Principles of HuR-RNA targeting, interaction dynamics, and functional outcomes

by

Neelanjan Mukherjee

University Program in Genetics and Genomics
Duke University

Date:_______________________

Approved:

___________________________
Jack Keene, Supervisor

___________________________
Christopher Counter

___________________________
Philip Benfey, Chair

___________________________
Jen-Tsan Chi

___________________________
Uwe Ohler

Dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy in the University Program in Genetics and Genomics
in the Graduate School
of Duke University

2010
ABSTRACT
Principles of HuR-RNA targeting, interaction dynamics, and functional outcomes
by
Neelanjan Mukherjee
University Program in Genetics and Genomics
Duke University

Date:_______________________
Approved:

___________________________
Jack Keene, Supervisor

___________________________
Christopher Counter

___________________________
Philip Benfey, Chair

___________________________
Jen-Tsan Chi

___________________________
Uwe Ohler

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the University Program in Genetics and Genomics
in the Graduate School
of Duke University

2010
Abstract

In recent years, the pervasiveness and importance of post-transcriptional regulation has reshaped the principles underlying the organizational logic of gene expression. RNA-binding proteins (RBPs) and non-coding RNAs are the regulatory molecules primarily responsible for interaction with target mRNAs and thereby regulating post-transcriptional processes, eventually influencing characteristics of the encoded protein. Many of the mRNA targets of RBPs encode functionally related proteins, which can participate in post-transcriptional operons capable of coordinating specific cellular processes or macromolecular complexes. Thus, identifying RNA targets, precise binding sites, and the dynamics of these interactions will reveal how these important regulatory factors contribute to gene regulatory networks.

ELAV family of human RBPs consist of 4 members, which all have 3 RRM (RNA-recognition motif) domains the last separated by a hinge region. Its predominant role is to positively regulate the stability and translation of target mRNAs through binding to ARE (AU-rich elements) in the 3’ UTR (untranslated region) of protein coding transcripts. In response to certain stimuli, HuR is subject to post-translational modifications and changes subcellular localization, which impacts its regulatory capacity. In this study on a transcriptome-wide level, we interrogate the RNA targets, precise binding sites, as well as the remodeling of these interactions in response to stimuli.

First we developed methods to quantitatively analyze RBP-mRNA interactions from RIP-chip (ribonucleoprotein immunoprecipitation on microarray) data. We utilized Gaussian mixture modeling as a discriminative model, treating RBP-association as a discrete variable (target or not target) and as a generative model, treating RBP-
association as a continuous variable (probability of association). We were able to calculate LOD scores, a continuous metric representing probability of association that can be compared across RIP-chips, and other data sets to identify subtle, yet coordinated patterns.

We utilized two complementary methods, RIP-chip and PAR-CLIP (photoactivatable ribonucleoside cross-linking and immunoprecipitation), to identify targets of HuR and high-resolution binding sites on a transcriptome-wide scale. We discovered that HuR-mRNA interactions are not restricted to the 3’ UTR and there are thousands of intronic binding sites. A significant proportion of intronic binding sites are contained in the poly-pyrimidine tract near 3’ splice sites. Binding sites in the 3’ UTR and intron are often approximately 30 nucleotides apart. HuR can bind to both AU-rich and U-rich sequences, the former more prevalent in 3’ UTRs and the latter more prevalent at the 3’ splice site.

We then integrated the binding data with transcriptomics of HuR siRNA mediated knockdown. We found that the degree of binding is proportional to the degree of HuR-dependent stabilization. Moreover the ability to stabilize mRNA is not restricted to 3’ UTR binding sites, as intronic binding sites also exhibited the binding degree correlated stabilization. Furthermore, the spatial pattern of HuR binding sites relative to exons influenced exon usage decisions. Specifically, binding sites upstream of the exon promote exclusion, while binding sites downstream of the exon promote inclusion.

We utilized T-cell activation as a model to investigate stimulus induced remodeling of HuR-mRNA interactions. We performed RIPs on endogenous ribonucleoprotein complexes containing HuR and PABP throughout a T-cell activation time course and identified the associated mRNA population using microarrays. HuR
interacts with different populations of mRNAs during T-cell activation. These populations encode functionally related proteins that are members of the Wnt pathway and proteins mediating T-cell receptor signaling pathways.

The role of other RBPs and microRNAs and post-translational modifications were investigated as potential mechanisms driving the HuR-mRNA interaction dynamics using computational tools as well as PAR-CLIP data. The mRNA targets and interaction sites of HuR were found to overlap with the targets and interaction sites of other posttranscriptional regulatory factors, indicating combinatorial interdependence of posttranscriptional regulatory networks and modules following activation. We utilized a model of Chk2-mediated HuR phosphorylation in response to oxidative stress. Modulating a single serine residue results in changes in target specificity. These differences were enriched for mRNAs encoding proteins important to oxidative stress response, consistent with our initial model and previous studies.

Finally, applying HuR mRNA dynamics as a quantitative phenotype in the drug-gene-phenotype Connectivity Map, we identified candidate small molecule effectors of HuR and T-cell activation. We show that one of these candidates, resveratrol, exerts T-cell activation dependent post-transcriptional effects that are rescued by HuR. Thus, we describe a strategy to systematically link an RBP and condition-specific post-transcriptional effects to small molecule drugs.

During the course of this work we developed new computational tools for analysis of RIP-chip and PAR-CLIP data. These were used to identify HuR RNA targets and their precise interaction sites on a transcriptome-wide scale. We determined that HuR regulates the stability of its mRNA targets as a function of the number of binding sites.
Additionally, many intronic binding sites were identified upstream of exons. The decision to include or exclude a specific exon was influenced by whether the HuR binding sites were either upstream or downstream of exons. We observed dramatic remodeling of HuR-RNA interactions during T-cell activation, which were likely driven by other RBPs and microRNAs, as well as post-translational modifications. The HuR-mRNA dynamics were utilized to identify and validate resveratrol as a small molecule effector of HuR. In summary, we revealed novel functions of the RBP HuR and developed systematic strategies of exploring RBP-RNA interactions and their consequences.
Contents

Abstract ................................................................................................................................. iv

List of Tables ........................................................................................................................ xi

List of Figures ........................................................................................................................ xii

Acknowledgements ............................................................................................................... xvii

1. Introduction ....................................................................................................................... 1
   1.1 Post-transcriptional regulation of gene expression ......................................................... 1
   1.2 Biochemical methods for global identification RNA-binding protein targets ............ 2
   1.3 ELAV family RNA-binding proteins ............................................................................ 12
   1.4 Goals of this thesis ...................................................................................................... 14

2. Quantitative analysis of RIP-chip data ........................................................................... 16
   2.1 LOD scores .................................................................................................................. 17
   2.2 Percentile rank ............................................................................................................ 20
   2.3 Comparison of various enrichment metrics ................................................................. 21
   2.4 Discussion ................................................................................................................... 23
   2.5 Methods ..................................................................................................................... 23

3. Identifying HuR targets and precise binding sites ......................................................... 25
   3.1 PAR-CLIP .................................................................................................................... 25
      3.1.1 Defining binding sites ............................................................................................ 26
      3.1.2 Characteristics of binding sites ............................................................................ 27
      3.1.3 HuR binding motif ................................................................................................. 31
   3.2 RIP-chip ....................................................................................................................... 33
   3.3 Comparison of techniques .......................................................................................... 34
   3.4 Discussion ................................................................................................................... 39
3.5 Methods ........................................................................................................... 44

4. Relationship between targeting and functional outcome .................................... 47
   4.1 HuR knockdown ............................................................................................... 47
   4.2 mRNA stability ............................................................................................... 49
   4.3 mRNA splicing ............................................................................................... 52
   4.4 Discussion ........................................................................................................ 54
   4.5 Methods ......................................................................................................... 57

5. Coordinated RNP dynamics during T-cell activation ........................................... 59
   5.1 Global Identification of HuR target mRNAs During T cell Activation .......... 60
   5.2 Functional Characteristics of HuR targets During Activation ...................... 63
      5.2.1 Sequence Characteristics and Motifs ....................................................... 63
      5.2.2 Common Functional Groups Enriched in HuR mRNA Targets ............. 66
   5.3 RNP Dynamics During Activation .................................................................. 69
      5.3.1 Functional Characteristics of Ribonomic and Transcriptomic Dynamics During Activation ................................................................. 69
      5.3.2 Functional Characteristics of HuR RNP State Dynamics During Activation 71
   5.3 Discussion ...................................................................................................... 72
   5.4 Methods ........................................................................................................ 74

6. Mechanisms driving HuR-mRNA interaction dynamics .................................... 77
   6.1 Combinatorial interactions .......................................................................... 77
      6.1.1 Other ARE-RBPs .................................................................................... 77
      6.1.2 microRNAs ............................................................................................ 79
   6.2 Post-translational modifications ................................................................... 81
      6.2.1 Chk2-mediated phosphorylation as a model ......................................... 82
      6.2.2 Phosphorylation state specific targeting ............................................... 83
6.3 Discussion ...................................................................................................................... 85
6.4 Methods ......................................................................................................................... 87

7. Systematic discovery of small molecule effectors of HuR ................................. 88
  7.1 Identification of candidate small molecule effectors .............................................. 88
  7.2 Validation of resveratrol as a small molecule effector of HuR ......................... 90
    7.2.1 Resveratrol effects subcellular localization of HuR ................................ 90
    7.2.2 Resveratrol antagonizes HuR-dependent posttranscriptional regulation .... 92
  7.3 Discussion ................................................................................................................. 94
  7.4 Methods ..................................................................................................................... 96

8. Conclusions and future directions ........................................................................... 98

References ....................................................................................................................... 99

Biography ......................................................................................................................... 121
List of Tables

Table 1: Ribonomic analyses of RNPs. Authors’ results are reported in appropriate columns. Motifs represent the single most significant motif reported. For the column ‘Notes’: E, endogenous; T, tagged; S, sequencing; H, high-throughput sequencing; M, microarray; and X, cross-linking. 5

Table 2: In vitro binding affinity of HuR towards both U-rich and AU-rich RNA. 43

Table 3: HuR-Associated mRNAs Encode Proteins Critical to the Immunological Synapse. Different colors represent the following classes: Orange = TCR Receptor and Membrane, Yellow = Adapters, Green = Regulators of Cytoskeletal Dynamics, and Blue = Regulators of Signaling. 66

Table 4: GSEA Results for Dynamic and Correlation Comparisons. GSEA analysis was conducted on ΔLOD HuR for both 0hr to 4hr and 4hr to 12hr. Additionally, GSEA was conducted incorporating data from all time points using correlations between HuR-association profiles and PABP association profiles, as well as correlations between HuR-association profiles and transcriptomic profiles. 71

Table 5 HuR RNP Enriched for Predicted Targets of microRNAs. GSEA analysis was conducted on LOD HuR scores for each time point using the following three classes of gene sets for pathways, transcription factor targets (TF), and predicted microRNA (miRNA) target. Positively enriched gene sets were rank ordered and plotted by FWER p-values. Only results for 4hr are shown; 0hr and 12hr results are similar. 78

Table 6: Candidate Small Molecule HuR Effectors Derived Using CMAP. Sign of enrichment scores (ES) represents whether changes in HuR association across the time course were either positively or negatively correlated with transcriptomic changes. Significant p-values (<0.05) listed in red. ** Known effector of HuR. Common classes of compounds are highlighted, Hsp90 inhibitors (blue), HDAC inhibitors (yellow), PI3K inhibitors (orange), and COX inhibitors (green). 89
List of Figures

Figure 1: Role of RBPs in coupling and regulating steps of gene expression. Depicted are coupled processes regulated by RBPs and in some cases non-coding RNAs. ........................ 2

Figure 2: States of HuR. Depicted are A) protein domains, associated proteins, and all known post-translational modifications and B) known signaling events and their consequences. .............................................................................. 13

Figure 3: Overview of ribonomic analysis. (A) Isolation of RNP complexes, in this case HuR (blue), in parallel with IgG negative control, followed by extraction of mRNAs and hybridization to microarray (RIP-chip). (B) Gaussian mixture modeling, of RNP IP versus negative IP t-scores for 3 biological replicates of RIP-Chip, to identify and quantify biochemically enriched populations of mRNAs represented by probes. (C) LOD scores representing a condition-specific probability of HuR association per probe. (D) Discrete approach, LOD HuR > 0, defines a subset of probes associated with HuR in a given biological condition. Continuous approach, LOD score values to compare HuR mRNA dynamics across different biological conditions, with other data types or databases, and identify small molecule effectors of HuR. Both approaches identify common functional groups or motifs associated with HuR. This framework has been utilized for different RBPs, cell lines, and array platforms................................................................. 18

Figure 4: Identification of DRBP76-associated mRNAs. Deviation of data from the diagonal quantile:quantile plot indicates the non-Gaussian nature of the DRBP76 IP vs. mock IP distribution, corresponding with transcripts specifically enriched in the DRBP76 IP. B. Distribution of the average percentile ranks (APR) of 5 biological replicates for DRBP76 RIPs and mock RIPs. Transcripts considered DRBP76 RNP-associated are indicated by black boxes. ............................................................................................................. 21

Figure 5: Comparison of RIP-chip analysis metrics. Enrichment was assessed by using A) KS (kolmogorov-smirnov) tests for the whole distribution, and by B) determining the percent of genes containing a motif for the top 500, 1000, and 500-1000, mRNAs ranked by the respective metric. Results were averaged across 0, 4, and 12 hr HuR RIP-chip during T-cell activation. Error bars represent standard deviation. ......................................................... 22

Figure 6: HuR PAR-CLIP. A) A radioactive band representing directly bound RNA fragments at expected size of the epitope tagged HuR. B) Kernel density estimates of conversions versus non-conversions were utilized to map high resolution binding sites. C) Distribution of HuR binding sites shows most clusters are well below the 40-nucleotide clusters (dashed line) identified by the original analytical method. D) Binding sites are more likely to be near each other than random (dashed lines represent (1st, 50th, and 99th percentiles of the null distribution). ................................................................................................................................. 26

Figure 7: Cluster quality across all genomic locations. A) Proportion of densities and B) comparison of distributions are displayed separately for the each type of genomic location a cluster by their respective CLI. ................................................................. 28
Figure 8: Clusters mapping to repetitive sequences. All repetitive elements except for low complexity and unknown repeats had statistically significant lower CLI scores and were discarded. ................................................................. 29

Figure 9: HuR binding sites in mRNAs. A) The proportion of densities for clusters derived from each region of an mRNA. B) Distribution of the number of HuR binding sites per mRNA. .................................................................................. 30

Figure 10: HuR binding motifs. A) Highest ranking motif from cERMIT analysis of PAR-CLIP data ranked by CLI. B) Highest ranking motif from MEME analysis of clusters found within the first 50 nucleotides upstream of introns, a location overrepresented for HuR binding sites........................................................................... 32

Figure 11: HuR RIP-chip. A) QRT-PCR of signal of controls from biological triplicate RIPs. **TNF was not detected in mock IP and therefore the ct was set to 40, which is the total number of cycles. This value was used to calculate enrichment as well. B) Significant enrichment of positive controls over both mock and totals............................................ 34

Figure 12: Comparison of PAR-CLIP and Rip-chip. A) Venn diagram of the overlap between mRNAs defined as HuR targets using PAR-CLIP and RIP-chip. Comparison of the abundance mRNAs detected by Rip-chip only, PAR-CLIP only, and both using B) microarray data or C) sequencing data. D) Comparison of the HuR LOD score distributions of PAR-CLIP data separated by number of binding sites per transcript. E) Comparison of PAR-CLIP binding site distribution of targets identified by PAR-CLIP only, PAR-CLIP, and RIP-chip and PAR-CLIP................................................................................................................. 36

Figure 13: Location of PAR-CLIP sites in RIP-chip data. A) Comparison of the number of binding sites in distinct regions of the mRNA for targets identified with PAR-CLIP only, or PAR-CLIP and RIP-chip. B) Comparison of the distribution of HuR LOD scores for mRNA targets identified by RIP-chip only, or RIP-CHIP and PAR-CLIP. C) Pie chart of percent of RIP-targets containing either intronic or 3' UTR binding sites (green)........... 38

Figure 14: Summary of comparisons between mRNA targets identified using PAR-CLIP versus RIP-chip................................................................................................................................. 42

Figure 15: HuR targets destabilized in HuR depleted cells. A) A variety of concentrations and times of HuR siRNA knockdown were tested previous to transcriptomic analysis. B) Abundance of known HuR targets decrease 72 hrs after HuR depletion. C) Distribution of differential expression after HuR knockdown for RIP-chip and PAR-CLIP targets as compared to all expressed mRNAs............................................................... 48

Figure 16: Relationship between degree of binding and functional outcome. A) Transcripts with more HuR binding sites exhibit greater decrease in response to HuR knockdown. B) Transcripts with Higher LOD scores exhibit greater decrease in response to HuR knockdown. C) Transcripts with binding sites only in introns, 3' UTRs, and CDS were significantly destabilized by HuR knockdown. D) More intronic binding sites lead to more destabilization after HuR knockdown. .................................................................................. 50
Figure 17: Distribution of HuR Binding sites in 3' UTRs. Red line indicates distribution of observed 3' UTR HuR binding sites. Gray lines indicate distribution of null model (1st, 10th, 50th, 90th, and 99th percentile). .................................................................................................................... 51

Figure 18: HuR binding sites and alternative splicing. A) Overrepresentation of intronic HuR binding sites in the first 25 nucleotides upstream of introns. B) Differential exon inclusion/exclusion preferences based on location of intronic HuR binding sites. .......... 54

Figure 19: Proposed mechanisms by which HuR regulates mRNA splicing and stability. .................................................................................................................................................. 56

Figure 20: HuR targets during T-cell activation. (A) Distribution of HuR IP versus negative IP t-scores at 0, 4, and 12 hours post-activation. HuR-associated (blue curve), background (green curve) and sum of HuR-associated and background (red curve) probability distributions are shown, as defined by Gaussian mixture modeling. (B) Venn diagram representing all probes (n=1219) that had a LOD HuR > 0 at any time point. (C) Upper triangular matrix of Spearman correlation coefficients for all pair-wise comparisons of LOD HuR and Totals S2N values for 0, 4, and 12 hours post-activation. 61

Figure 21: AU-content and length of HuR-associated mRNAs. Null distribution of average length and AU-content of 3' UTRs of randomly chosen messages from total expressed population (10,000 random sets) each with same # of UTRs as the HuR-RNP associated set. The black arrows mark the average length and AU content for HuR-RNP associated mRNAs, indicating a significant (p<0.0001) bias in the length and AU-content of 3' UTRs of HuR-RNP associated mRNAs. .................................................................................................................. 64

Figure 22: Enrichment of HuR COVE model in HuR-associated mRNAs. The same random sampling procedure used in SF1 was conducted on various HuR COVE model statistics. Null distributions represent averages for all listed statistics for random sets (light grey) and di-nucleotide shuffled 3' UTR of HuR-associated mRNAs (dark grey). The average values for 3' UTR of HuR-associated mRNAs indicated by black arrow. ... 64

Figure 23: HuR RNP 3' UTRs are Enriched for HuR COVE Motif. GSEA analysis of HuR LOD scores using HuR COVE motif gene set. Normalized Enrichment Scores (NES) and the FamilyWise-Error Rate (FWER) p-value are shown. LOD scores > 0 are to the left of the vertical red line in the plots.................................................................................................................................................. 65

Figure 24: Coordination of HuR RNP Associated Functional Modules During T Cell Activation. Analysis of HuR-associated mRNAs (LOD > 0) at each time point was conducted using Panther and InnateDB. Relative enrichment of functionally related groups (must have a Bonferroni corrected p-value < 0.05 for at least one time point) are represented as profiles and heatmap. ** Functional groups that represent regulation of DNA and RNA. .................................................................................................................................................. 68

Figure 25: Gene Expression Dynamics Demonstrate Unique Changes in mRNA Subpopulations. Upper triangular matrix of Spearman correlation coefficients for all pair-wise comparisons of ΔLOD HuR, ΔLOD PABP and Totals t-scores for 0hr to 4hr and 4hr to 12hr dynamics. .................................................................................................................................................. 70
Figure 26: HuR RNP Enriched for Predicted Targets of microRNAs. GSEA analysis was conducted on LOD HuR scores for each time point using the following three classes of gene sets for pathways, transcription factor targets (TF), and predicted microRNA (miRNA) target. Positively enriched gene sets were rank ordered and plotted by FWER p-values. Only results for 4hr are shown; 0hr and 12hr results are similar.

Figure 27: Combinatorial regulation by HuR and Ago. Distribution of distances between HuR and Ago sites identified using PAR-CLIP. Red line indicates distribution of distances between observed HuR and Ago binding sites. Gray lines indicate distribution of null model (1st, 10th, 50th, 90th, and 99th percentile). Blue box zooms in on distances < 50 nucleotides.

Figure 28: Chk2-mediated phosphorylation of HuR during oxidative stress.

Figure 29: Ser100 from HuR may make contact with HuR. A crystal structure of closely related HuD was made with c-fos RNA with the following properties color coded: (white) the first two RRMs of HuD, (red) Ser118 of HuD maps to Ser100 of HuR, (blue) nine nucleotide stretch of c-Fos mRNA interacting with HuD, and (yellow) eighth uracil in c-Fos binding site.

Figure 30: Flag-tagged HuR variant constructs. Four individual cell lines were created each with a single epitope-tagged HuR variant. Ser100 was replaced with alanine (phospho-deficient), aspartic acid or glutamic acid (phospho-mimetic) residues. Since we used the Flip-IN T-Rex system, the cell lines contain the different variants in the same genomic location under control of a tetracycline responsive promoter.

Figure 31: A) Differential targeting between HuR phospho-mutants. Enrichment of gene sets representing pathways were performed on mRNAs detected in WT, S100A, and S100D/E libraries using PantherDB. B) UCSC genome browser depiction of p21 (CDKN1A) binding sites from various libraries.

Figure 32: Resveratrol Affects Subcellular Localization of HuR During Activation. (A) Representative immunoblot of subcellular localization of HuR protein 0, 4, and 12 hrs post-activation of Jurkat cells using PMA and PHA with or without resveratrol pretreatment. Proteins were analyzed by immunoblot using anti-HuR, anti-β-tubulin as a cytosolic (Cyto) marker, anti-TRAP-α as an endoplasmic reticulum/outer nuclear envelope (ER/ONE) marker, and anti-histone H4 as a Nuclear (Nuc) marker. Four fold more Cyto and ER/ONE fraction extract was loaded compared to Nuc fraction extract. (B) Quantification of relative fold change in HuR following resveratrol pre-treatment. HuR levels were normalized to the appropriate loading control for each fraction. Normalized HuR levels for resveratrol pre-treated samples were compared to mock pretreated samples. The means and standard error of the means (SEM) are represented from 3 independent experiments. Paired t-tests were used to calculate p-values (* P <0.05).

Figure 33: Activation Dependent Effects of Resveratrol and HuR on 3'UTR Luciferase Reporters. A) Firefly to renilla ratio of each reporter normalized to the control 3'UTR reporter, pLuc, for each condition. Activation significantly increases reporter expression...
through specific sequences. Furthermore, resveratrol significantly blunts this increase. B) The fold change of the normalized firefly to renilla ratio for the indicated condition (Activation + Resveratrol or Activation + Resveratrol + HuR) compared to the normalized firefly to renilla ratio for its matched mock treated control. As shown in A) resveratrol inhibits activation mediated increases in reporter expression, however B) shows that HuR overexpression antagonizes resveratrol-dependent inhibition.
Acknowledgements

I would like to thank Matt Friedersdorf and Marshall Thompson for major contributions to HuR T-cell activation dynamics study; Matt Friedersdorf for putting up with me and the long potentially random conversations as an office mate over these years; Adam Morris, Kyle Mansfield, Jeff Blackinton, and Jordan Komisarow for being great colleagues, friends, and along with Jack create an amazing graduate school experience; committee members for their guidance and time; Shelton Bradrick for helpful discussions; the Gromeier and Nicchitta labs for plasmids and antibodies; David Corcoran and David Reid for contributions to the PAR-CLIP studies, Uwe Ohler for opportunities and significant advice with computational approaches; Jeff Nusbaum, Manuel Ascano Jr, and Markus Hafner, and Tom Tuschl for discussions and training with PAR-CLIP; Jack Keene for being a fantastic advisor and role model; my parents, Shyama and Rina, for being loving and encouraging; my brother, Sayan, for always having time and also countless conversations about data and biology; and my loving wife Katie, for marrying me and for her unwavering support and love.
1. Introduction

1.1 Post-transcriptional regulation of gene expression

The phenotypic fate of a cell is dependent on its ability to execute defined gene expression programs in response to environmental variability in the face of intracellular stochasticity. Phenotypic fates are not governed by individual molecules, but rather arise from complex interactions between various types of cellular molecules. Achieving the appropriate phenotypic outcome is dependent upon proper execution of gene expression programs that coordinate dynamic remodeling of molecular interactions spatially and temporally. Therefore, global investigation of these complex interactions and their dynamics is critical to identifying and understanding the gene regulatory networks activated by the particular gene expression program regulating phenotypic fate.

While dynamics in chromatin structure and transcription factors activity determine production of mRNAs, RNA-binding proteins (RBP) and small-noncoding RNAs are post-transcriptional regulatory factors (PTRF) that determine the fate of each mRNA species and are therefore essential for determining phenotypic fate (Figure 1). PTRFs interact with mRNA to form dynamic multi-component ribonucleoprotein (RNP) complexes that govern critical events ranging from splicing and transport to stability and translation[1-3]. In fact, there is evidence for RBPs coupling these distinct aspects of gene expression, for example co-transcriptional splicing [4]; mRNA degradation and nuclear export [5]; and the obvious dependency between mRNA stability and translation. Ultimately, regardless of the mechanisms that govern transcriptional regulation, the outcome of gene expression is determined at the level of translation.
Figure 1: Role of RBPs in coupling and regulating steps of gene expression. Depicted are coupled processes regulated by RBPs and in some cases non-coding RNAs.

Earlier studies have lead to the formation of the post-transcriptional operon model, in which RBPs coordinate the expression of transcripts encoding functionally related proteins through combinatorial and dynamic interactions. There are now many studies supporting this model.

1.2 Biochemical methods for global identification RNA-binding protein targets

Biochemical, genetic, and computational approaches have been utilized, alone and in combination with one another, to identify and validate regulatory elements and their cognate RBPs or, in some cases, non-coding RNAs. These investigations have proved to be challenging due to the degenerate and/or short nature of many regulatory sequences, as well as the availability of definitive biochemical techniques. Table 1 (below) lists many studies utilizing biochemical methods for identifying RBP targets.
RNPs are the molecular machines carrying out post-transcriptional processes. The first developed and most widely used genomic method to identify targets of a given RBP is immunoprecipitation (IP) of the RNP in conditions preserving endogenous RNA-protein interactions, followed by detecting the bound RNAs using microarrays (RIP-chip) [6]. This is typically accompanied by a mock IP or total mRNA arrays, which serve as specificity controls. RIP-chip studies across many organisms have uncovered putative regulatory elements as well as coherent functional relationships between targets of individual RBPs (reviewed in [7]).

The primary deficiency of RIP-chip is the inability to precisely identify the site of the RNA-protein interaction. A subsequent method, named CLIP (cross-linking and immunoprecipitation), which utilizes short wave UV (254 nm) to cross-link RNA-protein complexes followed by RNA digestion and detection of the protected RNA fragments [8], was intended to address the deficiency of RIP-chip. However, CLIP is limited by poor cross-linking efficiency, non-specific cross-links, and photo-damage to RNA and protein.

A recent technique, PAR-CLIP (photoactivatable ribonucleoside cross-linking and immunoprecipitation) utilizes long wave UV (365 nm) to cross-link photoactive thiolated-uridylates, which have been incorporated into nascent RNA, resulting in a 100-1000 fold more efficient cross-linking reaction [9]. This cross-linking reaction is very specific and efficient and does not produce photo-damage [10]. After cross-linking there is a light RNaseT1 digestion, followed by end-labeling the RNA fragments, immunoprecipitating the desired RBP, and then another RNase T1 digestion to trim RNA fragments down to a reasonable size. The IP material is run on an SDS gel and there should be a radioactively labeled RNA band migrating at the expected molecular weight of the RBP of interest. After cutting out the band of interest, the sample is subject
to protease treatment followed by RNA extraction. Sequencing adapters are ligated to the ends of the purified RNA. During reverse transcription to cDNA, thiolated-uridylates cross-linked to proteins exhibit altered hydrogen bond acceptor/donor properties and basepair with adenosine rather than guanine. Therefore, when creating the opposite strand (the original strand), cytosine (C) is incorporated across from guanine (G) rather than a thymine (T). Importantly, since these T to C conversions in cDNAs derived from protected cross-linked RNAs are diagnostic of the RNA-protein interactions, they allow for discrimination of signal versus noise and delineation of precise binding sites. Conveniently, other modified nucleosides or combinations thereof can be utilized to circumvent the potential sequence bias, while retaining increased cross-linking efficiency.
Table 1: Ribonomic analyses of RNPs. Authors’ results are reported in appropriate columns. Motifs represent the single most significant motif reported. For the column ‘Notes’: E, endogenous; T, tagged; S, sequencing; H, high-throughput sequencing; M, microarray; and X, cross-linking.

<table>
<thead>
<tr>
<th>Cell Type/Line</th>
<th>Additional Biological Conditions</th>
<th>Name</th>
<th>RNA Target(s) Summary</th>
<th>Motif</th>
<th>Notes</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUMAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293</td>
<td>-</td>
<td>Ago1</td>
<td>~ 600 mRNAs</td>
<td>-</td>
<td>E, S</td>
<td>[11]</td>
</tr>
<tr>
<td>HEK 293</td>
<td>-</td>
<td>Ago1</td>
<td>Mostly miRNAs, but also snoRNA, mRNAs, noncoding RNAs, transposons</td>
<td>-</td>
<td>E, H</td>
<td>[12]</td>
</tr>
<tr>
<td>HEK 293</td>
<td>-</td>
<td>Ago2</td>
<td>~ 600 mRNAs</td>
<td>-</td>
<td>T, M</td>
<td>[13]</td>
</tr>
<tr>
<td>HEK 293</td>
<td>-</td>
<td>Ago2</td>
<td>Mostly miRNAs, but also snoRNA, mRNAs, noncoding RNAs, transposons</td>
<td>-</td>
<td>E, H</td>
<td>[12]</td>
</tr>
<tr>
<td>HEK 293</td>
<td>miR122 overexpression</td>
<td>Ago2</td>
<td>-</td>
<td>-</td>
<td>T, M</td>
<td>[13]</td>
</tr>
<tr>
<td>HEK 293</td>
<td>miR-124 Overexpression</td>
<td>Ago2</td>
<td>miRNAs, 56 miR-1 targets, 288 miR-124 Targets</td>
<td>miR-1 and miR-124 Seeds</td>
<td>T, M</td>
<td>[15]</td>
</tr>
<tr>
<td>Jurkat</td>
<td>-</td>
<td>Ago2</td>
<td>~75% miRNA, 0.7% mRNA</td>
<td>-</td>
<td>E, H</td>
<td>[16]</td>
</tr>
<tr>
<td>HEK 293</td>
<td>-</td>
<td>Ago3</td>
<td>~88% miRNA, 2.3% tRNA, 1.5% mRNA</td>
<td>-</td>
<td>E, H</td>
<td>[16]</td>
</tr>
<tr>
<td>Jurkat</td>
<td>-</td>
<td>Ago4</td>
<td>-</td>
<td>-</td>
<td>T, M</td>
<td>[13]</td>
</tr>
<tr>
<td>K562</td>
<td>-</td>
<td>alphaCP2</td>
<td>Cytoskeleton, Transcription Factors, Proto-oncogenes, and Signaling</td>
<td>CU-rich</td>
<td>E, M</td>
<td>[17]</td>
</tr>
<tr>
<td>HeLa</td>
<td>-</td>
<td>AUF1</td>
<td>58% of Targets were AUF1 Targets</td>
<td>-</td>
<td>E, M</td>
<td>[18]</td>
</tr>
<tr>
<td>HeLa</td>
<td>-</td>
<td>AUF1</td>
<td>Pre-mRNA and mRNAs</td>
<td>Long AU-Rich Stem with 4 internal Bulges</td>
<td>E, M</td>
<td>[19]</td>
</tr>
<tr>
<td>MCF10A</td>
<td>MCT-1</td>
<td>AUF1</td>
<td>Cell Cycle and MAPK Signaling in Nontransformed</td>
<td>-</td>
<td>E, M</td>
<td>[20]</td>
</tr>
<tr>
<td>Cell Type</td>
<td>Overexpression</td>
<td>Gene</td>
<td>Function</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------------</td>
<td>---------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293</td>
<td>-</td>
<td>Dicer</td>
<td>Mitochondrial, Oxidative Damage, and PI3K/AKT Pathway</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dopaminergic Neuroblastoma Cells</td>
<td>-</td>
<td>DJ1</td>
<td>GG/CC Rich</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hESC</td>
<td>-</td>
<td>Fox2</td>
<td>~7% of Genes in hESCs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HeLa</td>
<td>-</td>
<td>GW182</td>
<td>18 mRNAs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HeLa</td>
<td>-</td>
<td>HuR</td>
<td>57% of Targets were AUF1 Targets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCF10A</td>
<td>-</td>
<td>MCT-1</td>
<td>Cell Cycle and MAPK Signaling Upon Transformation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCF7</td>
<td>-</td>
<td>MCT-1</td>
<td>Ras, PI3K, Angiogenesis, and Hypoxia Pathways</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RKO</td>
<td>ActinomycinD</td>
<td>HuR</td>
<td>U-Rich 12- to 20-nt long Stem Loop</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293</td>
<td>-</td>
<td>IGF2BP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alphaT3-1</td>
<td>-</td>
<td>KSRP</td>
<td>12 mRNAs (many RBPs)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HeLa</td>
<td>-</td>
<td>KSRP</td>
<td>~11% genes expressed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T98G</td>
<td>Co-cultivated</td>
<td>PABP</td>
<td>Human T98G Specific Translational Profile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293</td>
<td>-</td>
<td>PABPC4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HeLa</td>
<td>-</td>
<td>PTB</td>
<td>Intracellular Transport, Vesicle Trafficking and Apoptosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HeLa</td>
<td>-</td>
<td>Pum1</td>
<td>Cell Cycle and Transcription Factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HeLa</td>
<td>-</td>
<td>Pum1</td>
<td>Nucleic Acid Binding and p53 pathway</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HeLa</td>
<td>-</td>
<td>Pum1</td>
<td>Cell Cycle and Transcription Factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293</td>
<td>-</td>
<td>QK1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3T3</td>
<td>-</td>
<td>SAM68</td>
<td>Cell Migration, Cell Cycle, and Transcriptional Regulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293</td>
<td>-</td>
<td>SFRS1</td>
<td>RNA Processing Factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293</td>
<td>-</td>
<td>SF2/ASF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell Line</td>
<td>Gene</td>
<td>Functional Role</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>----------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HeLa-SLBP</td>
<td>Replication Dependent Histones</td>
<td>Stem Loop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RKO-TIA1</td>
<td>U-Rich 30-37 nt bipartite element</td>
<td>T, M [35]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RKO-TIAR</td>
<td>Many Targets Dissociate upon UVC</td>
<td>C-Rich with stem and loop with additional loop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293-TNRC6A</td>
<td>-</td>
<td>T, M [36]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293-TNRC6B</td>
<td>-</td>
<td>T, M [36]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293-TNRC6C</td>
<td>-</td>
<td>T, M [36]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human Dendritic Cells-/- LPS and CD40L-TTP</td>
<td>Protein Synthesis</td>
<td>AAUAAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293-TNRC6A</td>
<td>-</td>
<td>E, M [37]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293-TNRC6B</td>
<td>-</td>
<td>E, M [37]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293-TNRC6C</td>
<td>-</td>
<td>E, M [37]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human Dendritic Cells-/- LPS and CD40L-TTP</td>
<td>Protein Synthesis</td>
<td>AAUAAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293-TNRC6A</td>
<td>-</td>
<td>E, M [37]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293-TNRC6B</td>
<td>-</td>
<td>E, M [37]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEK 293-TNRC6C</td>
<td>-</td>
<td>E, M [37]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human Dendritic Cells-/- LPS and CD40L-TTP</td>
<td>Protein Synthesis</td>
<td>AAUAAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Mouse

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Gene</th>
<th>Functional Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>P19-EIF4E</td>
<td>A Subset of Expressed mRNAs</td>
<td>-</td>
</tr>
<tr>
<td>ES Cells-Esg1</td>
<td>Cell Cycle, Chromatin Remodelling, and Transcription Factors</td>
<td>-</td>
</tr>
<tr>
<td>Brain-FMRP</td>
<td>G-Quartet</td>
<td>E, M [38]</td>
</tr>
<tr>
<td>Brain-HuB</td>
<td>Early Response Genes</td>
<td>UUUUUU</td>
</tr>
<tr>
<td>Brain-HuR</td>
<td>-</td>
<td>E, M [38]</td>
</tr>
<tr>
<td>Brain-JERKY</td>
<td>Ribosomal Proteins and Cellular Stress</td>
<td>-</td>
</tr>
<tr>
<td>Brain-Nova</td>
<td>Also bound 3'UTRs to regulate polyadenylation</td>
<td>YCA rich regions</td>
</tr>
<tr>
<td>Brain-Nova</td>
<td>Neuronal Synapse, Neuronal Inhibition</td>
<td>YCA rich regions</td>
</tr>
<tr>
<td>Brain-PABP</td>
<td>A Subset of Expressed mRNAs</td>
<td>-</td>
</tr>
<tr>
<td>PY4.1-Co-cultivated with T98G and 4TI-PABP</td>
<td>Mouse Endothelial PY4.1 Specific Translational Profiles</td>
<td>-</td>
</tr>
<tr>
<td>24 CNS Cell Populations-Rpl10a</td>
<td>Translational Profiles for CNS Cell Populations</td>
<td>-</td>
</tr>
<tr>
<td>Striatonigral and Striatopallidal Medium Spiny Neurons-Cocaine Stimulated-Rpl10a</td>
<td>GABA-A Receptors Upon Cocaine Stimulation</td>
<td>-</td>
</tr>
<tr>
<td>RAW 264.7-LPS activation-TTP</td>
<td>Development</td>
<td>UUUAAUU, AUUUA</td>
</tr>
</tbody>
</table>

### References

1. [35]
2. [36]
3. [37]
4. [13]
5. [29]
6. [38]
7. [39]
8. [40]
9. [41]
10. [39]
11. [42]
12. [43]
13. [44]
14. [39]
15. [28]
16. [45]
17. [46]
18. [47]
<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>RNA Structure</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aco1</td>
<td>Mitochondrial Encoded RNAs</td>
<td>-</td>
<td>[49]</td>
</tr>
<tr>
<td>Atg8</td>
<td>-</td>
<td>-</td>
<td>[49]</td>
</tr>
<tr>
<td>Bfr1</td>
<td>Proteins Localized to Nucleus and Nucleus</td>
<td>UGAAAAWUUUY</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Bud27</td>
<td>Component of RNA Polymerase I Complex</td>
<td>-</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Cbc2</td>
<td>Intron-containing Transcripts and Potential CUTs</td>
<td>-</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Cbf5</td>
<td>H/ACA snoRNAs and potentially some mRNAs</td>
<td>-</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Dhh1</td>
<td>-</td>
<td>-</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Gbp2</td>
<td>Nucleolus and Chromosome, Ribosome Biogenesis, and Chromatin Modification</td>
<td>UGAAAAWUUUY</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Hrb1</td>
<td>Intron-containing Transcripts</td>
<td>-</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Idh1</td>
<td>-</td>
<td>-</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Khd1</td>
<td>Cell Wall and Bud Tip</td>
<td>(6) CNN repeats</td>
<td>T, M [50]</td>
</tr>
<tr>
<td>Lhp1</td>
<td>RNA Pol II Transcripts and Ribosomal Proteins</td>
<td>-</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Mex67</td>
<td>Integral Membrane Proteins, Cell Wall, RNA-binding Protein, Ribosomal Proteins</td>
<td>-</td>
<td>T, M [52]</td>
</tr>
<tr>
<td>Msl5</td>
<td>Intron-containing Transcripts</td>
<td>-</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Myo4</td>
<td>Bud localized, Stress Sensing/Response</td>
<td>4 Loops Separated by Short Stem</td>
<td>T, M [53]</td>
</tr>
<tr>
<td>Myo4</td>
<td>Ash1, IST2 and other Bud localized mRNAs</td>
<td>-</td>
<td>T, M [54]</td>
</tr>
<tr>
<td>Nab2</td>
<td>Nucleus, Nucleic Acid Metabolism</td>
<td>DRARAMGMD</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Nab2</td>
<td>Transcription</td>
<td>AAAAG</td>
<td>T, M [55]</td>
</tr>
<tr>
<td>Nab3</td>
<td>snoRNAs, Potential CUTs</td>
<td>-</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Nab4</td>
<td>Amino Acid Metabolism</td>
<td>UAUAUAA</td>
<td>T, M [55]</td>
</tr>
<tr>
<td>Nab6</td>
<td>Cell Wall</td>
<td>-</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Nce102</td>
<td>-</td>
<td>-</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Nop13</td>
<td>5.8S and 35S ribosomal RNAs</td>
<td>-</td>
<td>T, M [49]</td>
</tr>
<tr>
<td>Nop56</td>
<td>C/D Box snoRNAs and Potentially Many mRNAs</td>
<td>UUUUCUUUUU</td>
<td>T, M [49]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Npl3</td>
<td>Cytosolic and Mitochondrial Ribosome, Oxidative Phosphorylation, Intron-containing Transcripts</td>
<td>-</td>
<td>T, M</td>
</tr>
<tr>
<td>Npl3</td>
<td>Ribosomal Proteins</td>
<td>-</td>
<td>T, M</td>
</tr>
<tr>
<td>Nrd1</td>
<td>snoRNAs, Potential CUTs</td>
<td>UUCUGUW</td>
<td>T, M</td>
</tr>
<tr>
<td>Nrp1</td>
<td>-</td>
<td>-</td>
<td>T, M</td>
</tr>
<tr>
<td>Nsr1</td>
<td>Ribosomal Proteins, Ribosomal Internal Transcribed Regions (ITS)</td>
<td>GGGWAACGW</td>
<td>T, M</td>
</tr>
<tr>
<td>Pab1</td>
<td>Ribosomal Proteins and Intron-containing Transcripts</td>
<td>WUAUAUAU</td>
<td>T, M</td>
</tr>
<tr>
<td>Pbp2</td>
<td>-</td>
<td>-</td>
<td>T, M</td>
</tr>
<tr>
<td>Pin4</td>
<td>Novel Sequence Element in 3'-UTRs</td>
<td>UUAAUGA</td>
<td>T, M</td>
</tr>
<tr>
<td>Pub1</td>
<td>Ribosomal Proteins</td>
<td>U-Rich Degenerate Stabilize Elements (STE)</td>
<td>T, M</td>
</tr>
<tr>
<td>Pub1</td>
<td>Localized ER, Vacuole and Plasma Membranes, Cell Wall, Intron-containing Transcripts</td>
<td>HUUUUAH</td>
<td>T, M</td>
</tr>
<tr>
<td>Pub1</td>
<td>Membrane-Associated Proteins</td>
<td>CHUGUAWAUA</td>
<td>T, M</td>
</tr>
<tr>
<td>Pu3</td>
<td>Mitochondrial Proteins</td>
<td>WHUGUAHUA</td>
<td>T, M</td>
</tr>
<tr>
<td>Pu4</td>
<td>Nucleolar Ribosomal RNA-Processing Factors</td>
<td>UGUAYAWUA</td>
<td>T, M</td>
</tr>
<tr>
<td>Pu5</td>
<td>Chromatin Modifiers and Spindle Pole Body</td>
<td>-</td>
<td>T, M</td>
</tr>
<tr>
<td>Pu5</td>
<td>Several tRNAs</td>
<td>-</td>
<td>T, M</td>
</tr>
<tr>
<td>Rna15</td>
<td>-</td>
<td>-</td>
<td>T, M</td>
</tr>
<tr>
<td>Scp160</td>
<td>Cell Wall, Plasma Membrane, Nuclear Membrane, ER Proteins, and Ribosome Biogenesis and Transport</td>
<td>UGAAAAUUU</td>
<td>T, M</td>
</tr>
<tr>
<td>Scp160</td>
<td>~1% of Yeast Genes</td>
<td>-</td>
<td>T, M</td>
</tr>
<tr>
<td>Sgn1</td>
<td>RPS20</td>
<td>-</td>
<td>T, M</td>
</tr>
<tr>
<td>She2</td>
<td>Bud localized, Stress Sensing/Response</td>
<td>2 Loops Separated by Short Stem</td>
<td>T, M</td>
</tr>
<tr>
<td>She2</td>
<td>Ash1, IST2 and other Bud localized mRNAs</td>
<td>-</td>
<td>T, M</td>
</tr>
<tr>
<td>She3</td>
<td>Bud localized, Stress Sensing/Response</td>
<td>3 Loops Separated by Short Stem</td>
<td>T, M</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>She3</td>
<td>Ash1, IST2 and other Bud localized mRNAs</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Ski2</td>
<td>Ski3 and Ski7 of the Exosome</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Sof1</td>
<td>C/D Box snoRNAs Including U3 snoRNA</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Ssd1</td>
<td>Proteins Localized to the Cell Wall and Bud</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Tdh3</td>
<td>Mitochondrial-encoded mRNAs, Many Nuclear-encoded mRNAs</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Upf1</td>
<td>-</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Vts1</td>
<td>Carbohydrate Transport</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Yli032c</td>
<td>Transport, Development, Cell Signaling</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Ypl184c</td>
<td>Proteins Localized to the Cell Wall, ER, Plasma Membrane, mRNAs Containing uORFs and IRES</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Yra1</td>
<td>Cell Wall, RNA-binding Proteins, Ribosomal Proteins, Carbohydrate Metabolism</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>Yra2</td>
<td>-</td>
</tr>
</tbody>
</table>

**Arabidopsis**

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Leaf</td>
<td>-</td>
<td>RPL18</td>
<td>-</td>
<td>-</td>
<td>T, M</td>
<td>[59]</td>
<td></td>
</tr>
<tr>
<td>Leaf</td>
<td>-</td>
<td>RPL23a</td>
<td>-</td>
<td>-</td>
<td>T, M</td>
<td>[59]</td>
<td></td>
</tr>
</tbody>
</table>

**Maize**

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroma</td>
<td>-</td>
<td>CRP1</td>
<td>Photosystem I Core Complex</td>
<td>Identical 7-mer and 11-mer separated by 51 nts</td>
<td>T, M</td>
<td>[60]</td>
<td></td>
</tr>
<tr>
<td>Leaf</td>
<td>-</td>
<td>PPR4</td>
<td>rps12 pre-mRNA</td>
<td>-</td>
<td>T, M</td>
<td>[60]</td>
<td></td>
</tr>
<tr>
<td>Leaf</td>
<td>-</td>
<td>PPR5</td>
<td>Intron of unspliced plastid tRNA precursor pre-trnG-UCC</td>
<td>-</td>
<td>E, M</td>
<td>[61]</td>
<td></td>
</tr>
</tbody>
</table>

**C. Elegans**

|          |          |          |                                                                                           |          |          |          |          |
|----------|----------|----------|------------------------------------------------------------------------------------------|          |          |          |          |
| Intestine| Intestinal Expression of Pab1 | Pab1 | Intestinal Translational Profiles, Chromosomal Clustering                                  | -        | T, M     | [62]     |
| Muscle   | Muscle Expression of Pab1    | Pab1 | Muscle Translational Profiles, Chromosomal Clustering                                     | -        | T, M     | [63]     |

**Drosophila**

|          |          |          |                                                                                           |          |          |          |          |
|----------|----------|----------|------------------------------------------------------------------------------------------|          |          |          |          |
| S2       | miR1-KO  | Ago1     | 89 and 108 mRNAs respectively                                                            | ~ 30% contained miR-1 seed | T, M     | [64]     |
|          | PABP     | Different Classes of Photoreceptor Genes                                                  | -        | T, M     | [65]     |
| Adult Ovary | - | Pum | Transcriptional Regulation, V-ATPase, Cyclins | UGUAAUA | T, M | [66] |
| Embryo     | - | Pum | Germ Cell Development andAnterior-Posterior Axis Patterning | UGUAAUA | T, M | [66] |
| Embryo     | - | U2AF | Nuclear Export of Intronless mRNAs | - | E, M | [67] |
| S2         | - | Hrp36 | Combinatorial Regulation by Hrp38 and Hrp40 | - | E, M | [68] |
| S2         | - | Hrp38 | Combinatorial Regulation by Hrp36 and Hrp40 | - | E, M | [68] |
| S2         | - | Hrp40 | Combinatorial Regulation by Hrp36 and Hrp38 | - | E, M | [68] |
| S2         | - | Hrp48 | Distinct subset from Hrp36, Hrp38 and Hrp40 | - | E, M | [68] |

**Trypanosome**

| CL Brener | - | UBP1 | Metabolism and Glycoproteins | 25-30 bp stem loop | E, M | [69] |
| CL Brener | - | RBP3 | Ribosomal Proteins | 25-30 bp stem loop | E, M | [69] |
| 427 T. brucei | - | DRBD3 | Membrane Proteins | U-Rich | E, M | [70] |
1.3 ELAV family RNA-binding proteins

HuR (ELAVL1) is a ubiquitously essential protein member of the ELAV family of RBPs. HuR is necessary for proper embryonic development and the immune response [71, 72]. Additionally, it is known to be crucial in aging, stress-response, and cancer. ELAV proteins contain three highly conserved (>75% identity) RRMs (RNA recognition motifs), which are the RNA-binding domains of the proteins. Investigation of RNA-targeting by Hu proteins, both *in vitro* and *in vivo*, suggest that the binding sites are single-stranded short stretches of uridylates often separated by adenosines. Thus, these proteins are part of a broader class of RBPs, which interact with RNA regulatory elements known as AREs (AU-rich elements) discovered in the mid 1980s [73, 74]. Although the mRNA targets of HuR have been identified in numerous cell-types and conditions [75-77], there has not been a definitive global study identifying the precise RNA-binding sites *in vivo*.

HuR positively regulates the stability and translation of its target mRNAs [78]. This positive regulatory effect by HuR is a distinctive feature compared to other RBP-ARE interactions, as well as post-transcriptional regulation in general given the prevalence of miRNAs. Additionally, mammalian HuR and other ELAV proteins have been implicated in other aspects of RNA processing, such as poly-adenylation [79], splicing [80], and nuclear export [81], however it is based on very few specific cases and not shown to be a generalized function of Hu proteins.
Figure 2: States of HuR. Depicted are A) protein domains, associated proteins, and all known post-translational modifications and B) known signaling events and their consequences.

Specific posttranslational (PTL) modifications of HuR are consequent to distinct cellular processes (Figure 2). These PTL modifications include phosphorylation [82-84], methylation [85], ubiquitination [86], and cleavage [87]. Concomitant with these PTL modifications is regulation of the subcellular localization of HuR. The cellular process initiating these changes range from response to specific stimuli (oxidative stress, immune activation, heat shock, ATP levels, and others) to basic cellular processes (cell cycle and differentiation). Interestingly, some of these phosphorylation events may alter the target sequence specificity of HuR. Therefore, most RIP experiments are actually enriching a heterogeneous population of HuR RNPs, which are not likely to be in
functionally equivalent states due to differences in post-translational modifications and/or subcellular localizations.

1.4 Goals of this thesis

It is my contention that it is important to analyze high-throughput biochemical RBP-RNA interaction data in a manner preserving quantitative differences between targets and across cellular conditions. I hypothesize the quantitative difference exhibited in these genomic data underlie the mechanisms by which RBP-RNA interactions result in specific functional outcomes for target mRNAs. By considering RNAs as either targets or not targets, such mechanisms would be obscured. For example, it probably does matter if a mRNA has 1 or 10 binding sites for a given RBP; or if 95% of the copies of a given species of mRNA are occupied by an RBP rather than only 5%; or if across conditions there are differences in the proportion of an RBP existing in non-functionally equivalent phosphorylation states. Quantitative examination of these data will further our understanding of such complicated problems, for which current methods do not provide the necessary level of resolution.

Finally, subtle coordinated differences in biology, gene expression, and post-transcriptional regulation can have major effects on how a cell, organism, or system behaves. There are numerous widely accepted examples of this, from SIR (susceptible-infected-recovered) models of disease epidemics to control theory applied to metabolic pathways (below).

“Metabolic control theory suggests that small adjustments in many sequential steps of a metabolic pathway can lead to a substantial change in the total flux through the pathway, whereas large changes in a single enzyme might have no measurable effects” –Mootha et al, Nat Gen 2003
This quote is from the first paper utilizing GSEA (gene set enrichment analysis), which is a tool to identify coordinated changes in a priori defined gene sets. The study found that mRNAs involved in oxidative stress were coordinately differentially regulated by a transcription factor in individuals with type 2 diabetes mellitus [88]. This theoretical basis may be due to what is described as the “butterfly effect”, in which subtle changes can drastically perturb a system depending on the initial conditions of the system. However, there exists a belief in molecular biology that subtle changes, anything less than a 2-fold change, are not meaningful. It is incongruous to recognize that subtle changes in epidemic or metabolic models can have large effects, but not for gene expression. This is particularly important for post-transcriptional regulation, which is often thought to be fine-tuning. It will be interesting to see to what extent combinatorial regulation by PTRFs capitalize on the initial conditions of a system and through subtle coordinated changes across mRNAs encoding functionally related proteins drive important physiological processes such as differentiation and development, response to stimuli, and even progression or maintenance of disease.
2. Quantitative analysis of RIP-chip data

Ribonomic profiling has been widely used to identify mRNAs associated with a given RBP [89, 90], however the overwhelming majority of these studies utilized RIP-Chip experiments from a single condition of growth or perturbation. This is in part due to the lack of analytical approaches for modeling RIP-chip data to determine targets and assign values of condition-specific RNP-association that allow systematic comparisons across physiological conditions. Therefore, the development of probabilistic models of RNP-association will allow a more thorough and systematic investigation of the contribution of RNP dynamics to the molecular networks activated during development or in response to perturbations.

Most methods for analyzing RIP-chip data generate a list of mRNAs, designated as either “target” or “not a target” using an arbitrary cut-off, all of which are subsequently considered equivalent in their likelihood of being associated with the RBP of interest. Furthermore, two problems not sufficiently addressed by these current approaches are the objective determination of thresholds of protein-mRNA association and the assignment of values that reflect the relative association of mRNAs across conditions [91-95]. If a value representing a condition-specific probability of RBP-association could be assigned to each mRNA, one could assess RNP remodeling by comparing these quantitative values across physiological conditions, rather than comparing “all or none” values (i.e. target or not target), in which case information will be lost. The development of more objective approaches that quantify condition-specific association and changes in RNP-association would allow a more detailed investigation of the contribution of RNP dynamics to molecular networks activated in response to perturbations.
2.1 **LOD scores**

After normalization of microarray data and preprocessing for expression, t-scores for RNP-IP versus mock-IP were calculated using GSEA. A histogram of the t-score distribution indicated two populations of mRNAs (Figure 3A), an enriched population representing RNP-associated mRNAs and a non-enriched population representing background mRNAs.
Figure 3: Overview of ribonomic analysis. (A) Isolation of RNP complexes, in this case HuR (blue), in parallel with IgG negative control, followed by extraction of mRNAs and hybridization to microarray (RIP-chip). (B) Gaussian mixture modeling, of RNP IP versus negative IP $t$-scores for 3 biological replicates of RIP-Chip, to identify and quantify biochemically enriched populations of mRNAs represented by probes. (C) LOD scores representing a condition-specific probability of HuR association per probe. (D) Discrete approach, LOD HuR > 0, defines a subset of probes associated with HuR in a given biological condition. Continuous approach, LOD score values to compare HuR mRNA dynamics across different biological conditions, with other data types or databases, and identify small molecule effectors of HuR. Both approaches identify common functional groups or motifs associated with HuR. This framework has been utilized for different RBPs, cell lines, and array platforms.
Gaussian mixture modeling (GMM) was first devised in 1894 by Karl Pearson to discriminate genetic subpopulations of prawns based on carapace size [96]. We applied GMM to identify and quantify the biochemically-enriched population of mRNAs associated with RNPs in a particular context of time or condition of treatment. GMM estimated the parameters of the Gaussian curve (the mean, standard deviation, and weight) representing each population or component within the distribution. Each component of the mixture model corresponded to a conditional probability of a t-score (the continuous variable) given class membership, “target” or “not target” (the discrete variable). Given the relative measure of enrichment, the t-score, we assigned class membership, “target” or “not target”, based on the log of odds ratio of the corresponding inferred mixtures. First, we treated RNP-association as a discrete variable by modeling the continuous variable, the t-score, conditioned on the class membership to “target” or “not target” and second, we treated RNP-association as a continuous variable by providing a conditional probability for class membership given a t-score. Invariably, the model with the highest log-likelihood demonstrated excellent fit to the data and discriminated the RNP-associated population (Figure 3B, blue curve) from background populations (Figure 3B, green curves).

We then generated values representing the probability of RNP association for each probe at each time point by calculating a log of odds (LOD) ratio comparing the weighted probability density function for the RNP-associated distribution to the sum of all background distributions based on a given probe’s t-score (Figure 3C). The RNP LOD scores provide a continuous variable representing the condition-specific probability of RNP association for each probe, allowing us to determine and to compare changes in the likelihood of RNP association throughout the activation. Further, to gain insight into the similarities and differences of various layers of gene expression, we compared these
LOD scores with other data types, such as transcriptomic data, and incorporated other datasets, such as other published RIP-Chip experiments, into our analysis (Figure 3D). This approach has been successfully employed across RBPs (Pumilio-1, HuR, and PABP), cell lines (HeLa, Jurkat, and HEK293), and cellular conditions (basal, T-cell activation).

### 2.2 Percentile rank

The RNP-IP versus mock IP t-score distribution is not always clearly multimodal. In many cases RIPs are not performed with matching mock IPs, but rather as a standalone or, in some cases, only compared to total mRNA levels. An alternative method developed and utilized for the RBPs TTP and SLBP, is the median percentile rank \([97, 98]\). In this approach, a percentile rank transformation is applied to individual samples and then one can use either the median or mean of the transformed values across all replicate samples. This process can result in an enriched population, similar to that found above using t-scores, which can be used to determine a cut-off for RBP targets. In our case we utilized an average percentile rank for DRBP76 RIP-chip data (Figure 4B). This was done because the t-score comparison of DRBP76 RIP versus mock RIP did not exhibit a bimodal nature (Figure 4A). We conclude there is no single right way to analyze RIP-chip data and multiple approaches should be examined for any particular data set.
Figure 4: Identification of DRBP76-associated mRNAs. Deviation of data from the diagonal quantile:quantile plot indicates the non-Gaussian nature of the DRBP76 IP vs. mock IP distribution, corresponding with transcripts specifically enriched in the DRBP76 IP. B. Distribution of the average percentile ranks (APR) of 5 biological replicates for DRBP76 RIPs and mock RIPs. Transcripts considered DRBP76 RNP-associated are indicated by black boxes.

2.3 Comparison of various enrichment metrics

Though there are many RIP-chip studies, the methods for analysis of these data are quite diverse. For example, some studies calculate the enrichment of immunoprecipitated mRNAs over a mock IP, while others calculate enrichment over total mRNA levels. Many different metrics are used to calculate enrichment; these include the t-score, median percentile rank, and fold enrichment. Surprisingly, there has not been a systematic comparison of these methods to one another. Furthermore, each method relies on certain assumptions and has different inherent biases. Here we examine the performance of current methodologies for the analysis of RIP-chip data.

In order to compare various metrics, we utilized our HuR T-cell activation dataset, for which we had high confidence in the quality of the data. In addition to average percentile rank, t-scores and log fold enrichment metrics were used to compare both RNP-IP to mock IP and RNP-IP to total mRNA. We looked for enrichment of
mRNAs containing independently motifs derived from earlier studies, specifically the COVE-model for HuR.

![Figure 5: Comparison of RIP-chip analysis metrics. Enrichment was assessed by using A) KS (kolmogorov-smirnov) tests for the whole distribution, and by B) determining the percent of genes containing a motif for the top 500, 1000, and 500-1000, mRNAs ranked by the respective metric. Results were averaged across 0, 4, and 12 hr HuR RIP-chip during T-cell activation. Error bars represent standard deviation. Based on the results, we conclude that comparing RNP-IPs versus mock IPs rather than total mRNA results in superior signal to noise (Figure 5). All other metrics gave similar results, but only the distributions of t-score for the RNP-IP versus mock IP was bimodal allowing for GMM and calculation of LOD scores.](image-url)


2.4 Discussion

For well-performed RIP-chips, in which RNP-IPs are compared to mock IPs and not total mRNA, there is minimal difference between the top results and rank-ordered results of the many diverse approaches to analyze RIP-chip data [91-95]. However, the advantage of probabilistic mixture modeling, such as GMM, is the full specification of the distribution generating the data. In our case, this results in the specification of a Gaussian distribution for each mixture and the probability of an mRNA belonging to each mixture. Using this model we can discriminate HuR associated mRNAs from background mRNAs, as well as generate condition-specific probabilities of association. This is especially useful for assessing condition-specific differences in the likelihood of RNP-association for all mRNAs detected. Such a probabilistic framework is particularly appealing given the stochasticity in gene expression among individual cells and the consequent heterogeneity within a population of cells implicit in most biological experiments [99, 100].

2.5 Methods

Deriving LOD scores. GSEA was used to calculate t-scores comparing the RNP-IP to matching the IgG IP. Gaussian mixture modeling was performed multiple times on the t-score distributions to estimate the mean, standard deviation and weight of each component using the Mixtools package in R [101]. The number of components was determined by visual inspection. Since this implementation of GMM used expectation maximization, which is prone to convergence on local optimum, multiple runs of GMM were conducted that initialized at different points. The parameters from the model with the highest likelihood were used to create LOD scores of RNP-association by comparing the weighted probability density functions of the RNP-associated versus the background
distribution or in the case of multiple “non-enriched” populations the sum of the background distributions.

**Log fold enrichment.** First, the average was calculated for triplicate array samples of RNP-IPs, mock IPs, and total mRNA. Since the data was in log2 space, average mock IP values or average total mRNA values were subtracted from average RNP-IP values per probe.

**Average percentile rank.** First, the percentile rank transformation was applied to each individual replicate of the RNP-IP. Next the values for each probe were averaged across biological triplicates.

**Enrichment.** Gene sets were made for all mRNAs containing a HuR COVE motif and the Pum1 motif. GSEA using the classic enrichment metric, which reduces to a kolmogorov-smirnov test, was utilized to compare the motifs in all the various rank metrics.

**Percent overlap.** The percentage of genes containing a motif in the top 500, 1000, and 500-1000 ranked genes were calculated for lists ranked by all metrics described.
3. Identifying HuR targets and precise binding sites

We utilized two complementary methods, RIP-chip and PAR-CLIP, to identify targets of HuR and high-resolution binding sites on a transcriptome-wide scale. These experiments were performed in normally grown HEK293 cells under basal conditions. Specifically, we could induce the expression of an epitope-tagged HuR with tetracycline. This allowed this data to be utilized downstream in the investigation of specifically engineered HuR variants.

3.1 PAR-CLIP

We applied the recently developed PAR-CLIP technique, as described in [9], to precisely identify RNA elements interacting with HuR in vivo. All experiments were performed in the same cell line as the initial PAR-CLIP study (HEK 293), thereby allowing data derived in that study to be integrated. HuR bound RNA fragments were successfully identified (Figure 6A), converted into cDNA after sequencing adapter ligation and subsequently sequenced using Illumina.
Figure 6: HuR PAR-CLIP. A) A radioactive band representing directly bound RNA fragments at expected size of the epitope tagged HuR. B) Kernel density estimates of conversions versus non-conversions were utilized to map high resolution binding sites. C) Distribution of HuR binding sites shows most clusters are well below the 40-nucleotide clusters (dashed line) identified by the original analytical method. D) Binding sites are more likely to be near each other than random (dashed lines represent 1st, 50th, and 99th percentiles of the null distribution).

3.1.1 Defining binding sites

We applied a probabilistic approach using kernel density estimation (KDE) to determine the precise positions of HuR binding sites based on the occurrence of T-C conversions, which indicate protein-RNA cross-linking sites (Fig 1B). Read data from Illumina libraries were mapped to the genome, grouped by overlaps, and then binding sites were derived using KDE for conversion events and non-conversion events with the
group (Figure 6B). We first found regions for which the density estimates for the conversion were higher than the non-conversion for at least 5 nucleotides. Next three adjacent nucleotides were added to both ends, to produce a minimal cluster size consistent with known HuR biochemistry, unless there was no supporting read evidence, in which case the cluster would end. The average length of each cluster was 21 nucleotides (Figure 6C), with very few clusters likely to contain more than one binding site. The higher resolution, relative to the existing PAR-CLIP analysis, which yields 40 nucleotide clusters, was important because otherwise we would not have noticed that HuR clusters are often about 30 nts from adjacent HuR clusters (Figure 6D). Therefore, this new analytical approach utilizing the T-C conversions represents an improvement in the resolution binding events.

3.1.2 Characteristics of binding sites

We scored each cluster for evidence of cross-linking by calculating a cross-linking index, CLI = \(\log_{10}(\text{observed conversions for all reads in cluster})\). Importantly, clusters with high CLIs not only have many conversions, but they are also comprised of many reads and thus are well represented in the PAR-CLIP library. Of the \(~78,362\) clusters identified, the majority mapped to 3’ UTRs, introns, and intergenic regions. Comparing the distribution of CLIs for clusters mapping to different regions of the transcriptome revealed a very obvious hierarchy, with regions representing mature mRNA being the highest (3’ UTR, CDS, and 5’ UTR), followed by introns, and finally intergenic regions (Figure 7).
Figure 7: Cluster quality across all genomic locations. A) Proportion of densities and B) boxplots of cluster CLI scores are displayed by genomic location.

A closer examination of the clusters, particularly the intergenic clusters overlapping with repetitive elements, suggest most repetitive elements had significantly lower evidence of crosslinking and were filtered out (Figure 8) leaving 73,599 clusters. The exceptions were the low complexity and unknown repeats. The low complexity repeats were over-represented for T-rich and AT-rich sub-categories of repetitive
elements. While there were intergenic clusters that had similar characteristics to genic HuR targets, for the rest of this study we focused on clusters mapping to mRNA.

Figure 8: Clusters mapping to repetitive sequences. All repetitive elements except for low complexity and unknown repeats had statistically significant lower CLI scores and were discarded.

Mapping clusters to mRNA regions, revealed that 3’ UTRs and introns were the transcript regions primarily bound by HuR with considerably more intronic binding sites than anticipated (Figure 9A). Clusters with high CLIs mapped to the 3’ UTR at a much higher proportion than to introns; whereas, clusters with average or low CLIs mapped at a higher proportion to intron and intergenic regions, respectively. This is
consistent with current hypotheses that the preferred mode for HuR targeting is via the 3’ UTR to execute its primary function as a regulator of mRNA stability and/or translation. However, the surprising degree of intronic binding indicates a substantial role for HuR in pre-mRNA processing.

Figure 9: Landscape of HuR binding sites in mRNAs. A) The proportion of densities for HuR binding clusters derived from each region of an mRNA as a function of cross-linking index (CLI). B) Histogram of the distribution of the number of HuR binding sites per mRNA. C) Pie-chart depicting mRNAs with specific regions of HuR binding sites.
The distribution of binding sites across individual transcripts provided several insights into HuR mRNA targeting. First, the average number of binding sites per transcript was ~6, with some transcripts having well over 100 sites (Figure 9B). Also, binding sites tended to occur much closer to each other than random, suggesting possible multimerization of HuR molecules at specific locations within a transcript (Figure 6D), an observation previously noted for other ELAV proteins. Almost all targeted transcripts had a binding site in either the 3’ UTR or an intron (>99%), while ~38% of transcripts had binding sites in both introns and 3’ UTRs (Figure 9C), which suggests HuR may couple splicing and stability for a subset of mRNAs as discussed below.

3.1.3 HuR binding motif

To determine the nucleotide composition of the HuR regulatory element we used cERMIT, an evidence-ranked motif discovery tool designed to investigate quantitative genome-wide binding evidence. We ranked clusters by the CI as the evidence to define the in vivo RNA binding site for HuR. The most enriched sequences contained stretches of 3-5 U’s separated by A or C (Figure 10A). Long stretches of only U were enriched in clusters with low CI scores (Figure 10B). We separately investigated the motif for intronic HuR binding sites significantly overrepresented for occurring directly upstream of the 3’ splice site (see below Figure 18A), and found it was U-rich and separated by C’s more often than A’s (Figure 10C). This poly-pyrimidine tract (Py-tract) compatible motif was compelling considering the substantial number of observed binding sites proximal to the 3’ splice site.
Figure 10: HuR binding motif. A) cERMIT was used to analyze HuR binding sites ranked by conversion index. Sequences, cERMIT scores, and number of targets are shown for the top ranked cluster of 6mers identified. B) Histogram of the distribution of UUUUUUUUs in clusters ranked by CI. C) Weeder-derived motif of HuR binding sites located within the first 50 nucleotides upstream of exons.
3.2 RIP-chip

RIP-chip was also employed to identify mRNA targets of HuR in vivo. We confirmed significant enrichment of positive control HuR targets in HuR RIPS compared to Mock RIPS with QRT-PCR as well as enough signal to detect targets by microarray (Figure 11A,B). RNA from these samples was interrogated with Affymetrix Human Exon 1.0 ST Arrays to obtain transcriptomic binding data complementary to PAR-CLIP data. After normalization and preprocessing, t-scores for HuR RIPS versus Mock RIPS were calculated at the whole transcript level. Similar to previous RIP-chips [77, 102], visual inspection of the t-score distribution indicated two populations of mRNAs: an enriched population representing HuR-associated mRNAs and a non-enriched population representing background mRNAs (Figure 11C). We performed Gaussian mixture modeling to calculate log odds ratios (LOD) representing the probability of an mRNA being associated with HuR (Figure 11D). Transcripts with HuR LOD scores greater than zero (higher probability of being associated with HuR compared to background) were considered a discrete population of HuR-bound mRNAs.
Figure 11: HuR RIP-chip. A) QRT-PCR of signal of controls from biological triplicate RIPs. **TNF was not detected in mock IP and therefore the cT was set to 40, which is the total number of cycles. This value was used to calculate enrichment as well. B) Significant enrichment of positive controls over both mock and totals.

3.3 Comparison of techniques

Next we evaluated the mRNA targets of HuR identified by the two methods. After collapsing data to the transcript level, 7,839 and 1,856 HuR bound transcripts were identified using PAR-CLIP and RIP-chip, respectively (Figure 12A). Over 70% of the HuR targets identified by RIP-chip were also identified by PAR-CLIP (1,337/1,856).
Conversely, only ~17% of the HuR targets identified by PAR-CLIP were also identified by RIP-chip (1,337/7,839). Compared to HuR targets identified by either method alone, targets identified by both methods were significantly more enriched than PAR-CLIP only targets based on the number of binding sites (Figure 12E), but were not significantly more enriched in the RIP-chip than RIP-only targets based on LOD scores (Figure 13 A,B).
Figure 12: Comparison of PAR-CLIP and RIP-chip HuR interaction data. A) Venn diagram of the overlap between mRNAs defined as HuR targets using PAR-CLIP and RIP-chip. Comparison of the abundance mRNAs detected by RIP-chip only (blue), PAR-CLIP only (green), and both using B) microarray data or C) RNA sequencing data. D) Comparison of the HuR LOD score distributions of PAR-CLIP data separated by number of binding sites per transcript. E) Comparison of PAR-CLIP binding site distribution of targets identified only by PAR-CLIP versus by both PAR-CLIP and RIP-chip.
Since we did not utilize cross-linking in our RIP protocol, the 519 HuR targets identified by the RIP-chip but not the PAR-CLIP were investigated to determine if they were true-positives. A U-rich motif that was found in the overlapping targets was also enriched in the 519 targets. Furthermore, these targets were functionally responsive (described later Chapter 4.2). We found that the HuR targets identified only by RIP-chip were significantly overrepresented for low abundance mRNAs determined by microarray or RNA-seq (Figure 12B,C), suggesting we did not reach a sequencing depth with the PAR-CLIP analysis necessary to identify all HuR targets. We conclude that the 519 transcripts identified by RIP-chip alone were functional HuR targets of lower abundance mRNAs.
Figure 13: Location of PAR-CLIP sites in RIP-chip data. A) Comparison of the number of binding sites in distinct regions of the mRNA for targets identified with PAR-CLIP only, or PAR-CLIP and RIP-chip. B) Comparison of the distribution of HuR LOD scores for mRNA targets identified by RIP-chip only, or RIP-CHIP and PAR-CLIP. C) Comparison of the distribution of HuR LOD scores for mRNAs containing binding sites only in 3' UTRs, only in introns, and in both 3' UTRs and introns, as well as those without binding sites.
HuR targets identified by PAR-CLIP but not RIP-chip could be explained by differences in the IP protocols, different detection strategies, and/or the kinetic stabilities of the RNA-protein interactions. Unlike PAR-CLIP, there is no RNase digestion step in the RIP-chip protocol. Thus, the RIP-chip may be biased against identifying intronic binding events due to the difficulty in immunoprecipitating long pre-mRNAs coated with mRBPs. Indeed mRNAs containing binding sites only in introns were significantly less enriched in the RIP-chip than those mRNAs containing binding sites in the intron and 3’ UTR or 3’ UTR only (Figure 13C). The use of microarrays rather than deep sequencing may contribute to these differences; however this is unlikely due to the extensive probe coverage for known mRNAs by the whole transcript arrays.

We investigated the hypothesis that RIP preferentially enriches for kinetically stable interactions assuming that more binding sites per transcript would result in a more kinetically stable interaction of an mRNA with HuR. Transcripts with more HuR binding sites were significantly more enriched by the RIP-chip than mRNAs with fewer binding sites. (Figure 12D,E). On average RIP-chip targets contained ~1.5 more binding sites than targets identified by PAR-CLIP only. This supports the hypothesis that RIP-chip enriches for more kinetically stable interactions, while PAR-CLIP identifies both kinetically stable as well as transient interactions, conceivably attributable to the high cross-linking efficiency.

3.4 Discussion

The two techniques employed, PAR-CLIP and RIP-chip, are synergistic in that combined they provide a more detailed understanding of the biochemical aspects underlying quantitative differences in mRNA targeting. The difference in the number of
targets defined by each method could be due to differences in the protocol, detection strategy, and analytical method for designating a target. After taking these into account, we propose that the kinetic stability of HuR-mRNA interactions at the level of populations of mRNAs and cells is the primary driving force for the differences between the two techniques (Figure 14).

The major differences between the two methods are a cross-linking step and digestion step. The UV 365 nm cross-linking of protein to thiolated-uridylates is exceedingly efficient and thus even transient interactions can be captured. Also, there is a pre-IP light RNase digestion step not present in the RIP-chip, but employed in the PAR-CLIP. This effectively breaks up large mRNP complexes assembled on a transcript and thereby allows identification of a single binding site in long pre-mRNAs coated with likely hundreds of other proteins. Such HuR-mRNA interactions would be difficult to find using the RIP protocol without digestion. In both cases above, the knowledge provided by the difference in the protocols are descriptive of interactions that are more or less likely to be stable and functionally relevant, which we would otherwise not have known without utilizing both methods.

The difference in detection strategy, deep sequencing and microarray also influences the differences between the observed targets. The coverage difference is not likely to contribute much for mRNAs since these Affymetrix arrays tile across the whole transcript with a minimum of 4 probes per exon and many intronic probes. High-throughput sequencing is known to be more sensitive than microarrays. However, we observed that RIP-chip only targets tend to be enriched for low abundance mRNAs. This suggests that we did not reach saturation in the PAR-CLIP, probably due to insufficient sequencing depth. These low abundance mRNAs have a better chance of hybridizing to a specific probe rather than having enough reads in the top ~10 million the library.
Altogether, we believe that the differences due to detection technology contribute little compared to differences in the kinetic stability of the interactions.
Lastly, the analytical methods for determining RNA targets may also influence the observed differences. To ensure accurate KDEs, we required at least 10 reads in the initial group and 5 read depth at any specific position being analyzed. Therefore it is possible that some of the RIP-chip only targets fall below this threshold. However, with much less stringent thresholds, we still see mRNAs in the RIP-chip and not the PAR-CLIP, again indicating that saturation was not achieved. As we had speculated before (Chapter 2) it is likely that the LOD > 0 cutoff is conservative. This is very likely to be the case (Figure 12 D), however even reducing the cut-off to a LOD score > -2 results in ~4,000 targets, still half of what was identified by PAR-CLIP. Indeed the PAR-CLIP data may be quite useful in devising more informed methods for RIP-chip analysis in the future.

Figure 14: Summary of comparisons between mRNA targets identified using PAR-CLIP versus RIP-chip.
The combined results of the two methods suggest that the HuR-mRNA targeting is best described by a quantitative spectrum, rather than a binary decision. Features including the number of binding sites per mRNA, the accessibility of the binding site, and the affinity of HuR toward the sequence of a given binding site together determine the kinetic stability of the interaction between HuR and a given mRNA. Our data clearly demonstrate that the more HuR binding sites per transcript the more likely it is enriched in the RIP-chip. Given that HuR preferentially binds to single-stranded RNA, it is very likely that many of the PAR-CLIP sites may in fact more often retain more secondary structure in vivo and therefore are not as enriched in the RIP-chip. This is consistent with previous biochemical and computational studies of HuR-mRNA interactions [103-105] and will be examined in this dataset.

Table 2: In vitro binding affinity of HuR towards both U-rich and AU-rich RNA. (Personnel communication M.G. and G.W.)

<table>
<thead>
<tr>
<th>Protein</th>
<th>RNA</th>
<th>ka(1/Ms)</th>
<th>kd(s-1)</th>
<th>kD (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HuR</td>
<td>U-Rich</td>
<td>1x10^7</td>
<td>5x10^-4</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>AU-Rich</td>
<td>4x10^6</td>
<td>8x10^-4</td>
<td>0.2</td>
</tr>
</tbody>
</table>

In vitro studies of both U-rich and AU-rich RNA probes have shown that the HuR binding affinities are less than 1 nM for both sequences (Table 2). HuR may bind to U-rich sequences with higher affinity than AU-rich sequences, however the interaction with AU-rich sequences once formed is more stable. While these different characteristics probably influence the differences between the results from the two methods, given the high affinity toward both sequences, this number of binding sites and accessibility probably contribute more.

It is very clear that methods to investigate HuR-RNA targeting exhibit quantitative differences that reflect the mode of binding of HuR in vivo. Specific features,
most prominently the number of binding sites per mRNA, exist as a continuous spectrum reflecting the overall strength of a HuR-mRNA interaction buried within the genomic data. These quantitative differences are also important for functional outcome as shown in chapter 4.

### 3.5 Methods

**Cell Culture.** HEK 293 Flip-In T-REx containing a N-terminal FLAG and HA tagged HuR were created using manufacturer protocol (Invitrogen). Cells were cultured in DMEM supplemented with 10% tet-reduced FBS, appropriate selection antibiotics. For induction, cells were treated overnight ~ 16hrs with 1 uL/ml doxycycline (Sigma).

**PAR-CLIP.** Protocol was performed as detailed in [9].

**RIP-chip.** Protocol was performed as described in (RIP-methods) except 50 uL of anti-FLAG M2-conjugated beads or mouse IgG-conjugated beads were used for immunoprecipitation reaction. Appropriate amounts of total RNA from biological replicates of HuR and Mock RIPS were submitted to the Microarray Facility at the Duke University IGSP. RNA was applied to GeneChip Human Exon 1.0 ST Array according to manufacturer’s protocol (Affymetrix).

**Microarray data processing.** Microarray data was normalized with PLIER using Affymetrix Power Tools. Probes with DABG p-values > 0.05 in all three of three HuR RIP replicates were excluded. T-scores were calculated with GSEA and LOD scores were computed using GMM as described in chapter 2.

**PAR-CLIP sequence library processing**

**Mapping reads.** The reads from the deep sequencing library were first stripped of the 3’ adapter sequence using the FASTX [106] toolkit. Reads that were less than 13nt in length or contained an ambiguous nucleotide were discarded. The remaining reads
were aligned to the human genome (hg19), with up to 3 mismatches allowed, by the Bowtie [107] algorithm. Mapped locations were only reported for the optimal mismatch-stratum for each read. All T to C mismatches between the small RNA and the genome were subtracted from the mismatch count for each mapped location. Only reads that mapped to a single genomic location with the minimum number of mismatches after that subtraction were used for further analysis. The location that a read mapped, relative to a known transcript, was determined based on the ENSEMBL v57 [108] database. If a read mapped to a location that could be placed in multiple categories, the read was reported to belong to the category based on the following order of preference: 3’UTR, coding sequence, 5’UTR, intron, non-coding RNA, intergenic.

**Defining HuR binding sites.** Overlapping reads or reads within 5nt of each other are grouped together for further analysis. A group must have contained ≥10 reads with ≥ 5 conversions cumulative at two or more locations. Kernel density estimates were calculated for conversions and non-conversions separately, as long as there were 5 nucleotides of read depth at a position to calculate a robust density estimate. Binding sites were defined as regions for which the conversion density was greater than the non-conversion density for at least 5 consecutive nucleotides. Three nucleotides were added to each end so the minimum cluster length was 11 nucleotides, based on known HuR *in vitro* binding studies. Clusters were not extended if there was no supporting read evidence.

**Distance between binding sites.** Using custom python scripts, each cluster's nearest companion cluster was calculated and compared to a null in which the clusters are randomly distributed in their genetic region. Observed density estimates and density estimates of the 1st, 50th, and 99th percentile of the null were displayed.
Motif finding. All clusters mapping to mRNA were ranked using CLI scores. cERMIT was utilized to determine motifs enriched at the top of the list ranked by CLI scores. This was repeated for clusters mapping to 3’ UTRs and introns separately. MEME was utilized to derive a motif for clusters mapping within the first 50 nucleotides upstream of exons.

Comparisons of distributions. JMP 7.0 (SAS) was utilized for all comparisons of distribution and densities for various categories and metrics. These include cumulative distribution fraction plots, comparison of densities, and proportion of densities.
4. Relationship between targeting and functional outcome

In order to relate the previously identified mRNA targets and binding sites, we depleted endogenous HuR from these cells and conducted transcriptomics on the mRNA using Affymetrix Human Exon 1.0 ST Arrays. As such, we would be able to identify alterations in mRNA abundance and isoform usage.

4.1 HuR knockdown

We utilized siRNA-mediated HuR knockdown combined with whole transcript expression profiling as a third independent genomic approach to identify HuR target mRNAs. We conducted transcriptomic analysis of cells treated with HuR siRNA and negative control scramble siRNA, which resulted in the most depletion of HuR protein (Figure 15A). The abundance of known targets of HuR, such as VEGF, TNF-alpha, and BRF1, were significantly lower after HuR depletion (Figure 15B). RNA from these samples was interrogated with Affymetrix Human Exon 1.0 ST Arrays. Importantly, these data allowed us to assess the functional significance of HuR targets identified by both PAR-CLIP and RIP-chip. Indeed, targets identified by both methods were significantly enriched for decreasing in response to HuR knockdown (Figure 15C). This is consistent with many studies demonstrating that HuR is a positive regulator of mRNA stability. Since the RIP-chip targets were functionally responsive to HuR depletion (Figure 15C), these data contradict a study producing post-lysis reassortment of HuR-mRNA interactions, which utilized quite different RIP conditions. This result demonstrates that the likelihood of post-lysis reassortment is negligible in our RIP-chip data. Interestingly, RIP-chip targets were more responsive than PAR-CLIP targets, reinforcing the notion that RIP-chip enriches for kinetically stable interactions and thus functionally responsive targets.
Figure 15: HuR targets destabilized in HuR depleted cells. A) A variety of concentrations and times of HuR siRNA knockdown were tested previous to transcriptomic analysis. B) Abundance of known HuR targets decrease 72 hrs after HuR depletion. C) Distribution of differential expression after HuR knockdown for RIP-chip and PAR-CLIP targets as compared to all expressed mRNAs.
4.2 mRNA stability

Indeed, HuR targets identified by both biochemical methods significantly enriched for positive regulation by HuR (Figure 15C). We examined if the degree of binding was correlated with the degree of HuR-dependent stabilization. We discovered a correlation between the cumulative number of binding sites per transcript and the degree of HuR-dependent stabilization (Figure 16A). Higher HuR LOD scores were also correlated with more HuR-dependent stabilization (Figure 16B). Therefore, both PAR-CLIP and RIP-chip provide quantitative binding data, which strongly correlates with the degree of regulation by HuR.
Figure 16: Relationship between degree of binding and functional outcome. A) Transcripts with more HuR binding sites exhibit a greater decrease in response to HuR knockdown. B) Transcripts were divided into 4 categories based on their percentile of decrease after knockdown, 75-100th representing mRNAs that decreased the most. Transcripts with higher HuR LOD scores correlated with a greater decrease in mRNA levels following HuR knockdown. C) Transcripts with binding sites only in introns, 3’ UTRs, and CDS were significantly destabilized by HuR knockdown. D) More intronic HuR binding sites exhibit greater destabilization after HuR knockdown. E) Transcripts with HuR binding sites in 3’ UTRs and introns are significantly more destabilized by HuR knockdown than transcripts with HuR binding sites only in introns or 3’ UTRs.
Since many RBPs and miRNAs exert their functional effects through interactions in the 3’ UTR of target mRNAs, it is a major platform for regulating mRNA stability. Therefore, we examined positional patterns between HuR binding sites and 3’ UTR landmarks that can impact the regulation of mRNA stability, such as the start, end, polyadenylation sites, and miRNA sites. We did not observe a significant enrichment for HuR binding sites across the 3’ UTR (Figure 17).

![Figure 17: Distribution of HuR Binding sites in 3' UTRs. Position of HuR binding sites within the 3’UTR were normalized by 3’ UTR length. (0 indicates close to beginning of 3’ UTR and 1 indicates close to end of 3’ UTR).](image)

In light of the substantial number of intronic binding events identified, we tested the extent to which these events contributed to the stability of the mRNA targets. It was necessary to restrict this analysis to transcripts containing binding sites exclusive to a certain transcript region to not conflate effects potentially mediated through different regions of the transcript. Surprisingly, transcripts containing only intronic binding sites were at least as responsive to HuR knockdown as transcripts containing only 3’ UTR...
binding sites (Figure 16C). This result was dependent upon the degree of binding because a transcript needed to contain more than one HuR binding site to exhibit significant stabilization (Figure 16D). Thus, the positive regulation of stability can be accomplished through pre-mRNA sites not limited to the mature mRNAs, and specifically the 3’ UTR. Furthermore, mRNAs containing both intronic and 3’ UTR binding sites were significantly more stabilized than mRNAs with binding sites only in the 3’ UTR or intron (Figure 16E). This supports the existence of a mechanism by which HuR couples pre-mRNA processing and mRNA stability.

4.3 mRNA splicing

While it was evident that 3’ UTR binding sites are a major mode of HuR-dependent regulation of mRNA stability, the purpose of the staggering amount of intronic binding sites remained ambiguous. In addition to the above-described contribution to mRNA stability, it was possible these intronic sites could reflect a non-canonical function of HuR in pre-mRNA splicing. Consequently, the significant overrepresentation of intronic HuR binding sites within the first 50 nts of the 3’ splice sites was revealing (Figure 18A). This strongly suggests that HuR is involved in splicing, which is consistent with a study demonstrating HuR can influence Fas exon definition [80]. The position of these intronic binding sites and underlying sequence was consistent with the Poly T/C motif (Fig 10B), suggesting that HuR interacts with poly-pyrimidine stretches between the 3’ splice site and the branchpoint.

We analyzed the HuR knockdown exon array data to identify HuR-dependent alternative processing events related to the observed intronic binding sites. We utilized AltAnalyze to identify splicing events in response to HuR knockdown and found both exon-inclusion and -exclusion events. We categorized exons based on if they had binding sites upstream or downstream and inspected the Firma score for the different
category of alternative exon usage events. We found that upstream binding site tended to correlate with exon exclusion (negative Firma score), whereas downstream binding sites were correlated with exon inclusion (positive Firma score) (Fig 18B). Importantly, a number of the alternative processing events, particularly those without intronic HuR binding sites, could be downstream effects since many functionally responsive HuR targets were RBPs and, specifically, splicing factors. Taken together, we conclude that HuR can influence many splicing decisions, either through direct interfacing with splicing machinery or through regulating the expression of splicing factors.
Figure 18: HuR binding sites and alternative splicing. A) Overrepresentation of intronic HuR binding sites in the first 25 nucleotides upstream of introns. B) Differential exon inclusion/exclusion preferences based on location of intronic HuR binding sites.

4.4 Discussion

There has been controversy regarding the potential for post-lysis reassortment when employing our RIP-chip protocol, therefore suggesting the absolute necessity of cross-linking when conducting such experiments. Past studies from many labs, including ours, have demonstrated specificity of the RIP-chip many ways including the lack of correlation between measured HuR association and mRNA abundance; the lack
of correlation between changes in HuR association and changes in mRNA abundance in response to stimuli; and even consistent results when employing 254nm cross-linking for individual targets. Here we demonstrate that RIP-chip targets are even more functionally responsive than PAR-CLIP targets and so, not surprisingly, the likelihood of post-lysis reassortment is extremely low. Obviously, these results do not excuse investigators from determining appropriate RIP conditions for their particular application, especially since the only study demonstrating post-lysis reassortment with HuR utilized very different RIP conditions, for example sonication of lysates. However, we reject the notion that cross-linking is a de facto assay necessary to claim an mRNA to be a “bona fide” target.

Our data support the findings that HuR is a positive regulator of mRNA stability. We observe that the number of HuR binding sites per mRNA is proportional to the degree of HuR-dependent stabilization. Fascinatingly, we observed this trend to be true even for transcripts with only intronic HuR binding sites. One speculative possibility would be that HuR influences deposition of the EJC and thereby downstream mRNA quality control mechanisms linked to mRNA stability. While individual cases exist where one site is enough to confer regulation at the level of mRNA stability, these are the exception. It is yet to be determined if this observation holds true for HuR-dependent translational activation.

We discovered a significant role for HuR in pre-mRNA processing, specifically mRNA splicing. There is one study demonstrating that HuR can directly regulate the inclusion of exon 6 of Fas ligand, but this is the only example. Our study suggests that splicing is an important aspect of the function of HuR. This is consistent with the large amount of nuclear localized HuR, which is predominantly nuclear at basal states. By
sequence homology, HuR is very closely related to splicing factors, including the fly Sxl, which also has a similar structure.

Particularly interesting is the possibility that HuR may be involved in coupling splicing and stability, since ~3000 (~40%) of the mRNA targets contained both intronic and 3’ UTR binding sites. The positive effect on stability of specific isoforms promoted by HuR would be advantageous to that particular isoform relative to other possible transcript isoforms. It will be important to explore the precise mechanisms of this potential coupling.

Figure 19: Proposed mechanisms by which HuR regulates mRNA splicing and stability.

We were able to detect many HuR-dependent alternative mRNA processing events utilizing HuR knockdown combined with Affymetrix exon array. While there
were many alternatively regulated exons, we focused on ones with adjacent HuR binding sites since they would more likely represent primary effects of HuR. We found that binding sites upstream of exons were more likely to result in exon exclusion, while binding sites downstream of exons were more likely to result in exon inclusion. CLIP-seq studies of the splicing factor Fox2 combined with RT-PCR exhibit a similar pattern of spatial-dependency linked to exon choice. However, the precise mechanism of how HuR interfaces with the splicing machinery is not known and is extremely important. Additionally, HuR stabilizes the mRNA encoding many RBPs and specifically splicing factors. We hypothesize this regulation of a regulator is responsible for the alternative exon usage, not associated with HuR binding sites. Altogether, we conclude that splicing represents a significant aspect of the function of HuR.

4.5 Methods

HuR knockdown. Flag-HA HuR Flip-In T-Rex HEK 293 cells were grown as described in chapter 3.5. Transfections were performed using RNAiMAX (Invitrogen) using the manufacture protocol for a 6-well plate. For collection, cells were washed with PBS and then lysed directly on the plate using TRIzol for RNA or laemmli buffer for protein. Biological triplicates of RNA samples from samples treated with 50 nM scramble siRNA or HuR siRNA for 72 hrs were submitted for analysis on Affymetrix exon arrays.

Microarray data processing. Microarray data was normalized with PLIER using Affymetrix Power Tools. Probes with DABG p-values > 0.05 in all three of three HuR RIP replicates were excluded.

Microarray Analysis. Normalized gene level data was compared using t-scores and log-fold change in GSEA. Splicing analysis was performed using AltAnalyze and specifically, FIRMA for calculation of alternative exon usage.
Comparisons of distributions. As described in chapter 3, JMP 7.0 (SAS) was utilized for all comparisons of distribution and densities for various categories and metrics. These include cumulative distribution fraction plots and comparison of densities.

Distance between binding sites and transcript landmarks. Using custom python scripts, the distance of each cluster from the 5’ and 3’ of each genetic region was compared to a null in which the clusters are randomly distributed in their genetic region. In the case of the intronic binding sites the median density of the null permutations was then subtracted from the observed density. Otherwise, observed density estimates and density estimates of the 1st, 10th, 50th, 90th, and 99th percentile of the null were displayed.
5. Coordinated RNP dynamics during T-cell activation

While messenger RNP complexes are highly dynamic cellular environments [109], very few studies have focused on global RNA dynamics of RNPs across different physiological conditions [92, 110, 111]. Even though ribonomic profiling has been widely used to identify mRNAs associated with a given RBP [89, 90], the overwhelming majority of these studies utilized RIP-Chip experiments from a single condition of growth or perturbation. This is in part due to the lack of analytical approaches for modeling RIP-chip data to determine targets and assign values of condition-specific RNP-association that allow systematic comparisons across physiological conditions. Therefore, the development of probabilistic models of RNP-association will allow a more thorough and systematic investigation of the contribution of RNP dynamics to the molecular networks activated during development or in response to perturbations.

Global studies of T cell activation, a model for the engagement of the T cell receptor (TCR) complex by presented antigens, have demonstrated extensive posttranscriptional regulation, specifically alteration of mRNA stability [112] and alternative splicing [113]. One study found that more than half of the transcriptomic changes that occur during T cell activation are regulated at the level of mRNA stability and not accompanied by any transcriptional change [114]. Furthermore, HuR [115] and ARE containing mRNAs [116], many of which are critical immune regulators, respond dynamically during T cell activation. For example, engagement of LFA-1, a β2 integrin that is important for TCR complex signaling, results in HuR nuclear export and stabilization of cytokine mRNAs [117].
5.1 Global Identification of HuR target mRNAs During T cell Activation

We used our established RIP-Chip protocol to identify the mRNAs associated with RNP complexes containing the RNA-binding proteins HuR and PABP at 0, 4, and 12 hours post-activation of Jurkat cells with phorbol 12-myristate 13-acetate (PMA) and phytohemagglutinin-A (PHA). RNP immunoprecipitation assays (RIPs) for mRNAs associated with HuR and PABP (a quality control for the RIP-chip procedure) utilized previously described methods and antibodies [92, 118, 119]. We conducted RIPs with PABP, a housekeeping RBP, as another indication for the biochemical specificity of the HuR IP, the primary focus of the study. Briefly, antibody coated Protein A-Sepharose beads were incubated with cell lysates, thoroughly washed, and RNA was extracted from the pellets. For each time point, three biological replicates each of HuR, PABP, negative (IgG) RIP-Chip pellets, and total cellular RNA (totals) samples were analyzed using oligo microarrays that interrogated 35K genes. To qualify for subsequent analysis and to be treated as “expressed”, a probe had to be twofold above local background for all replicates in any of the IPs or the totals at any time point. For all probes expressed (n = 14,789), t-scores for HuR-IP versus negative-IP were calculated using Gene Set Enrichment Analysis (GSEA) at each time point. Visual inspection of the t-score distributions indicated two populations of mRNAs (Figure 20A), an enriched population representing HuR-associated mRNAs and a non-enriched population representing background mRNAs.

We employed GMM and calculated LOD scores as described in chapter 2. The model with the highest log-likelihood demonstrated excellent fit to the data (Figure 20A, red curves) and discriminated the HuR-associated population (Figure 20A, blue curves) from background populations (Figure 20A, green curves) at all time points. The HuR LOD scores provide a continuous variable representing the condition-specific
probability of HuR RNP association for each probe, allowing us to determine and to compare changes in the likelihood of HuR RNP association throughout the activation.

Figure 20: HuR targets during T-cell activation. (A) Distribution of HuR IP versus negative IP t-scores at 0, 4, and 12 hours post-activation. HuR-associated (blue curve), background (green curve) and sum of HuR-associated and background (red curve) probability distributions are shown, as defined by Gaussian mixture modeling. (B) Venn diagram representing all probes (n=1219) that had a LOD HuR > 0 at any time point. (C) Upper triangular matrix of Spearman correlation coefficients for all pair-wise comparisons of LOD HuR and Totals S2N values for 0, 4, and 12 hours post-activation.

Probes with HuR LOD scores greater than zero, thus having a higher likelihood of being within the HuR RNP-associated population in comparison with the background population, were considered to be a discrete population of mRNAs associated with HuR. Downstream analysis of the enrichment of a previously reported and independently derived HuR binding element (see HuR COVE motif below) suggested that the LOD > 0 cut-off was apt. Moreover, this threshold was substantiated in the ribonomic analysis of human Pum1 [120]. Of the 14,789 probes expressed in the Jurkat cells, the number representing HuR targets increased from 599 (4.05%) to 800 (5.41%) to 61
924 (6.25%) probes at 0, 4, and 12 hrs post-activation, respectively (Figure 2B).

Altogether, 1219 probes (the actual number of genes is lower since multiple probes map to the same gene) were determined to be HuR targets for at least one of the time points. Of these ~1/3 (405) were HuR targets at all time points. This demonstrates the plasticity of HuR-association during T cell activation.

Complementarily to the comparisons of discrete values (target or not a target) above, we examined the quantitative differences in the mRNA content of HuR RNPs using continuous values representing condition-specific HuR RNP association for each probe (HuR LOD). As expected, for a given condition, there was very high correlation among the 3 independent biological replicates (average r = 0.89, 0.92 and 0.93 for comparisons within 0hr, 4hr and 12hr, respectively), demonstrating the reproducibility of the RIP-Chip method. Moreover, there were marked differences in the 0hr, 4hr and 12 hr HuR LOD scores as evidenced by the relatively low correlations between them (r = 0.24 – 0.42), demonstrating HuR-associated mRNA population dynamics (Figure 2C, black numbers). In comparison, the total mRNA levels were substantially more similar during the activation (r = 0.82 – 0.86) (Figure 2C, white numbers), highlighting the dramatic remodeling of HuR RNPs during activation, as compared with the transcriptome. A more detailed analysis of HuR RNP dynamics is presented below.

We examined the relationships between HuR-association, PABP-association, and total mRNA expression level for probes considered HuR targets at any time point. The steady-state PABP-association resembled the transcriptome more than HuR-association (Figure 2C, blue numbers) at each time point. Notably, there was little to no correlation at 0, 4 or 12 hr between the probability of HuR association and mRNA abundance (Figure 2C, red numbers), indicating that RIP-Chip successfully isolated specific
subsets of mRNAs that were not quantitatively representative of total mRNA, consistent with our previous studies [92, 120].

5.2 Functional Characteristics of HuR targets During Activation

To identify the salient characteristics of the mRNA components of HuR RNPs at 0, 4, and 12 hrs post-activation, we analyzed the following: 1) common sequence characteristics and motifs, and 2) functional relationships among the proteins encoded by the HuR associated mRNAs.

5.2.1 Sequence Characteristics and Motifs

Since HuR typically bind elements in the 3’ UTR of transcripts, we searched for unique characteristics common among the 3’UTR of HuR targets. Indeed, these transcripts have exceedingly long (1.54 kb, p<0.0001) and AU rich (62.6%, p<0.0001) 3’ UTRs compared to randomly selected sets of UTRs (1.00kb, 57.3%) (Figure 21). Also, though HuR-associated mRNAs were enriched for the presence of computationally identified subclasses of AREs [121], there was no difference in propensity toward any subclass. Given that our results show that 3’ UTR length and AU content are good predictors of HuR-association, and that HuR has no preference for any ARE subclass, combining global ARE RBP-target interaction data combined with computational approaches that use sequence and structural features may improve classification of AREs.
Figure 21: AU-content and length of HuR-associated mRNAs. Null distribution of average length and AU-content of 3’ UTRs of randomly chosen messages from total expressed population (10,000 random sets) each with same # of UTRs as the HuR-RNP associated set. The black arrows mark the average length and AU content for HuR-RNP associated mRNAs, indicating a significant (p<0.0001) bias in the length and AU-content of 3’ UTRs of HuR-RNP associated mRNAs.

<table>
<thead>
<tr>
<th></th>
<th>Avg HuR-RNP UTRs</th>
<th>Avg Random Set UTRs</th>
<th>Avg DINuc Shuffle HuR-RNP UTRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction with ≥ 1 Hit</td>
<td>0.46</td>
<td>0.26</td>
<td>0.45</td>
</tr>
<tr>
<td># of Hits</td>
<td>0.83</td>
<td>0.43</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Figure 22: Enrichment of HuR COVE model in HuR-associated mRNAs. The same random sampling procedure used in SF1 was conducted on various HuR COVE model statistics. Null distributions represent averages for all listed statistics for random sets (light grey) and di-nucleotide shuffled 3’ UTR of HuR-associated mRNAs (dark grey). The average values for 3’ UTR of HuR-associated mRNAs indicated by black arrow.

As noted above, a previous study discovered a potential structural RNA motif for HuR binding by using covariance modeling (COVE) on 3’ UTR sequences of mRNAs.
identified from a HuR RIP-Chip experiment in a different cell line and condition [94].

We observed significant enrichment for all COVE-based metrics in the HuR associated mRNAs at all time points (Figure 22). Surprisingly, the enrichment of COVE-based metrics for 3’ UTR sequences of HuR associated mRNAs that had been randomized while preserving di-nucleotide frequencies were not different from the enrichment for actual 3’ UTR sequences of HuR associated mRNAs (Figure 22). This indicates that the HuR COVE motif does not identify a unique RNA structure, but rather strongly preferred sequence characteristics of 3’ UTRs that are capable of HuR binding.

Figure 23: HuR RNP 3’ UTRs are Enriched for HuR COVE Motif. GSEA analysis of HuR LOD scores using HuR COVE motif gene set. Normalized Enrichment Scores (NES) and the FamilyWise-Error Rate (FWER) p-value are shown. LOD scores > 0 are to the left of the vertical red line in the plots.

Next, we used GSEA to determine how mRNAs that contain at least one HuR COVE motif in their 3’ UTR are distributed in the 0hr, 4hr and 12hr lists ranked by HuR LOD scores. Enrichment profiles demonstrated the utility of the COVE model for identifying elements in common to HuR RNP mRNAs (Figure 23), consistent with the results above. Moreover, the running enrichment score of the COVE motif peaked after LOD = 0 for all time points (Figure 23), supporting the validity of the LOD > 0 cut-off and indicating that it may be somewhat conservative.
5.2.2 Common Functional Groups Enriched in HuR mRNA Targets

Since our experimental model is T cell activation, we examined the HuR associated mRNAs for encoded proteins with functions known to be critical in T cell receptor (TCR) engagement and local signaling, which involves adapter molecules, signal transduction, and cytoskeletal remodeling at the immunological synapse. HuR associated with 26 mRNAs critical to each of the aspects listed above (Table 3). Moreover, all 26 mRNA had at least one HuR COVE motif in its 3' UTR. These data predict that HuR may help coordinate dynamic events directly downstream of TCR engagement following T cell activation.

Table 3: HuR-Associated mRNAs Encode Proteins Critical to the Immunological Synapse. Different colors represent the following classes: Orange = TCR Receptor and Membrane, Yellow = Adapters, Green = Regulators of Cytoskeletal Dynamics, and Blue = Regulators of Signaling

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD3G</td>
<td>T cell surface glycoprotein CD3 gamma chain</td>
<td>Member of TCR complex; deficiency in humans can lead to SCID [122]</td>
</tr>
<tr>
<td>TRG V3</td>
<td>T cell receptor gamma variable 3</td>
<td>Member of TCR complex; variable gamma chain [123]</td>
</tr>
<tr>
<td>BCL10</td>
<td>B-cell lymphoma leukemia 10</td>
<td>Bc10-/- lymphocytes are defective in antigen receptor or PMA/ionomycin-induced activation [124]</td>
</tr>
<tr>
<td>CD2AP</td>
<td>CD2-associated protein</td>
<td>Required for formation of c-SMAC [125]</td>
</tr>
<tr>
<td>ITGB1</td>
<td>Integrin beta 1</td>
<td>TCR engagement results in an increase in T cell adhesion mediated by β1 integrins [126]</td>
</tr>
<tr>
<td>FYB</td>
<td>Adhesion and degranulation promoting adaptor protein</td>
<td>Important for coupling TCR-mediated actin cytoskeletal rearrangement with activation of integrin function [127]</td>
</tr>
<tr>
<td>SLP-76</td>
<td>Lymphocyte cytosolic protein 2</td>
<td>Adapter protein required for TCR mediated signaling [128]</td>
</tr>
<tr>
<td>VAV2</td>
<td>Vav 2 oncogene</td>
<td>Family of Rho–guanine nucleotide exchange factors important for proper TCR mediated signaling [129]</td>
</tr>
<tr>
<td>VAV3</td>
<td>Vav 3 oncogene</td>
<td></td>
</tr>
<tr>
<td>WASF1</td>
<td>Wiskott-Aldrich syndrome protein family member 1</td>
<td>Member of Wave complex which is a regulator of TCR-mediated actin dynamics [130]</td>
</tr>
<tr>
<td>ACTB</td>
<td>Beta actin</td>
<td>Drives formation of SMAC at immunological synapse [131]</td>
</tr>
<tr>
<td>ARP2</td>
<td>Actin-related protein 2</td>
<td></td>
</tr>
<tr>
<td>ARP C4</td>
<td>Actin related protein 2/3 complex, subunit 4</td>
<td>Member of Arp2/3 complex which is required for the reorganization of the actin cytoskeleton at immunological synapse in T cells [132, 133]</td>
</tr>
<tr>
<td>ARP C5</td>
<td>Actin-related protein 2/3 complex subunit 5</td>
<td></td>
</tr>
<tr>
<td>ARF 6</td>
<td>ADP-ribosylation factor 6</td>
<td>Forced expression in Jurkat T cells prevented TCR-mediated reorganization of actin [134]</td>
</tr>
<tr>
<td>CAPZA1</td>
<td>F-actin-capping protein subunit alpha-1</td>
<td>Mediates interaction between c-SMAC and actin filaments [135]</td>
</tr>
<tr>
<td>CAPZA2</td>
<td>F-actin-capping protein subunit alpha-2</td>
<td></td>
</tr>
<tr>
<td>CDC42</td>
<td>Cell division cycle 42</td>
<td>Small Rho GTPase important for reorganization of cytoskeleton upon TCR complex engagement [136]</td>
</tr>
</tbody>
</table>
To determine if the proteins encoded by HuR associated mRNAs were functionally related we used Panther, InnateDB, and GSEA, which explore known relationships among a list of genes. InnateDB and Panther analysis revealed significantly enriched pathways (Figure 24) vital to cellular function, for many of which regulation by HuR has been demonstrated for individual members. These pathways include “Wnt signaling” [144, 145], “metabotropic glutamate receptor group 1” [146] and “p53 feedback loop 2” [147]. GSEA on HuR LOD scores also identified pathways and perturbations in which HuR has known roles, such as aging [148], induced UVC stress [149] and HCMV infection [150]. Biological processes that HuR has been previously shown to be involved in, such as the cell cycle [151], and cell proliferation and differentiation [115], were also significantly enriched (Figure 24). “Hedgehog signaling” and the “circadian clock” represent novel pathways that may involve regulation by HuR. HuR-associated messages were also enriched for “transcription”, “other transcription factors”, “mRNA processing”, and “other RNA-binding proteins” (Figure 24); a defining characteristic of these categories is that they represent proteins that have important regulatory consequences for gene expression [89, 152].
Figure 24: Coordination of HuR RNP Associated Functional Modules During T Cell Activation. Analysis of HuR-associated mRNAs (LOD > 0) at each time point was conducted using Panther and InnateDB. Relative enrichment of functionally related groups (must have a Bonferroni corrected p-value < 0.05 for at least one time point) are represented as profiles and heatmap. ** Functional groups that represent regulation of DNA and RNA.

We examined the hypothesis that HuR functions as a regulator of regulators in Jurkat cells, specifically of the group of adaptive mRNA subset-specific regulatory RBPs [89, 119, 153, 154]. We used a convenient catalog of RBPs that were compiled by Silver and colleagues identifying RBPs based on the presence of known RNA-binding domains, primarily RRM (RNA recognition motif) and KH (hnRNP K homology) domains [155]. GSEA of HuR LOD scores demonstrated that mRNAs encoding these RBPs were significantly enriched in HuR RNPs at all time points (shown in table X chapter 6.1.1). Further, more than one half of the mRNAs encoding RBPs that associate with HuR were unique to at least one time point. The potential to regulate and
coordinate mRNAs of regulatory RBPs indicates that HuR has a substantial role in the homeostasis and modulation of posttranscriptional regulatory networks [152].

5.3 RNP Dynamics During Activation

Since the RNP LOD scores are condition-specific, we calculated the difference between 0hr and 4hr and 4hr and 12 hr for both HuR and PABP LOD scores to generate values representing changes in RNP association of these mRNAs. To assess changes in total mRNA, we generated t-scores comparing 0hr to 4hr and 4hr to 12hr time points. We excluded probes that had a low probability of being associated with HuR (LOD < 0 at all time points), as they would confound interpretation of changes across conditions.

5.3.1 Functional Characteristics of Ribonomic and Transcriptomic Dynamics During Activation

We examined the similarity between HuR RNP, PABP RNP, and transcriptomic dynamics for 0hr to 4hr (early) or 4hr to 12hr (late) activation intervals. As expected, these values did not show strong correlations for early or late changes (r = -0.12 – 0.20) (Figure 25, black circles), demonstrating that each sample had unique overall gene expression dynamics. Importantly, the changes in HuR association were unique and were not explained by changes in mRNA abundance. Combined with the earlier observation of cell-type and condition specific differences in HuR association, we have no evidence that potential adventitious re-association [156] was a concern in these data. As noted previously [118], published criticisms [157] subsequently directed at this RIP-chip procedure were inappropriately applied by exploiting results from a different method that showed re-association of a HuR target from one cell to another in an incubated extract [156]. Indeed, RIP-chip has been used many laboratories without cross-linking, reviewed in [158]. Within each data type, particularly PABP, there were reciprocal patterns of mRNA dynamics between early and late intervals (Figure 25,
white circles). This may suggest a recovery from the activation and that increased temporal resolution would be informative. While intriguing, the dynamics and the contrast of PABP-association profiles with the mRNA abundance profiles are outside the scope of the current study.

![Figure 25: Gene Expression Dynamics Demonstrate Unique Changes in mRNA Subpopulations. Upper triangular matrix of Spearman correlation coefficients for all pair-wise comparisons of ΔLOD HuR, ΔLOD PABP and Totals t-scores for 0hr to 4hr and 4hr to 12hr dynamics.](image)

Next, we identified common functional groups exhibiting HuR association profiles that are either very similar to or very different from the transcriptomic profiles using rank-ordered pair-wise correlation profiles from all time points. We found that “translation factors”, was significantly enriched for negative correlation (Table 4). Therefore, mRNAs encoding translation factors had HuR association profiles that were the opposite of their transcriptomic profiles across the T cell activation interval.
Table 4: GSEA Results for Dynamic and Correlation Comparisons. GSEA analysis was conducted on ΔLOD HuR for both 0hr to 4hr and 4hr to 12hr. Additionally, GSEA was conducted incorporating data from all time points using correlations between HuR-association profiles and PABP association profiles, as well as correlations between HuR-association profiles and transcriptomic profiles.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>NAME</th>
<th>NES</th>
<th>NOM p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>HuR:Transcriptome Correlations</td>
<td>TRANSLATION_FACTORS</td>
<td>-2.3316872</td>
<td>0</td>
</tr>
<tr>
<td>HuR:PABP Correlations</td>
<td>STRESS_TPA_SPECIFIC_UP</td>
<td>2.4617658</td>
<td>0</td>
</tr>
<tr>
<td>HuR:PABP Correlations</td>
<td>HINATA_NFKB_UP</td>
<td>2.29813</td>
<td>0.00204918</td>
</tr>
<tr>
<td>Early HuR Dynamics</td>
<td>STRESS_TPA_SPECIFIC_UP</td>
<td>-2.6012323</td>
<td>0</td>
</tr>
<tr>
<td>Late HuR Dynamics</td>
<td>STEMCELL_NEURAL_UP</td>
<td>2.4550486</td>
<td>0.002</td>
</tr>
<tr>
<td>Late HuR Dynamics</td>
<td>V$LEF1_Q2</td>
<td>-2.4301183</td>
<td>0</td>
</tr>
</tbody>
</table>

We made the analogous comparison between HuR association profiles and PABP association profiles. We observed two gene sets that were significantly enriched (Table 4) and positively correlated, which was interesting since both of these RBPs have been shown to be positive regulators of translation. One gene set represented transcripts that were up-regulated by the NF-kappa B transcription factor, a molecule critical to T cell activation. The other gene set represented transcripts that were up-regulated at 4 hours after PMA treatment and that discriminate PMA from other stress agents, as would be expected. This gene set was also enriched when examining early HuR RNP dynamics.

5.3.2 Functional Characteristics of HuR RNP State Dynamics During Activation

Next, we identified functionally related mRNAs that exhibited common HuR RNP dynamics. GSEA analysis of HuR RNP dynamics was carried out for early changes (0hr to 4hr) and late changes (4hr to 12hr). For the early dynamics, the only significantly enriched gene set was the PMA induced gene set, exhibiting increased HuR association from 0hr to 4hr (Table 4). For late dynamics, two gene sets were significantly enriched (Table 4): 1) genes enriched in mouse neural stem cells compared to differentiated brain and bone marrow cells, which decreased in HuR association from 4hr to 12 hr, and 2)
genes with *LEF1* promoter elements which demonstrated increased HuR association from 4hr to 12hr. Interestingly, HuR LOD scores for *LEF1* went from -0.29 at 0hr, to 1.37 at 4hr, to 1.13 at 12hr and the HuR promoter contains a predicted *LEF1* binding site. The latter result raised the fascinating possibility of a translationally and transcriptionally coupled regulatory loop between HuR and *LEF1*.

### 5.3 Discussion

Our results demonstrate dramatic remodeling of the HuR RNP and PABP RNP during T cell activation, especially in comparison to transcriptomic expression dynamics (Figure 20B,C). Similar RNP remodeling was observed previously with HuB, another member of the ELAV/Hu family, following retinoic acid induction of neuronal differentiation using P19 embryonic carcinoma cells, however, continuous metrics of mRNA-association were not derived [92]. More importantly, coordinate changes of mRNAs in induced P19 cells encoded functionally related proteins known to be involved in neuronal differentiation leading, in part, to the posttranscriptional RNA operon (PTRO) model [159]. In this study we uncovered temporally coordinated changes in populations of HuR-associated mRNAs whose encoded proteins are functionally related and necessary for T cell activation (Table 3) and Wnt signaling (Figure 24), consistent with the PTRO model.

Our data predict that HuR has a role in coordinating posttranscriptional events imminent to T cell receptor (TCR) signaling (Table 3). This Jurkat cell activation model uses PMA and PHA to mimic the consequences of antigen presentation, which results in the engagement of the T cell receptor (TCR) complex. Signaling elicited by TCR complex engagement results in the formation of supramolecular activation clusters (SMACs) at the immunological synapse. TCR complex engagement stimulates extensive remodeling of the cytoskeleton and plasma membrane to facilitate the formation of the SMACs.
Proteins involved in this process can be roughly classified into four categories, TCR receptor and membrane, adapters, regulators of cytoskeletal dynamics, and regulators of signaling. In our experiments, HuR was associated with 26 mRNAs that encode multiple proteins from each of the four categories of SMAC formation that are critical for this spatially and temporally dynamic process that drives adaptive immunity (Table 3). Significantly, all 26 of the mRNAs have at least one HuR COVE motif in their 3' UTR. Our prediction was validated by the finding of TCR signaling defects obstructed activation driven positive selection in a thymus-specific knockout of HuR in mouse [71]. Further corroborating the importance of HuR in T cell activation, chemical inhibition of HuR-mRNA interaction has been shown to inhibit nucleo-cytoplasmic redistribution of HuR and to block T cell activation [160].

The Wnt signaling pathway is a key regulator of T cell development in the thymus[161], however its role in formation and functioning of the immunological synapse during T cell activation has not been well studied. Our data indicate interdependence between HuR and the Wnt pathway during T cell activation. Previous studies demonstrated regulation of members of the Wnt pathway by HuR, for example B–catenin mRNA is a known HuR target and is regulated at the level of stability in an ARE-dependent and splicing-dependent manner [144, 162]. Wnt signaling is also known to induce stabilization of downstream target mRNAs via the Pitx2 transcription factor by a mechanism requiring HuR recruitment [145]. Similarly, we found that targets of LEF1, another transcription factor downstream of Wnt signaling, increase in HuR association from 4hr to 12hr (Table 4). The enrichment of mRNAs encoding proteins of the Wnt pathway (Figure 24) is consistent with previous reports leading to the elucidation of functionally related mRNAs associated with an RNP that is inherent in the PTRO model and warrants further investigation in cellular and animal models.
5.4 Methods

Cell Culture. Jurkat cells were cultured in RPMI 1640 supplemented with 10% FBS (GIBCO). For activation, cells were treated with 50 ng/ml PMA and 2 µg/ml PHA (Calbiochem). Cells were pretreated with 50 µM resveratrol for 2 hours and then subject treated with PMA/PHA.

Immunoprecipitation assays and RNA isolation. Lysates were prepared from samples collected at 0, 4, and 12 hours post-activation as described, with the addition of 10% glycerol to the polysome lysis buffer (PLB) and resuspension of harvested cells in PLB. RIP of endogenous HuR and PABP RNP complexes were used to assess association of endogenous target mRNAs. Assays were performed as described [92, 119]. RIPs utilized 100 µL pre-swollen and packed Protein-A Sepharose beads (Sigma) loaded with 30 µg of anti-HuR (3A2), anti-PABPC1 serum and anti-PABPC4 serum (sera generated in Penalva et al. 2004), and mouse IgG1. Antibody loaded beads were incubated with 3 mg cell lysate for four hours at 4ºC, washed 4 times with ice-cold NT2 buffer (50mM Tris pH 7.4/150mM NaCl/1 mM MgCl2/0.05% Nonidet P-40) followed by 3 washes with ice-cold NT2 supplemented with 1M Urea. Extraction of associated RNA was performed as described, and total RNA was isolated using the Trizol (Invitrogen).

Microarray Analysis. Arrays were printed at the Duke Array Facility using the Genomics Solutions OmniGrid300 Arrayer and contained Human Operon v3.0.2 oligo set (Oligo Source) consisting of ~35k unique 70-mers. RNA quality was checked using an Agilent2100 bioanalyzer (Agilent technologies) for total RNA samples only. For all arrays RNA was assayed using direct labeling of experimental samples (Cy 3) and Stratagene Universal Human Reference RNA (Cy 5). Array data were submitted to the GEO, GSE11989. All arrays were subject to loess normalization within each array and scale normalization across arrays using the Array Magic [163]. Replicate probes were
collapsed to the median value. To be considered for subsequent analysis, probes had to be 2x > local background in all biological replicates for any of the RIPs or the totals at any time point.

**GMM and LOD scores.** These were conducted as described in chapter 2.5.

**Ribonomic-Transcriptomic Comparisons.** For values representing mRNA abundance we calculated a signal-to-noise ($S_2N$) ratio (to account for variance across replicates) for the three biological replicates per time point. The Spearman correlation coefficient between HuR-association, PABP-association and mRNA abundance profiles across all time points were calculated per probe. GSEA was used to calculate t-scores per probe representing differential expression between 0hr and 4hr and 4hr and 12hr. Upper triangular matrix color maps were made using JMP 7.0 (SAS).

**Sequence Characteristics.** We used a local pipeline to retrieve high quality 3’ UTR sequence for all transcripts expressed [164]. The AU-content and length of each UTR was calculated. We mapped the ARED 3.0 database to refseqs to determine what transcripts contained either class I or class II AREs. COVE-LS was used to search sequences using the HuR COVE model and the following statistics were calculated: at least one match, number of matches, maximum score, sum of all scores, and the average of scores. Significance of the enrichment of each HuR COVE model statistic was tested using random sampling. Null distributions were created for each characteristic listed above by calculating the average of randomly chosen sets from total expressed population (10,000 random sets, with the same # of UTRs as HuR-associated set) and compared to the average value for the HuR-associated mRNAs to determine statistical significance. Null distribution for assessing the contribution of secondary structure to HuR COVE model statistics were created by calculating the average of randomly generated di-nucleotide shuffled sequence from 3’ UTRs associated with HuR (1,000 sets) and compared to the
average value calculated above for the actual 3’ UTR sequence of HuR-associated mRNAs to determine statistical significance.

**Functional Enrichment.** GSEA [165], Panther [166] and InnateDB [167] were used for enrichments. A gene set had an false discovery rate (FDR) q-value < 0.05 or family-wise error rate (FWER) < 0.1 to be considered significant for all GSEA analysis. For non-Gaussian data the classic enrichment statistic was used. For Panther and InnateDB analysis gene sets were required to have a Bonferroni corrected p-value < 0.05.
6. Mechanisms driving HuR-mRNA interaction dynamics

We hypothesize that combinatorial interactions from other RBPs or non-coding RNA, changes in subcellular localization, and PTL modifications could all be mechanisms driving HuR-mRNA interactions. These possibilities were explored using cellular fractionation, RIP-chip, PAR-CLIP, and RBP targets from other datasets.

6.1 Combinatorial interactions

We next explored mutual relationships between HuR RNPs and the ribonome, specifically focusing on regulatory RBPs and microRNAs. We asked the following questions: 1) Is there a bias for mRNAs of RBPs among the population of mRNAs associated with HuR RNPs, implicating HuR as a regulator of PTRFs? 2) Is the population of mRNAs associated with HuR enriched for known targets of ARE-RBPs and predicted targets of microRNAs, indicating that these mRNAs may be subject to combinatorial regulation?

6.1.1 Other ARE-RBPs

The potential influence of other ARE-RBPs on HuR-associated mRNAs was assessed by creating gene sets for targets of TTP in activated mouse macrophage cells (RAW264.7) [91], and TIAR [168] and HuR [94] in human colon carcinoma cells (RKO). Since HuR, TTP, and TIAR are all ARE-RBPs, and therefore putatively utilize similar cis-regulatory elements, we expected significant enrichment of lists ranked by HuR LOD scores. This was the case for RKO HuR RNP mRNAs (Table 5), which were significantly enriched (NES = 2.03, FDR = 0.0075) in the 12 hr HuR RNPs and demonstrated strong, but not statistically significant enrichment (NES = 1.62 and 1.54, FDR = 0.0550 and 0.0809) in the 0hr and 4hr HuR RNPs (Table 5). As anticipated, TTP targets were significantly enriched (NES = 2.53 – 2.89, FDR = 0 – 0.0006) in the Jurkat HuR RNPs,
suggesting that TTP RNPs and HuR RNPs are likely to contain many of the same mRNA species, and that TTP and HuR may in some cases co-occur on the same individual transcript. In contrast, the significant depletion (NES = -1.54 – -2.29, FDR = 0.0015 – 0.0657) of TIAR targets (Table 5) was consistent with the possibility of a C-rich motif identified in TIAR target mRNAs, rather than AREs as initially described [169].

**Table 5** HuR RNP Enriched for Predicted Targets of microRNAs. GSEA analysis was conducted on LOD HuR scores for each time point using the following three classes of gene sets for pathways, transcription factor targets (TF), and predicted microRNA (miRNA) target. Positively enriched gene sets were rank ordered and plotted by FWER p-values. Only results for 4hr are shown; 0hr and 12hr results are similar.

<table>
<thead>
<tr>
<th>Gene Sets Analyzed</th>
<th>LOD HuR 0hr</th>
<th>LOD HuR 4hr</th>
<th>LOD HuR 12hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENRICHED</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulatory RBPs</td>
<td>NES 3.03</td>
<td>NES 3.11</td>
<td>NES 2.83</td>
</tr>
<tr>
<td>TTP RNP (Activated RAW)</td>
<td>NES 2.53</td>
<td>NES 2.64</td>
<td>NES 2.89</td>
</tr>
<tr>
<td>HuR RNP (RKO)</td>
<td>NES 1.62</td>
<td>NES 1.54</td>
<td>NES 2.03</td>
</tr>
<tr>
<td>DEPLETED</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIAR RNP (RKO)</td>
<td>NES -2.29</td>
<td>NES -1.54</td>
<td>NES -2.00</td>
</tr>
</tbody>
</table>

A potential caveat to this comparative analysis was that each experiment was performed under different conditions and in different cell types, thus these observations could be explained by variations in condition-specific mRNA-RBP association. However, the consistency and strength of the enrichment for RKO HuR RNP mRNAs indicated that condition-specific association, although evident, did not confound interpretation of these RNP enrichment results. This indicates that cell-type and condition-specific differences in association do not explain the significant depletion of the TIAR associated mRNAs, since the TIAR RIP-Chip experiments were also performed using RKO cells. Thus, the dichotomy between global TIAR mRNA targets and the more similar HuR and TTP mRNA targets provides insight into the organization of posttranscriptional regulatory networks.
6.1.2 microRNAs

In addition to the analysis of RBP targets, we used gene sets corresponding to predicted targets of microRNAs (from MSigDb, http://www.broad.mit.edu/gsea/msigdb/) to obtain a comprehensive evaluation of the potential targeting of microRNAs to mRNAs associated with HuR RNPs. At all three time points, over 90 different microRNA target gene sets (Figure 26) were significantly enriched in HuR RNPs. The enrichment of microRNA target gene sets is especially striking when compared to the number enriched gene sets representing transcription factor targets or curated gene sets (representing signaling pathways or experimental perturbations). Importantly, the biased enrichment of microRNA target gene sets compared to other classes of gene sets in the HuR IP was not recapitulated in either the PABP IP or the transcriptome (data not shown). Therefore, these analyses demonstrate that both microRNAs and regulatory RBPs have a high potential for combinatorial regulation of HuR RNP mRNAs, suggesting that the population of HuR-associated mRNAs represent a class of highly regulated transcripts. Interestingly, miR-181 targets are one of the most enriched microRNA target gene sets for HuR association and the most depleted for PABP association. Similar to HuR, miR-181a has been shown to modulate TCR signaling and T cell selection [170].
Figure 26: HuR RNP Enriched for Predicted Targets of microRNAs. GSEA analysis was conducted on LOD HuR scores for each time point using the following three classes of gene sets for pathways, transcription factor targets (TF), and predicted microRNA (miRNA) target. Positively enriched gene sets were rank ordered and plotted by FWER p-values. Only results for 4hr are shown; 0hr and 12hr results are similar.

Due to the availability of PAR-CLIP data for both HuR and Ago we were able to compare actual binding sites for HuR and miRNAs. Previous studies have indicated combinatorial regulation by HuR and Ago, hence we inspected the positional relationship between HuR and Ago binding sites, which were derived from the initial PAR-CLIP study [9]. Of the 5,496 mRNAs for which PARalyzer detected Ago binding sites in 3’ UTRs, 3,238 mRNAs also contained HuR binding sites in the 3’ UTR. We found a significant enrichment for overlapping/adjacent HuR and Ago binding sites in 3’ UTRs of targeted mRNAs (Figure 27), with 227 pairs of HuR and Ago sites (in 218 3’ UTRs) that were within 20 nts of one another.
Currently, there are two models of combinatorial regulation by HuR and miRs: one model indicating that HuR can antagonize miRNA-mediated repression and the other indicating that HuR promotes miRNA-mediated repression. We examined the fate of these mRNAs in response to HuR depletion to investigate the functional outcome of combinatorial regulation by HuR and Ago. Although there was a slight trend for targets bound by HuR and Ago to respond more in the knockdown than targets bound by HuR alone, these differences were not significant (data not shown). Thus our data do not clearly favor either antagonism or synergism more than the other.

6.2 Post-translational modifications

HuR is subject to PTL modifications by many kinases, including Chk2, PKC-a and d, and CDK. Both Chk2 and PKC-d phosphorylation have been shown to disrupt or enhance association between specific mRNA targets. Therefore, we investigated PTL
modifications of HuR as a mechanism driving HuR-mRNA interaction dynamics using Chk2-mediated phosphorylation in response to oxidative stress as a model.

6.2.1 Chk2-mediated phosphorylation as a model

Chk2 is phosphorylated in response to oxidative stresses that produce high levels of free radicals. Phospho-Chk2 phosphorylates HuR at Ser88, Ser100, and Thr118, which results in differential association and expression of target mRNAs. This phosphorylation event results in the dissociation of the HuR-Sirt1 RNP complex and a decrease in the expression of Sirt1 mRNA and protein. A phosphorylation deficient mutant at the Ser100 site (S100A) can rescue the changes in HuR-association and expression level of Sirt1. Intriguingly, conversely to Sirt1 mRNA, HuR-association with PTMA mRNA increases in response to oxidative stress. Thus, we propose a model by which Chk2-phosphorylated HuR is important for regulating a subset of mRNAs encoding proteins important for the oxidative stress response (Figure 28).

![Figure 28: Chk2-mediated phosphorylation of HuR during oxidative stress](image)

A potential explanation for such divergent behavior is that different states of HuR exhibit different target specificity. Supporting this model is the fact that the phosphorylated residues occur in the region including the first and second RNA-
recognition motifs (RRM), which could change the conformation of these known RNA binding domains and alter binding properties. In fact, the Ser100 maps to a serine in HuD, which actually makes contact with the RNA ligand in crystal structures of HuD complexed with either c-fos or TNF RNA elements (Figure 29). However, there is no direct evidence demonstrating different states of HuR exhibit differential target sequence specificity.

![Figure 29](image)

**Figure 29: Ser100 from HuR may make contact with HuR.** A crystal structure of closely related HuD was made with c-fos RNA with the following properties color coded: (white) the first two RRM of HuD, (red) Ser118 of HuD maps to Ser100 of HuR, (blue) nine nucleotide stretch of c-Fos mRNA interacting with HuD, and (yellow) eighth uracil in c-Fos binding site.

### 6.2.2 Phosphorylation state specific targeting

In order to discriminate between specific forms of HuR, we created cell lines capable of expressing various FLAG tagged HuR variants (Figure 30). These included a WT variant, a phospho-deficient variant (S100A), and phospho-mimetic variants (S100D and S100E). These cell lines are inducible tet-on cell lines to avoid selecting for different
properties resulting from constitutive expression of the mutant HuR constructs. We also created a FLAG tagged EGFP as an additional control for all experiments related to the project.

![Diagram of HuR variant constructs]

Figure 30: Flag-tagged HuR variant constructs. Four individual cell lines were created each with a single epitope-tagged HuR variant. Ser100 was replaced with alanine (phospho-deficient), aspartic acid or glutamic acid (phospho-mimetic) residues. Since we used the Flip-IN T-Rex system, the cell lines contain the different variants in the same genomic location under control of a tetracycline responsive promoter.

We utilized PAR-CLIP of these tagged HuR variants to determine if phosphorylation at these sites could indeed change HuR-mRNA interactions, potentially through changes in target sequence specificity. Preliminary results of this analysis show that the phospho-mimetic mutants have a higher preference for binding mRNAs encoding proteins important for the oxidative stress response than the phospho-deficient mutant and WT HuR (Figure 31A). Additionally, we found p21, an important regulator of cell-cycle checkpoint and progression, to be a phospho-mimetic specific target (Figure 31B). While preliminary, these data suggest that PTL modification can indeed modify HuR-mRNA interactions.
Discussion

Our data establish that HuR-associated mRNAs are significantly enriched for predicted targets of over 90 microRNAs and TTP targets (Table 5, Figure 26). Furthermore, we found many hundred overlapping HuR and Ago binding sites. Since HuR can promote mRNA stability and translation, the presence of microRNAs and RBPs such as TTP, which promote mRNA degradation and/or translational repression, in HuR RNPs suggests competition with HuR, resulting in opposing functional outcomes. Indeed, competition between HuR and TTP has been demonstrated for individual
mRNAs, specifically IL-3 [171] and TNF-alpha [172]. While HuR was shown to be essential for the relief of microRNA mediated repression of the CAT-1 mRNA in stressed cells [173], it also has been implicated to recruit let-7 to destabilize c-MYC mRNA. Our functional HuR knockdown data does not support one model over the other. However, to truly assess which model is more prevalent, it would be necessary to have loss of function of Ago, and HuR and Ago combined. Furthermore, other RBPs will likely need to be accounted for in the complex models of combinatorial regulation.

Initial results investigating phosphorylation of HuR suggest that differential targeting of mRNAs can be achieved through modulating the phosphorylation state of HuR. We observed that Chk2 phospho-mimetic HuR mutants are more likely to target mRNA encoding proteins involved in oxidative stress response, as compared to WT or the phospho-deficient mutant. 3’ UTR reporter assay experiments are underway with candidate targets exhibiting differential targeting. Additionally, since Chk2 phosphorylates three residues of HuR, we have created triple mutants modeling these. It is likely this will provide a stronger signal with regards to binding and reporter expression studies.

Our data suggest that the targeting of these functionally antagonistic mechanisms is more widespread than currently believed, yet specific to subsets of transcripts. This is consistent with the PTRO model that predicts combinatorial interactions by RBPs and microRNAs that either compete or cooperate to determine the final functional outcomes that are shared by a subset of functionally related mRNAs [89, 174]. Additionally, it will be important to account for the post-translational state of individual RBPs, as they are not functionally equivalent as related to RNA processing and gene expression. Given the results described and the lack of correlation between
RBP-RNA association and mRNA abundance, other PTRFs and PTL modifications are mechanisms very likely to drive our observed HuR-RNA association dynamics.

6.4 Methods

Enrichment of RBP and microRNA targets. Gene sets were created for RBP targets from other data sets. MSigDB gene sets representing miRNA targets were also utilized. GSEA was conducted using the gene sets described above to calculate the significance of enrichments.

Comparing HuR and Ago binding sites. Ago clusters were analyzed using a local pipeline with parameters specific to Ago. Nearest neighbors analysis was conducted as described in chapter 3.

Site directed mutagenesis. FLAG-HA-HuR plasmids were the template utilized for site-directed PCR using the following primers:

Comparison of PAR-CLIP libraries. Each library was collapsed to gene symbols and then enrichment of pathways was calculated using PantherDB. Binding sites for individual mRNAs were depicted on the UCSC genome browser.
7. Systematic discovery of small molecule effectors of HuR

We hypothesized that HuR RNP dynamics could serve as a quantitative phenotype to explore the link between HuR and the physiological state of these cells. To test this possibility, we used the Connectivity Map [175] (CMAP) to identify small molecules that could potentially modify functional states of HuR. The CMAP is a tool that begins with a biological state of interest, specifically an *a priori* defined gene expression signature, and scans a database of perturbagen-induced transcriptomic profiles to connect the query signatures with small molecules based on correlation of gene expression dynamics.

In summary, the results of our analysis of HuR RNP dynamics using the CMAP led to the following conclusions: 1) Quantification of HuR RNP dynamics from ribonomic profiling identified effectors capable of modulating HuR; 2) Since we defined our biological state of interest based on HuR RNP dynamics rather than transcriptomic signature, this represents a novel application of the CMAP; 3) Small molecule drugs can have posttranscriptional consequences for cells that are largely unknown and uninvestigated.

7.1 Identification of candidate small molecule effectors

We utilized the 50 most dynamic mRNAs, those demonstrating the greatest increase in HuR association and greatest decrease in HuR association, as a quantitative phenotype of HuR RNP mRNA dynamics to be compared with the CMAP profiles. The small molecule candidates that were identified for both early and late HuR RNP dynamics (Table 6) may either mimic HuR functionality or act through HuR by yet to be determined mechanisms. Utilizing the 100 most dynamic mRNAs yielded very similar results (data not shown).
Table 6: Candidate Small Molecule HuR Effectors Derived Using CMAP. Sign of enrichment scores (ES) represents whether changes in HuR association across the time course were either positively or negatively correlated with transcriptomic changes. Significant p-values (<0.05) listed in red. ** Known effector of HuR. Common classes of compounds are highlighted, Hsp90 inhibitors (blue), HDAC inhibitors (yellow), PI3K inhibitors (orange), and COX inhibitors (green).

<table>
<thead>
<tr>
<th>Small Molecule</th>
<th>Mechanism</th>
<th>ΔHuR-RNP LOD scores</th>
<th>0hr to 4hr</th>
<th>4hr to 12hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetylsalicylic acid</td>
<td>COX inhibitor</td>
<td>-0.709</td>
<td>0.0492</td>
<td>0.301</td>
</tr>
<tr>
<td>Sirolimus</td>
<td>mTOR Inhibitor</td>
<td>-0.574</td>
<td>0.0014</td>
<td>-0.417</td>
</tr>
<tr>
<td>17-AAG</td>
<td>Hsp90 Inhibitor</td>
<td>-0.557</td>
<td>0.0001</td>
<td>0.0594</td>
</tr>
<tr>
<td>Resveratrol</td>
<td>COX Inhibitor</td>
<td>-0.511</td>
<td>0.1007</td>
<td>-0.562</td>
</tr>
<tr>
<td>LY-294002</td>
<td>PI3K Inhibitor</td>
<td>-0.446</td>
<td>0.0015</td>
<td>0.33</td>
</tr>
<tr>
<td>Deferoxamine</td>
<td>Iron Chelator</td>
<td>-0.415</td>
<td>0.5649</td>
<td>0.0133</td>
</tr>
<tr>
<td>Valproic acid</td>
<td>HDAC Inhibitor</td>
<td>-0.414</td>
<td>0.0021</td>
<td>0.0617</td>
</tr>
<tr>
<td>Wortmannin</td>
<td>PI3K Inhibitor</td>
<td>-0.353</td>
<td>0.2172</td>
<td>0.49</td>
</tr>
<tr>
<td>Trichostatin A **</td>
<td>HDAC Inhibitor</td>
<td>-0.315</td>
<td>0.1549</td>
<td>0.747</td>
</tr>
<tr>
<td>Monorden</td>
<td>Hsp90 Inhibitor</td>
<td>-0.189</td>
<td>0.802</td>
<td>0.451</td>
</tr>
<tr>
<td>Rofecoxib</td>
<td>COX Inhibitor</td>
<td>0</td>
<td>1</td>
<td>-0.552</td>
</tr>
<tr>
<td>Vorinostat</td>
<td>HDAC Inhibitor</td>
<td>0</td>
<td>1</td>
<td>0.943</td>
</tr>
<tr>
<td>Geldanamycin</td>
<td>Hsp90 Inhibitor</td>
<td>0.326</td>
<td>0.4547</td>
<td>0.728</td>
</tr>
<tr>
<td>Raloxifene</td>
<td>SERM/Estrogen Inhibitor</td>
<td>0.547</td>
<td>0.2355</td>
<td>0.753</td>
</tr>
<tr>
<td>Ikarugamycin</td>
<td>Endocytosis Inhibitor</td>
<td>0.779</td>
<td>0.0221</td>
<td>0.684</td>
</tr>
</tbody>
</table>

Many of the candidate HuR effectors fell into drug classes that have overlapping mechanisms of action, including PI3K, COX, HDAC, and Hsp90 inhibitors. HDAC inhibitors, which exert global effects on gene expression through chromatin remodeling, showed significant negative correlation with both early and late HuR-RNP dynamics. A negative correlation indicates that when the small molecule was tested in the set of cell lines used to prepare the CMAP, it induced changes in mRNA levels that were in the opposite direction of our observed changes in HuR association. Trichostatin A (TSA), a reversible HDAC inhibitor, has been shown to induce cell cycle arrest in Jurkat cells [176]. Importantly, the identification of TSA represents an independent validation of our approach, since it has been shown to affect mRNA stability by decreasing the amount of cytoplasmic HuR in MCF7 [177] and RKO cells [178], breast and colon cancer lines respectively, without affecting the total amount of HuR in the cell.
7.2 Validation of resveratrol as a small molecule effector of HuR

To further validate the outcomes of the CMAP analysis, we extensively examined a novel candidate from our list, resveratrol. We examined if resveratrol could modulate the subcellular localization of HuR as well as HuR-dependent gene expression.

7.2.1 Resveratrol effects subcellular localization of HuR

First, we examined the effects of pretreatment with resveratrol on the subcellular distribution of HuR 0, 4, and 12 hours post-activation. Regardless of pretreatment with resveratrol, overall levels of HuR protein did not change during the activation (data not shown). However, pretreatment with resveratrol resulted in an accumulation of endoplasmic reticulum/outer nuclear envelope (ER/ONE) localized HuR and a concomitant depletion of nuclear localized HuR 12 hrs post activation (Figure 32). Thus, resveratrol pretreatment modulates the subcellular distribution of HuR during T cell activation.
Figure 32: Resveratrol Affects Subcellular Localization of HuR During Activation. (A) Representative immunoblot of subcellular localization of HuR protein 0, 4, and 12 hrs post-activation of Jurkat cells using PMA and PHA with or without resveratrol pretreatment. Proteins were analyzed by immunoblot using anti-HuR, anti-β-tubulin as a cytosolic (Cyto) marker, anti-TRAP-α as an endoplasmic reticulum/outer nuclear envelope (ER/ONE) marker, and anti-histone h4 as a Nuclear (Nuc) marker. Four fold more Cyto and ER/ONE fraction extract was loaded compared to Nuc fraction extract. (B) Quantification of relative fold change in HuR following resveratrol pre-treatment. HuR levels were normalized to the appropriate loading control for each fraction. Normalized HuR levels for resveratrol pre-treated samples were compared to mock pretreated samples. The means and standard error of the means (SEM) are represented from 3 independent experiments. Paired t-tests were used to calculate p-values (* P <0.05).
7.2.2 Resveratrol antagonizes HuR-dependent posttranscriptional regulation

Next, we tested if resveratrol could have effects on posttranscriptional gene expression. We designed luciferase reporter constructs containing regions of the 3’ UTR from several of the top 50 most dynamic mRNAs that were used in the CMAP analysis (TNF-α, CSF2, ANP32A, and NAB2). Activation of the Jurkat cells induced a two-fold or greater increase in expression (Figure 33A) for three of the four reporters (TNF-α, CSF2, and ANP32A). Pretreatment with resveratrol resulted in a ~20-30% decrease in TNF-α, CSF2, and ANP32A reporter expression only in activated cells (Figure 33A). This suggests that resveratrol suppresses activation-induced increases in gene expression through one or more posttranscriptional mechanisms. However, resveratrol pretreatment did not suppress reporter expression of activated Jurkat cells when co-transfected with HuR (Figure 33B). Therefore, these data indicate that HuR can partially rescue resveratrol-mediated posttranscriptional suppression of reporter expression in activated Jurkat cells, suggesting that HuR has a role in the effects of resveratrol on these cells.
Figure 33: Activation Dependent Effects of Resveratrol and HuR on 3’UTR Luciferase Reporters. A) Firefly to renilla ratio of each reporter normalized to the control 3’UTR reporter, pLuc, for each condition. Activation significantly increases reporter expression through specific sequences. Furthermore, resveratrol significantly blunts this increase. B) The fold change of the normalized firefly to renilla ratio for the indicated condition (Activation + Resveratrol or Activation + Resveratrol + HuR) compared to the normalized firefly to renilla ratio for its matched mock treated control. As shown in A) resveratrol inhibits activation mediated increases in reporter expression, however B) shows that HuR overexpression antagonizes resveratrol-dependent inhibition.
7.3 Discussion

Based on HuR RNP dynamics we identified classes of small molecule candidates that can modulate HuR functionality using the Connectivity Map (Table 6). We show for the first time that RNP dynamics can be used as a quantitative phenotype to systematically identify compounds that exert posttranscriptional effects and modulate the corresponding RBP, in this case, HuR (Figures 32 and 33).

Hsp90 inhibitors are a class of small molecule drugs for which multiple candidate compounds were identified, specifically geldanamycin, monorden, and 17-AAG. It was shown previously that during LPS activation of macrophages, treatment with geldanamycin resulted in decreased production of inhibitory cytokines [179] by negatively affecting stability and translation of cytokine mRNAs, including those known to be regulated by HuR.

Many of the compounds identified as candidate small molecule effectors of HuR have been shown to have anti-cancer properties. HuR has been shown to be directly involved in colon carcinogenesis [144] and inappropriate cytoplasmic expression of HuR protein has been associated with many cancers [180-183]. Since our data is from Jurkat cells, a human T cell leukemia line, our identification of the HDAC inhibitor Voronistat (Table 5) is especially noteworthy because it is used for the treatment of cutaneous T-cell lymphoma. Additionally, coadministration of 17-AAG with Voronistat may represent a promising anti-leukemic strategy [184]. Furthermore, exploration of the connection between the effects of those identified compounds and HuR RNP dynamics could further our understanding of both HuR and cancer. Therefore, a more general application of RNP dynamics to identify tools that can be used to modulate a posttranscriptional process provides a new avenue for investigating biological mechanisms as well as the relationship between RBPs and disease.
We found that resveratrol, a COX inhibitor that exhibits anti-inflammatory and chemopreventive effects, modulates the subcellular localization of HuR during activation (Figure 32). Also, our data show that resveratrol can suppress activation induced gene expression (Figure 33). Similarly, a previous study demonstrated that resveratrol suppresses TNF-dependent activation of transcription factors in PMA treated Jurkat cells [185]. However, our results reveal a posttranscriptional component to the effects of resveratrol. Interestingly, since we show resveratrol suppresses activation-induced increase in TNF reporter expression, this could be a mechanism working upstream of TNF-dependent activation of transcription factors during T cell activation. Furthermore, we observed that HuR antagonized the resveratrol-mediated effects on gene expression. Therefore, we can systematically identify compounds that modulate RBP function, and have posttranscriptional consequences for gene expression, and whose effects can be rescued by the RBP being examined.

Additionally, resveratrol, which is found in red wine, exhibits anti-aging properties putatively through activation of sirtuin-1 (SIRT1), a known HuR target. A connection between HuR and aging was demonstrated in studies by Gorospe and colleagues using multiple models of senescence, that showed correlated decreases in both HuR levels and the stability of target mRNAs involved in aging, including sirtuin-1 (SIRT1) [82, 148]. Indeed, HuR is functionally important to this process as overexpression of HuR in senescent cells restores a “younger” phenotype. In our study, we not only found SIRT1 as a target, but also found that genes reported to have reduced expression in the brain of humans after the age of 40 [186] were significantly enriched as HuR targets. Further evidence for this connection comes from our systematic use of HuR RNP dynamics to identify resveratrol as a candidate HuR effector. Resveratrol, a compound found in red wine, is a SIRT1 activator that has been shown to have anti-
aging properties and is under extensive clinical investigation. These data suggest that the mechanism of action of resveratrol may have posttranscriptional effects that act in some part through HuR. Therefore, it will be important to determine if there are HuR dependent effects of resveratrol. This hypothesis can be tested by genetically or pharmacologically perturbing HuR, and determining if the known physiological effects of resveratrol are altered. The results shown here that led to a predicted connection between HuR and aging via resveratrol demonstrated one of the powerful implications of using RNP dynamics in a functionally perturbed cellular model. Furthermore, this strategy could be used to discover novel connections between other RBPs and disease states or to provide tools to study uncharacterized RBPs that have been implicated in a disease, such as RNA binding motif protein 17 identified from a genome-wide association study of type 1 diabetes [187].

7.4 Methods

Plasmids. The Firefly-UTR reporters used for this study were generated by cloning the UTR fragments into the NotI and ApaI sites of pCDNA3-Luc. Fragments of the UTRs were created using the following primers:

- TNF ARE-Fwd TCCAGATTTTCCAGACTTC
- TNF ARE-Rev TGAGCCAAGGCACTCTCTAC
- CSF2-Fwd TGATACAGGCATGCAGAAG
- CSF2-Rev TACGGTAAAAACATCTTGAATAATATG
- ANP32A-Fwd AGTGGAATAACCTATTTTGAAAAATTC
- ANP32A-Rev CATTCTTTTTCAATACAGACAAAAACAA
- NAB2-Fwd AGGTTTGGACTGGTGTCTTC
- NAB2-Rev GCCATAAAAAATTTATCCAAA

purl, pCDNA3-HuR, and pCDNA3 were used in this study as well.
**Transfections.** Transfections were performed using Lipofectamine 2000 (Invitrogen) using the standard protocol. Briefly, 1.6 μg of total DNA was diluted in 100 μl of Opti-MEM I (Gibco) and mixed with 4μl of Lipofectamine 2000 diluted in 100 μl of Opti-MEM I and incubated at room temperature for 30 minutes. 1 x 10^6 Jurkats were plated in fresh media in 12-well plates. The re-plated cells were then immediately mixed with the DNA/Lipofectamine 2000 complexes. The Luciferase reporter plasmids were transfected in equimolar amounts, 20 pmoles for each of the Firefly-UTR constructs and 10 pmoles for the Renilla construct. 0.5 μg of pCDNA3-HuR was co-transfected in indicated experiments and the remainder of the transfection mix was brought to 1.6μg using the pCDNA3 vector.

**Cell Fractionation.** Cells were collected, washed with PBS and then subject to a previously described fractionation protocol [115]. Cytoplasmic, endoplasmic reticulum/outer nuclear envelope, and nuclear fraction control antibodies included anti B–Tubulin (Harlan Sera-Lab), anti-TRAP-α (kindly provided by Chris Nicchitta), and Histone H4 (Abcam), respectively. Protein bands were quantified using GelEval (Frog Dance Software).

**Luciferase Assay.** Luciferase assays were performed using the Dual-Luciferase Reporter Assay System (Promega). Transfected cells were pre-treated with 50nM Resveratrol for 2 hours and then activated with 50 ng/ml PMA and 2 μg/ml PHA for 4 hours. The cells were then collected and washed with PBS and then lysed with 100 μl of 1x Passive Lysis Buffer. For both cell fractionation and luciferase experiments, paired T-tests were performed in GraphPad Prism (GraphPad Software).
8. Conclusions and future directions

Much progress has been made in the development of both experimental and computational methods, which set the stage for large strides to be made towards understanding mechanisms and the importance of post-transcriptional regulation. This will still be challenging due to the highly dynamic and combinatorial nature of post-transcriptional processes. In order for these issues to be adequately addressed advances in RIP efficiency and detection sensitivity will be necessary. Specifically, being able to perform many RIPs in parallel from the same lysate as well as sequential RIPs do demonstrate co-occurrence of RBPs on a given transcript. PTL modifications present a heterogeneity issue for RIP procedures, however it provides a direct mechanism to immediately respond to the specific condition or stimuli, effectively linking gene expression and signaling while bypassing transcription.

The staggering amount of non-coding RNA, in addition to miRNAs, from the vast intergenic regions of the genome will inevitably feed into this layer of gene expression. Lastly, it will be necessary to connect these interactions of functional outcomes for the mRNA. Deep sequencing coupled to classic biochemical protocols has already started to make this a reality for mRNA processing, abundance, and translation. However, there is no high-throughput assay for localization, which represents the most challenging and quite fascinating regulatory capability of RBPs.

Integration of these large data sets will require quantitative modeling of both interactions and functional outcomes. Probabilistic modeling is an attractive approach given the subtlety of most regulation; rarely are events all or none. Finally, such modeling approaches will allow for integration with large datasets of transcriptional control, such as transcription factor binding and chromatin structure, to gain top down view of coordinated regulation throughout gene expression.
References

Parts of chapters 1, 2, 5, 6, and 7 also appear in the following articles:


185. Manna, S.K., A. Mukhopadhyay, and B.B. Aggarwal, Resveratrol suppresses TNF-induced activation of nuclear transcription factors NF-kappa B, activator protein-1, and


Biography

Neelanjan Mukherjee was born August 29th, 1979 in Columbus, Ohio. Neel graduated from UCSD with a Bachelor of Sciences degree in General Biology.

Publications at Duke:


