library prep was performed according to the manufacturer's instructions using 300ng of alkylated RNA and the LEXOGEN 3‘ mRNA-Seq Library Prep Kit plus the PCR Add-on Kit for Illumina and UMI Second Strand Synthesis Module for Quant- Seq FWD (Illumina, Read 1). The reverse transcriptase at this step will introduce a Guanine instead of an Adenine at each alkylated 4-thiouridine modification. Deep sequencing was performed using the HiSeq4000 platform (Illumina) at the UCSF Center for Advanced Technologies. Nascent transcripts were distinguished from existing transcripts by identifying Thymine to Cytosine transitions.

ATAC-seq

The assay for transposase-accessible chromatin using sequencing (ATAC-seq) was performed according to the optimized Omni-ATAC protocol (Corces et al. 2017; Grandi et al. 2022). Briefly, samples containing 50,000 cells as input were pelleted, lysed, washed, and re-pelleted using the lysis and wash buffers specified in the Omni-ATAC protocol. A transposition mix containing Tn5 was then added to the samples and the transposition reaction was carried out for 30 minutes at 37C in a thermomixer with 1000 rpm mixing. After transposition, the transposed DNA was purified

using the DNA Clean & Concentrator-5 Kit (Zymo Research, cat. no. D4014). The samples underwent two PCR steps. First a pre-amplification was performed for 3 cycles to attach unique barcode adapters to the transposed DNA. The concentration of each pre-amplified sample was quantified via qPCR using the NEBNext Library Quant Kit (New England Biolabs, cat. no. E7630). Afterward, samples underwent a second PCR amplification step to obtain the desired DNA concentration of 8 nM in 20 ul. DNA cleanup and qPCR quantification were performed again, and final libraries were diluted down to 8 nM using sterile water. Samples were sequenced using paired-end 75 bp sequencing on an Illumina NextSeq sequencer.

ChIP-qPCR

ChIP-qPCR was performed as described in (Rossi et al. 2018).

Human chronic myelogenous leukemia K562 cells were grown at 37 °C and 5% CO2 in RPMI-1640 medium supplemented with 10% FBS, glucose (2 g/L), L-glutamine (2 mM), 25 mM HEPES, penicillin (100 units/mL), streptomycin (100 µg/mL) and amphotericin B (1 µg/mL) (Gibco). 20 million cells per sample were washed with PBS, pelleted, and cross-linked with 1% paraformaldehyde for 10 minutes at room temperature. Glycine at a final concentration of 125mM was added to the samples and incubated at room temperature for 5 minutes to quench the paraformaldehyde. Samples were washed with PBS, pelleted, flash-frozen, and stored at -80. Samples were thawed, lysed in 200 µl Membrane Lysis Buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.5% IGEPAL CA-630, 1X protease inhibitors), and incubated on ice for 10 minutes. Samples were centrifuged at 4°C at 2500 g for 5 minutes, resuspended in 200 µl Nuclei Lysis Buffer (50 mM Tris pH 8.0, 10 mM EDTA, 0.32% SDS, 1X protease inhibitors), and incubated on ice for 10 minutes. 120 µl of IP Dilution Buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150

mM NaCl, 1% Triton X-100, 1X protease inhibitors) was added to the samples, and the samples were sonicated using the Bioruptor UCD-200 sonicator for 7 minutes with 30 second on/off intervals for a total of 3 times. Samples were centrifuged to clear the lysate, and the supernatant containing the chromatin was stored at -80.

230 µl IP Dilution Buffer was added to 270 µl chromatin along with 3 µg ZNF800 or QKI antibody or same species IgG, and the samples were incubated overnight at 4°C. The next day, the ChIP samples were spun down and the supernatant was transferred onto 20 µl of washed Protein A/G beads (Pierce). Samples were incubated for 2 hours at 4°C.

The ChIP material was washed once with 500 µl of cold FA lysis low salt buffer (50 mM Hepes-KOH pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% Triton-X 100, 0.1% sodium deoxycholate), twice with cold NaCl high salt buffer (50 mM Hepes-KOH pH 7.5, 500 mM NaCl, 2 mM EDTA, 1% Triton-X 100, 0.1% sodium deoxycholate), once with cold LiCL buffer (100 mM Tris-HCl pH 8.0, 500 mM LiCl, 1% IGEPAL CA-630, 1% sodium deoxycholate), and twice with cold 10 mM Tris 1 mM EDTA pH 8.0. Samples were eluted in 300 µl of Proteinase K reaction mix (20 mM Tris pH 8, 300 mM NaCl, 10 mM EDTA, 5 mM EGTA, 1% SDS, 60 µg Proteinase K) and incubated at 65°C for 1 hour. The supernatant was transferred to phase lock tubes (VWR), purified via phenol chloroform extraction, and eluted in 30 µl 10 mM Tris pH 8.0.

qPCR was performed using PerfeCTa SYBR Green SuperMix (QuantaBio) per the manufacturer's instructions. HPRT1 was used as endogenous control.

Reanalysis of enhanced CLIP ENCODE data

To reliably identify RNA targets of RBPs in K562 cells, we started with the raw eCLIP FASTQ files of 'released' K562 experiments for 120 RBPs that were available in the ENCODE database.

The analysis was performed as follows: (1) the reads were preprocessed in the same way as in (Van Nostrand et al. 2016), including adapter trimming with *cutadapt* (v1.18) (Martin 2011), (2) preprocessed reads were mapped to the hg38 genome assembly with GENCODE v31 comprehensive annotation using *hisat2* (v.2.1.0) (Kim et al. 2019), (3) the aligned reads were deduplicated using the *barcodecollapsepe.py* script (https://github.com/YeoLab/eclip/tree/master/bin) as in (Van Nostrand et al. 2016), (4) properly paired and uniquely mapped second reads were extracted using *samtools* (v.1.9, with -f 131 -q 60 parameters) (Li et al. 2009), (5) gene-level read counts were obtained with *plastid* (v.0.4.8) by counting 5' ends of the reads (Dunn and Weissman 2016), (6) analysis of specific enrichment against size-matched control experiments was performed with *edgeR* (v.3.18.1) for each RBP separately, considering only genes passing 2 cpm in at least 2 of 3 samples (Robinson et al. 2010). Reliable RNA targets of each RBP were defined as those passing 5% FDR and $\log_2$(Fold Change) > 0.5. eCLIP target scores (TSs) used in datasets integration were estimated as -$\log_{10}$(p-values)*sign(logFC) for every RBP-gene pair separately.

Gene set enrichment analysis of RBP RNA targets

A joint set of 22471 genes detected at 2 cpm in at least two samples of one eCLIP experiment was used as the background for further analysis. RBPs preferences to bind RNAs of a particular type were assessed using one-sided Fisher's exact test. The following types of RNAs were selected based upon GENCODE annotation: miRNA, lncRNA, protein_coding, snRNA, snoRNA, rRNA. For each RBP separately, the P-values were adjusted for multiple testing using FDR correction for the number of tested RNA types.

GO enrichment analysis of ZNF800 and QKI targets (according to eCLIP or ATAC-seq) was performed with *goana* function from *limma* package (v.3.42.2) and the resulting P-values were FDR-adjusted for the total number of tested GO terms (Ritchie et al. 2015). 15253 eCLIP-detectable genes with Entrez IDs provided in the GENCODE annotation were considered as the background. Visualization of the eCLIP, RNA-seq, and ATAC-seq profiles generated using *bedtools genomecov* (v.2.27.1) was performed with *svist4get* (v.1.2.24) (Quinlan and Hall 2010; Egorov et al. 2019).

Functional annotation of RBPs

To annotate the RBPs based upon preys identified in BioID experiments, target scores (TSs) were estimated as $-\log_{10}(P\text{-value})*\text{sign}(\log\text{-FoldChange})$ for every bait-prey pair separately. Next, for each prey, TSs were converted to Z-scores by estimating mean and average across baits. The preys were ranked by Z-scores and *Fgsea* R package (v.1.12.0) was applied to perform gene set enrichment analysis with 100000 permutations and three GO terms annotation sets (BP, MF, and CC, each taken separately) (Korotkevich et al. 2021). The annotation sets were generated with the *go.gsets* function of *gage* R package (v.2.36.0) (Luo et al. 2009). Lists of 2865 Entrez ids of preys were used in *fgsea* analysis for each RBP of the total set of fifty. GO terms with NES > 2 for at least one RBP were considered when plotting **Figure 3** and **Figure S3** (related GO terms were merged manually)**,** negative NES were zeroed for clarity and easier interpretability of the consequent clusterization, see complete data in Table S3). Ward.D2 clusterization along with cosine distance (1 - cosine similarity) were used to generate the heatmaps using the *heatmap.2* function of the *gplots* R package (v.3.1.1) (Warnes et al. 2009).

To check the consistency between predicted and known RBP annotations, the same procedure was performed excluding the Z-scoring step to avoid penalizing common generic GO terms e.g. “organelle”, “cell”, etc. The resulting GSEA P-values and NESs were used to calculate the <RPB, GO term> scores as -$\log_{10}$(*P*-value)*sign(log-FoldChange) for each RBP and GO term separately. RBPs’ “true” annotations were extracted from the same GO BP, CC, or MF annotation set as used in GSEA. Finally, all data were merged to generate the ROC curve with PRROC (v.1.3.1) roc.curve function (Grau et al. 2015).

Dataset integration

The functional similarity of RBPs was estimated by joint analysis of eCLIP, BioID, and Perturb-seq data (**Figure 1**, **Figure S1**). First, TS Z-scores were calculated for every gene across RBPs separately for each type of experimental data (eCLIP, BioID, or Perturb-seq) in the same way as preys of the BioID data, see above and **Figure S1a**. Next, cosine distance was computed for all 7140 pairs of different RBPs followed by ranking and calculation of empirical *P*-values defined as a fraction of RBP pairs with the cosine distance less than the score of the tested pair, see **Figure S1b**. The empirical *P*-values were aggregated with *logitp* function from the *metap* R package (v.1.4), see **Figure S1c** and **S1d**, raw (non-aggregated) P-values were used for the RBP pairs assessed in a single type of experiment (George and Mudholkar 1983). *Heatmap.2* function of the *gplots* R package (v.3.1.1) with cosine distance and Ward’s (*ward.D2*) clusterization was used to generate the integration heatmap shown in **Figure 2**.

STRING-based RBP interaction heatmap was generated using protein links’ combined scores (STRING v.11.5) and the same RBP clusterization received from the integration procedure (Szklarczyk et al. 2018).

Alternative splicing events analysis

RNA-seq data was processed as follows: (1) to fulfill MISO requirements (see below), the reads obtained with different sequencing lengths were truncated to 75bps with *cutadapt* (v.2.10) *-l* option, (2) the truncated reads were mapped to the human hg38 genome assembly with GENCODE v38 comprehensive gene annotation using *STAR* (v.2.7.9) with options *--outFilterScoreMinOverLread* and *--outFilterMatchNminOverLread* both set to 0.25 (Dobin et al. 2013), (3) non-unique alignment were filtered and the replicates were merged, (4) the insert size distribution was estimated for each merged bam file separately using *pe_utils –compute-insert-len* from *MISO* (v.0.5.4), constitutive exons were retrieved using *exon_utils* with *--get-const-exons* and *--min-exon-size 1000* (Katz et al. 2010), (5) alternative splicing events were identified using *miso --run* with *--read-len* set to 75 and *--paired-end* set to the previously estimated insert size parameters. For clarity, only cases with non-zero counts are shown in Figure 4.

Ribosome profiling analysis

To process the reads, the Ribo-seq reads were first trimmed using cutadapt (v2.3) to remove the linker sequence AGATCGGAAGAGCAC. The fastx_barcode_splitter script from the Fastx toolkit was then used to split the samples based on their barcodes. Since the reads contain unique molecular identifiers (UMIs), they were collapsed to retain only unique reads. The UMIs were then removed from the beginning and end of each read (2 and 5 Ns, respectively) and appended to the name of each read. Bowtie2 (v2.3.5) was then used to remove reads that map to ribosomal RNAs and tRNAs, and the remainder of reads were then aligned to mRNAs (we used the isoform

with the longest coding sequence for each gene as the representative). Subsequent to alignment, umitools (v0.3.3) was used to deduplicate reads.

The adapters in the sequencing reads are removed using cutadapt with options '--trimmed-only -m 15 -a AGATCGGAAGAGCAC '. The PCR duplicates in the reads were collapsed using CLIPflexR v0.1.19. The Unique Molecule Identifiers (UMIs) for each read are extracted using umi_tools v1.1.1 with the options 'extract --bc-pattern=NN' for the 5' end and options 'extract --3prime --bc-pattern=NNNNN' for the 3' end.

Reads corresponding to rRNA and other non-nuclear mRNA are removed by aligning out the reads using Bowtie2 on a depletion reference (rRNA, tRNA and mitochondrial RNA sequences). This depletion reference was built from a non-coding transcriptome for Mus musculus (Ensembl release v96). The reads that did not align to the depletion reference are aligned to the Mus Musculus mRNA transcriptome using Bowtie2 with options '--sensitive --end-to-end --norc'. The mRNA transcriptome was built using the cDNA longest CDS reads of Mus Musculus downloaded from Ensembl release v96. The resulting reads were converted to bam files and then sorted using samtools v1.11. The duplicate reads in the sorted files were removed using umi_tools v1.1.1 with options 'dedup'.

The quality check and downstream processing of the processed reads was performed using Ribolog v0.0.0.9 (Navickas et al. 2021). To distinguish stalling peaks from stochastic sequencing artifacts, we followed a multi-step procedure. We calculated P-site offsets and identified the codon at the ribosomal A-site for each RPF read using the riboWaltz package. A loess smoother was used to de-noise codon-wise RPF counts. The loess span parameter varied depending on the transcript length and allowed borrowing information from approximately 5 codons on either side of the A-site. We calculated an excess ratio at each codon position by dividing the loess-smoothed count by

the transcript's background translation level (median of no-zero loess-smoothed counts). After median normalization of the corrected counts and removal of transcripts with 0 counts, the translational efficiency testing was performed using logistic regression in Ribolog.

SLAM-seq analysis

The SLAM-seq reads were aligned to an hg38 reference using salmon to obtain the gene counts. The reads were then used to produce a special reference of the sequences present in the data. This allows us to re-align the data to a targeted reference. The targeted alignments were then processed using samtools mpileup to obtain the pileup of certain nucleotides for each genomic position. The pileups were then used to count the number of C and T nucleotides. The genomic positions with very low C and T counts were filtered out. The C and T counts were then modeled against the gene counts to analyze the differential expression of relative C and T ratios using DESeq2. The relative measures were then used to calculate the half-life of the transcripts.

Perturb-seq analysis

Cell Ranger (version 3.0.1, 10X Genomics) with default parameters was used to align reads and generate digital expression matrices from single-cell sequencing data. To assign cell genotypes, a bwa reference (Li and Durbin 2009) database was created containing all guide sequences present in the library using *bwa index* command. The barcode-enrichment libraries were mapped to this database to establish the guide identities; to detect the cell barcodes, the barcode correction scheme used in Cell Ranger was used (the mapping of uncorrected to corrected barcodes was extracted from Cell Ranger analysis run of the whole transcriptome libraries; this mapping was then applied to the reads of barcode-enrichment libraries). UMI correction was performed by merging the UMIs

within the hamming distance of 1 from each other. For each UMI, the guide assignment was done by choosing the guide sequence most represented among the reads containing the given UMI. To make the final assignment of a guide to cell barcodes, we only considered the barcodes that were represented by at least 5 different UMIs, with >80% UMIs representing the same guide.

Data filtering was performed using scanpy package (Wolf et al. 2018). Data were denoised using a modification of scvi autoencoder (Gayoso et al. 2022) with loss function penalizing for the similarity between cells having different RBPs knocked down. The distance between transcriptome profiles of individual RBP knockdowns was calculated by applying the t-test to individual gene counts across the cells that were assigned the respective guide sequence.

MS data analysis (BioID2 mass spectrometry data)

Data were quantified and queried against a Uniprot human database (January 2013) using MaxQuant MaxLFQ command (Cox et al. 2014). Data normalization was performed in Perseus (Tyanova et al. 2016) (version 1.6.2.1). For batch correction, Brent Pedersen's implementation (Pedersen) of ComBat function from sva package (Leek et al. 2012) was used. The protein abundances in "experiment" (biotin +) and "control" (biotin -) samples were compared using t-test for each protein individually.

ATAC-seq analysis

ENCODE ATAC-seq pipeline (Lee et al. 2016) with default parameters was used for sequencing data processing and analysis. The differentially accessible peaks were identified with DESeq2 package (Love et al. 2014) and annotated with ChIPseeker package (Yu et al. 2015).

MS data analysis (TAF15 KD proteomic quantification)

Quantitative analysis of the TMT experiments was performed simultaneously to protein identification using Proteome Discoverer 2.5 software. The precursor and fragment ion mass tolerances were set to 10 ppm, 0.6 Da, respectively), enzyme was Trypsin with a maximum of 2 missed cleavages and Uniprot Human proteome FASTA file and common contaminant FASTA file used in SEQUEST searches. The impurity correction factors obtained from Thermo Fisher Scientific for each kit was included in the search and quantification. The following settings were used to search the data; dynamic modifications; Oxidation / +15.995Da (M), Deamidated / +0.984 Da (N, Q), Acetylation /+42.011 Da (N-terminus) and static modifications of TMT6plex / +229.163 Da (N-Terminus, K), MMTS / +45.988 Da (C).

Scaffold Q+ (version Scaffold_5.0.1, Proteome Software Inc., Portland, OR) was used to quantitate TMT Based Quantitation peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 78.0% probability to achieve an FDR less than 1.0% by the Percolator posterior error probability calculation (Käll et al. 2008). Protein identifications were accepted if they could be established at greater than 5.0% probability to achieve an FDR less than 1.0% and contained at least 1 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al. 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. Channels were corrected by the matrix [0.000,0.000,0.931,0.0689,0.000]; [0.000,0.000,0.933,0.0672,0.000]; [0.000,0.00750,0.931,0.0619,0.000]; [0.000,0.0113,0.929,0.0593,0.000]; [0.000,0.0121,0.934,0.0532,0.000934]; [0.000,0.0148,0.923,0.0499,0.0120]; [0.000,0.0251,0.931,0.0438,0.000];

[0.000,0.0206,0.936,0.0431,0.000]; [0.000,0.0291,0.937,0.0337,0.000]; [0.000,0.0776,0.892,0.0303,0.000] in all samples according to the algorithm described in i-Tracker (Shadforth et al. 2005). Normalization was performed iteratively (across samples and spectra) on intensities, as described in Statistical Analysis of Relative Labeled Mass Spectrometry Data from Complex Samples Using ANOVA (Oberg et al. 2008). Means were used for averaging. Spectra data were log-transformed, pruned of those matched to multiple proteins, and weighted by an adaptive intensity weighting algorithm. Of 22889 spectra in the experiment at the given thresholds, 20372 (89%) were included in quantitation. Differentially expressed proteins were determined by applying T-Test with unadjusted significance level $p < 0.05$ corrected by Benjamini-Hochberg.

Chapter 2: a systematic search for RNA structural switches across the human transcriptome

INTRODUCTION

Gene expression is regulated at the RNA level in all kingdoms of life. The two oldest groups of RNA-based regulatory mechanisms are ribozymes (catalytically active RNA molecules) and RNA structural switches (or riboswitches). RNA switches are regulatory elements that control gene expression by direct binding of a small-molecule ligand or other *trans*-acting factor. In bacteria, RNA switches are one of the most widely observed mechanisms for gene expression control. Classic bacterial riboswitches bind small molecule ligands directly, inducing a conformational change, leading to modulation of expression of the host transcript (Serganov and Nudler 2013). While such ligand-binding riboswitches have not been observed in higher eukaryotes, examples of protein-dependent RNA switches have been described (Liu et al. 2015) (Ray et al. 2009). Given the paucity of such examples, it is unclear how widely this regulatory mechanism impacts gene expression in higher eukaryotes. To address this, we introduce a systematic approach for the identification of RNA structural switches across large transcriptomes.

While several RNA switch detection software packages have been developed, most identify new switch sequences based on their homology to one of the 40 known RNA switch families (Kalvari et al. 2021). Several tools for *de novo* discovery of RNA switches have been developed; however, their predictions often lack experimental verification of both structure and function (Barsacchi et al. 2016; Manzourolajdad and Arnold 2015). Therefore, there is a need for scalable methods of

detecting eukaryotic RNA switches and assessing the extent to which they carry out regulatory functions in gene expression control.

Here, we describe a platform for the systematic discovery and characterization of RNA structural switches across transcriptomes of interest. Our approach relies on integrating multiple computational and experimental methods, where RNA switches are first predicted *in silico*, then functionally characterized *in vivo*, which in turn informs the next iteration of *in silico* predictions. First, we developed a computational model for *de novo* RNA switch detection, named SwFinder. We showed that SwFinder identifies RNA switches from novel families with higher accuracy than the existing models. We applied SwFinder to the human transcriptome to select putative RNA switches. We then used massively parallel assays *in vivo* to interrogate both the structure and function of these candidate elements *en masse*. By iteratively improving the SwFinder predictions with experimental data, we reported ~250 high-confidence and functional RNA structural switches. We then selected the top scoring switch, located in the 3'UTR of the RORC transcript, for further analysis and dissection. We used DMS-MaPseq structural probing and single-particle cryogenic electron microscopy (cryo-EM) to confirm that the predicted switch populates alternate molecular conformations. We then performed a genome-wide CRISPRi screen, which revealed that only one of these conformations reduces gene expression through activation of the non-canonical nonsense-mediated decay (NMD) pathway. Taken together, our framework provides new insights into the role of RNA structural switches in shaping the transcriptome in human cells and outlines a broader approach for future characterization of RNA switches regulating eukaryotic gene expression control across cell types.

RESULTS

Systematic annotation of human RNA structural switches

Our first goal was to identify RNA sequences that can exist in two secondary structural conformations. For this, we developed SwFinder, a tool that predicts whether an RNA sequence contains putative RNA structural switches. We designed SwFinder to satisfy several criteria: first, it should predict if a given RNA sequence contains a potential RNA switch and suggest the two mutually exclusive folding conformations (examples shown in Fig 1A, Suppl. Fig 1A). Secondly, it should be able to effectively capture a more generalizable definition of RNA switches in order to find instances beyond the 40 known RNA switch families (Kalvari et al. 2021). Thirdly, it should allow for seamless integration of experimental data to improve predictions. This is especially important as mRNA secondary structure in the cell is shown to be highly dynamic (Mortimer, Kidwell, and Doudna 2014) and compartment-dependent (Sun et al. 2019); therefore, the predictions can be greatly improved with *in vivo* secondary structure probing data.

To discover new families of RNA switches, we aimed to design an approach that does not rely on known sequence motifs, which has been the case for most published software (Wheeler and Eddy 2013; Nawrocki and Eddy 2013; Bengert and Dandekar 2004; Abreu-Goodger and Merino 2005; Chang et al. 2009; Mukherjee and Sengupta 2016). Instead, SwFinder uses the sequence to generate an ensemble of secondary structures and their corresponding energy landscape using the Boltzmann ensemble concept (Ding and Lawrence 2003). It then prioritizes those sequences that show RNA switch-like features, such as having two local minima in close proximity with a relatively small barrier in between (Fig. 2.1B). This approach ensures that RNA switches are

described in a generalizable and family-agnostic way. We demonstrated this point by holding individual Rfam families out of the training set and testing whether SwFinder would predict the switches correctly (Fig. 2.1C). We observed high performance metrics across all held-out families as measured by the Area Under the Receiver Operating Characteristic curve (AUROC) values (Fig. 2.1D). We further compared the performance of SwFinder to SwiSpot, the state-of-the-art method for family-agnostic riboswitch prediction (Barsacchi et al. 2016), and observed significant improvement of performance across most RNA switch families (Fig. 2.1E). By relying on biophysical features of the folding energy landscape as opposed to sequence features, SwFinder captures a wider variety of RNA switches compared to the existing methods.

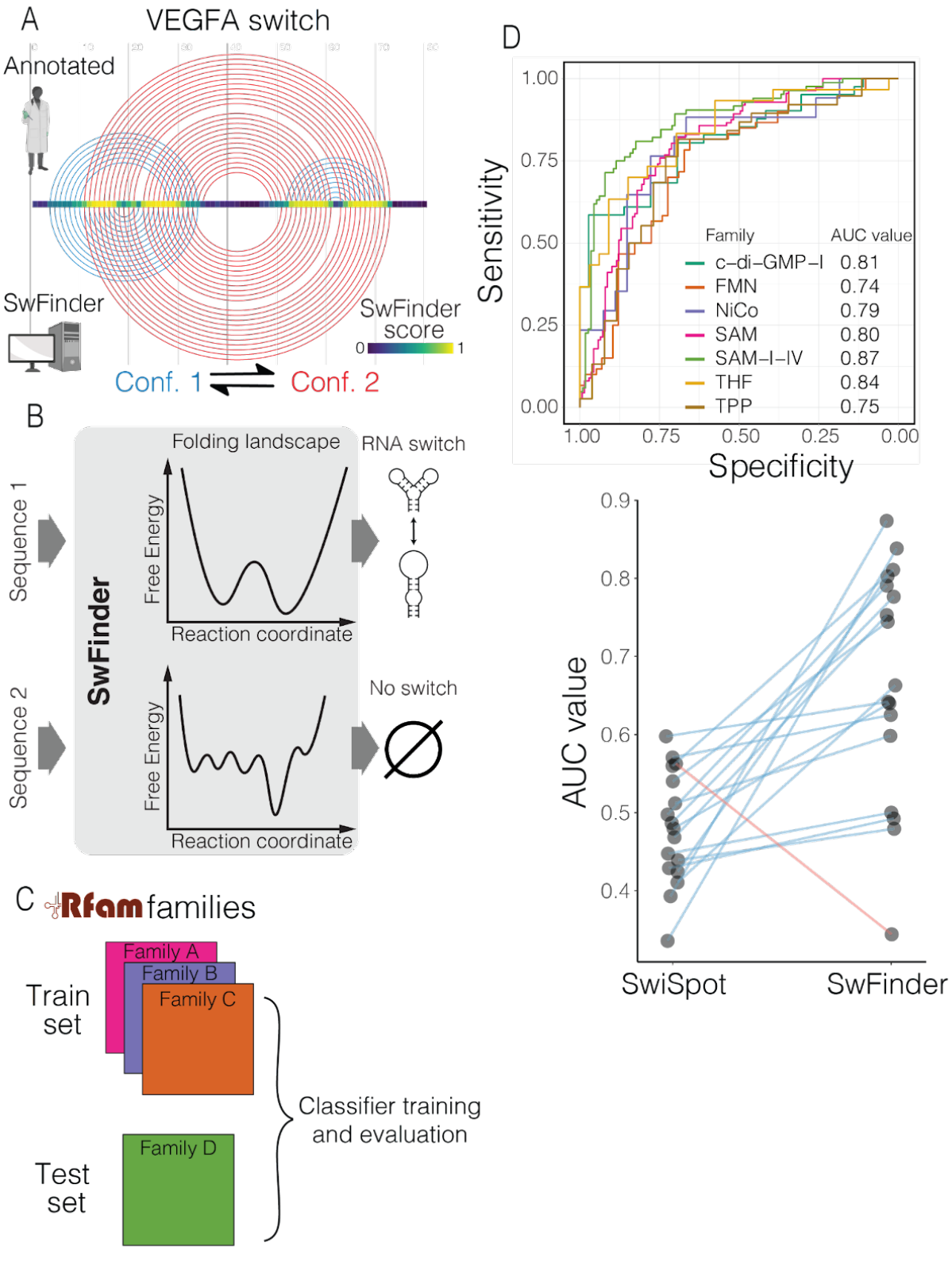
A
VEGFA switch
Annotated
SwFinder
SwFinder score
0
1
Conf. 1
Conf. 2
B
Folding landscape
RNA switch
Sequence 1
SwFinder
Free Energy
Reaction coordinate
Sequence 2
No switch
C
Rfam families
Family A
Family B
Family C
Train set
Classifier training and evaluation
Family D
Test set
D
Sensitivity
Specificity
Family
AUC value
c–di–GMP–I 0.81
FMN 0.74
NiCo 0.79
SAM 0.80
SAM–I–IV 0.87
THF 0.84
TPP 0.75
AUC value
SwiSpot
SwFinder

Figure 2.1: SwFinder identifies candidate RNA switches in the human genome

(A) Example of SwFinder locating the RNA switch within the VEGFA mRNA sequence. Top: arc representation of the RNA base pairs that change between the two conformations of VEGFA RNA switch, as in (Ray et al. 2009). The two conformations are shown in red and blue, respectively. Bottom: the two conformations of the VEGFA RNA switch as predicted by SwFinder. Middle: SwFinder score reflecting the likelihood of a given nucleotide to be involved in two mutually exclusive base pairings.
(B) Scheme of SwFinder model
(C) The setup for evaluating the ability of a model to find RNA switches from novel families
(D) ROC curves of SwFinder predictions of RNA switches from the common RFAM families
(E) AUC values of RNA switch predictions across the RFAM families for two models: SwFinder and SwiSpot (Barsacchi et al. 2016). Each dot represents one RFAM family. The lines show the change in accuracy between the two models. The families that have higher AUC values for SwFinder are shown with blue lines; the ones that have higher AUC values for SwiSpot are shown in red.

In addition to primary sequence, we use RNA secondary structure probing data, when available, to improve SwFinder predictions by updating the energy terms of the model. Eukaryotic genomes are large, therefore the models for RNA structural element prediction have to show very high specificity. Such specificity is difficult to reach by relying on *in silico* RNA folding alone, since RNAs can fold differently depending on the cell state (Beaudoin et al. 2018). However, it is possible to achieve higher specificity if the RNA secondary structure is first probed *in vivo* and the model is then guided by this data. Therefore, we designed SwFinder to have an option to update the energy landscape based on RNA secondary structure data (see Methods). SwFinder can use two approaches for modeling the RNA folding energy landscape, relying either on *in silico* RNA folding tools alone, or in tandem with experimental RNA secondary structure probing via methods like SHAPE-seq or DMS-MaPseq (REFS).

We used this functionality of SwFinder to improve our RNA switch predictions iteratively. First, we applied the SwFinder model, using the naive *in silico* folding, to the entirety of the 3' untranslated regions (3'UTR) of the human transcriptome, and chose the top set of 3,750 top

candidate switches (of the length <= 186 nucleotides) as putative elements. We then performed RNA secondary structure probing for 3,750 candidate RNA switches *in vivo* using DMS-MaPseq applied to a massively parallel reporter system. We used the resulting data to improve the initial SwFinder predictions. We then carried forward the top 1,454 high-confidence RNA structural switches from the second round of predictions for further functional and biochemical validation *in vivo* (Fig. 2.2A).

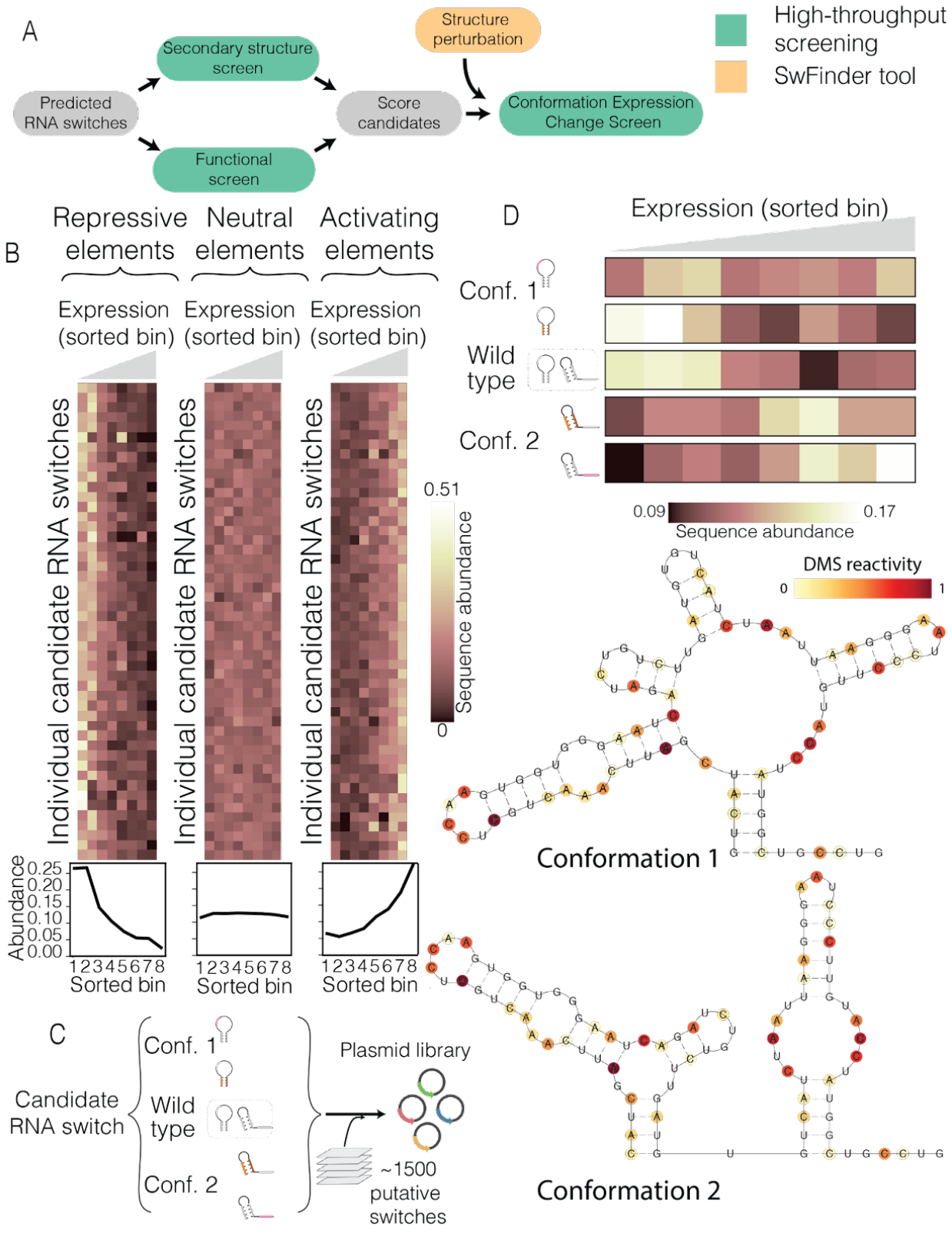
A
Predicted RNA switches
Secondary structure screen
Functional screen
Score candidates
Structure perturbation
Conformation Expression Change Screen
High-throughput screening
SwFinder tool
B
Repressive elements
Neutral elements
Activating elements
Expression (sorted bin)
Individual candidate RNA switches
Abundance
Sorted bin
C
Candidate RNA switch
Conf. 1
Wild type
Conf. 2
Plasmid library
~1500 putative switches
D
Expression (sorted bin)
Conf. 1
Wild type
Conf. 2
0.51
0
Sequence abundance
0.09
0.17
Sequence abundance
DMS reactivity
0
1
Conformation 1
Conformation 2

Figure 2.2: Massively parallel reporter assay captures the functional difference between the conformations of candidate RNA switches

(A) Overview of the platform for RNA switch identification
(B) Examples of regulatory elements identified by Functional Screen. Each row represents a single candidate RNA switch, each column represents a single bin defined by the reporter gene expression (eGFP fluorescence, normalized by mCherry fluorescence). The value in each cell is the relative abundance of the given RNA switch in the given bin, normalized across the 8 bins. The three plots show examples of candidate switches with repressive, neutral, and activating effects on gene expression. The plots below show cumulative abundances across all the candidate switches within each group.
(C) The setup of the Conformation Expression Change Screen. For each candidate RNA switch, we design 4 mutated sequence variants. Two of them lock the switch into the conformation 1, and the other two lock it in the conformation 2. A sequence library is then generated, where each candidate RNA switch is represented by the 4 mutated sequence variants, along with the wild type sequence.
(D) Example of a high-confidence candidate RNA switch identified by the Conformation Expression Change Screen. Bottom: two alternative conformations as predicted by SwFinder. The RNA secondary structure probing data collected with the Structure Screen is shown in color. Top: the effect of the candidate RNA switch locked in one or another conformation on reporter gene expression. Each column corresponds to a single sequence variation that locks the RNA switch in one of the two conformations.

Discovery of RNA switches with regulatory function in the human transcriptome

In order to identify the RNA switches that are both functional and structurally bistable in the cell, we performed two high-throughput *in vivo* screens, which we have named the Structure Screen and the Functional Screen, respectively. The Structure Screen differentiates the RNAs that exist as an ensemble of two mutually exclusive conformations from those that reside only in a single conformation. The Functional Screen measures the effect of candidate RNA switches on the expression of a reporter gene. Integrating data from the two screens allowed us to identify the putative RNA switches that are regulatory active and act as switches in vivo.

Large-scale RNA secondary structure probing for improved RNA structural switch predictions

In order to improve the SwFinder predictions of RNA switches, we performed an *in vivo* RNA Structure Screen, where we probed the secondary structure of 3,750 candidate switches *in vivo* using DMS-MaPseq (Zubradt et al. 2017). DMS preferentially modifies unpaired nucleotides resulting in substitutions during the reverse transcription process. Once the cDNA library is sequenced, the substitution frequency at a given position provides an estimate of nucleotide accessibility. Paired nucleotides typically have lower accessibility values compared to unpaired nucleotides. We applied this method to cells expressing the library of candidate RNA switches, enabling pooled and targeted accessibility measurement with single-nucleotides resolution across all 3,750 candidate RNA switches.

We used the Structure Screen data to identify bi-stable RNA structures. DMS-MaPseq measures the base reactivities of adenines and cytosines across all the coexisting RNA conformations *in vivo*. The reactivity of a single nucleotide is a population average of multiple RNA molecules that represent different minima in the RNA folding conformation ensemble. If one conformation dominates within the ensemble, it dominates the DMS-MaPseq reactivity profile; however, if multiple conformations co-exist, they all contribute to the reactivity profile (Morandi et al. 2021; Tomezsko et al. 2020). SwFinder exploits this difference to find the RNA switches that coexist in a balanced state between the two conformations. Therefore, SwFinder, informed by nucleotide accessibility data, predicts not just the RNAs with the potential to act as RNA switches, but rather the RNAs that do act as RNA switches *in vivo*.

Massively parallel reporter assays identify regulatory RNA structural switches

We next sought to explore the potential role of the identified RNA switches in regulating gene expression. We implemented a massively parallel reporter assay (MPRA) to functionally interrogate RNA switches *in vivo* (Functional Screen). For this, we tested whether a given RNA switch placed in a 3' UTR can affect expression of its host mRNA (in this case the eGFP ORF), compared to a control scrambled sequence. We cloned a library of 3,750 candidate RNA switch sequences into a dual eGFP-mCherry fluorescent reporter vector, directly downstream of the eGFP ORF. We used eGFP fluorescence to measure the effect of candidate RNA switches on gene expression, and we used mCherry fluorescence as an endogenous control. We transduced HEK293 cells with this synthetic library, used flow-cytometry to sort cells by eGFP/mCherry expression ratio, and sequenced the genomic DNA and RNA from the resulting eight pools of cells (see Methods). Of the candidate RNA switches tested, 536 (14%) showed significant downregulation relative to their scrambled control, and 538 (14%) showed a significant upregulation. We have included representative candidates with repressive, neutral, or activating function in Fig. 2.2B.

To annotate a high-confidence set of RNA switches with regulatory potential in the human transcriptome, we performed a second iteration of SwFinder predictions, guided by the *in vivo* RNA structure probing data. To test the performance of this procedure, we compared the fraction of regulatory active candidate RNA switches among the first and the second iterations of SwFinder prediction, using the Functional Screen data. We observed a higher fraction of regulatory active RNA switches among the second iteration of SwFinder predictions compared to the first iteration (16% up- and 16% down- regulation versus 14% up- and 14% down- regulation). This supports the hypothesis that incorporating the *in vivo* RNA structure probing data improves its performance.

We then integrated the high-confidence RNA switches with the massively parallel reporter data. Together these analyses resulted in 1,454 elements that were significant in both screens (see methods).

Massively parallel mutagenesis analysis identifies conformation-specific RNA switch activities

Having identified the candidate RNA switches that affect gene expression, we aimed to assess the degree to which the two stable conformations show divergent regulatory function. For this, we extended our MPRA to include targeted mutations designed to shift the equilibrium between the two conformations of each riboswitch. This additional screen allowed us to identify *bona fide* RNA switches with strong conformation-dependent activity. Starting with the 1,454 high-confidence RNA switches described above, we engineered mutated variants that would lock RNA switches in one of their two predicted conformations. This was achieved by either disrupting or strengthening conformation-specific stem loops. We then performed a massively parallel reporter assay in which each candidate RNA switch was represented by four additional conformation-specific variants (i.e., two vs. two) (Fig. 2.2C). We observed a total of 245 RNA switches that differentially regulated reporter gene expression when locked in a specific structural conformation. An example candidate switch (located in the 3'UTR of TCF7) is shown in Fig. 2.2D. The TCF7 RNA switch landscape has two local minima, corresponding to two alternative conformations supported by *in vivo* DMS-MaPseq data (Fig. 2.2D, right). Two mutations in different parts of the switch sequence that disfavor conformation 2 resulted in lower expression of the eGFP reporter. Conversely, two mutations that disfavor conformation 1 increased eGFP expression. This observation indicates that the two conformations of the TCF7 RNA switch elicit divergent regulatory functions.

Describing a bi-stable RNA switch in the 3'UTR of RORC

Among the hundreds of conformation-specific RNA switches, we chose the top performing element for further analysis. This RNA switch is located in the 3'UTR of the RORC mRNA. SwFinder predicted a bistable secondary structure for this sequence (Fig. 2.3A). In this RNA switch, the 5' region can pair either with the 3' or with the middle region, leading to two mutually exclusive conformations (Fig. 2.3A). Our measurements indicate that this RNA switch exists in an equilibrium state between the two conformations *in vivo*, and that these two conformations have different effects on the expression of RORC.

In order to further confirm that the RORC RNA switch exists as an ensemble of two conformations, we performed targeted mutagenesis experiments and used *in vitro* RNA SHAPE as the read-out. We designed mutation-rescue pairs of sequences that first shift the equilibrium towards one conformation (mutation), and then shift it towards the other conformation (rescue) (Fig. 2.3B, Suppl. Table 1). We then measured the accessibility of individual nucleotides using the *in vitro* SHAPE assay (Fig. 2.3C). We observed that mutating the 3' region (117-AC), which is expected to stabilize conformation 2, reduced 1M7 accessibility of the middle region. Conversely, the rescue mutation (65-GT,117-AC) of the 5' region restored its wild-type accessibility (Fig. 2.3C). Complementary experiments using the mutation (77-GA) to stabilize conformation 1, and the rescue mutation (63-TC,77-GA) to stabilize conformation 2, had a similar outcome. Even though we did not observe a significant decrease in accessibility of the 3' region upon the 77-GA mutation, the rescue significantly increased its accessibility (Fig. 2.3C). These findings support the role of the three highlighted regions in forming an ensemble of states in which the middle and the 3' region compete for base pairing to the 5' region.

To extend our *in vitro* observations to living cells, we performed high-coverage DMS-MaPseq of the RORC switch. We used a DMS concentration sufficient to cause multiple modifications to the same RNA molecule. This enabled us to cluster reads originating from alternative secondary structures using a state-of-the-art unsupervised computational approach, named DRACO (Morandi et al. 2021). In both biological replicates, DRACO identified two clusters, each representing one of the two conformations, at the approximate ratio of 27% to 73% (Fig. 2.3D).

To confirm that the DRACO clusters truly represent distinct tertiary structures, we applied single-particle cryo-EM to the wild-type RORC RNA switch, as well as the conformation 1 mutant (77-GA) and the conformation 2 mutant (117-AC). Particles from wild-type RORC RNA separate into three structural classes labeled A, B, and C, with the Class B structure absent in the (77-GA), and Class A absent in (117-AC) (Fig. 2.4A). The Class B structure represents Conformation 2, while Class A represents Conformation 1. Interestingly, Class C is present in all three datasets and may represent a folding intermediate lacking the tertiary interactions made by nucleotides 77 and 117.

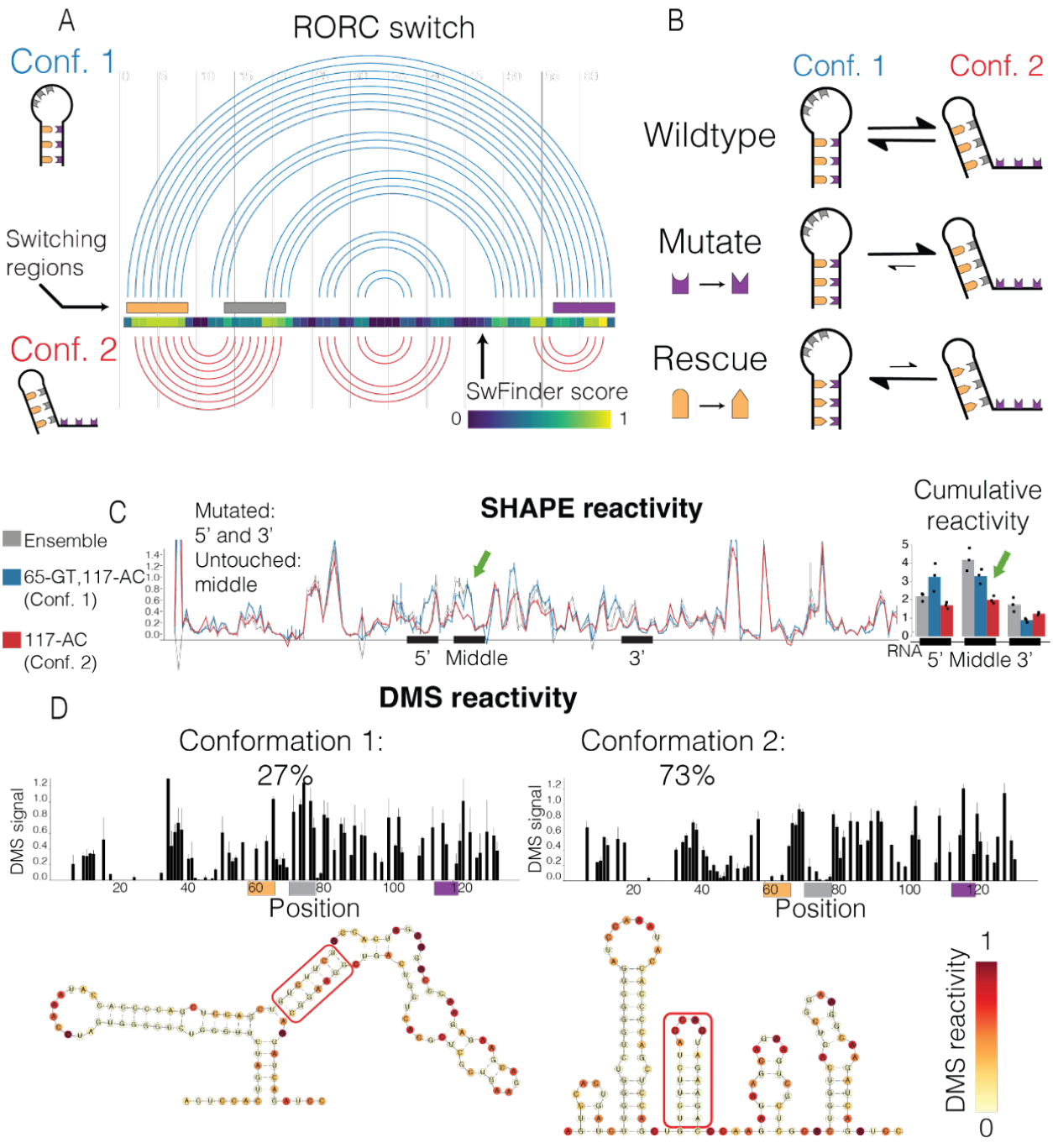
A
RORC switch
Conf. 1
Switching regions
Conf. 2
SwFinder score
0
1
B
Conf. 1
Conf. 2
Wildtype
Mutate
Rescue
C
Ensemble
65-GT,117-AC (Conf. 1)
117-AC (Conf. 2)
Mutated: 5' and 3'
Untouched: middle
SHAPE reactivity
5' Middle 3'
Cumulative reactivity
RNA
5' Middle 3'
D
DMS reactivity
Conformation 1: 27%
Conformation 2: 73%
DMS signal
Position
DMS reactivity
1
0

Figure 2.3: A fragment of RORC 3'UTR forms an ensemble of two alternative structures

(A) Arc representation of the two alternative conformations of the RORC RNA switch as predicted by SwFinder. The two conformations are shown in blue and red, respectively. The 3 switching regions are color coded: the 5' region is shown in orange, the middle region - in gray, and the 3' region - in purple. Middle: SwFinder score reflecting the likelihood of a given nucleotide to be involved in two mutually exclusive base pairings. Left: the schematic representations of the two conformations, as used throughout the article.
(B) The setup of mutation-rescue experiments. The 3 switching regions are color coded as in (A). A-U and C-G base pairing is shown with compatible shapes (triangle and half-circle). The two conformations of the switch reside in the equilibrium state. Mutation of the 3' switching region disrupts the base pairing between the 5' and the 3' regions. This causes a shift of the equilibrium towards the conformation 2. Rescue mutation of the 5' switching region restores the base pairing between the 5' and the 3' regions, but at the same time it disrupts the base pairing between the 5' and the middle regions. Therefore, the equilibrium shifts towards conformation 1.
(C) In vitro SHAPE reactivity of the RORC RNA switch sequence *in vitro*. Left: SHAPE reactivity profiles for the wild type sequence (in gray) and for the mutation-rescue pair of sequences (blue - "65-GT,117-AC", red - "117-AC".). Shown is the average for 3 replicates with the respective error bars. The three switching regions are labeled. The SHAPE reactivity changes in the non-mutated regions are highlighted in bold arrows. Right: barplots of cumulative SHAPE reactivity within the switching regions. The color scheme for the conformations is the same as in the left panel. N replicates = 3.
(D) DMS reactivity of the RORC RNA switch *in vivo*. Top: DMS reactivities of the two clusters identified by the DRACO unsupervised deconvolution algorithm (Morandi et al. 2021). The algorithm was run on two replicates independently, and identified the same clusters in both of them. The ratios of the clusters reported by DRACO are 22% to 78% in replicate 1 and 32% to 68% in replicate 2. The ratio shown in the Figure 2. is an average between the two replicates. The switching regions are shown in color. Bottom: secondary structures of the two conformations of RORC RNA switch predicted by RNAstructure algorithm (Reuter and Mathews 2010) guided by the DMS reactivity data. DMS reactivity is shown in color.

These 3D structures demonstrate RNA-like tertiary features, including apparent double-stranded helical segments with a discernible major groove, and typical RNA hairpin elements. The resolution of these reconstructions is limited to ~10 Å, due to the extreme flexibility evinced by the raw micrographs and 2D class averages, but is sufficient for recognition of RNA folds. Comparing our cryo-EM structures to the DRACO secondary structure clusters (Fig. 2.4B), we propose that the complex tertiary fold at the center of Class A is reflective of the multiple stem-loop motif at the center of Conformation 2 (Fig. 2.4C). We further posit that the two legs of the

hairpin-like Class B correspond to the two long helical segments in Conformation 1. Finally, the relatively simple inverted-L shape of Class C could reflect the helix detected at the leading and trailing ~20 residues in some members of either cluster. As these residues are identical in WT and mutant sequences, this hypothetical assignment is consistent with the appearance of Class C in all three cryo-EM datasets.

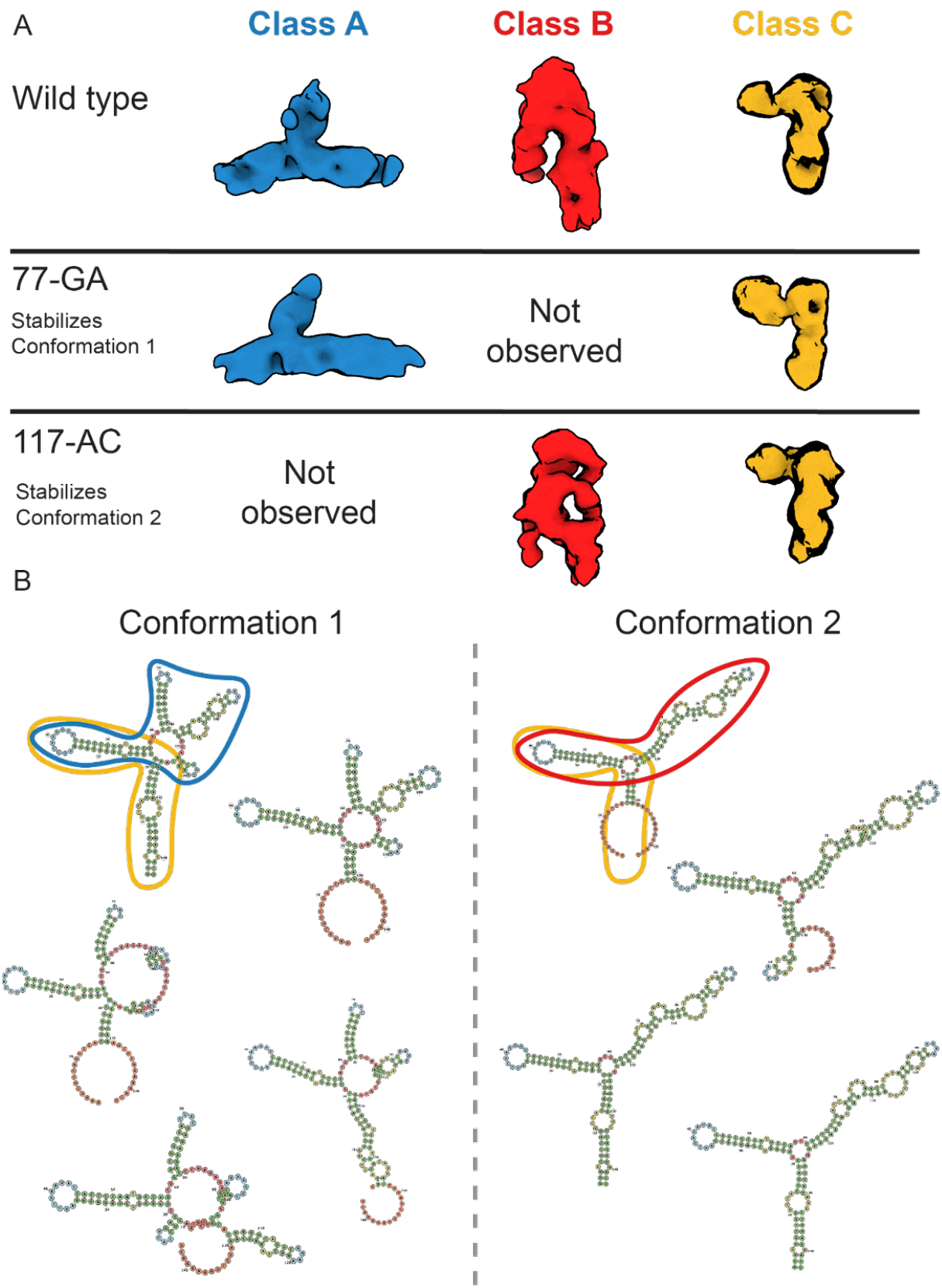
A
Class A
Class B
Class C
Wild type
77-GA
Stabilizes
Conformation 1
Not
observed
117-AC
Stabilizes
Conformation 2
Not
observed
B
Conformation 1
Conformation 2

Figure 2.4. Cryo-EM of RORC 3' mRNA is consistent with dynamic exchange in a shallow energy landscape.

(A) Cryo-EM of the refolded RORC 3' mRNA element reveals three classes with RNA-like features (top). Class A is presented in red, Class B in blue, and Class C in yellow. Further cryo-EM imaging and 3D classification of the 77-GA mutant (middle) and 117-AC mutant (bottom) indicate that Class A is present in WT and 77-GA samples but absent from the 117-AC sample, and Class B is conversely present in WT and 117-AC but absent from the 77-GA mutant. Class C is common to all three samples. We thus putatively assign Class A as the conformation 1 state, and Class B as the conformation 2 state. We propose Class 3 to represent a partly folded intermediate which is not disrupted in the mutated constructs.
(B) Secondary structure models for conformation 1 and conformation 2 clusters inferred by DMS-Seq/DRACO. Solid outlines indicate potential associations with cryo-EM classes, colors as in (A). The complex multiple stem-loop structure detected within the conformation 2 state representative is identified with the central fold and prominences of Class A (red outline), and the two long helices of conformation 1 with the two linear elements of Class 2 (blue outline). The shorter, L-shaped motif of Class C might be identified with multiple structures within either model as the RNA dynamically refolds (yellow outlines).

The alternative conformations of the *RORC* RNA switch play divergent roles in gene regulation

Having discovered the RORC RNA switch as an ensemble of two conformations, we set out to further characterize the divergent regulatory activity of its two states. For this, we generated HEK293 cell lines stably expressing a reporter with the conformation mutant sequences cloned into the 3'UTR of the eGFP ORF. We then measured the changes in eGFP expression of each mutant by flow cytometry. We employed two parallel strategies to lock the RNA switch in conformation 1: first, we mutated the middle region so that it cannot pair with the 5' region, and second, we introduced complementary mutations in both 5' and 3' regions, thus preventing base pairing between the 5' and middle regions. These two orthogonal strategies achieve the same stabilization of secondary structures while modifying different parts of the sequence. Strikingly, both sets of modifications led to similar changes in eGFP expression: locking the RNA switch in Conformation 1 increased reporter gene expression compared to wild-type (Fig. 2.5A).

Furthermore, we applied the same two strategies to instead lock the RNA switch in conformation 2 and observed the expected opposite effect: both modifications led to decrease of the reporter gene (Fig. 2.5A). Therefore, we concluded that the two conformations play divergent functional roles.

We next tested whether the secondary structure, rather than sequence composition, is the major determinant of the observed modulation in gene expression. To do so, we generated cell lines (as described above) stably expressing the mutant sequences from the rescue-mutation experiments (Fig. 2.3B), and measured the effect of each mutant on eGFP reporter gene expression. In total, we tested 3 mutation-rescue pairs. In all three cases, we observed lower eGFP expression of the Conformation 2 mutant (117-AC), as compared to (77-GA) which favors Conformation 1 (Fig. 2.5B). Taken together, the reciprocal mutation-rescue experiments provide evidence that the RNA secondary structure is critical for the functional effects of the RORC RNA switch.

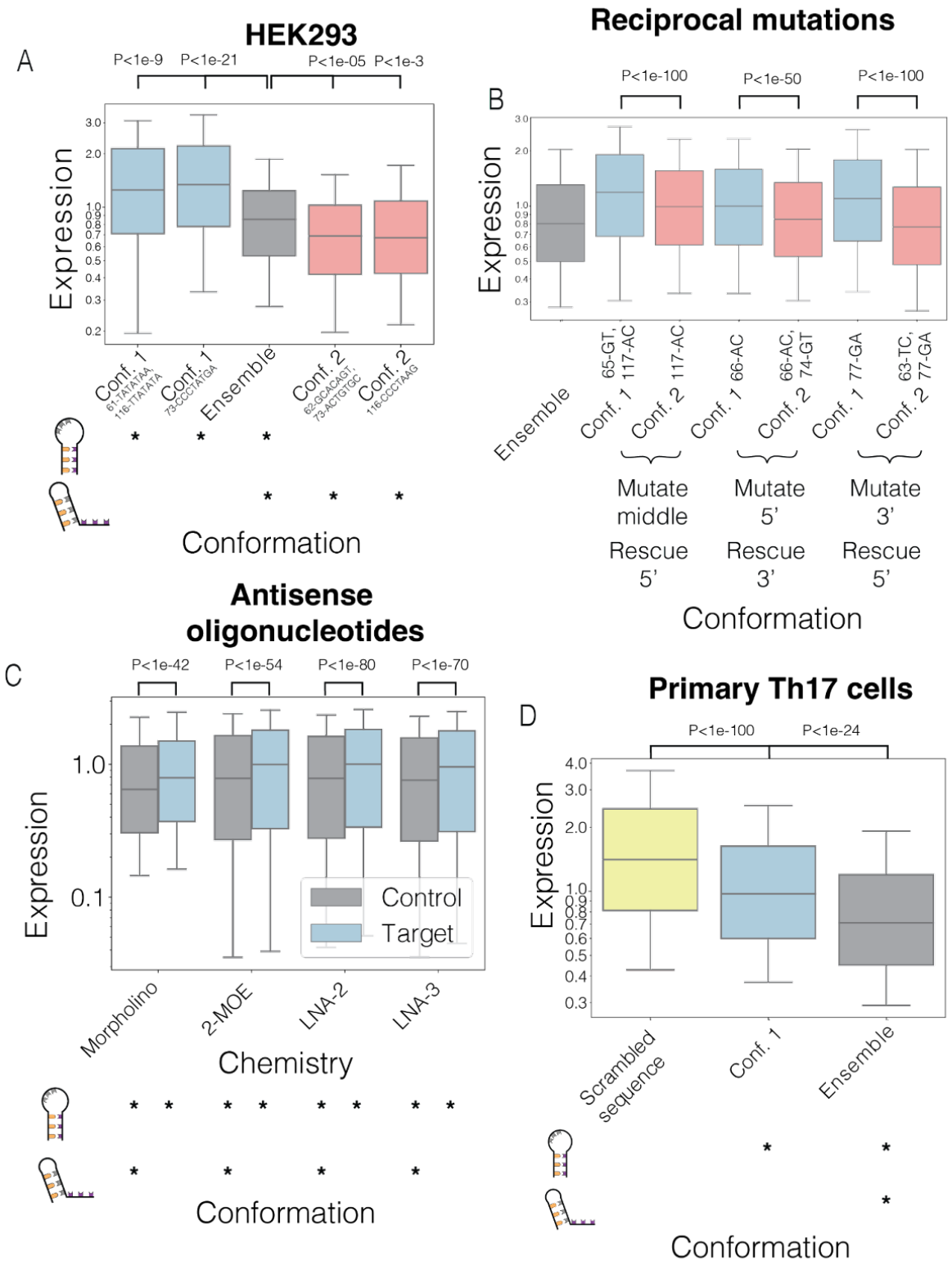
A
HEK293
P<1e-9
P<1e-21
P<1e-05
P<1e-3
Expression
Conf. 1
61-TATATAA,
116-TTATATA
Conf. 1
73-CCGTATGA
Ensemble
Conf. 2
62-GCACAGT,
73-ACTGTGC
Conf. 2
116-CCCTAAG
Conformation
B
Reciprocal mutations
P<1e-100
P<1e-50
P<1e-100
Expression
Ensemble
Conf. 1
65-GT,
117-AC
Conf. 2
117-AC
Conf. 1
66-AC
Conf. 2
66-AC,
74-GT
Conf. 1
77-GA
Conf. 2
63-TC,
77-GA
Mutate
middle
Rescue
5'
Mutate
5'
Rescue
3'
Mutate
3'
Rescue
5'
Conformation
C
Antisense
oligonucleotides
P<1e-42
P<1e-54
P<1e-80
P<1e-70
Expression
Control
Target
Morpholino
2-MOE
LNA-2
LNA-3
Chemistry
Conformation
D
Primary Th17 cells
P<1e-100
P<1e-24
Expression
Scrambled
sequence
Conf. 1
Ensemble
Conformation

Figure 2.5: The two alternative conformations of the RORC RNA switch have opposing effects on target gene expression.

(A-D) Bar plots of relative expression of the reporter construct. The relative expression values are calculated as the ratios of eGFP and mCherry fluorescence values for individual cells, as measured on a flow cytometer. Confidence intervals for the expression value are shown as calculated from the distribution of N>10.000 individual cells. The dominating conformations of RORC RNA switch are shown in color: conformation 1 - in blue, conformation 2 - in red, wild type - in gray, scrambled sequence with the same dinucleotide content - in yellow. P-values were calculated using the Student's T test.
(A) The effect of locking the RORC RNA switch in one or another conformation on reporter gene expression. Bottom: the scheme of the RNA secondary structure ensemble within each sample. The two alternative conformations are depicted as simplified stem-loops; their presence in each sample is shown by an asterisk. Each sample is a measurement of a HEK293 cell line with the reporter construct stably integrated in the genome. The mutations left to right: "61-TATATAA,116-TTATATA", "73-CCCTATGA", wildtype, "62-GCACAGT,73-ACTGTGC", "116-CCCTAAG".
(B) The effect of shift in the equilibrium between two conformations of the RORC switch on reporter gene expression. The mutation-rescue experiments were performed as shown in Fig. 2.3B. Bottom: the direction of mutation-rescue experiments is shown schematically. Top: the relative expression of the reporter upon the RORC switch sequence modifications. The dominating conformation of the switch within a sample is shown in color and labeled under the plot. The mutations left to right: wildtype, "65-GT,117-AC", "117-AC", "66-AC", "66-AC,74-GT", "77-GA", "63-TC,77-GA"
(C) The effect of antisense oligonucleotides (ASOs) on the reporter expression. A HEK293 cell line with stably integrated reporter construct encoding the RORC RNA switch in the 3'UTR was transfected by ASOs of 4 different chemistries. The targeting ASOs are complementary to the switching region of the RNA switch in the way that would result in a shift towards conformation 1 within the ensemble; the control ASOs have the same nucleotide composition but do not target the RORC RNA switch sequence. Bottom: the RNA switch conformations present in each sample. Top: the relative expression of the target gene upon the transfection of either targeting or control ASOs.
(D) The effect of locking the RORC RNA switch in conformation 1 in primary CD4+ T cells. Human CD4+ T cells were infected with lentiviral constructs carrying one of the three sequences in the reporter gene's 3'UTR. The dominating conformations within each sample are shown in color as described above. The mutations left to right: scrambled RORC RNA switch, "77-GA", wildtype.

Next, we investigated whether shifting the equilibrium between the two conformations by trans-acting agents, rather than sequence mutations, would have an effect on gene expression. We hypothesized that adding an antisense oligonucleotide (ASO) complementary to a part of the RNA

switch sequence could shift the equilibrium between the two conformations to change reporter gene expression. We designed a set of ASOs targeting the middle region of the switch. These ASOs were designed to shift the equilibrium towards conformation 1 and thereby increase eGFP expression. We transfected the stable cell line expressing the RNA switch-containing reporter with ASOs and measured the changes in eGFP expression by flow cytometry. To ensure this effect is not specific to a single ASO design, we varied the oligonucleotide chemistry by using either morpholino, or 2'-O-(2-Methoxyethyl) (2-MOE) oligoribonucleotides, or locked nucleic acids (LNA) as the key modifications. Additionally, we varied the sequence length and the frequency of modifications. In all cases, transfecting cells with an ASO targeting the middle region of the RNA switch resulted in higher eGFP expression compared to a non-targeting ASO of the same chemistry and nucleotide composition (Fig. 2.5D). Thus, the repressive activity of the RORC RNA switch can be alleviated with the use of trans agents such as ASOs.

The RORC gene encodes nuclear receptor ROR-gamma, and has two protein isoforms that differ by a short N-terminal sequence. The shorter isoform, RORγ, is expressed in many tissues, and is involved in circadian rhythms. The longer isoform, RORγt, is expressed in several subsets of T cells and some lymphoid cells, and is a key driver of Th17 cell type differentiation (Eberl 2017). Therefore, we also measured the activity of the RORC RNA switch in its endogenous cellular context. We assessed whether the conformation-dependent regulatory function of the RORC RNA switch is observed in Th17 cells. To do this, we infected primary CD4+ cells with lentivirus carrying the reporter and a sequence of interest in the eGFP 3’UTR. The cells were then differentiated into the Th17 cell type (Montoya and Ansel 2017). The presence of the wild type RORC RNA switch strongly decreased eGFP reporter expression compared to a scrambled version

of the same sequence (Fig. 2.5D). On the other hand, the 77-GA mutant decreases the strength of the repression, confirming the activity of the RORC RNA in Th17 cells.

Genome-wide screens reveal pathways downstream of the RORC RNA switch

To explore the molecular mechanisms underlying the effect of the RORC RNA switch on gene expression, we performed genome-wide CRISPRi screens with two distinct eGFP reporters (Fig. 2.6A). The first reporter was designed to identify *trans* factors epistatic to the repressive function of the RORC switch. For this, we created an eGFP construct carrying the wild type RORC RNA switch in its 3'UTR. The second reporter was engineered to assess conformation-dependent activity by inserting the 77-GA RORC switch mutant (favoring conformation 1) in the eGFP 3'UTR. Considering the importance of RORC in T cell biology, we chose the Jurkat T cell leukemia line as the model system for this screen. We infected both reporter cell lines with a genome-wide CRISPRi sgRNA library (Gilbert et al. 2014), sorted cells on a flow cytometer by eGFP expression, and collected the 25% of cells with highest and lowest eGFP expression (Fig. 2.6A, Fig. 2.S6) (de Boer et al. 2020). We hypothesized that knockdown of genes important for the repressive function of the RORC RNA switch would result in higher expression of the reporter gene. Similarly, genes involved in modulating the switch functionality of the two conformations would result in higher reporter expression for the wildtype switch compared to the 77-GA mutant.

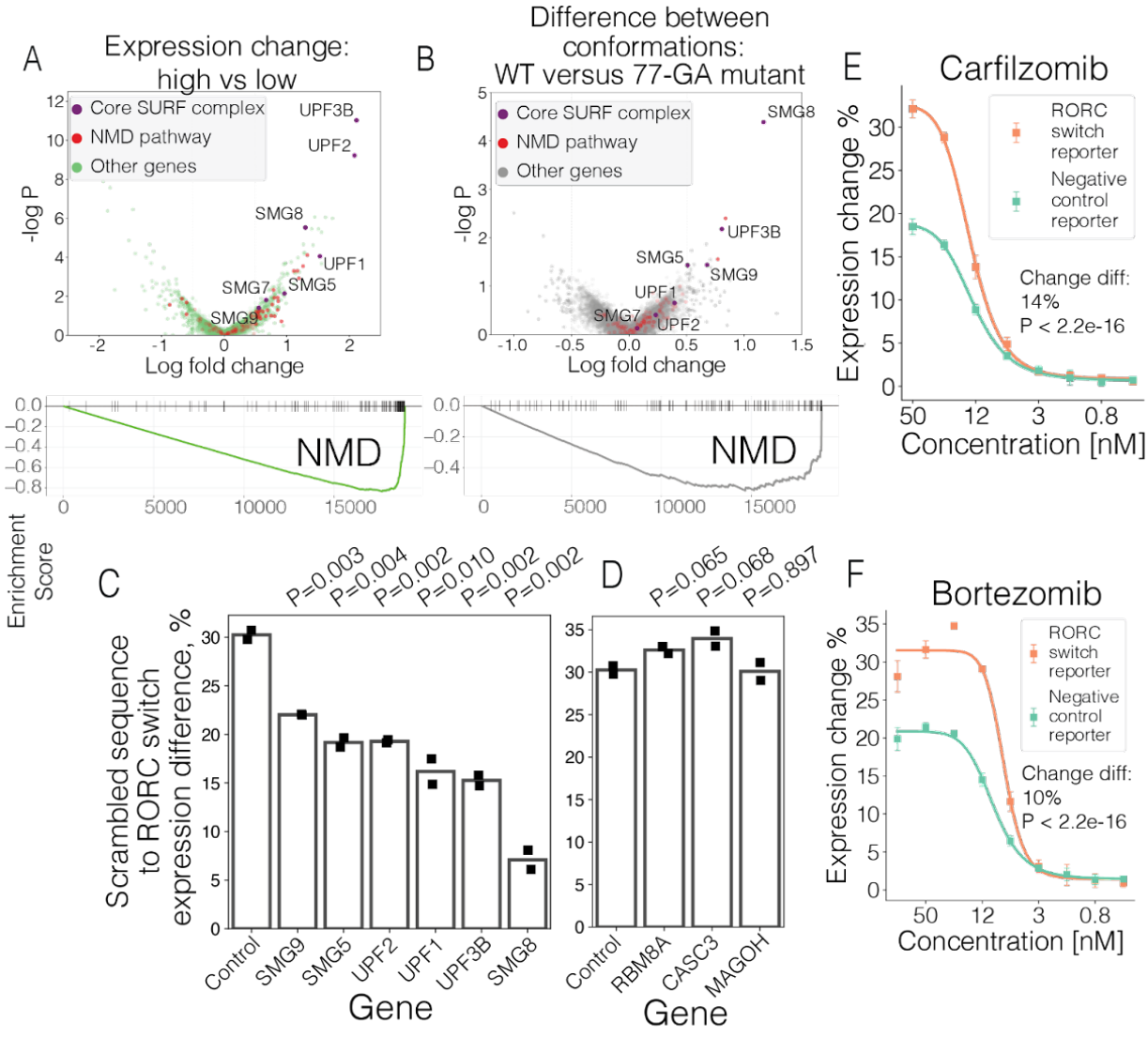
A
Expression change: high vs low
Core SURF complex
NMD pathway
Other genes
UPF3B
UPF2
SMG8
UPF1
SMG7
SMG5
SMG9
-log P
Log fold change
B
Difference between conformations: WT versus 77-GA mutant
Core SURF complex
NMD pathway
Other genes
SMG8
UPF3B
SMG5
SMG9
UPF1
SMG7
UPF2
-log P
Log fold change
NMD
NMD
Enrichment Score
C
P=0.003
P=0.004
P=0.002
P=0.010
P=0.002
P=0.002
Scrambled sequence to RORC switch expression difference, %
Control
SMG9
SMG5
UPF2
UPF1
UPF3B
SMG8
Gene
D
P=0.065
P=0.068
P=0.897
Control
RBM8A
CASC3
MAGOH
Gene
E
Carfilzomib
RORC switch reporter
Negative control reporter
Change diff: 14%
P < 2.2e-16
Expression change %
Concentration [nM]
F
Bortezomib
RORC switch reporter
Negative control reporter
Change diff: 10%
P < 2.2e-16
Expression change %
Concentration [nM]

Figure 2.6: Genome wide CRISPRi screen identifies SURF complex as acting downstream of the RORC RNA switch

(A) The comparison between the high expression and low expression quartiles. Top: volcano plot for gene-wise comparison. The comparison of sgRNA representation between the bottom and the top quantiles by expression (across both cell lines) is represented as a volcano plot. Genes, annotated as part of the nonsense mediated decay (NMD) pathway by GO, are colored in red. The core components of the canonical NMD pathway are colored in purple and labeled. Bottom: Gene Set Enrichment Analysis (GSEA) plot for NMD pathway for the above comparison.
(B) Comparison of ratios between top and bottom expression quantiles for the two cell lines. Higher values on the X axis indicate that sgRNAs targeting this gene have a stronger effect on reporter gene expression in the WT cell line compared to the RC cell line. Top: "ratio of ratios" comparison ("DESeq2 Testing Ratio of Ratios (RIP-Seq, CLIP-Seq, Ribosomal Profiling)" n.d.) represented as a volcano plot. Same color scheme is applied as in the C panel. Bottom: GSEA plot for NMD pathway for the above comparison.
(C) (C and D) The effect of knockdown of SURF and EJC complex member proteins on the expression change upon the conformation equilibrium shift. The individual genes were knocked down using the CRISPRi system in both "wild type" and "scrambled" cell lines, then the change of reporter gene expression was measured by flow cytometry (N replicates = 2). The bar plots demonstrate the expression ratios of WT to the scrambled sequence of RORC RNA switch. P-values were calculated using the Student's T test.
(E) The effect of EJC complex knockdown on the expression change upon the conformation equilibrium shift.
(E and F) The effect of proteasome inhibitors bortezomib and carfilzomib on the RNA switch-mediated expression change. K562 cell lines expressing GFP/mCherry double reporter either with or without RORC RNA switch in the 3'UTR of GFP were treated with the drugs for 24 hours, then the reporter gene expression (GFP signal normalized by mCherry signal) was measured by flow cytometry (N replicates = 4).
(H) The effect of proteasome inhibitor bortezomib on the RNA switch-mediated expression change.

To identify factors responsible for the repressive function of RORC RNA switch, we compared the abundance of sgRNAs in the cells with high reporter expression relative to those with lower expression (Fig. 2.6B). We also performed gene-set enrichment analysis (Korotkevich et al. 2021) to identify the key pathways involved. The most highly enriched pathway was nonsense-mediated decay (NMD), and several core NMD factors, such as SMG8, UPF1, UPF2, UPF3B were among the highest scoring hits of the screen (Fig. 2.6C, Fig. 2.S7. As expected, among other enriched pathways, many were associated with general gene expression such as translation, ribosome biogenesis, and endoplasmic reticulum stress. Next, we asked which factors are responsible for the

divergent activity of the two conformations. For this, we performed the ratio of ratios test ("DESeq2 Testing Ratio of Ratios (RIP-Seq, CLIP-Seq, Ribosomal Profiling)" n.d.), by comparing the ratios of sgRNA abundance in low and high expression cells between the two reporter screens: wildtype versus 77-GA mutant. Interestingly, this comparison also highlighted the NMD pathway as the key contributor (Fig. 2.6D). Among the highest scoring hits were several core NMD factors; all of them are part of the SURF complex, which is thought to recognize stalled ribosomes when a premature termination codon (PTC) occurs upstream of the exon-junction complex (EJC) (Yamashita 2013). Therefore, we hypothesized that the RORC RNA switch affects gene expression by acting through the EJC-independent NMD pathway, and that one conformation of the switch is more likely to act through this mechanism than the other.

To confirm this observation, we used CRISPRi to knock down the identified core NMD factors individually in the wild type and 77-GA mutant reporter lines, along with cells expressing a scrambled RORC RNA switch control ("scrambled" cell line). We first asked whether knocking down the NMD factors would affect the repressive function of the RNA switch. We compared the expression of eGFP in the wild type and scrambled reporter lines. We observed that silencing the core members of the SURF complex (Fig. 2.6E), but not of the EJC complex (Fig. 2.6F), affected the repressive function of the RORC RNA switch sequence on gene expression. Next, we asked whether knockdown of the NMD proteins would reduce the functional difference between the two switch conformations. by comparing eGFP expression in the wild type and the 77-GA mutant reporter lines. Consistently, we observed that knocking down the core members of the SURF complex (Fig. 2.S6E), and not of the EJC complex (Fig. 2.S5F), reduced this difference. This data demonstrates that NMD acts preferentially on the Conformation 2 of the RORC RNA switch.

The canonical NMD pathway causes the proteins translated from aberrant mRNA to be degraded by the proteasome (Kuroha, Tatematsu, and Inada 2009). Therefore, we expect the RORC RNA switch to target its gene product for degradation by the proteasome through NMD pathway recruitment. To confirm this, we treated two dual reporter cell lines with the proteasome inhibitor bortezomib, one line expressing the RORC RNA switch (the "wild type" line), and the other expressing the scrambled control (see Fig. 2.6D). As expected, proteasome inhibition resulted in a significantly larger change in eGFP expression in the "wild-type" line relative to the "scrambled" cell line. To verify that the observed effect is due to proteasome inhibition rather than a side effect of bortezomib, we treated cells with carfilzomib, another proteasome inhibitor which acts through a different mechanism, and observed the same effect as with bortezomib. Therefore, our data suggests that proteasomal degradation is involved in the repressive activity of the RORC RNA switch on gene expression.

DISCUSSION

Historically, RNA switch discovery has been tackled by one of two methods: comparative genomics analysis or biochemical experimentation. Comparative genomics analysis searches for conserved positions within non-coding RNA regions; it works well for identifying cis-regulatory elements in bacteria, such as RNA switches and transcription factor binding sites (Rodionov 2007). The biochemical approach involves measuring the affinity of a putative RNA switch to its ligands and analyzing the conformational change caused by the binding event. Both approaches were used to discover the first known RNA switches in bacteria (Epshtein, Mironov, and Nudler 2003) (Winkler, Nahvi, and Breaker 2002). However, no algorithms have been specifically designed to

search for RNA switches in eukaryotic transcriptomes. In Eukaryotes, mRNA secondary structure is highly dynamic. Multiple studies have shown that RNA structure vastly differs when measured *in vitro* vs *in vivo* (Rouskin et al. 2014), and that multiple cellular processes can rearrange mRNA secondary structures (Sun et al. 2019). Several studies have shown the functional importance of individual RNA structure rearrangements, such as RNA thermosensors (Shamovsky et al. 2006); however, the extent to which structural switches control gene expression in eukaryotes remains largely unexplored. There are several reasons why models trained on bacteria cannot be readily applied to higher eukaryotes. First, the sequence search space is much larger in eukaryotic transcriptomes compared to their bacterial counterparts, hindering the application of pre-trained models due to high false-positive counts (Ureta-Vidal, Ettwiller, and Birney 2003). Second, poor sequence conservation of many eukaryotic RNA regulatory elements limits the applicability of the comparative genomics analyses (Backofen et al. 2018; Leypold and Speicher 2021). Hence, the primary approach used for finding eukaryotic switches has been low-throughput biochemical characterization of candidate sequences (Breaker 2011; Serganov and Nudler 2013). Here, we have described an integrative and comprehensive platform for studying RNA switches in eukaryotic transcriptomes.

Recent advances in genomic technologies was a key contributor in our ability to carry out this systematic search for RNA switches. The development of RNA secondary structure probing techniques, such as DMS-seq and SHAPE-seq, has enabled researchers to move from only measuring the averaged structures of the folding ensemble to sampling multiple alternative conformations (Tomezsko et al. 2020; Morandi et al. 2021). Moreover, recent advances in single-particle cryo-EM and computational modeling have enabled structure determination of 3D folds

of some RNA molecules (Kappel et al. 2020), despite their small size and intrinsic flexibility. This opens up a prospect of studying the functional difference between alternative RNA conformations and their role in gene expression control. Our DMS-MapSeq and cryo-EM data suggest that the RORC 3' mRNA element inhabits a shallow energy landscape with two rugged minima linked to two major molecular conformations (Fig 2.7A). The kinetic barrier separating these conformations might itself contain one or more metastable plateaus reflecting partially folded intermediates in which some structural elements are transiently absent, perhaps explaining the presence of an additional tertiary conformation (Class 3). Mutations that stabilize one conformation or another, such as 77-GA and 117-AC, alter the energy landscape and thus the distribution of states. Experimental structure determination thus validates the SwFinder approach to identification of RNA molecules with these types of bistable energy landscapes. We also functionally characterized the RORC RNA switch using CRISPRi, finding that RORC switch control of gene expression is mediated by the SURF complex of the EJC-independent NMD. We propose that the SURF complex recognizes switch Conformation 2, but not Conformation 1, and that recruitment of the SURF complex by Conformation 2 in turn leads to decreased gene expression through proteasome-mediated degradation of any translation products, and mRNA decay preventing repeated rounds of translation (Fig. 2.7B). We show that point mutants which modulate the conformational equilibrium likewise cause small, but significant, incremental changes in the degree of gene repression by the switch. One consequence is that switch sequence variation can likely provide nearly continuous modulation of SURF recruitment and NMD activation.

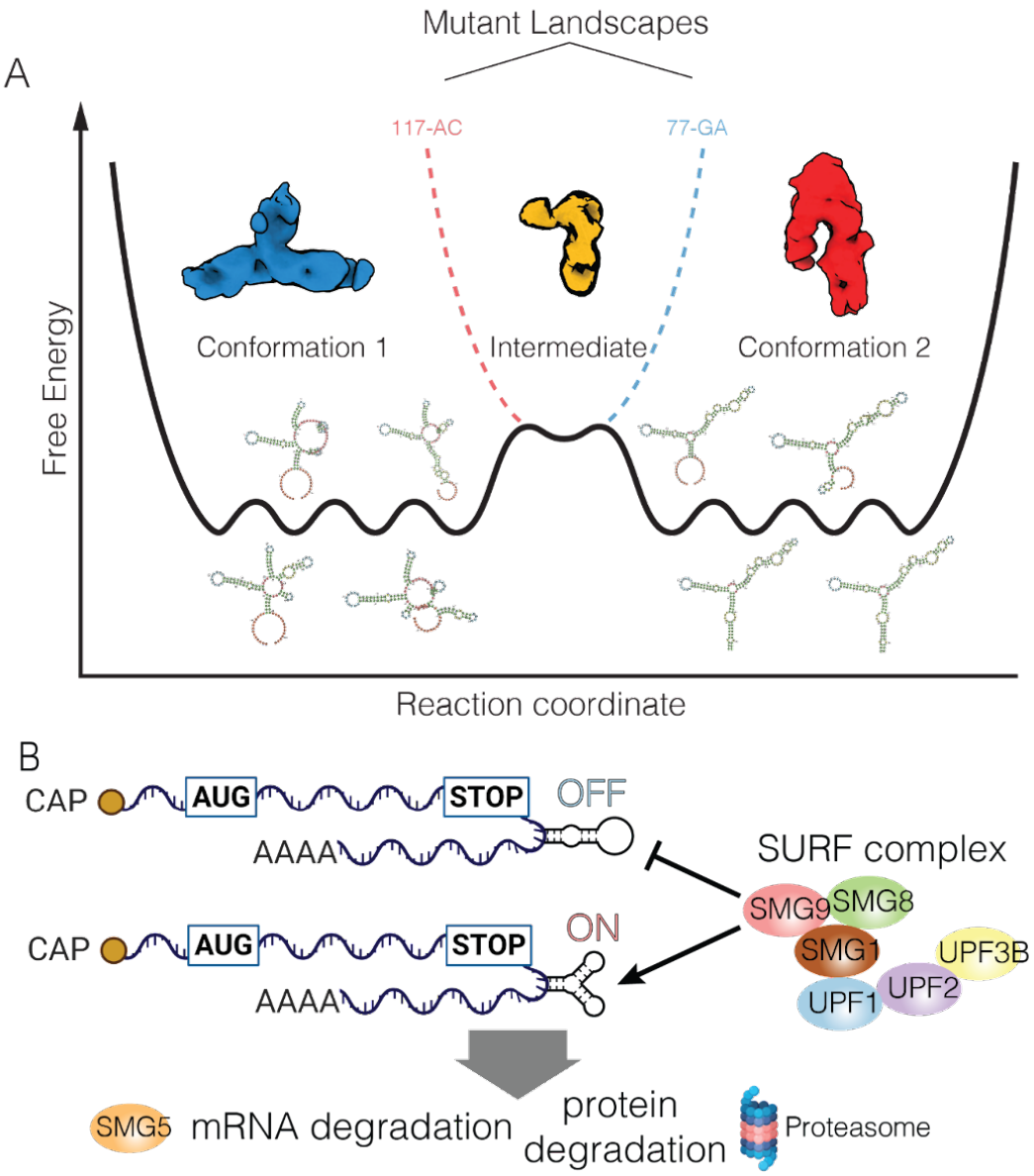
A
Mutant Landscapes
117-AC
77-GA
Conformation 1
Intermediate
Conformation 2
Free Energy
Reaction coordinate
B
CAP
AUG
STOP
OFF
AAAA
SURF complex
SMG9
SMG8
SMG1
UPF3B
UPF1
UPF2
CAP
AUG
STOP
ON
AAAA
SMG5
mRNA degradation
protein degradation
Proteasome

Figure 2.7: The proposed mechanism of RORC RNA switch functioning

(A) Schematic for a shallow energy landscape of the RORC 3' mRNA element. Shallow global minima characterizing the conformation 1 (cryo-EM Class A) and conformation 2 (cryo-EM Class B) structures themselves comprise multiple local minima in which various secondary structure elements fold or unfold while preserving overall tertiary structure and biological activity. These local minima are illustrated by secondary structure models for various DRACO cluster members. The two global minima are separated by a kinetic barrier that represents a partially folded intermediate (cryo-EM Class 3). The two dashed lines indicate alterations to the global landscape exhibited by the mutant sequences, red for the 77-GA mutant and blue for the 117-AC mutant. These altered landscapes eliminate one of the global minima without disrupting the intermediate.
(A) Proposed mechanism of RORC RNA switch. The RNA switch exists in an ensemble of two states. One of them is recognized by the SURF complex; such recognition triggers mRNA degradation (likely mediated by SMG5) and protein degradation (mediated by proteasome), thus affecting gene expression.

We have observed that a large number of regulatory elements in the human transcriptome act as RNA switches. Our findings also demonstrate that conformation-dependent modes of gene expression control may be ubiquitous in the human transcriptome. Regulatory information encoded in dynamic RNA structural elements adds an underexplored avenue of gene expression control with fundamental roles in health and disease, and capturing the regulatory grammar of RNA switches across the transcriptome is a key step towards a more complete understanding of post-transcriptional control of gene expression. In our study, we chose restrictive criteria for selection of RNA switches. We required these elements to be bistable *in vivo*, i.e. to populate two mutually exclusive structural conformations. However, this condition need not apply to all functional RNA switches, which could be bistable only under specific conditions or in specific cell types. RNA switches also likely function through a variety of different mechanisms. The known examples of human RNA switch mechanisms include mutually exclusive binding of RBPs by two different conformations (Ray et al. 2009) and m6A modification-based switching (Liu et al. 2015). We demonstrated a new RNA switch that acts through the NMD pathway. We speculate that other RNA switches similarly tap into the repertoire of known RNA metabolic pathways. For example,

we have recently shown that specific RNA structures cause aberrant splicing in metastatic cancers through binding SNRPA1 (Fish et al. 2021). Thus, future studies will likely find a wider variety of RNA switches than those discovered in the present study, under steady-state conditions. Identifying the regulatory programs that govern RNA secondary structure switching will lead to a mechanistic understanding of gene expression control. We expect this work will provide a basis for future studies of RNA switches in other contexts ranging from development and differentiation to various models of human disease.

METHODS

Cell culture

All cells were cultured in a 37°C 5% CO2 humidified incubator. The 293T cells (ATCC CRL-3216) were cultured in DMEM high-glucose medium supplemented with 10% FBS, glucose (4.5 g/L), L-glutamine (4 mM), sodium pyruvate (1 mM), penicillin (100 units/mL), streptomycin (100 μg/mL) and amphotericin B (1 μg/mL) (Gibco). The Jurkat cell line was cultured in RPMI-1640 medium supplemented with 10% FBS, glucose (2 g/L), L-glutamine (2 mM), 25 mM HEPES, penicillin (100 units/mL), streptomycin (100 μg/mL) and amphotericin B (1 μg/mL) (Gibco). All cell lines were routinely screened for mycoplasma with a PCR-based assay.

CRISPRi-mediated gene knockdown

Jurkat cells expressing dCas9-KRAB fusion protein were constructed by lentiviral delivery of pMH0006 (Addgene #135448) and FACS isolation of BFP-positive cells.

The lentiviral constructs were co-transfected with pCMV-dR8.91 and pMD2.D plasmids using TransIT-Lenti (Mirus) into 293T cells, following the manufacturer's protocol. Virus was harvested

48 hours post-transfection and passed through a 0.45 μm filter. Target cells were then transduced overnight with the filtered virus in the presence of 8 μg/mL polybrene (Millipore).

Guide RNA sequences for CRISPRi-mediated gene knockdown were cloned into pCRISPRia-v2 (Addgene #84832) via BstXI-BlpI sites. After transduction with sgRNA lentivirus, Jurkat cells were selected with 2 μg/mL puromycin (Gibco).

Cryo-EM Sample Preparation and Data Collection

3.5 μl of target mRNA at an approximate concentration of 1.5 mg/mL was applied to gold, 300 mesh TEM grids with a holey carbon substrate of 1.2/1.3 μm spacing (Quantifoil). The grids were blotted with #4 filter papers (Whatman) and plunge frozen in liquid ethane using a Mark IV Vitrobot (Thermo Fisher), with blot times of 4 - 6 s, blot force of -2, at a temperature of 8˚C and 100% humidity. All grids were glow discharged in an easiGlo (Pelco) with rarefied air for 30 s at 15 mA, no more than 1 hour prior to preparation. Duplicate WT and mutant RNA specimens were imaged under different conditions on several microscopes as per Table Y; all were equipped with K3 direct electron detector (DED) cameras (Gatan), and all data collection was performed using SerialEM[2].

Cryo-EM Image processing

Dose-weighted and motion-corrected sums were generated from raw DED movies on-the-fly during data collection using UCSF MotionCor2[3]. Images from super-resolution datasets were downsampled to the physical pixel size before further processing. CTF estimation was performed in CTFFIND4[4], followed by neural-net based particle picking in EMAN2[5]. 2D classification, *ab initio* 3D classification, and gold-standard refinement were done in cryoSPARC[6]. CTFs were then

re-estimated in cryoSPARC and particles repicked using low-resolution (20 Å) templates generated from chosen 3D classes. Extended datasets were pooled where appropriate, and particle processing was repeated through gold-standard refinement as before.

Reporter vector design and library cloning

First, mCherry ORF was cloned into the BTV backbone (Addgene #84771). Then, the vector was digested with MluI-HF and PacI restriction enzymes (NEB), with the addition of rSAP (NEB). The digested vector was purified with Zymo DNA Clean and Concentrator-5 kit.

DNA oligonucleotide libraries (one for Functional Screen and one for Conformation Expression Change Screen) consisting of 7500 sequences total were synthesized by Agilent. The second strand was synthesized as follows: the library DNA was digested with BmtI, then a UMI-containing primer (sequence CTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNCTAG) was used to initiate the second strand synthesis by Klenow Fragment (3′→ 5′ exo-) (NEB). The library was digested with MluI-HF and PacI restriction enzymes (NEB) and run on a 6% TBE polyacrylamide gel. The band of the corresponding size was cut out and the gel was dissolved in the DNA extraction buffer (10 mM Tris pH 8, 300 mM NaCl, 1 mM EDTA). The DNA was precipitated with isopropanol. The digested DNA library and the digested vector were ligated with T4 DNA ligase (NEB). The ligation reaction was precipitated with isopropanol and transformed into MegaX DH10B T1R Electrocompetent Cells (Thermo Fisher). The library was purified with ZymoPURE II Plasmid Maxiprep Kit (Zymo). The representation of individual sequences in the library was verified by sequencing the resulting library on MiSeq instrument (Illumina).

Massively Parallel Reporter Assay

The DNA library was co-transfected with pCMV-dR8.91 and pMD2.D plasmids using TransIT-Lenti (Mirus) into 293T cells, following the manufacturer's protocol. Virus was harvested 48 hours post-transfection and passed through a 0.45 μm filter. HEK293 cells were then transduced overnight with the filtered virus in the presence of 8 μg/mL polybrene (Millipore); the amount of virus used was optimized to ensure the infection rate of ~20%. The infected cells were selected with 2 μg/mL puromycin (Gibco). Cells were harvested at 90%–95% confluency for sorting and analysis on a BD FACSaria II sorter. The distribution of mCherry to GFP ratios was calculated. For sorting a library into subpopulations, we gated the population into 8 bins each containing 12.5% of the total number of cells. A total of 1.2 million cells were collected for each bin to ensure sufficient representation of sequence in the population in two replicates each. For each subpopulation, we extracted gDNA and total RNA with the Quick-DNA/RNA Miniprep kit. gDNA was amplified by PCR with Phusion polymerase (NEB) using primers CAAGCAGAAGACGGCATACGAGAT - i7 - GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCACTGCTAGCTAGATGACTAAACGCG and AATGATACGGCGACCACCGAGATCTACAC - i5 - ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGGTCTGGATCCACCGGTCC. Different i7 indices were used for 8 different bins, and different i5 indices were used for the two replicates. RNA was reverse transcribed with Maxima H Minus Reverse Transcriptase (Thermo Fisher) using primer CTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNTGGTCTGGATCCACCGGTCCGG. The cDNA was amplified with Q5 polymerase (NEB) using primers CAAGCAGAAGACGGCATACGAGAT - i7 -

GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCCTGCTAGCTAGATGACTAAACGC and CAAGCAGAAGACGGCATACGAGAT - i5 - GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTACCCGTCATTGGCTGTCCA. Different i7 indices were used for 8 different bins, and different i5 indices were used for the two replicates. The amplified DNA libraries were size purified with the Select-a-Size DNA Clean & Concentrator MagBead Kit (Zymo). Deep sequencing was performed using the HiSeq4000 platform (Illumina) at the UCSF Center for Advanced Technologies.

DMS-MaPseq

Briefly, HEK293 cells were incubated in culture with 1.5% DMS (Sigma) at room temperature for 7 minutes, the media was removed, and DMS was quenched with 30% BME. Total RNA from DMS-treated cells and untreated cells was then isolated using Trizol (Invitrogen). RNA was reverse transcribed using TGIRT-III reverse transcriptase (InGex) and target-specific primers. PCR was then performed to amplify the desired sequences and to add Illumina compatible adapters. The libraries were then sequenced on a MiSeq instrument using MiSeq micro kit v2, 300 cycles (Illumina). See Supplementary Table for oligo sequences used in library preparation.

Pear (v0.9.6) was used to merge the paired reads into a single combined read. The UMI was then removed from the reads and appended to read names using UMI tools (v1.0). The reads were then reverse complemented (fastx toolkit) and mapped to the amplicon sequences using bwa mem (v0.7). The resulting bam files were then sorted and deduplicated (umi_tools, with method flag set to unique). The alignments were then parsed for mutations (CTK). The mutation frequency at every position was then reported.

SHAPE chemical probing of RNAs

Chemical probing and mutate-and-map experiments were carried out as described previously (Palka et al. 2020). Briefly, 1.2 pmol of RNA was denaturated at 95°C in 50 mM Na-HEPES, pH 8.0, for 3 min, and folded by cooling to room temperature over 20 min, and adding MgCl2 to 10 mM concentration. RNA was aliquoted in 15 µL volumes into a 96-well plate and mixed with nuclease-free H2O (control), or chemically modified in the presence of 5 mM 1-methyl-7-nitroisatoic anhydride (1M7) (Turner, Shefer, and Ares 2013), for 10 min at room temperature. Chemical modification was stopped by adding 9.75 µL quench and purification mix (1.53 M NaCl, 1.5 µL washed oligo-dT beads, Ambion), 6.4 nM FAM-labeled, reverse-transcriptase primer (/56-FAM/AAAAAAAAAAAAAAAAAAAAGTTGTTCTTGTTGTTTCTTT), and 2.55 M Na-MES. RNA in each well was purified by bead immobilization on a magnetic rack and two washes with 100 µL 70% ethanol. RNA was then resuspended in 2.5 µL nuclease-free water prior to reverse transcription.

RNA was reverse-transcribed from annealed fluorescent primer in a reaction containing 1× First Strand Buffer (Thermo Fisher), 5 mM DTT, 0.8 mM dNTP mix, and 20 U of SuperScript III Reverse Transcriptase (Thermo Fisher) at 48°C for 30 min. RNA was hydrolyzed in the presence of 200 mM NaOH at 95°C for 3 min, then placed on ice for 3 min and quenched with 1 volume 5 M NaCl, 1 volume 2 M HCl, and 1 volume 3 M sodium acetate. cDNA was purified on magnetic beads, then eluted by incubation for 20 min in 11 µL Formamide-ROX350 mix (1000 µL Hi-Di Formamide (Thermo Fisher) and 8 µL ROX350 ladder (Thermo Fisher). Samples were then transferred to a 96-well plate in “concentrated” (4 µL sample + 11 µL ROX mix) and “dilute” (1 µL sample + 14 µL ROX mix) for saturation correction in downstream analysis. Sample plates were sent to Elim Biopharmaceuticals for analysis by capillary electrophoresis.

CRISPRi screen

Reporter screens were conducted using established flow cytometry screen protocols (Gilbert et al. 2014)(Horlbeck et al., 2016; Sidrauski et al., 2015). Jurkat cells with previously verified CRISPRi activity were used (Horlbeck et al., 2018). An optimal MOI (<0.3) of the collected library virus to the jurkat cells was determined by lentiviral titration.The CRISPRi-v2 (5 sgRNA/TSS, Addgene: Cat#83969) sgRNA library was then transduced into Jurkat cells at an MOI < 0.3 (BFP+ cell percentages were ~30%). For the flow based CRISPRi screen with the Jurkat cells, the sgRNA library virus was transfected at an average of 500x coverage after transduction (Day 0). Puromycin (1 µg/mL) selection for positively-transduced cells was performed 48 hours (Day 2) and 72 hours (Day 3) post transduction (Day 3). On Day 11, cells were collected in PBS and sorted with the BD FACSAria™ Fusion cell sorter. Cells were gated into the 25% of cells with the highest ratio between GFP and mCherry fluorescence intensity and 25% of cells with the lowest ratio. The screens were performed with two conditions: cells with a wildtype RORC Riboswitch-GFP reporter and a mutated Riboswitch reporter. Screens were additionally performed in duplicate. After sorting, genomic DNA was harvested (Macherey-Nagel Midi Prep kit) and amplified using NEB Next Ultra II Q5 master mix and primers containing TruSeq Indexes for NGS analysis. Sample libraries were prepared and sequenced on a HiSeq 4000. Guides were then quantified with the published ScreenProcessing (https://github.com/mhorlbeck/ScreenProcessing) method and phenotypes generated with an in-house processing pipeline, iAnalyzer (https://github.com/goodarzilab/iAnalyzer). Briefly, iAnalyzer relies on fitting a generalized linear model to each gene. Coefficients from this GLM were z-score normalized to the negative control guides and finally the largest coefficients were analyzed as potential hits.

T cell isolation, transduction, and Th17 cells differentiation

Th17 cells were derived as described previously (Montoya and Ansel 2017). Plates were coated with 2 µg/mL anti-human CD3 (UCSF monoclonal antibody core, clone: OKT-3) and 4 µg/mL anti-human CD28 (UCSF monoclonal antibody core, clone: 9.3) in PBS with calcium and magnesium for at least 2 h at 37 °C or overnight at 4 °C with plate wrapped in parafilm. Human CD4+ T cells were isolated from human peripheral blood using EasySep human CD4+ T cell isolation kit (17952; STEMCELL) and stimulated in ImmunoCult-XF T cell expansion medium (10981; STEMCELL) supplemented with 10 mM HEPES, 2 mM L-glutamine, 100 µM 2-ME, 1 mM sodium pyruvate, and 10 ng/ml TGF-β. 24 h after T cell isolation and initial stimulation on a 96-well plate, 7 ul of lentivirus was added to each sample. After 24 h, the media was removed from each sample without disturbing the cells and replaced with 200 µl fresh media. After 48 h, cells were stimulated with 1.2 µM ionomycin, 25 nM PMA, and 6 µg/ml brefeldin-A, resuspended by pipetting, incubated for 4 h at 37 °C, and harvested for analysis.

Analysis of capillary electrophoresis data with HiTRACE

Capillary electrophoresis runs from chemical probing and mutate-and-map experiments were analyzed with the HiTRACE MATLAB package (Yoon et al. 2011). Lanes were aligned together, bands fit to Gaussian peaks, background subtracted using the no-modification lane, corrected for signal attenuation, and normalized to the internal hairpin control. The end result of these steps is a numerical array of “reactivity” values for each RNA nucleotide that can be used as weights in structure prediction.

References

Åberg, Karolina, Peter Saetre, Niclas Jareborg, and Elena Jazin. 2006. “Human QKI, a Potential Regulator of mRNA Expression of Human Oligodendrocyte-Related Genes Involved in Schizophrenia.” Proceedings of the National Academy of Sciences 103 (19): 7482–87.

Abreu-Goodger, Cei, and Enrique Merino. 2005. “RibEx: A Web Server for Locating Riboswitches and Other Conserved Bacterial Regulatory Elements.” Nucleic Acids Research 33 (Web Server issue): W690–92.

Adamson, Britt, Thomas M. Norman, Marco Jost, Min Y. Cho, James K. Nuñez, Yuwen Chen, Jacqueline E. Villalta, et al. 2016. “A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response.” Cell 167 (7): 1867–82.e21.

Agrawal, Saumya, Tanvir Alam, Masaru Koido, Ivan V. Kulakovskiy, Jessica Severin, Imad Abugessaisa, Andrey Buyan, et al. 2021. “Functional Annotation of Human Long Noncoding RNAs Using Chromatin Conformation Data.” bioRxiv. https://doi.org/10.1101/2021.01.13.426305.

Antonicka, Hana, Zhen-Yuan Lin, Alexandre Janer, Mari J. Aaltonen, Woranontee Weraarpachai, Anne-Claude Gingras, and Eric A. Shoubridge. 2020. “A High-Density Human Mitochondrial Proximity Interaction Network.” Cell Metabolism. https://doi.org/10.1016/j.cmet.2020.07.017.

Antonicka, Hana, and Eric A. Shoubridge. 2015. “Mitochondrial RNA Granules Are Centers for Posttranscriptional RNA Processing and Ribosome Biogenesis.” Cell Reports 10 (6): 920–32.

Backofen, Rolf, Jan Gorodkin, Ivo L. Hofacker, and Peter F. Stadler. 2018. “Comparative RNA Genomics.” In Comparative Genomics: Methods and Protocols, edited by João C. Setubal, Jens Stoye, and Peter F. Stadler, 363–400. New York, NY: Springer New York.

Ballarino, M., L. Jobert, D. Dembélé, P. de la Grange, D. Auboeuf, and L. Tora. 2013. “TAF15 Is Important for Cellular Proliferation and Regulates the Expression of a Subset of Cell Cycle Genes through miRNAs.” Oncogene 32 (39): 4646–55.

Barak, Tanyeri, Emma Ristori, A. Gulhan Ercan-Sencicek, Danielle F. Miyagishima, Carol Nelson-Williams, Weilai Dong, Sheng Chih Jin, et al. 2021. “PPIL4 Is Essential for Brain Angiogenesis and Implicated in Intracranial Aneurysms in Humans.” Nature Medicine 27 (12): 2165–75.

Barsacchi, Marco, Eva Maria Novoa, Manolis Kellis, and Alessio Bechini. 2016. “SwiSpot: Modeling Riboswitches by Spotting out Switching Sequences.” Bioinformatics 32 (21): 3252–59.

Basak, Anindita, Mathias Munschauer, Caleb A. Lareau, Kara E. Montbleau, Jacob C. Ulirsch, Christina R. Hartigan, Monica Schenone, et al. 2020. “Control of Human Hemoglobin Switching by LIN28B-Mediated Regulation of BCL11A Translation.” Nature Genetics. https://doi.org/10.1038/s41588-019-0568-7.

Beaudoin, Jean-Denis, Eva Maria Novoa, Charles E. Vejnar, Valeria Yartseva, Carter M. Takacs, Manolis Kellis, and Antonio J. Giraldez. 2018. “Analyses of mRNA Structure Dynamics Identify Embryonic Gene Regulatory Programs.” Nature Structural & Molecular Biology 25 (8): 677–86.

Bengert, Peter, and Thomas Dandekar. 2004. “Riboswitch Finder--a Tool for Identification of Riboswitch RNAs.” Nucleic Acids Research 32 (Web Server issue): W154–59.

Bertolotti, A., Y. Lutz, D. J. Heard, P. Chambon, and L. Tora. 1996. “hTAF(II)68, a Novel RNA/ssDNA-Binding Protein with Homology to the pro-Oncoproteins TLS/FUS and EWS Is Associated with Both TFIID and RNA Polymerase II.” The EMBO Journal. https://doi.org/10.1002/j.1460-2075.1996.tb00882.x.

Boer, Carl G. de, John P. Ray, Nir Hacohen, and Aviv Regev. 2020. “MAUDE: Inferring Expression Changes in Sorting-Based CRISPR Screens.” Genome Biology 21 (1): 134.

Box, Joseph K., Nicolas Paquet, Mark N. Adams, Didier Boucher, Emma Bolderson, Kenneth J. O’Byrne, and Derek J. Richard. 2016. “Nucleophosmin: From Structure and Function to Disease Development.” BMC Molecular Biology 17 (1): 19.

Breaker, Ronald R. 2011. “Prospects for Riboswitch Discovery and Analysis.” Molecular Cell 43 (6): 867–79.

Chang, Tzu-Hao, Hsien-Da Huang, Li-Ching Wu, Chi-Ta Yeh, Baw-Jhiune Liu, and Jorng-Tzong Horng. 2009. “Computational Identification of Riboswitches Based on RNA Conserved Functional Sequences and Conformations.” RNA 15 (7): 1426–30.

Chen, Xinyun, Ying Liu, Chen Xu, Lina Ba, Zhuo Liu, Xiuya Li, Jie Huang, et al. 2021. “QKI Is a Critical Pre-mRNA Alternative Splicing Regulator of Cardiac Myofibrillogenesis and Contractile Function.” Nature Communications 12 (1): 89.

Chen, Xinyun, Jianwen Yin, Dayan Cao, Deyong Xiao, Zhongjun Zhou, Ying Liu, and Weinian Shou. 2021. “The Emerging Roles of the RNA Binding Protein QKI in Cardiovascular Development and Function.” Frontiers in Cell and Developmental Biology 9 (June): 668659.

Cho, Kelvin F., Tess C. Branon, Namrata D. Udeshi, Samuel A. Myers, Steven A. Carr, and Alice Y. Ting. 2020. “Proximity Labeling in Mammalian Cells with TurboID and Split-TurboID.” Nature Protocols 15 (12): 3971–99.

Cooke, Amy, Thomas Schwarzl, Ina Huppertz, Gertjan Kramer, Panagiotis Mantas, Anne-Marie Alleaume, Wolfgang Huber, Jeroen Krijgsveld, and Matthias W. Hentze. 2019. “The RNA-Binding Protein YBX3 Controls Amino Acid Levels by Regulating SLC mRNA Abundance.” Cell Reports. https://doi.org/10.1016/j.celrep.2019.05.039.

Corces, M. Ryan, Alexandro E. Trevino, Emily G. Hamilton, Peyton G. Greenside, Nicholas A. Sinnott-Armstrong, Sam Vesuna, Ansuman T. Satpathy, et al. 2017. “An Improved ATAC-Seq Protocol Reduces Background and Enables Interrogation of Frozen Tissues.” Nature Methods 14 (10): 959–62.

Cox, Jürgen, Marco Y. Hein, Christian A. Luber, Igor Paron, Nagarjuna Nagaraj, and Matthias Mann. 2014. “Accurate Proteome-Wide Label-Free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ.” Molecular & Cellular Proteomics: MCP 13 (9): 2513–26.

Datlinger, Paul, André F. Rendeiro, Christian Schmidl, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda C. Schuster, Amelie Kuchler, Donat Alpar, and Christoph Bock. 2017. “Pooled CRISPR Screening with Single-Cell Transcriptome Readout.” Nature Methods 14 (3): 297–301.

DeJong, Caitlin S., Darwin S. Dichmann, Cameron R. T. Exner, Yuxiao Xu, and Richard M. Harland. 2021. “The Atypical RNA-Binding Protein Taf15 Regulates Dorsoanterior Neural Development through Diverse Mechanisms in Xenopus Tropicalis.” Development 148 (15). https://doi.org/10.1242/dev.191619.

Deng, Liping, Ruotong Ren, Zunpeng Liu, Moshi Song, Jingyi Li, Zeming Wu, Xiaoqing Ren, et al. 2019. “Stabilizing Heterochromatin by DGCR8 Alleviates Senescence and Osteoarthritis.” Nature Communications 10 (1): 3329.

“DESeq2 Testing Ratio of Ratios (RIP-Seq, CLIP-Seq, Ribosomal Profiling).” n.d. Accessed June 16, 2022. https://support.bioconductor.org/p/61509/.

Ding, Ye, and Charles E. Lawrence. 2003. “A Statistical Sampling Algorithm for RNA Secondary Structure Prediction.” Nucleic Acids Research 31 (24): 7280–7301.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. “STAR: Ultrafast Universal RNA-Seq Aligner.” Bioinformatics 29 (1): 15–21.

Dotu, Ivan, William A. Lorenz, Pascal Van Hentenryck, and Peter Clote. 2010. “Computing Folding Pathways between RNA Secondary Structures.” Nucleic Acids Research 38 (5): 1711–22.

Dunn, Joshua G., and Jonathan S. Weissman. 2016. “Plastid: Nucleotide-Resolution Analysis of next-Generation Sequencing and Genomics Data.” BMC Genomics 17 (1): 958.

Eberl, G. 2017. “RORγt, a Multitask Nuclear Receptor at Mucosal Surfaces.” Mucosal Immunology 10 (1): 27–34.

Egorov, Artyom A., Ekaterina A. Sakharova, Aleksandra S. Anisimova, Sergey E. Dmitriev, Vadim N. Gladyshev, and Ivan V. Kulakovskiy. 2019. “svist4get: A Simple Visualization Tool for Genomic Tracks from Sequencing Experiments.” BMC Bioinformatics 20 (1): 113.

Epshtein, Vitaly, Alexander S. Mironov, and Evgeny Nudler. 2003. “The Riboswitch-Mediated Control of Sulfur Metabolism in Bacteria.” Proceedings of the National Academy of Sciences of the United States of America 100 (9): 5052–56.

Escobar-Hoyos, Luisa F., Alex Penson, Ram Kannan, Hana Cho, Chun-Hao Pan, Rohit K. Singh, Lisa H. Apken, et al. 2020. “Altered RNA Splicing by Mutant p53 Activates Oncogenic RAS Signaling in Pancreatic Cancer.” Cancer Cell 38 (2): 198–211.e8.

Fiesel, Fabienne C., and Philipp J. Kahle. 2011. “TDP-43 and FUS/TLS: Cellular Functions and Implications for Neurodegeneration.” The FEBS Journal 278 (19): 3550–68.

Fish, Lisa, Matvei Khoroshkin, Albertas Navickas, Kristle Garcia, Bruce Culbertson, Benjamin Hänisch, Steven Zhang, et al. 2021. “A Prometastatic Splicing Program Regulated by SNRPA1 Interactions with Structured RNA Elements.” Science 372 (6543). https://doi.org/10.1126/science.abc7531.

Fish, Lisa, Albertas Navickas, Bruce Culbertson, Yichen Xu, Hoang C. B. Nguyen, Steven Zhang, Myles Hochman, et al. 2019. “Nuclear TARBP2 Drives Oncogenic Dysregulation of RNA Splicing and Decay.” Molecular Cell 75 (5): 967–81.e9.

Fish, Lisa, Nora Pencheva, Hani Goodarzi, Hien Tran, Mitsukuni Yoshida, and Sohail F. Tavazoie. 2016. “Muscleblind-like 1 Suppresses Breast Cancer Metastatic Colonization and Stabilizes Metastasis Suppressor Transcripts.” Genes & Development 30 (4): 386–98.

Folco, Eric G., Chung-Sheng Lee, Kobina Dufu, Tomohiro Yamazaki, and Robin Reed. 2012. “The Proteins PDIP3 and ZC11A Associate with the Human TREX Complex in an ATP-Dependent Manner and Function in mRNA Export.” PloS One 7 (8): e43804.

Fuchs, Gilad, Yoav Voichek, Michal Rabani, Sima Benjamin, Shlomit Gilad, Ido Amit, and Moshe Oren. 2015. “Simultaneous Measurement of Genome-Wide Transcription Elongation Speeds and Rates of RNA Polymerase II Transition into Active Elongation with 4sUDRB-Seq.” Nature Protocols 10 (4): 605–18.

Gayoso, Adam, Romain Lopez, Galen Xing, Pierre Boyeau, Valeh Valiollah Pour Amiri, Justin Hong, Katherine Wu, et al. 2022. “A Python Library for Probabilistic Analysis of Single-Cell Omics Data.” Nature Biotechnology 40 (2): 163–66.

George, E. O., and G. S. Mudholkar. 1983. “On the Convolution of Logistic Random Variables.” Metrika 30 (1): 1–13.

Georges, Amedee des, Vidya Dhote, Lauriane Kuhn, Christopher U. T. Hellen, Tatyana V. Pestova, Joachim Frank, and Yaser Hashem. 2015. “Structure of Mammalian eIF3 in the Context of the 43S Preinitiation Complex.” Nature 525 (7570): 491–95.

Gilbert, Luke A., Max A. Horlbeck, Britt Adamson, Jacqueline E. Villalta, Yuwen Chen, Evan H. Whitehead, Carla Guimaraes, et al. 2014. “Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation.” Cell 159 (3): 647–61.

Gingras, Anne-Claude, Kento T. Abe, and Brian Raught. 2019. “Getting to Know the Neighborhood: Using Proximity-Dependent Biotinylation to Characterize Protein Complexes and Map Organelles.” Current Opinion in Chemical Biology 48 (February): 44–54.

Go, Christopher D., James D. R. Knight, Archita Rajasekharan, Bhavisha Rathod, Geoffrey G. Hesketh, Kento T. Abe, Ji-Young Youn, et al. 2021. “A Proximity-Dependent Biotinylation Map of a Human Cell.” Nature 595 (7865): 120–24.

Goodarzi, Hani, Olivier Elemento, and Saeed Tavazoie. 2009. “Revealing Global Regulatory Perturbations across Human Cancers.” Molecular Cell 36 (5): 900–911.

Grandi, Fiorella C., Hailey Modi, Lucas Kampman, and M. Ryan Corces. 2022. “Chromatin Accessibility Profiling by ATAC-Seq.” Nature Protocols, April. https://doi.org/10.1038/s41596-022-00692-9.

Grau, Jan, Ivo Grosse, and Jens Keilwagen. 2015. “PRROC: Computing and Visualizing Precision-Recall and Receiver Operating Characteristic Curves in R.” Bioinformatics 31 (15): 2595–97.

Hill, Andrew J., José L. McFaline-Figueroa, Lea M. Starita, Molly J. Gasperini, Kenneth A. Matreyek, Jonathan Packer, Dana Jackson, Jay Shendure, and Cole Trapnell. 2018. “On the Design of CRISPR-Based Single-Cell Molecular Screens.” Nature Methods 15 (4): 271–74.

Hinrichs, A. S., D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, et al. 2006. “The UCSC Genome Browser Database: Update 2006.” Nucleic Acids Research 34 (Database issue): D590–98.

Hung, Victoria, Namrata D. Udeshi, Stephanie S. Lam, Ken H. Loh, Kurt J. Cox, Kayvon Pedram, Steven A. Carr, and Alice Y. Ting. 2016. “Spatially Resolved Proteomic Mapping in Living Cells with the Engineered Peroxidase APEX2.” Nature Protocols 11 (3): 456–75.

Hung, Victoria, Peng Zou, Hyun-Woo Rhee, Namrata D. Udeshi, Valentin Cracan, Tanya Svinkina, Steven A. Carr, Vamsi K. Mootha, and Alice Y. Ting. 2014. “Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging.” Molecular Cell 55 (2): 332–41.

Ibrahim, Fadia, Manolis Maragkakis, Panagiotis Alexiou, Margaret A. Maronski, Marc A. Dichter, and Zissimos Mourelatos. 2013. “Identification of in Vivo, Conserved, TAF15 RNA Binding Sites Reveals the Impact of TAF15 on the Neuronal Transcriptome.” Cell Reports 3 (2): 301–8.

Ingolia, Nicholas T., Gloria A. Brar, Silvia Rouskin, Anna M. McGeachy, and Jonathan S. Weissman. 2012. “The Ribosome Profiling Strategy for Monitoring Translation in Vivo by Deep Sequencing of Ribosome-Protected mRNA Fragments.” Nature Protocols 7 (8): 1534–50.

Jobert, Laure, Manuela Argentini, and László Tora. 2009. “PRMT1 Mediated Methylation of TAF15 Is Required for Its Positive Gene Regulatory Function.” Experimental Cell Research 315 (7): 1273–86.

Käll, Lukas, John D. Storey, and William Stafford Noble. 2008. “Non-Parametric Estimation of Posterior Error Probabilities Associated with Peptides Identified by Tandem Mass Spectrometry.” Bioinformatics 24 (16): i42–48.

Kalvari, Ioanna, Eric P. Nawrocki, Nancy Ontiveros-Palacios, Joanna Argasinska, Kevin Lamkiewicz, Manja Marz, Sam Griffiths-Jones, et al. 2021. “Rfam 14: Expanded Coverage of Metagenomic, Viral and microRNA Families.” Nucleic Acids Research 49 (D1): D192–200.

Kapeli, Katannya, Gabriel A. Pratt, Anthony Q. Vu, Kasey R. Hutt, Fernando J. Martinez, Balaji Sundararaman, Ranjan Batra, et al. 2016. “Distinct and Shared Functions of ALS-Associated Proteins TDP-43, FUS and TAF15 Revealed by Multisystem Analyses.” Nature Communications 7 (July): 12143.

Kappel, Kalli, Kaiming Zhang, Zhaoming Su, Andrew M. Watkins, Wipapat Kladwang, Shanshan Li, Grigore Pintilie, et al. 2020. “Accelerated Cryo-EM-Guided Determination of Three-Dimensional RNA-Only Structures.” Nature Methods 17 (7): 699–707.

Katz, Yarden, Eric T. Wang, Edoardo M. Airoldi, and Christopher B. Burge. 2010. “Analysis and Design of RNA Sequencing Experiments for Identifying Isoform Regulation.” Nature Methods. https://doi.org/10.1038/nmeth.1528.

Kent, W. J., A. S. Zweig, G. Barber, A. S. Hinrichs, and D. Karolchik. 2010. “BigWig and BigBed: Enabling Browsing of Large Distributed Datasets.” Bioinformatics 26 (17): 2204–7.

Kim, Daehwan, Joseph M. Paggi, Chanhee Park, Christopher Bennett, and Steven L. Salzberg. 2019. “Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype.” Nature Biotechnology 37 (8): 907–15.

Kim, Dae In, Samuel C. Jensen, Kyle A. Noble, Birendra Kc, Kenneth H. Roux, Khatereh Motamedchaboki, and Kyle J. Roux. 2016. “An Improved Smaller Biotin Ligase for BioID Proximity Labeling.” Molecular Biology of the Cell 27 (8): 1188–96.

Korotkevich, Gennady, Vladimir Sukhov, Nikolay Budin, Boris Shpak, Maxim N. Artyomov, and Alexey Sergushichev. 2021. “Fast Gene Set Enrichment Analysis.” bioRxiv. https://doi.org/10.1101/060012.

Kugel, Jennifer F., and James A. Goodrich. 2009. “In New Company: U1 snRNA Associates with TAF15.” EMBO Reports.

Kuroha, Kazushige, Tsuyako Tatematsu, and Toshifumi Inada. 2009. “Upf1 Stimulates Degradation of the Product Derived from Aberrant Messenger RNA Containing a Specific Nonsense Mutation by the Proteasome.” EMBO Reports 10 (11): 1265–71.

Lai, William K. M., Luca Mariani, Gerson Rothschild, Edwin R. Smith, Bryan J. Venters, Thomas R. Blanda, Prashant K. Kuntala, et al. 2021. “A ChIP-Exo Screen of 887 Protein Capture Reagents Program Transcription Factor Antibodies in Human Cells.” Genome Research 31 (9): 1663–79.

Lee, Jin, Grey Christoforo, Grey Christoforo, C. S. Foo, Chris Probert, Anshul Kundaje, Nathan Boley, kohpangwei, Daniel Kim, and Mike Dacre. 2016. Kundajelab/atac_dnase_pipelines: 0.3.0. https://doi.org/10.5281/zenodo.156534.

Leek, Jeffrey T., W. Evan Johnson, Hilary S. Parker, Andrew E. Jaffe, and John D. Storey. 2012. “The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments.” Bioinformatics 28 (6): 882–83.

Leypold, Nicole A., and Michael R. Speicher. 2021. “Evolutionary Conservation in Noncoding Genomic Regions.” Trends in Genetics: TIG 37 (10): 903–18.

Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform.” Bioinformatics 25 (14): 1754–60.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence Alignment/Map Format and SAMtools.” Bioinformatics 25 (16): 2078–79.

Lin, Xiaoyan, Jill W. Miller, Ami Mankodi, Rahul N. Kanadia, Yuan Yuan, Richard T. Moxley, Maurice S. Swanson, and Charles A. Thornton. 2006. “Failure of MBNL1-Dependent Post-Natal Splicing Transitions in Myotonic Dystrophy.” Human Molecular Genetics 15 (13): 2087–97.

Li, Peipei, Jingjing Li, Li Wang, and Li-Jun Di. 2017. “Proximity Labeling of Interacting Proteins: Application of BioID as a Discovery Tool.” Proteomics 17 (20). https://doi.org/10.1002/pmic.201700002.

Liu, Nian, Qing Dai, Guanqun Zheng, Chuan He, Marc Parisien, and Tao Pan. 2015. “N(6)-Methyladenosine-Dependent RNA Structural Switches Regulate RNA-Protein Interactions.” Nature 518 (7540): 560–64.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” Genome Biology 15 (12): 550.

Luo, Weijun, Michael S. Friedman, Kerby Shedden, Kurt D. Hankenson, and Peter J. Woolf. 2009. “GAGE: Generally Applicable Gene Set Enrichment for Pathway Analysis.” BMC Bioinformatics 10 (May): 161.

Mallory, Michael J., Sean P. McClory, Rakesh Chatrikhi, Matthew R. Gazzara, Robert J. Ontiveros, and Kristen W. Lynch. 2020. “Reciprocal Regulation of hnRNP C and CELF2 through Translation and Transcription Tunes Splicing Activity in T Cells.” Nucleic Acids Research 48 (10): 5710–19.

Mankodi, Ami, Masanori P. Takahashi, Hong Jiang, Carol L. Beck, William J. Bowers, Richard T. Moxley, Stephen C. Cannon, and Charles A. Thornton. 2002. “Expanded CUG Repeats Trigger Aberrant Splicing of ClC-1 Chloride Channel Pre-mRNA and Hyperexcitability of Skeletal Muscle in Myotonic Dystrophy.” Molecular Cell 10 (1): 35–44.

Manzourolajdad, Amirhossein, and Jonathan Arnold. 2015. “Secondary Structural Entropy in RNA Switch (Riboswitch) Identification.” BMC Bioinformatics 16 (April): 133.

Martin, Marcel. 2011. “Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads.” EMBnet.journal 17 (1): 10–12.

McGlincy, Nicholas J., and Nicholas T. Ingolia. 2017. “Transcriptome-Wide Measurement of Translation by Ribosome Profiling.” Methods. https://doi.org/10.1016/j.ymeth.2017.05.028.

Montoya, Misty M., and K. Mark Ansel. 2017. “Small RNA Transfection in Primary Human Th17 Cells by Next Generation Electroporation.” Journal of Visualized Experiments: JoVE, no. 122 (April). https://doi.org/10.3791/55546.

Moore, D. A. 2016. “FUS, RNA and the Nucleolus.” core.ac.uk. 2016. https://core.ac.uk/download/pdf/74226236.pdf.

Morandi, Edoardo, Ilaria Manfredonia, Lisa M. Simon, Francesca Anselmi, Martijn J. van Hemert, Salvatore Oliviero, and Danny Incarnato. 2021. “Genome-Scale Deconvolution of RNA Structure Ensembles.” Nature Methods 18 (3): 249–52.

Mortimer, Stefanie A., Mary Anne Kidwell, and Jennifer A. Doudna. 2014. “Insights into RNA Structure and Function from Genome-Wide Studies.” Nature Reviews. Genetics 15 (7): 469–79.

Muhar, Matthias, Anja Ebert, Tobias Neumann, Christian Umkehrer, Julian Jude, Corinna Wieshofer, Philipp Rescheneder, et al. 2018. “SLAM-Seq Defines Direct Gene-Regulatory Functions of the BRD4-MYC Axis.” Science 360 (6390): 800–805.

Mukherjee, Sumit, and Supratim Sengupta. 2016. “Riboswitch Scanner: An Efficient pHMM-Based Web-Server to Detect Riboswitches in Genomic Sequences.” Bioinformatics 32 (5): 776–78.

Mukhopadhyay, Amitabha, Lalit Sehgal, Arunabha Bose, Anushree Gulvady, Parijat Senapati, Rahul Thorat, Srikanta Basu, et al. 2016. “14-3-3γ Prevents Centrosome Amplification and Neoplastic Progression.” Scientific Reports 6 (June): 26580.

Müller-McNicoll, Michaela, Valentina Botti, Antonio M. de Jesus Domingues, Holger Brandl, Oliver D. Schwich, Michaela C. Steiner, Tomaz Curk, Ina Poser, Kathi Zarnack, and Karla M. Neugebauer. 2016. “SR Proteins Are NXF1 Adaptors That Link Alternative RNA Processing to mRNA Export.” Genes & Development 30 (5): 553–66.

Nachmani, Daphna, Anne H. Bothmer, Silvia Grisendi, Aldo Mele, Dietmar Bothmer, Jonathan D. Lee, Emanuele Monteleone, et al. 2019. “Germline NPM1 Mutations Lead to Altered rRNA 2’-O-Methylation and Cause Dyskeratosis Congenita.” Nature Genetics 51 (10): 1518–29.

Navickas, Albertas, Hosseinali Asgharian, Juliane Winkler, Lisa Fish, Kristle Garcia, Daniel Markett, Martin Dodel, et al. 2021. “An mRNA Processing Pathway Suppresses Metastasis by Governing Translational Control from the Nucleus.” bioRxiv. https://doi.org/10.1101/2021.10.04.463118.

Nawrocki, Eric P., and Sean R. Eddy. 2013. “Infernal 1.1: 100-Fold Faster RNA Homology Searches.” Bioinformatics 29 (22): 2933–35.

Neph, Shane, M. Scott Kuehn, Alex P. Reynolds, Eric Haugen, Robert E. Thurman, Audra K. Johnson, Eric Rynes, et al. 2012. “BEDOPS: High-Performance Genomic Feature Operations.” Bioinformatics 28 (14): 1919–20.

Nesvizhskii, Alexey I., Andrew Keller, Eugene Kolker, and Ruedi Aebersold. 2003. “A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry.” Analytical Chemistry 75 (17): 4646–58.

Norman, Thomas M., Max A. Horlbeck, Joseph M. Replogle, Alex Y. Ge, Albert Xu, Marco Jost, Luke A. Gilbert, and Jonathan S. Weissman. 2019. “Exploring Genetic Interaction Manifolds Constructed from Rich Single-Cell Phenotypes.” Science 365 (6455): 786–93.

Oberg, Ann L., Douglas W. Mahoney, Jeanette E. Eckel-Passow, Christopher J. Malone, Russell D. Wolfinger, Elizabeth G. Hill, Leslie T. Cooper, et al. 2008. “Statistical Analysis of Relative Labeled Mass Spectrometry Data from Complex Samples Using ANOVA.” Journal of Proteome Research 7 (1): 225–33.

Palka, Christina, Nicholas M. Forino, Jendrik Hentschel, Rhiju Das, and Michael D. Stone. 2020. “Folding Heterogeneity in the Essential Human Telomerase RNA Three-Way Junction.” RNA 26 (12): 1787–1800.

Pedersen, Brent. n.d. Combat.py: Python / Numpy / Pandas / Patsy Version of ComBat for Removing Batch Effects. Github. Accessed May 10, 2022. https://github.com/brentp/combat.py.

Popow, Johannes, Anne-Marie Alleaume, Tomaz Curk, Thomas Schwarzl, Sven Sauer, and Matthias W. Hentze. 2015. “FASTKD2 Is an RNA-Binding Protein Required for Mitochondrial RNA Processing and Translation.” RNA 21 (11): 1873–84.

Quinlan, Aaron R., and Ira M. Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” Bioinformatics 26 (6): 841–42.

Ray, Partho Sarothi, Jie Jia, Peng Yao, Mithu Majumder, Maria Hatzoglou, and Paul L. Fox. 2009. “A Stress-Responsive RNA Switch Regulates VEGFA Expression.” Nature 457 (7231): 915–19.

Ren, Yue, Yue Huo, Weiqian Li, Manman He, Siqi Liu, Jiabin Yang, Hongmei Zhao, et al. 2021. “A Global Screening Identifies Chromatin-Enriched RNA-Binding Proteins and the

Transcriptional Regulatory Activity of QKI5 during Monocytic Differentiation." Genome Biology 22 (1): 290.

Reuter, Jessica S., and David H. Mathews. 2010. "RNAstructure: Software for RNA Secondary Structure Prediction and Analysis." BMC Bioinformatics 11 (March): 129.

Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." Nucleic Acids Research 43 (7): e47.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2010. "edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data." Bioinformatics 26 (1): 139–40.

Rodionov, Dmitry A. 2007. "Comparative Genomic Reconstruction of Transcriptional Regulatory Networks in Bacteria." Chemical Reviews 107 (8): 3467–97.

Rogelj, Boris, Laura E. Easton, Gireesh K. Bogu, Lawrence W. Stanton, Gregor Rot, Tomaž Curk, Blaž Zupan, et al. 2012. "Widespread Binding of FUS along Nascent RNA Regulates Alternative Splicing in the Brain." Scientific Reports 2 (August): 603.

Rossi, Matthew J., William K. M. Lai, and B. Franklin Pugh. 2018. "Simplified ChIP-Exo Assays." Nature Communications 9 (1): 2842.

Rouskin, Silvi, Meghan Zubradt, Stefan Washietl, Manolis Kellis, and Jonathan S. Weissman. 2014. "Genome-Wide Probing of RNA Structure Reveals Active Unfolding of mRNA Structures in Vivo." Nature 505 (7485): 701–5.

Roux, Kyle J., Dae In Kim, Brian Burke, and Danielle G. May. 2018. "BioID: A Screen for Protein-Protein Interactions." Current Protocols in Protein Science / Editorial Board, John E. Coligan ... [et Al.] 91 (February): 19.23.1–19.23.15.

Ruan, Xuelei, Jian Zheng, Xiaobai Liu, Yunhui Liu, Libo Liu, Jun Ma, Qianru He, et al. 2020. "lncRNA LINC00665 Stabilized by TAF15 Impeded the Malignant Biological Behaviors of Glioma Cells via STAU1-Mediated mRNA Degradation." Molecular Therapy. Nucleic Acids 20 (June): 823–40.

Sanjana, Neville E., Ophir Shalem, and Feng Zhang. 2014. "Improved Vectors and Genome-Wide Libraries for CRISPR Screening." Nature Methods 11 (8): 783–84.

Schneider-Lunitz, Valentin, Jorge Ruiz-Orera, Norbert Hubner, and Sebastiaan van Heesch. 2021. "Multifunctional RNA-Binding Proteins Influence mRNA Abundance and Translational Efficiency of Distinct Sets of Target Genes." PLoS Computational Biology 17 (12): e1009658.

Schwich, Oliver Daniel, Nicole Blümel, Mario Keller, Marius Wegener, Samarth Thonta Setty, Melinda Elaine Brunstein, Ina Poser, et al. 2021. "SRSF3 and SRSF7 Modulate 3'UTR Length through Suppression or Activation of Proximal Polyadenylation Sites and Regulation of CFIm Levels." Genome Biology 22 (1): 82.

Serganov, Alexander, and Evgeny Nudler. 2013. "A Decade of Riboswitches." Cell 152 (1-2): 17–24.

Shadforth, Ian P., Tom P. J. Dunkley, Kathryn S. Lilley, and Conrad Bessant. 2005. "I-Tracker: For Quantitative Proteomics Using iTRAQ." BMC Genomics 6 (October): 145.

Shamovsky, Ilya, Maxim Ivannikov, Eugene S. Kandel, David Gershon, and Evgeny Nudler. 2006. "RNA-Mediated Response to Heat Shock in Mammalian Cells." Nature 440 (7083): 556–60.

Shao, Zhengping, Ryan A. Flynn, Jennifer L. Crowe, Yimeng Zhu, Jialiang Liang, Wenxia Jiang, Fardin Aryan, et al. 2020. “DNA-PKcs Has KU-Dependent Function in rRNA Processing and Haematopoiesis.” Nature 579 (7798): 291–96.

Shin, Seula, Hao Zhou, Chenxi He, Yanjun Wei, Yunfei Wang, Takashi Shingu, Ailiang Zeng, et al. 2021. “Qki Activates Srebp2-Mediated Cholesterol Biosynthesis for Maintenance of Eye Lens Transparency.” Nature Communications 12 (1): 3005.

Shiohama, Aiko, Takashi Sasaki, Setsuko Noda, Shinsei Minoshima, and Nobuyoshi Shimizu. 2007. “Nucleolar Localization of DGCR8 and Identification of Eleven DGCR8-Associated Proteins.” Experimental Cell Research 313 (20): 4196–4207.

Slišković, Irena, Hannah Eich, and Michaela Müller-McNicoll. 2022. “Exploring the Multifunctionality of SR Proteins.” Biochemical Society Transactions 50 (1): 187–98.

Sternburg, Erin L., and Fedor V. Karginov. 2020. “Global Approaches in Studying RNA-Binding Protein Interaction Networks.” Trends in Biochemical Sciences 45 (7): 593–603.

Sun, Lei, Furqan M. Fazal, Pan Li, James P. Broughton, Byron Lee, Lei Tang, Wenze Huang, Eric T. Kool, Howard Y. Chang, and Qiangfeng Cliff Zhang. 2019. “RNA Structure Maps across Mammalian Cellular Compartments.” Nature Structural & Molecular Biology 26 (4): 322–30.

Szklarczyk, Damian, Annika L. Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, et al. 2018. “STRING v11: Protein–protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets.” Nucleic Acids Research 47 (D1): D607–13.

Thomas, Steven, Nathaniel D. Maynard, and John Gill. 2015. “DNA Library Construction Using Gibson Assembly®.” Nature Methods 12 (11): i – ii.

Tomezsko, Phillip J., Vincent D. A. Corbin, Paromita Gupta, Harish Swaminathan, Margalit Glasgow, Sitara Persad, Matthew D. Edwards, et al. 2020. “Determination of RNA Structural Diversity and Its Role in HIV-1 RNA Splicing.” Nature 582 (7812): 438–42.

Turner, Rushia, Kinneret Shefer, and Manuel Ares Jr. 2013. “Safer One-Pot Synthesis of the ‘SHAPE’ Reagent 1-Methyl-7-Nitroisatoic Anhydride (1m7).” RNA 19 (12): 1857–63.

Tyanova, Stefka, Tikira Temu, Pavel Sinitcyn, Arthur Carlson, Marco Y. Hein, Tamar Geiger, Matthias Mann, and Jürgen Cox. 2016. “The Perseus Computational Platform for Comprehensive Analysis of (prote)omics Data.” Nature Methods 13 (9): 731–40.

Ureta-Vidal, Abel, Laurence Ettwiller, and Ewan Birney. 2003. “Comparative Genomics: Genome-Wide Analysis in Metazoan Eukaryotes.” Nature Reviews. Genetics 4 (4): 251–62.

Van Nostrand, Eric L., Peter Freese, Gabriel A. Pratt, Xiaofeng Wang, Xintao Wei, Rui Xiao, Steven M. Blue, et al. 2020. “A Large-Scale Binding and Functional Map of Human RNA-Binding Proteins.” Nature 583 (7818): 711–19.

Van Nostrand, Eric L., Gabriel A. Pratt, Alexander A. Shishkin, Chelsea Gelboin-Burkhart, Mark Y. Fang, Balaji Sundararaman, Steven M. Blue, et al. 2016. “Robust Transcriptome-Wide Discovery of RNA-Binding Protein Binding Sites with Enhanced CLIP (eCLIP).” Nature Methods 13 (6): 508–14.

Wang, Hong, Fuxing Zeng, Qiao Liu, Huihui Liu, Zexian Liu, Liwen Niu, Maikun Teng, and Xu Li. 2013. “The Structure of the ARE-Binding Domains of Hu Antigen R (HuR) Undergoes Conformational Changes during RNA Binding.” Acta Crystallographica. Section D, Biological Crystallography 69 (Pt 3): 373–80.

Warnes, Gregory R., Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber, Andy Liaw, Thomas Lumley, et al. 2009. “Gplots: Various R Programming Tools for Plotting Data.” R Package Version 2 (4): 1.

Wheeler, Travis J., and Sean R. Eddy. 2013. “Nhmmer: DNA Homology Search with Profile HMMs.” Bioinformatics 29 (19): 2487–89.

Winkler, Wade, Ali Nahvi, and Ronald R. Breaker. 2002. “Thiamine Derivatives Bind Messenger RNAs Directly to Regulate Bacterial Gene Expression.” Nature 419 (6910): 952–56.

Wolf, Ashley R., and Vamsi K. Mootha. 2014. “Functional Genomic Analysis of Human Mitochondrial RNA Processing.” Cell Reports 7 (3): 918–31.

Wolf, F. Alexander, Philipp Angerer, and Fabian J. Theis. 2018. “SCANPY: Large-Scale Single-Cell Gene Expression Data Analysis.” Genome Biology 19 (1): 15.

Yamashita, Akio. 2013. “Role of SMG-1-Mediated Upf1 Phosphorylation in Mammalian Nonsense-Mediated mRNA Decay.” Genes to Cells: Devoted to Molecular & Cellular Mechanisms 18 (3): 161–75.

Yang, Ruiqing, Sergei A. Gaidamakov, Jingwei Xie, Joowon Lee, Luigi Martino, Guennadi Kozlov, Amanda K. Crawford, et al. 2011. “La-Related Protein 4 Binds poly(A), Interacts with the poly(A)-Binding Protein MLLE Domain via a Variant PAM2w Motif, and Can Promote mRNA Stability.” Molecular and Cellular Biology 31 (3): 542–56.

Yang, Xiaoping, Jesse S. Boehm, Xinping Yang, Kourosh Salehi-Ashtiani, Tong Hao, Yun Shen, Rakela Lubonja, et al. 2011. “A Public Genome-Scale Lentiviral Expression Library of Human ORFs.” Nature Methods 8 (8): 659–61.

Yoon, Sungroh, Jinkyu Kim, Justine Hum, Hanjoo Kim, Seunghyun Park, Wipapat Kladwang, and Rhiju Das. 2011. “HiTRACE: High-Throughput Robust Analysis for Capillary Electrophoresis.” Bioinformatics 27 (13): 1798–1805.

Younis, Shady, Wael Kamel, Tina Falkeborn, Hao Wang, Di Yu, Robert Daniels, Magnus Essand, Jorma Hinkula, Göran Akusjärvi, and Leif Andersson. 2018. “Multiple Nuclear-Replicating Viruses Require the Stress-Induced Protein ZC3H11A for Efficient Growth.” Proceedings of the National Academy of Sciences of the United States of America 115 (16): E3808–16.

Youn, Ji-Young, Wade H. Dunham, Seo Jung Hong, James D. R. Knight, Mikhail Bashkurov, Ginny I. Chen, Halil Bagci, et al. 2018. “High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies.” Molecular Cell 69 (3): 517–32.e11.

Yu, Guangchuang, Li-Gen Wang, and Qing-Yu He. 2015. “ChIPseeker: An R/Bioconductor Package for ChIP Peak Annotation, Comparison and Visualization.” Bioinformatics 31 (14): 2382–83.

Zhou, Xin, Seula Shin, Chenxi He, Qiang Zhang, Matthew N. Rasband, Jiangong Ren, Congxin Dai, et al. 2021. “Qki Regulates Myelinogenesis through Srebp2-Dependent Cholesterol Biosynthesis.” eLife 10 (May). https://doi.org/10.7554/eLife.60467.

Zhuo, Enqing, Changqing Cai, Wenzhe Liu, Kunsong Li, and Wenzhen Zhao. 2020. “Downregulated microRNA-140-5p Expression Regulates Apoptosis, Migration and Invasion of Lung Cancer Cells by Targeting Zinc Finger Protein 800.” Oncology Letters 20 (6): 1–1.

Zong, Feng-Yang, Xing Fu, Wen-Juan Wei, Ya-Ge Luo, Monika Heiner, Li-Juan Cao, Zhaoyuan Fang, et al. 2014. “The RNA-Binding Protein QKI Suppresses Cancer-Associated Aberrant Splicing.” PLoS Genetics 10 (4): e1004289.

Zubradt, Meghan, Paromita Gupta, Sitara Persad, Alan M. Lambowitz, Jonathan S. Weissman, and Silvi Rouskin. 2017. “DMS-MaPseq for Genome-Wide or Targeted RNA Structure Probing in Vivo.” Nature Methods 14 (1): 75–82.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

E8A90B2023044E5...

Author Signature

8/30/2022

Date