


Summer 8-15-2018

Characterization of the CELF6 RNA Binding Protein: Effects on Mouse Vocal Behavior and Biochemical Function

Michael A. Rieger

Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

 Part of the [Genetics Commons](#), [Molecular Biology Commons](#), and the [Neuroscience and Neurobiology Commons](#)

Recommended Citation

Rieger, Michael A., "Characterization of the CELF6 RNA Binding Protein: Effects on Mouse Vocal Behavior and Biochemical Function" (2018). *Arts & Sciences Electronic Theses and Dissertations*. 1649.
https://openscholarship.wustl.edu/art_sci_etds/1649

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS
Division of Biology and Biomedical Sciences
Molecular and Cell Biology

Dissertation Examination Committee:
Joseph D. Dougherty, Chair
Yehuda Ben-Shahar
Barak A. Cohen
Timothy E. Holy
Jean E. Schaffer

Characterization of the CELF6 RNA Binding Protein: Effects on Mouse Vocal
Behavior and Biochemical Function
by
Michael A. Rieger

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2018
St. Louis, Missouri

© 2018, Michael A. Rieger

Table of Contents

List of Figures	vi
List of Tables	viii
Acknowledgments	x
Dissertation Abstract	xii
1 CELF6 and ASD phenotyping in mice	1
Abstract	1
A Brief History of ASD Genetic Studies	2
Overview of Behavioral Phenotyping in Mouse Models of ASD	6
<i>Celf6</i> and its association to ASD and USV	9
Summary and Dissertation Overview	11
2 Characterization of <i>Celf6</i> KO vocalization	20
Abstract	20
Introduction	21
Results	23
<i>Celf6</i> ^{-/-} mice show reductions to overall rate of ultrasonic calling when separated from their mothers.	23
<i>Celf6</i> ^{-/-} knockout mice show similar temporal and spectral features of vocalization to wild-type mice	26
Mice with CELF6 loss in 5HT and DA neurons fail to show reductions to USV	29
Discussion	32
Methods	33
Animals	33
USV recording	34
USV processing	35
Immunofluorescent staining	35
Statistical Analysis	36
3 Analysis of USV variability within subjects	41
Abstract	41

Introduction	43
Results	47
Assessment of consistency of USV features across early postnatal development	47
Consistency of USV features within recording sessions	55
Consistency of features of USV in adult male-female C57BL/6J dyads	59
Discussion	63
Methods	67
Animals	67
USV Recording and Processing	68
USV recording - Pups	68
USV recording - Adult M-F dyads	69
White noise filtering in the frequency domain.	69
Spectrogram preparation and band-pass filtering & automated call detection	70
Call feature extraction.	71
Statistical Analysis	71
Supplemental Information	75
4 CELF Family of RBPs	88
Abstract	88
Introduction	89
Interactions with RNA through RNA Binding Domains	91
Identifying and defining RBP:RNA interactions	94
Regulation of mRNAs by CELF proteins	97
Summary	100
5 CELF6 Targets and Downstream Function	110
Abstract	110
Introduction	112
Results	114
Celf6 primarily associates with 3'UTRs of target mRNAs <i>in vivo</i>	114
CELF6-associated 3' UTRs are enriched for U-rich and UG-, CU- containing motifs	118
Massively Parallel Reporter Assay to define the function of CELF6 CLIP enriched motif sequences	121
CELF6 CLIP enriched motif sequences represent a set of repressive elements	123
CELF6 protein enhances repression in a sequence dependent manner	124
CELF3-5 show redundancy in ability to enhance repression of Celf6-CLIP enriched UTR elements	126
Motif strength and number alone can partially predict CELF6 impact.	131
Discussion	133
Methods	136

Animals	136
Cell culture	136
CLIP	137
CLIP-Seq Sequencing Library Preparation	139
CLIP-Seq Sequencing Data Processing	139
CLIP Motif Enrichment Analysis	141
PTRE-Seq Reporter Library Preparation	141
PTRE-Seq Reporter Library Transfection	143
PTRE-Seq Sequencing Library Preparation	143
PTRE-Seq Sequencing Library Data Processing	144
PTRE-Seq Sequencing Library Statistical Analysis	145
Data Accessibility	146
Supplemental Information	146
Conclusions And Future Directions	228
Summary	228
Future Directions	229
Does CELF6 enhance mRNA decay?	229
Does CELF6 result in target repression <i>in vivo</i> ?	230
Does loss of CELF6 impact neuronal function <i>in vivo</i> ?	230
What are the molecular and circuit contributions to variability in behavior?	231
Broader Impact and Significance	232
Appendices	235
1 Vocalization Data Processing	235
Installation & upgrades	235
Installing for the first time	236
Downloading the Archives	236
Installing the Rlike library	236
Installing the VocalizationFunctions library package	237
Upgrading to the newest version	238
Running the GUI	238
Preparing your data	238
Preparing a CSV for upload	239
Checking files for consistency	241
Starting the GUI	242
Setting options for processing	243
Project details	244
Audio File Parameters	244
White Noise Filtering	244
FFT Settings	245
Automated Call Scoring	245

	Spectrogram Spot-Check	247
	Inspecting output reports	248
	Adding new data to an existing project	249
2	Generalized RNA-Seq Lib Prep	251
	rRNA Probe Annealing	251
	RNase H Digestion	252
	DNase I Digestion	252
	Clean up with MyONE Silane Beads	253
	Dephosphorylation Reaction	254
	Clean up with MyONE Silane Beads (Abbreviated, see above for full protocol)	255
	A01m Ligation	255
	Clean up with MyONE Silane Beads (Abbreviated, see above for full protocol)	256
	Reverse Transcription	257
	Clean up with MyONE Silane Beads (Abbreviated, see above for full protocol)	258
	Rand103tr3 Ligation	259
	Clean up with MyONE Silane Beads	260
	Trial Library PCR	260
	Preparative Library PCR	261
	SPRI Purification	262
3	Translating Ribosome Affinity Purification - Optimized	263
	Protocol	263
	Bead Prep	263
	Homogenization and Lysis	264
	RNA Extraction	265
	Buffers and Reagents	267
	Products Numbers	267
	1% BSA/1XPBS	267
	0.1% BSA/1XPBS	267
	10% NP40	268
	300 mM DHPC	268
	100 mg/mL cycloheximide	268
	10X Roche cOmplete EDTA-free protease inhibitor	268
	Wash Buffer	268
	Homogenization Buffer	269
	High Salt Wash Buffer	269
4	PTRE-Seq Sequencing Library Preparation Protocol	270
	cDNA Synthesis	270
	ds cDNA amplification and plasmid DNA amplification	271
	AMPure 80/40 selection	271
	NheI/KpnI Digest	272
	AMPure 100/50 selection	272

Adapter Ligation	273
Preparative PCR	273

List of Figures

Chapter 2

- 1 *Celf6* knockout mice show reductions to overall rate of ultrasonic calling when separated from their mothers. 24
- 2 Spectral and temporal properties of ultrasonic calls are unperturbed in *Celf6* knockout mice. 27
- 3 Mice with CELF6 loss in 5HT and DA neurons fail to show reductions to USV. 30

Chapter 3

- 1 The Intra-Class Correlation (ICC) defined from a linear mixed model (LMM) reflects the level of behavioral consistency of individual animals across multiple measurements. 48
- 2 USV features under investigation include commonly measured features from the time and frequency domain. 50
- 3 Lack of Strong consistency across pup USV features in Pooled Cohort Study (PCS) 52
- 4 Lack of strong consistency across pup USV features in Time Course Study (TCS) 54
- 5 USV features show higher consistency within within sessions than between sessions. 60
- 6 Stronger consistency in some adult USV features across sessions. 62
- S1 Conditional knockout of *Celf6* in dopaminergic neurons does not significantly alter features of USV in Cohort 1 of PCS 75
- S2 Conditional knockout of *Celf6* in GABA-ergic neurons does not alter features of USV in Cohort 2 of PCS. 76
- S3 Global knockout of *Celf6* does not perturb USV features in adult male-female dyads. 77

Chapter 4

- 1 The CELF family of RNA binding proteins shows homologous structure with 3 RRM s and two distinct subclasses. 93

Chapter 5

1	Celf6 primarily associates with 3'UTRs of target mRNAs <i>in vivo</i>	115
2	CELF6-associated 3'UTRs are enriched for U-rich and UG-, CU- containing motifs	119
3	CELF6 CLIP enriched motif sequences represent a set of repressive elements	127
4	CELF3-5 show redundancy in ability to enhance repression of CELF6-CLIP enriched UTR elements	129
5	Motif match strength and number of matches can partially predict CELF6 impact.	132

List of Tables

Chapter 3

1	Values of the ICC and Confidence Intervals Computed in the PCS TCS. . .	55
2	Values of the ICC and Confidence Intervals Computed in the PCS across minute bins.	57
3	Values of the ICC and Confidence Intervals Computed in the TCS across minute bins.	58
4	ICC in the PCS and TCS across Minute Bins: Averages and Standard Deviations.	59
5	Spearman's Correlation and ICC computed for adult C57BL/6J data.	63

Chapter 5

S1	Oligos for CLIPSeq Library Preparation	146
S2	CLIP-Seq Aligned Read Summary	146
S3	CLIP Targets Defined from Piranha Peaks	147
S4	CLIP Targets Defined From Annotated Subgenic Regions	151
S5	Top Ranking Motif Cluster Matches Across UTR peaks	163
S6	PTRE-Seq - Effects on Expression	183
S7	PTRE-Seq - Effects on Translation Efficiency	200
S8	Oligos for PTRESeq Library Generation	219

Acknowledgments

There are many individuals owed acknowledgment for various aspects of this work. With respect to studies of ultrasonic vocalization, thanks to Timothy Holy, and especially to Terra Barnes who trained me on acquisition and analysis of vocalization, and with whom I have had endless stimulating conversation on this topic. In graduate school, I also found myself becoming a programmer, and she is owed a great deal of thanks for that. My strong training in signal processing I have gained from her has impacted the way I code and the way I look at all kinds of data.

I would also like to thank Afua "ChuChu" Akuffo who was my first friend in the Dougherty lab since the early days, who trained me in mouse work and also provided endless laughs and thought-provoking conversations, and to Susan Maloney who was a well-spring of information and insight regarding behavioral testing and statistical analysis. Before graduate school, I was trained as a molecular biologist, and these folks expanded my understanding of biology immensely. I would also like to thank others who contributed to data acquisition for the vocalization projects including Shyam Akula, Mengxi Zhang, Kellie Wilson, Satchel Siegel, whom I had the great pleasure of providing direct mentorship.

With respect to biochemical studies of CELF6, I would like to give a great deal of thanks to Jean Schaffer and Josh Langmade for training and consultation on CLIP. Also a big thank you to Kyle Cottrell and Sergej Djuranovic for consultation on implementing PTRE-Seq in our lab, and many thanks to Barak Cohen, Dana King, and Hemangi Chaudhari. Dana especially has been instrumental in getting our PTRE-Seq assay off the ground and has been great to consult with regarding how those data can be best understood. Also my sequencing studies have passed through the Center for Genome Sciences at Washington University in St. Louis, so great thanks are owed to Jessica Hoisington-Lopez.

Additionally, it is no secret that I have been tinkering with computers since I was a middle schooler in the 1990s. So in addition to the folks mentioned above, I'd like to thank Kyle Kniepkamp, our technology specialist, who always lets me know when a free piece of equipment is laying around ripe for recycling, and whom I have never once asked assistance in installing a network printer or ridding a computer of malware,

which I am certain has ensured that I am continually informed about said free equipment.

I would like to give thanks to all the members of my thesis committee: Timothy Holy, Jean Schaffer, Yehuda Ben-Shahar, Barak Cohen, and my mentor Joseph Dougherty. I assembled a truly diverse group of researchers across different fields of expertise. They have all provided me with excellent feedback and advice throughout my graduate career. I would also like to thank all other members of the Dougherty lab, especially Jasbir Dalal, Kristina Sakers, Nathan Kopp, Shyam Akula, David O'Brien, Allie Lake, Bernie Mulvey, and Rachel Rahn. Extra thanks to Rachel Rahn and Susan Maloney for providing some greatly needed copy-editing on this dissertation.

In addition to all these people, my life has been especially rich with friends of diverse experiences and circumstances. I would like to thank all my friends for keeping me sane through this process, including but not limited to Jonathon Edwards, Tirzah Abbott, Will Fruhwirth, Molly Pearson, Fitz Hartwig, Broccoli Pearson-Hartwig, Allison Harris, Christopher "Toph" Harris, Oscar Harris, Nate Lucena, Maureen Hanlon, the Rise Coffee crew, and everyone at Mynx Academy, especially Michelle Mynx who is a seemingly limitless source of positive encouragement and creativity, and countless others. My friends are lawyers, teachers, social workers, stage performers, programmers, musicians, therapists, artists, scientists, engineers, scattered locally and across the globe, and interacting with a diversely trained group of people enriches my life. Love and thanks as well to my parents and extended family for always providing support and encouragement.

Finally I would like to give great thanks to my mentor Joseph Dougherty. This work would not have been possible without his boundless support, encouragement, training, early morning and late night conversations. His mentorship has given me confidence, opportunities for growth, and I am forever thankful.

Original vocalization studies on Celf6 knockout mice was supported by National Institutes of Health Grant 4R00NS067239-03 to Joseph D. Dougherty and Michael A. Rieger was supported by NIH 5T32GM007067-38. Conditional knockout vocalization studies (Chapter 2-3) were additionally supported by the McDonnell Center for Systems Neuroscience, and NIMH (R01MH107515, U01MH109133), NINDS(5R00NS067239). Biochemical studies in Chapter 5 was supported by the NIH (5R01HG008687, 4R00NS067239, 5R21DA038458, and R01 GM092910 (Barak Cohen)). Joseph Dougherty is a NARSAD independent investigator of the Brain and Behavior Research Foundation.

Michael A. Rieger

Washington University in St. Louis

August 2018

Dissertation Abstract

Michael A. Rieger

Behavior in higher eukaryotes is a complex process which integrates signals in the environment, the genetic makeup of the organism, and connectivity in the nervous system to produce extremely diverse adaptations to the phenomenon of existence. Unraveling the subcellular components that contribute to behavioral output is important for both understanding how behavior occurs in an unperturbed state, as well as understanding how behavior changes when the underlying systems that generate it are altered. Of the numerous molecular species that make up a cell, the regulation of messenger RNAs (mRNAs), the coding template of all proteins, is of key importance to the proper maintenance and functioning of cells of the brain, and thus the synaptic signals and information integration which underlie behavior. RNA binding proteins, a class of regulatory molecules, associate with mRNAs and facilitate their maturation from pre-spliced nascent transcripts, their stabilization and degradation ensuring appropriate levels are maintained, as well as their translation and subcellular compartmentalization, which ensures that proteins are translated at the appropriate level and in the places where they are required to fulfill their cellular functions. Our laboratory identified polymorphisms in the gene coding for the CUGBP and ELAV-like Factor 6 (CELF6) RNA binding protein to be associated with Autism Spectrum Disorder risk in humans. ASD is a spectrum of disorders of early neu-

rodevelopment which present with lowered sociability and communication skills as well as restricted patterns of interests. When expression of the *Celf6* gene was ablated in mice, we found that they exhibited reductions to early communication as well as altered aspects of their exploratory behavior. In this dissertation, I explore the communication changes in young mouse pups with loss of CELF6 protein and identify that despite being able to produce vocalization patterns similar to their wild-type littermates, they nevertheless exhibit reduced response to maternal separation. Despite a history of literature on other CELF family proteins, the functions of the CELF6 protein in the brain have not been previously described. I provide characterization of the mRNA binding targets of CELF6 in the brain, and show that they share common UGU-containing sequence motifs which has been noted for other CELF proteins, and that CELF6 binding occurs primarily in the 3' untranslated regions (3' UTR) of mRNA. I hypothesized that this mode of interaction would result in regulation of mRNA degradation or translation efficiency as 3' UTR regions are known for providing binding sites for numerous regulators of such processes. In order to answer this question, I cloned sequence elements from the 3' UTRs of target mRNAs into a massively parallel reporter assay which has enabled me to test the effect of CELF6 expression on hundreds of binding targets simultaneously. When expressed *in vitro*, I found that CELF6 induced reduction to reporter library levels but exhibited few effects on translation efficiency, and I was able to rescue effects to reporter abundance mutation of binding motifs. Intriguingly, like CELF6, CELF3, CELF4, and CELF5 were all able to produce the same effect. CELF5 and CELF6 both showed similar, intermediate repression of reporter library mRNAs, while CELF3 and CELF4 exerted the strongest levels of repression. The level of repression under these conditions was somewhat predicted by number of motifs present per element, however a large amount of the variance in reporter levels is still unexplained and a mechanism for CELF6's action is unknown. Nevertheless, the work I present in this dissertation shows that CELF6 and other members of its family are

key regulators of mRNA abundance levels which has direct implications to downstream consequence in the cell. As several of CELF6 binding target mRNAs are known regulators of neuronal signaling and synaptic function, the information I present is crucial for future experimentation. This work will help lead us to understand how behavior is altered when this protein is absent, along the way uncovering important mechanistic steps connecting the molecular landscape of cells to the behavior of organisms.

Chapter 1

The *CELF6* gene and autism spectrum disorder phenotyping in mice

Abstract

Autism spectrum disorders (ASD) are developmental disorders with prevalence rates estimated 1-2.5% in the United States, defined by abnormal social behavior, restricted interests, and language delay. For the last four decades, ASD has been recognized to have a measurable heritable component, with that heritability stemming from some burden in common variation as well as rare/*de novo* genetic variation. Several Mendelian syndromes also present clinically with autistic features. The influence of genes believed to confer ASD risk has therefore been widely explored in research, with gene function commonly investigated in laboratory model systems such as rats and mice. Although rodents may not be expected to fully reproduce the features of the human condition, the biological function of risk genes in the brain can be explored vis-à-vis their relationship to behavioral changes in the expected domains of

sociability, repetitive behaviors, and altered communication. In particular, the ultrasonic vocalization (USV) communication system in mice has been widely used to evaluate models in which ASD risk genes are knocked out or mutated. Our lab discovered an association between the RNA binding protein CELF6, enriched for expression in serotonergic neurons, and ASD risk in human patient data. When we knocked out this gene in mice, the animals showed deficits to neonatal pup USV among other abnormalities in behavior. This association between the *CELF6* gene and changes to USV, among others, forms the basis for our pursued study of the *CELF6* gene in this dissertation work. Thus as a prelude to our work on *CELF6*, this chapter is devoted to a brief review of the history of ASD genetics research and modeling the disorder in mice, as well as our initial observations concerning the *CELF6* gene.

A Brief History of ASD Genetic Studies

Autism spectrum disorder (commonly "autism" or ASD) is a neurodevelopmental disorder diagnosed in early childhood typified by abnormalities in the domains of social interaction, stereotypies/restricted interests/repetitive behaviors, and language delay/communication difficulties, and is often comorbid with other medical problems or behavioral difficulties including epilepsy, ADHD, and intellectual disability [1]. ASD was recognized early on as a disorder with some measurable heritable component from concordance rates in diagnosis in monozygotic (MZ) vs. dizygotic (DZ) twins. An original twin study by Folstein and Rutter of 11 pairs of MZ & 10 pairs of DZ twins in 1977 showed a 36% concordance rate for ASD diagnosis in MZ twins with 0% concordance for ASD diagnosis in DZ twins, and an 82% concordance rate for cognitive abnormalities at all in MZ twins, and 10% in DZ twins [2]. At the time, the rates of ASD were considered uncommon at 2-4 in 10,000 and twins for

the Folstein study were selected to meet the then diagnostic criteria [3]:

- impairment in the development of social relationships (limited eye gaze, poor social responsiveness, impaired selective bonding, failure to go to parents for comfort, lack of empathy when older, lack of personal friendships, little group interaction);
- delayed and deviant language development (poor comprehension, limited gesture usage, echolalia, pronomial reversal, limited social use of language, repetitive utterances, flat or staccato speech, restricted play);
- stereotyped, repetitive/ritualistic play or interests (abnormal attachment to objects, resistance to change, rituals, repetitive behavior, unusual preoccupations, restricted interest patterns).

Diagnosis was carefully conducted to be as "blinded" as possible - randomizing the case histories when presented to the diagnoser so pairs were not diagnosed together, and separate diagnoses of cognitive impairment (e.g. IQ tests, verbal IQ) also performed. Almost 20 years later, a meta-analysis of two twin studies from the United Kingdom included 25 MZ twin pairs & 20 DZ twin pairs [4] found a 60% concordance rate in MZ twins vs. 0% DZ twins, and 92% of MZ twins were generally concordant more broadly for other cognitive and social impairments vs. 10% of DZ twins. Since the original Folstein and Rutter study in 1977, twin studies have attempted to control for obstetric hazards and other medical conditions, but of course these cannot be completely discounted. However the consensus at the time was that ASD was the result of multiple loci acting epistatically.

By the year 2000, a number of de novo chromosomal abnormalities such as deletions, translocations, and inversions, had been associated with ASD as well as other cognitive disorders in siblings [5]. A number of candidate genes were investigated, such as the GABA_A receptor cluster on chromosome 15, which was also implicated in epilepsy [6, 7]. Additionally, elevated peripheral serotonin levels in some autistic patients [8, 9] relative

to controls led to the investigation of genes in the serotonergic system. A particular focus was an association of a polymorphism in the promoter for the serotonin transporter gene *5-HTT/SLC6A4/SERT* with some contradiction in the literature as to whether the short variant was associated with ASD [10], or the long variant [11] of the same polymorphism. Other potential candidate genes emerging at this time included genes involved in syndromes with Mendelian inheritance patterns which presented clinical with autistic phenotypes, such as the neurofibromatosis type 1 (*NF1*) gene [12], the Rett syndrome gene *MECP2*[13], and the tuberous sclerosis genes *TSC1* and *TSC2* [14]. Some early reports at this time began to document the prevalence of ASD diagnosis & Fragile X syndrome (*FMR1*) [15, 16].

Also around this time, linkage studies were undertaken to identify ASD susceptibility genes in multiplex families (families with near relatives also sharing this diagnosis) such as the study by the International Molecular Genetic Study of Autism Consortium in 1998 [17] in 87 affected sibling pairs with significant linkage on chromosome 7q and 16p, genotyping several hundred markers across the genome. Another study by the Paris ASD Research International Sibpair Study of about 300 microsatellite markers in 51 families, found nominally significant ($p < 0.05$) linkage on chromosomes 2,4,5,6,7,10,15,16,18,19, & X, including the 7q and 16p previously associated loci [18]. However, these studies were performed in the pre-genomic sequence and pre-Next Generation Sequencing era. They were thus limited in their ability to precisely identify loci of interest, and retrospectively are considered somewhat underpowered.

By the early 21st century, prevalence rates for ASD had increased due to changes in awareness, and diagnostic testing/standardization. Initially the rate was estimated as 1 to 6 per 1000 people [19] but current estimates are an order of magnitude higher, near 2.5% in the United States[20] with similar estimates globally [21]. By 2008, a number of genome wide association studies (GWAS) were being performed on patients part of the

Autism Genetic Resource Exchange (AGRE) [22]. By 2010, three GWAS results were published on AGRE data [23–25] using unrelated unaffected individual samples as controls. Few variants reached genome wide significance leading to the hypothesis that common variants may contribute to ASD as a convergence of a number of genetic aberrations of small effect size. Alternatively, this could be explained by common variation having a limited role and that a large proportion of heritability could be explained by rare single nucleotide (SNV) or copy number variants (CNV) of moderate to large effect [26]. Indeed, by 2012, five key publications undertook exome sequencing of autistic probands and unaffected siblings, and these reports highlighted the effect of rare and *de novo* variation on the disorder[27–31].

The evolving picture of ASD genetics is one of clear influence of heritability, but heterogeneous genetic effects, with some contribution of common variants of small effect size, as well as an increased burden of rare and *de novo* variants of larger effect size. In order to determine the actionable approach to be taken in research, investigators often turn to model organisms to better understand the biological function of genes, variants of which exhibit risk in the human population. Model organisms such as mice have sequenced genomes, a molecular toolkit for experimentation, and a complex central nervous system to probe effects of ASD risk genes on the brain. ASD model organism research has focused on mice, looking at disruptions to behavior and physiological function in the central nervous system. Although an "autistic mouse" is a difficult construct to define, the biological function of specific genes can be explored, with the ability to then make generalizable hypotheses about their function in ASD in humans.

Overview of Behavioral Phenotyping in Mouse Models of ASD

Mouse models of ASD fall into two broad classes: (a) spontaneous models where a particular inbred strain or strains of laboratory mouse show phenotypes which mimic some aspects of human ASD, and (b) environmental or genetic models where experimenters have specifically engineered a manipulation in an inbred strain of mouse which is already implicated in human ASD risk, and then set about characterizing its behavior. The most well known classes of the spontaneous models are the BALB/c and BTBR strains which show reduced sociability in several assays, repetitive behaviors such as repetitive self-grooming, as well as altered ultrasonic vocalizations compared to other strains such as the common C57BL6 strain [32–35]. While these strains have proven to be exemplars in characterization of certain behavioral deficits, it is hard to isolate the effect of specific genes in comparison to control animals since control animals (e.g. C57BL6) will come from a different genetic background entirely. Furthermore, the identity of certain genes which may be causal to observed traits in these strains may not necessarily share identity with those contributing to ASD risk in humans. Therefore, the approach of our laboratory, as well as others, has been to manipulate specific ASD risk genes and determine their correlation to behavioral changes with respect to littermate control animals that lack these manipulations.

Regardless of the way the model is constructed, investigators then turn to determining whether the behavioral domains overlapping the core abnormalities in ASD are altered in model animals. Recently our group has reviewed behavioral paradigms used to phenotype mouse models of ASD [36]. To briefly summarize, as ASD is defined by deficits to the domains of sociability, restricted or repetitive interests, and communication, laboratory experiments focus on these activities in animals. Mice are pro-social animals, and social

assays in mice include full contact assays between conspecifics where different behaviors are measured such as fraction of assay time spent sniffing or engaging in other aspects of social investigation such as displays of aggression, locomotor play, allogrooming, and time spent in close proximity [37, 38]. The laboratory of Jacqueline Crawley developed a highly utilized three-chambered social approach assay which is amenable to automated scoring, and in which a mouse is able to choose between spending time with a novel conspecific or a familiar conspecific, or time spent investigation a conspecific or an object [39]. Repetitive behaviors are associated with several mouse models of ASD risk genes, such as repetitive grooming to the point of causing self-abrasions in *Shank3*-deficient mice [40]. In addition to repetitive grooming, repetitive digging and nestlet shredding in the homecage can also be quantified [41, 42]. Mice exhibit natural preferences for novel experiences, so resistance to change behavior can be assessed by quantifying the lack of preference in certain exploratory behaviors even in the presence of novel stimuli. One such assay is the T-maze, in which a reward presented in one of the arms of the maze is moved between arms in subsequent sessions, and the tendency of the mice to perseverate in exploring an arm even in the presence of a rewarding stimulus in an alternate arm is quantified, as well as the overall amount of spontaneous alternation between arms [43]. Additionally, mice naturally poke their noses in holes to engage in olfactory exploration of their environment, and mazes with multiple holes, charged with rewarded olfactory stimuli, can also be used to quantify how mice perseverate in their exploration of some holes versus others, even in the presence of changing stimuli [44, 45]. Again, although a truly "autistic mouse" is not an easily definable construct, laboratory mice have a repertoire of social and locomotor behaviors which can be exploited to determine whether changes to behavior occur after manipulation of ASD risk genes.

In addition to abnormal sociability and restricted interests, the modeling of communication abnormalities has been afforded a large amount of attention in model systems.

Human language development does not exist in mice and cannot be explored. However, vocal communication behavior is conserved across mammalian taxa, most commonly as the rudimentary communication between infants and their caregivers [46]. As other mammals, including humans, neonatal mice vocalize when separated from their parent. This behavior is a simple form of communication which is hypothesized to have evolved to facilitate reunion with an infant's caretaker in species such as mammals and birds which show extended periods of parental care [47]. This behavior in mice is "ultrasonic" (higher than the range of human hearing, > 20 kHz) and is correlated to search and retrieval behavior on the part of the parent [48]. In addition to neonatal vocalization behavior, mice also exhibit this ultrasonic vocalization (USV) in playful encounters as juvenile mice and during non-aggressive social exploration in adult mice [49]. In adult mice vocalization has primarily been studied in the context of male & female mating encounters [50, 51] which can be elicited with female urine or an anesthetized female. In fully awake male and female interactions, it was long thought that only the male is vocalizing, however it has since been learned that females are also vocalizing during these encounters [52]. In all these cases, vocalization appears to serve as a social attractant, facilitating interaction [49]. Because of the correlation of USV to other interactive behaviors in mice such as caregiving during infancy and social investigation later in life, deficits to this behavior may be important indicators of poor development of the social communication apparatus in mice, and thus USV is often assayed in mouse models of ASD [45, 53–57]. In addition to these groups, our group discovered variants in the RNA-binding protein gene *CELF6* associated with ASD risk. This gene was identified in a study of genes showing enriched expression in serotonergic neurons, and knockout of this gene in mice resulted in pronounced reductions to neonatal USV [45]. This finding forms the motivation for this thesis work, and thus in a final overview section, I will provide some of the background to our studies on the *CELF6* gene and our initial findings in the USV behavioral system.

***Celf6* and its association to ASD and USV**

In 2013, our group published a report in which we profiled serotonergic neurons using Translating Ribosome Affinity Purification (TRAP), which was work that began at The Rockefeller University under the laboratory of Nathaniel Heintz and continued at Washington University in St. Louis [45]. The serotonergic system, as described above, had previously been implicated in the etiology of ASD. As a candidate system approach, our group used TRAP, a method for profiling genes enriched for expression in a specific cell types [58], to determine previously unidentified transcripts with enriched expression in these cells, with the goal being to then determine whether these enriched genes showed association with risk of ASD diagnosis. Thus unlike a full GWAS endeavor, to conserve statistical power, this approach narrows the search space by first choosing a high priority candidate cell type, profiling its enriched transcripts, and determining whether the human homologues of these transcripts show evidence of association, rather than testing across the entire genome. TRAP requires the exogenous expression of an epitope-tagged ribosomal protein (RPL10a), and in our work this was performed using EGFP-RPL10a expressed under the control of the promoter for the serotonin transporter gene *Slc6a4* in transgenic mice. This approach generated a candidate list of 174 genes showing enriched expression in serotonergic neurons, 147 of which had clear human homologues, and were distributed across 136 noncontiguous regions in the genome.

These 136 regions were tested for association of single nucleotide polymorphisms to ASD diagnosis in data from the AGRE, with two SNPs achieving significance after multiple testing correction across the narrowed genomic search space of these 136 regions relevant to the serotonergic system. One of these variants is located in the 5' untranslated region (5UTR) of the *CELF6* gene near an alternative transcriptional start site. CELF6 is part of the CUGBP and ELAV-like Family (CELF) of RNA binding proteins, which I have reviewed more extensively in Chapter 4. Aside CELF1 (CUGBP1) (associated with my-

otonic muscular dystrophy [59]), these proteins are on the whole poorly studied, however they may act as regulators of alternative splicing [60, 61] and well as mRNA stability and translation regulation [62, 63]. In addition to the common variant association, we also identified a rare inherited premature stop codon in this gene in one autistic patient and no deleterious mutations in controls.

Due to these findings, we generated a knockout mouse with global loss of *Celf6* gene expression, in order to characterize behavioral changes that may be relevant to ASD, and to further explore the physiological and molecular consequences of loss of CELF6 protein. In characterizing *Celf6*-null (*Celf6*^{-/-}) mice, we found that *Celf6*^{-/-} mice showed almost 2 fold reduction to neonatal USV calls when assayed at post-natal day 8 compared to wild type littermates and we have since validated these findings in several replication cohorts of animals. In adult animals, *Celf6*^{-/-} showed normal performance on general locomotor activity tasks. Although they performed similarly to wild type littermates in the three-chambered social approach task described above, in an exploratory holeboard assay (used to test resistance to change) we found their exploratory behavior was not potentiated by exposure to an attracting stimulus (chocolate chips, later replicated as well with sweetened condensed milk). This exposure in wild type animals enhanced overall nose-poking over baseline throughout the apparatus, which may indicate some exploratory and novelty seeking behavior in *Celf6*^{-/-} animals, particularly after potentiation by a rewarding stimulus, was blunted.

Aside the behavioral findings, we also found that *Celf6*^{-/-} mice showed decreased levels of serotonin, dopamine, and norepinephrine in the brain as assayed by mass spectrometry. Indeed, in addition to the serotonergic system, we went on to determine that the *Celf6* gene shows enhanced expression in all of these neurotransmitter systems in the brain, and not only the serotonergic system [64]. We initially hypothesized that the behavioral changes we observed would be associated with actions of the CELF6 protein

in one or more of these systems. Additionally, aside a single publication on alternative splicing promoted by CELF6 protein *in vitro* [61], the overall molecular function of CELF6 has not been characterized. Thus, I set forth to:

- determine whether loss of CELF6 to key regions in the brain (neuromodulatory neurotransmitter systems) could phenocopy behavioral changes observed in the global knockout animal, thereby further implicating specific cell types in the mechanism of CELF6's relationship to behavior
- identify the mRNA binding targets of CELF6 in the brain and begin to assess how it may regulate them.

Using the advanced molecular toolkit available for mice, including Cre-Lox recombination [65], as well as other advances such as cross-linking immunoprecipitation and next-generation sequencing [66], we began to (a) assess the USV behavior of animals conditionally lacking CELF6 in specific cell populations, as well as (b) further understand how CELF6 may act in the brain to regulate downstream mRNA molecules.

Summary and Dissertation Overview

In summary, we began our studies of the *CELF6* gene by identifying its association to ASD risk, and investigating how it may be associated with ASD-related phenotyping in mice. We determined that certain domains of behavior, namely early communication and exploratory behavior, were altered in *Celf6*-null animals. Chapters 2 and 3, making up Part 1 of this dissertation, examine USV data in *Celf6* knockout animals. In Chapter 2, I present previously unpublished data regarding USV in *Celf6* knockout mice as neonatal pups, as well as neonatal USV in mice lacking *Celf6* expression in serotonergic and dopaminergic neurons. In Chapter 3, I present my work, as published in the journal *Frontiers in Behavioral Neuroscience*, consisting of a meta-analytic study of the development

of neonatal pup USV. This study shows that as pups this behavior is relatively unpredictable from animal to animal, but that as animals age, each animal demonstrates more characteristic individual behavior in certain aspects of USV production.

Part 2 of this dissertation is concerned with my initial work characterizing the molecular function of the CELF6 protein. In Chapter 4, I review the known work on the function of CELF RNA binding proteins and I review how RNA binding protein-RNA interactions are characterized. In Chapter 5, I present my original work identifying CELF6 binding targets and my initial experiments to assay CELF6 function in how it regulates these targets using a powerful massively parallel approach. It is my hope that the research I have presented here lays an important foundation in the study of this protein, which is likely a key regulator of mRNAs in the brain. Thus far, my work has represented parallel lines of inquiry: correlations of protein presence or absence to changes in behavior, and investigations of molecular function. While it is not yet possible for me to propose a direct mechanism by which the molecular function of Celf6 leads to changes in behavioral outcome, future research on this protein in our group and others will certainly shed important light on how this gene and its downstream regulatory pathways ultimately influence the larger phenotype of behavior.

Chapter 1 References

1. Edition, F. *Diagnostic and statistical manual of mental disorders* (Am Psychiatric Assoc, 2013).
2. Folstein, S. & Rutter, M. Infantile Autism: A genetic study of 21 twin pairs. *Journal of Child Psychology and Psychiatry* **18**, 297–321 (Sept. 1977).
3. Rutter, M. The development of infantile autism. *Psychological Medicine* **4**, 147 (May 1974).
4. Bailey, A. *et al.* Autism as a strongly genetic disorder: evidence from a British twin study. *Psychological Medicine* **25**, 63 (Jan. 1995).
5. Lamb, J. A., Moore, J., Bailey, A. & Monaco, A. P. Autism: recent molecular genetic advances. *Human Molecular Genetics* **9**, 861–868 (2000).
6. DeLorey, T. M. & Olsen, R. W. GABA and epileptogenesis: comparing Gabrb3 gene-deficient mice with Angelman syndrome in man. *Epilepsy research* **36**, 123–132 (1999).
7. Cook Jr, E. H. *et al.* Linkage-disequilibrium mapping of autistic disorder, with 15q11-13 markers. *The American Journal of Human Genetics* **62**, 1077–1083 (1998).
8. Schain, R. J. & Freedman, D. X. Studies on 5-hydroxyindole metabolism in autistic and other mentally retarded children. *The Journal of pediatrics* **58**, 315–320 (1961).

9. Piven, J. *et al.* Platelet serotonin, a possible marker for familial autism. *Journal of autism and developmental disorders* **21**, 51–59 (1991).
10. Cook Jr, E. H. *et al.* Evidence of linkage between the serotonin transporter and autistic disorder. *Molecular psychiatry* **2**, 247 (1997).
11. Kluck, S. M., Poustka, F., Benner, A., Lesch, K.-P. & Poustka, A. Serotonin transporter (5-HTT) gene variants associated with autism? *Human molecular genetics* **6**, 2233–2238 (1997).
12. Mbarek, O. *et al.* Association study of the NF1 gene and autistic disorder. *American Journal of Medical Genetics Part A* **88**, 729–732 (1999).
13. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature genetics* **23**, 185 (1999).
14. Baker, P., Piven, J. & Sato, Y. Autism and tuberous sclerosis complex: prevalence and clinical features. *Journal of autism and developmental disorders* **28**, 279–285 (1998).
15. Cohen, I. L. Behavioral profiles of autistic and nonautistic fragile X males. *Developmental Brain Dysfunction* (1995).
16. Turk, J. & Graham, P. Fragile X syndrome, autism and autistic features. *Autism* **1**, 175–197 (1997).
17. Of Autism Consortium, I. M. G. S. A full genome screen for autism with evidence for linkage to a region on chromosome 7q. *Human Molecular Genetics* **7**, 571–578 (1998).
18. Philippe, A. *et al.* Genome-wide scan for autism susceptibility genes. *Human molecular genetics* **8**, 805–812 (1999).

19. Fombonne, E. Epidemiological surveys of autism and other pervasive developmental disorders: an update. *Journal of autism and developmental disorders* **33**, 365–382 (2003).
20. Xu, G., Strathearn, L., Liu, B. & Bao, W. Prevalence of Autism Spectrum Disorder Among US Children and Adolescents, 2014-2016. *JAMA* **319**, 81 (Jan. 2018).
21. Baxter, A. J. *et al.* The epidemiology and global burden of autism spectrum disorders. *Psychological medicine* **45**, 601–613 (2015).
22. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nature reviews genetics* **9**, 341 (2008).
23. Weiss, L. A., Arking, D. E., of Johns Hopkins & the Autism Consortium, G. D. P., *et al.* A genome-wide linkage and association scan reveals novel loci for autism. *Nature* **461**, 802 (2009).
24. Wang, K. *et al.* Common genetic variants on 5p14. 1 associate with autism spectrum disorders. *Nature* **459**, 528 (2009).
25. Anney, R. *et al.* A genome-wide scan for common alleles affecting risk for autism. *Human molecular genetics* **19**, 4072–4082 (2010).
26. Berg, J. M. & Geschwind, D. H. Autism genetics: searching for specificity and convergence. *Genome biology* **13**, 247 (2012).
27. O’Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246 (2012).
28. Iossifov, I. *et al.* De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
29. Neale, B. M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242 (2012).

30. Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237 (2012).
31. Chahrour, M. H. *et al.* Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS genetics* **8**, e1002635 (2012).
32. McFarlane, H. G. *et al.* Autism-like behavioral phenotypes in BTBR T+ tf/J mice. *Genes, Brain and Behavior* **7**, 152–163 (2008).
33. Scattoni, M. L., Ricceri, L. & Crawley, J. N. Unusual repertoire of vocalizations in adult BTBR T+ tf/J mice during three types of social encounters. *Genes, Brain and Behavior* **10**, 44–56 (2011).
34. Brodtkin, E. BALB/c mice: Low sociability and other phenotypes that may be relevant to autism. *Behavioural Brain Research* **176**, 53–65 (Jan. 2007).
35. Panksepp, J. B. *et al.* Affiliative Behavior, Ultrasonic Communication and Social Reward Are Influenced by Genetic Variation in Adolescent Mice. *PLoS ONE* **2** (ed Crusio, W.) e351 (Apr. 2007).
36. Maloney, S. E., Rieger, M. A. & Dougherty, J. D. Identifying Essential Cell Types and Circuits in Autism Spectrum Disorders. *International Review of Neurobiology*, 61–96 (2013).
37. Pellis, S. M. & Pasztor, T. J. The developmental onset of a rudimentary form of play fighting in C57 mice. *Developmental psychobiology* **34**, 175–182 (1999).
38. Cheh, M. A. *et al.* En2 knockout mice display neurobehavioral and neurochemical alterations relevant to autism spectrum disorder. *Brain research* **1116**, 166–176 (2006).

39. Moy, S. *et al.* Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice. *Genes, Brain and Behavior* **3**, 287–302 (2004).
40. Peça, J. *et al.* Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* **472**, 437 (2011).
41. Kane, M. J. *et al.* Mice genetically depleted of brain serotonin display social impairments, communication deficits and repetitive behaviors: possible relevance to autism. *PloS one* **7**, e48975 (2012).
42. Thomas, A. *et al.* Marble burying reflects a repetitive and perseverative behavior more than novelty-induced anxiety. *Psychopharmacology* **204**, 361–373 (2009).
43. Silverman, J. L., Yang, M., Lord, C. & Crawley, J. N. Behavioural phenotyping assays for mouse models of autism. *Nature Reviews Neuroscience* **11**, 490 (2010).
44. Moy, S. S. *et al.* Development of a mouse test for repetitive, restricted behaviors: relevance to autism. *Behavioural brain research* **188**, 178–194 (2008).
45. Dougherty, J. D. *et al.* The disruption of *Celf6*, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors. *Journal of Neuroscience* **33**, 2732–2753 (2013).
46. Ehret, G. Development of sound communication in mammals. *Advances in the Study of Behavior* **11**, 179–225 (1980).
47. Lummaa, V., Vuorisalo, T., Barr, R. G. & Lehtonen, L. Why cry? Adaptive significance of intensive crying in human infants. *Evolution and human behavior* **19**, 193–202 (1998).
48. D’Amato, F. R. & Populin, R. Mother-offspring interaction and pup development in genetically deaf mice. *Behavior genetics* **17**, 465–475 (1987).

49. Portfors, C. V. Types and functions of ultrasonic vocalizations in laboratory rats and mice. *Journal of the American Association for Laboratory Animal Science* **46**, 28–34 (2007).
50. Holy, T. E. & Guo, Z. Ultrasonic Songs of Male Mice. *PLoS Biol* **3**, e386 (Nov. 2005).
51. Hammerschmidt, K., Radyushkin, K., Ehrenreich, H. & Fischer, J. Female mice respond to male ultrasonic songs with approach behaviour. *Biology letters* **5**, 589–592 (2009).
52. Neunuebel, J. P., Taylor, A. L., Arthur, B. J. & Egnor, S. R. Female mice ultrasonically interact with males during courtship displays. en. *eLife* **4**, e06203 (May 2015).
53. Maloney, S. E. *et al.* Characterization of early communicative behavior in mouse models of neurofibromatosis type 1. *Autism Research* **11**, 44–58 (2018).
54. Peñagarikano, O. *et al.* Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* **147**, 235–246 (2011).
55. Tsai, P., Chu, Y., Regehr, W., Sahin, M., *et al.* Autistic-Like Behavior and Cerebellar Dysfunction in Purkinje Cell Tsc1 Mutant Mice (S15. 001). *Neurology* **80**, S15–001 (2013).
56. De Filippis, B., Ricceri, L. & Laviola, G. Early postnatal behavioral changes in the Mecp2-308 truncation mouse model of Rett syndrome. *Genes, Brain and Behavior* **9**, 213–223 (2010).
57. Yang, M. *et al.* Reduced excitatory neurotransmission and mild autism-relevant phenotypes in adolescent Shank3 null mutant mice. *Journal of Neuroscience* **32**, 6525–6541 (2012).
58. Doyle, J. P. *et al.* Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135**, 749–762 (2008).

59. Wang, G.-S., Kearney, D. L., De Biasi, M., Taffet, G. & Cooper, T. A. Elevation of RNA-binding protein CUGBP1 is an early event in an inducible heart-specific mouse model of myotonic dystrophy. *The Journal of clinical investigation* **117**, 2802–2811 (2007).
60. Han, J. & Cooper, T. A. Identification of CELF splicing activation and repression domains in vivo. *Nucleic acids research* **33**, 2769–2780 (2005).
61. Ladd, A. N., Nguyen, N. H., Malhotra, K. & Cooper, T. A. CELF6, a member of the CELF family of RNA-binding proteins, regulates muscle-specific splicing enhancer-dependent alternative splicing. *Journal of Biological Chemistry* **279**, 17756–17764 (2004).
62. Louis, I. V.-S. & Bohjanen, P. R. Coordinate regulation of mRNA decay networks by GU-rich elements and CELF1. *Current opinion in genetics & development* **21**, 444–451 (2011).
63. Barreau, C., Paillard, L., Méreau, A. & Osborne, H. B. Mammalian CELF/Bruno-like RNA-binding proteins: molecular characteristics and biological functions. *Biochimie* **88**, 515–525 (2006).
64. Maloney, S. E., Khangura, E. & Dougherty, J. D. The RNA-binding protein Celf6 is highly expressed in diencephalic nuclei and neuromodulatory cell populations of the mouse brain. *Brain Structure and Function* **221**, 1809–1831 (2016).
65. Nagy, A. Cre recombinase: the universal reagent for genome tailoring. *genesis* **26**, 99–109 (2000).
66. Darnell, R. B. HITS-CLIP: panoramic views of protein - RNA regulation in living cells. *Wiley Interdisciplinary Reviews: RNA* **1**, 266–286 (2010).

Chapter 2

Characterization of deficit to ultrasonic vocalization in *Celf6* knockout mice

Abstract

We used Translating Ribosome Affinity Purification to identify genes whose expression is enriched in serotonergic neurons. The genomic loci containing these genes were assayed for associated to ASD diagnosis in case and control samples from the AGRE. A common variant in *CELF6* was found associated with ASD in the 5' untranslated region, as well as a premature stop codon observed in one autistic patient and no controls. Because of this association, we made a knockout mouse with global loss of CELF6 expression and subjected it to ASD behavioral phenotyping. *Celf6*^{-/-} mice showed >2-fold reduction to neonatal ultrasonic vocalizations compared to wild-type animals. Additionally, I assessed the temporal and spectral nature of the ultrasonic calls of *Celf6*^{-/-} mice. I found that the same reduction to total numbers of calls was also observed in total numbers of bouts. Differences in intercall pause times were not

detected, nor interbout intervals, nor call duration compared to wild-type littermates. *Celf6*^{-/-} and *Celf6*^{+/+} mice showed similar usage of the frequency spectrum and also showed similar proportions of calls containing jumps in pitch. Because CELF6 is enriched for expression in neuromodulatory neurotransmitter populations, I sought to determine whether cell-type specific knock-out in dopaminergic (DA) or serotonergic (5HT) neurons was sufficient to result in the same effect in vocalization as the global null mouse. I used a Cre-Lox recombination strategy to spatially restrict CELF6 loss to DA or 5HT neurons and confirmed loss of protein expression with immunofluorescent staining. Upon assay for vocalization in pups, however, mice with loss of CELF6 in DA or 5HT neurons did not reproduce the findings of the global null, but showed similar vocalization levels to CELF6+ littermates. Taken together, *Celf6*^{-/-} animals show reductions to vocalization but do not show gross changes to temporal structure or usage of the frequency spectrum. Future work on this protein will be needed to determine the population or populations which may be mediating reductions to vocalization in *Celf6*^{-/-} mice.

Introduction

As described in the preceding section, in order to narrow the search space for ASD-associated polymorphisms, we used a candidate cell type approach by profiling targeted a targeted cell type of interest. Cell types already implicated in ASD include the GABAergic system [1, 2] as well as the serotonergic system [3, 4]. By profiling the serotonergic system [5], we identified 174 transcripts showing enriched expression in these cells compared to the rest of the brain, using Translating Ribosome Affinity Purification. When testing these genomic loci for polymorphisms associated with ASD, we identified associ-

ations in the *CELF6* gene. We proceeded to make a knockout model of this gene and assess whether there were notable changes to behavior in an battery of behavioral tests related to ASD.

The *CELF6* gene codes for a protein which is a member of the CUGBP- and ELAV-Like Factor family of proteins containing 6 members CELF1-6. These proteins interact with RNA via RNA Recognition Motif (RRM) binding domains. Family members exhibit diverse regulatory capacities: regulating alternative splicing, mRNA degradation, and mRNA translation [6]. I discuss previous research on this family of proteins in greater detail in Chapter 4 of this dissertation. To our knowledge, our group's report on CELF6 in 2013 [5] is one of few studies linking a CELF protein to some change in behavior, in addition to CELF4 which has been shown to be associated with disruption of synaptic function and seizures [7, 8]. I found that *Celf6*^{-/-} mice showed reductions to neonatal ultrasonic vocalization (USV) [5].

USV is emitted by mice throughout the lifetime under different social contexts: as pups when separated from their caregiver[9], as juvenile mice during playful encounters [10], and as adults in both same sex and mating associated encounters [11]. Therefore, because USV is part of the normal repertoire of mouse social behavior, and as it is readily amenable to automated analysis pipelines [12, 13], it has thus become a standard phenotyping protocol across labs studying disruptions in ASD phenotyping. In this chapter, I present previously unpublished data characterizing the neonatal vocalization of *Celf6*^{-/-} mice. I found that although the overall rate of calling of these mice was reduced compared to wild-type littermates, there were not overt changes to the temporal and spectral structure of these calls. Additionally, because I aimed to characterize the cell populations which may be responsible for the deficit observed in knockout animals, I used a Cre-Lox recombination strategy to remove *Celf6*^{-/-} expression from serotonergic (5HT) and dopaminergic (DA) neurons. I found that deletion in either population was insufficient to

recapitulate the reduction to USV observed in the global null animal. Taken together, the *Celf6*^{-/-} phenotype is largely in the overall volume of response to maternal separation which is reduced in the knockout. Beyond this reduction in behavior, these animals are capable of emitting the same kinds of sounds as their wild-type littermates. We have so far been unable to identify a single cell type which can be more narrowly correlated to this change in behavior and future work will be needed to identify which cell subpopulation(s) or circuit(s) are responsible.

Results

***Celf6*^{-/-} mice show reductions to overall rate of ultrasonic calling when separated from their mothers.**

Ultrasonic vocalization emission is not a continuous phenomenon but rather is emitted in bouts of calling separated by longer pauses. An example trace demonstrating this structure is shown in Figure 1A. Empirically, a lower bound for the minimum pause between bouts can be established by inspecting the histogram of all pause times, which I use here as 300 msec. Calls vary in duration on the order of 20-100 msec, with pauses between these calls within bouts between 100-200 msec.

Total numbers of calls were reduced 2.4 fold in *Celf6*^{-/-} animals compared to wild-type animals ($p < 0.002$) and 2.8-fold reduced ($p < 0.02$) compared to *Celf6*^{+/-} animals (Figure 1B). When calls were segmented into bouts, we found a ≈ 2 -fold reduction in *Celf6*^{-/-} animals compared to wild-type (1.97-fold, $p < 0.03$) and *Celf6*^{+/-} animals (2.1-fold, $p < 0.008$) (Figure 1C). In either case (total calls or total numbers of bouts) we did not detect differences between wild-type and heterozygote animals, indicating that this phenotype did not

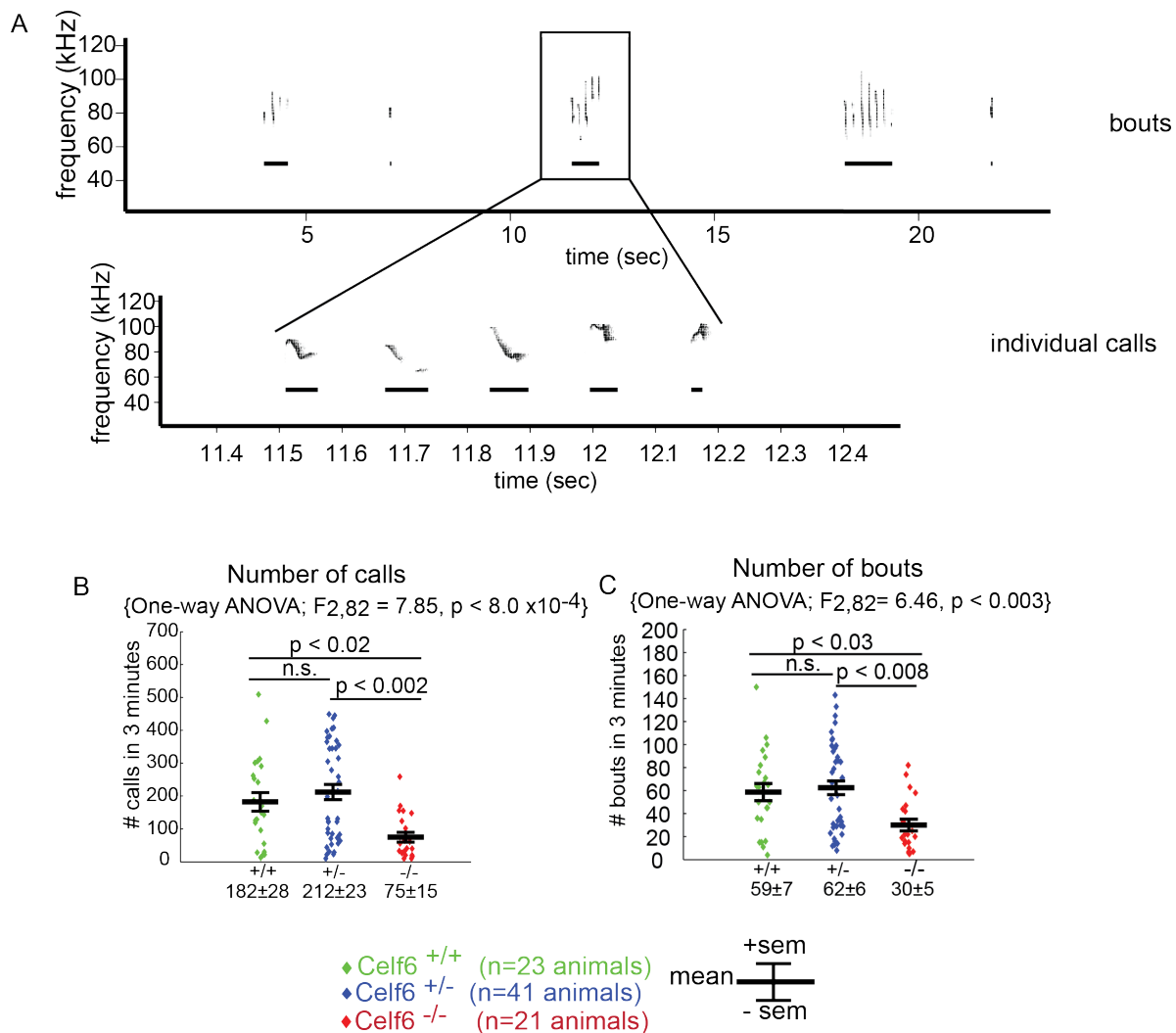


Figure 1: *Celf6* knockout mice show reductions to overall rate of ultrasonic calling when separated from their mothers.

(A) Example spectrogram showing a recording of neonatal pup USV on post-natal day 8. In lower pane, a single bout of USV is zoomed it to show the scale of individual calls. (B) Total numbers of calls per 3 minute recording in *Celf6* wild-type (N=23), heterozygote (+/-, N=41) and knockout (-/-, N=21) animals. Data points are shown with means \pm standard error. (C) Total numbers of calling bouts (minimum interbout gap length of 0.3 msec). Statistical testing is shown as ANOVA testing for main effect of genotype, with post hoc pairwise comparisons performed using t-tests with Bonferroni adjustment for multiple testing.

appear sensitive to *Celf6* gene dosage.

Although overall calling rate and number of bouts per recording are necessarily correlated features of vocalization, it might be possible that lowered numbers of total calls were still separated into an equal number of bouts per recording. This would be the case, for example, if the average pause length between calls were lengthened such that each bout were composed of fewer calls. A lack of correlation between call rate and bout rate then would indicate some change to temporal structure. Because segmentation of calls into bouts depends on the temporal structure of the calls and the intervals between them, I next set about analyze temporal features of vocalization, in addition to spectral features of calls produced by *Celf6*^{-/-} .

Celf6^{-/-} knockout mice show similar temporal and spectral features of vocalization to wild-type mice

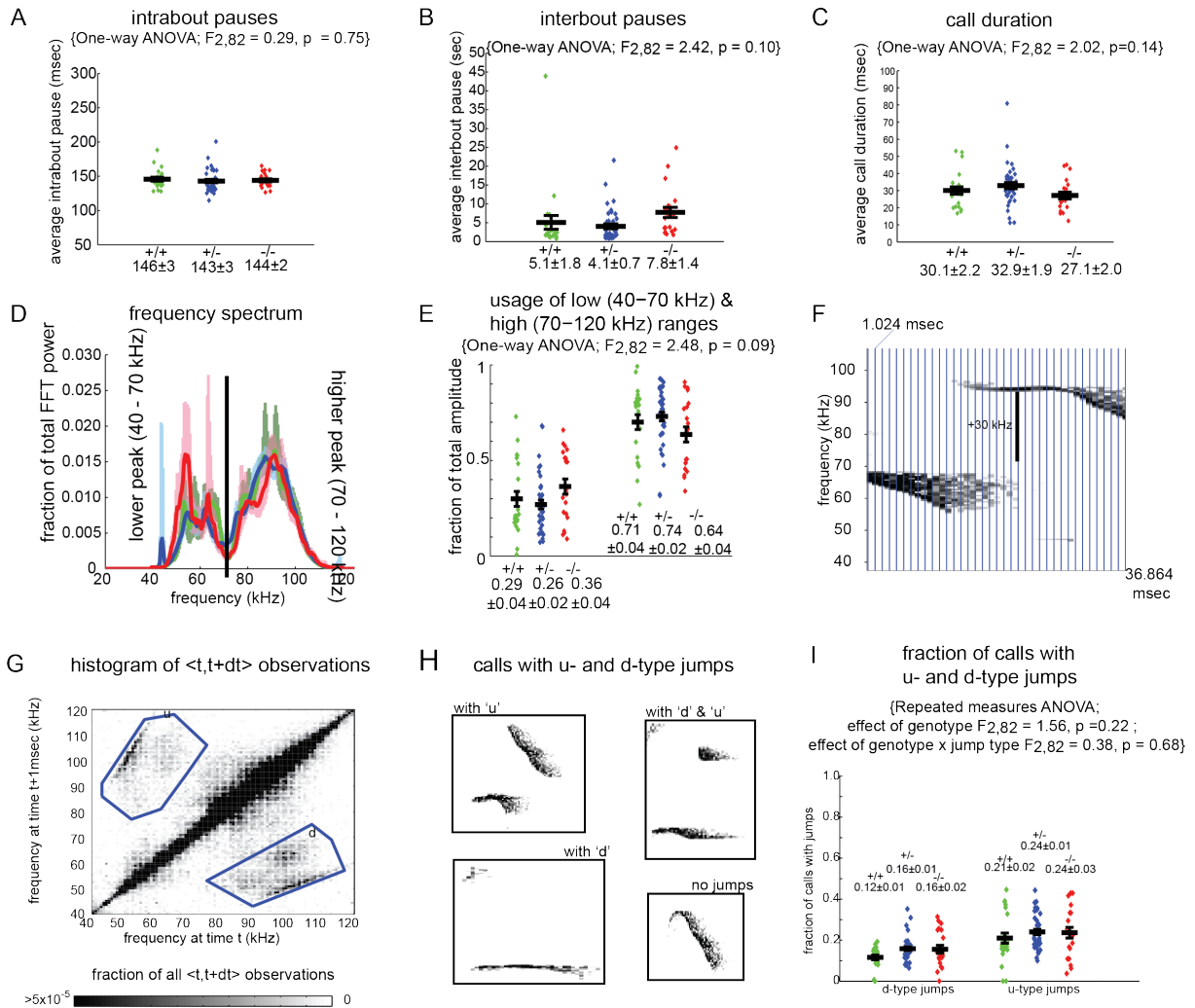


Figure 2: Spectral and temporal properties of ultrasonic calls are unperturbed in *Celf6* knockout mice.

As in Figure 1, all data are shown color-coded according to group: green: *Celf6*^{+/+} (N=23), blue: *Celf6*^{+/-} (N=41), red: *Celf6*^{-/-} (N=21). (A) Pauses between calls within bouts ("intra bout", < 0.3 msec), (B) Gaps between bouts ("interbout"), (C) Call duration. For features of USV measured in (A)-(C), data point represent the average across all calls per animal. All animals showed at least 10 calls per recording. (D) Frequency spectrum. Power summed over all calls was computed as FFT amplitude² and normalized to the sum across the spectrum. Spectra are shown as means ± standard error. (E) Area under each peak in (D) was summed and compared across groups. Because one sum is simply the exact complement of the other sum, statistical testing was only performed for the low peak data (40-70kHz). (F) Example call showing a jump in pitch. (G) In order to classify pitch jumps as either upward or downward, the two-dimensional histogram of all frequency transitions <t,t+1> was computed. "u" and "d" clusters were manually defined using ROI tools in MATLAB. (H) Calls were classified as containing at least one u-type jump, at least one d-type jump, both, or none. (I) The proportion of calls containing at least one u-type jump or one d-type jump is shown per animal. In cases (A-C) and (E), main effects of genotype were determined using one-way ANOVA. In (I) main effects of genotype, and genotype x call type interaction were computed using repeated measures ANOVA.

Temporal and spectral features of vocalization describe the repertoire of sounds that animals are able to produce in addition to acting as proxy measurements for the physical functioning of structures necessary in the vocalization apparatus. For example, in collaborative work with the laboratory of Paul Gray, I have found that pause time within bouts of vocalization is associated to the function of hind brain motor neurons controlling rhythmic breathing [14]. The relationship of intercall pause intervals and the duration of subsequent calls is also related to the respiratory apparatus of other species, e.g. the muscular output of the airsacs of zebra finches [15]. Additionally, the temporal nature of vocalization is related to its role in communication, as the duration of calls must exist within certain parameters to effectively elicit maternal search and retrieval behavior [16]. To determine whether these aspects of USV in *Celf6*^{-/-} mice was altered, I looked at pauses and duration times of calls and these data are shown in Figure 2. Pauses were separated into short pauses between calls of bouts of behavior ("intra bout pauses", 2A) and longer gaps between bouts ("interbout pauses", 2B). Pauses and duration times (Fig-

ure 2C) were averaged across all calls per animal, with all animals considered producing at least 10 calls per recording. No significant effects of genotype were detected across all features. Intra bout calls were on average 144 msec, with interbout gap times at about 6 seconds. Intra bout pauses exhibited a tight distribution (2A), while interbout pauses ranged to tens of seconds.

The neurons responsible for controlling the spectral content of vocalization are not known, however it has been shown that mice are able to discriminate between calls of different spectral subtypes [17], and populations in the auditory cortex and inferior colliculus respond to USVs of different frequency composition [18–20]. In the periphery, laryngeal muscle contraction is most proximally involved with the ability to produce sounds of differing frequency, as high frequency requires maintenance of a narrow aperture. Mice with transections to the laryngeal nerve cannot vocalize as well as mice with intact control over this musculature [21]. Thus producing ultrasound, as measured in the frequency spectrum, may act as a proxy for normal laryngeal function. In order to examine whether *Celf6*^{-/-} produced vocalizations spectrally similar to their wild-type and heterozygote littermates, I estimated usage of the spectrum across each recording by computer power per frequency across all calls (Figure 2D). Average traces largely overlapped and I was able to identify two main bands, between 40-70 kHz and 70-120 kHz. I quantified these bands as an area under the curve for all groups (Figure 2E) but did not detect significant differences (Figure 2E).

USV stereotypically presents with a number of calls containing instantaneous jumps in pitch (Example in Figure 2F, where I define pitch as frequency containing maximum power over time) and this has been used directly to classify calls [12, 22] as well as forming part of the criteria for other classification schemes [23]. I used the method of [12] by quantifying pairwise pitch estimates at time t and $t+1$ across all recordings (Figure 2G). Most pitch transitions are between the same values of frequency and exist along a diagonal, but

clear clusters of larger transitions upward and downward in pitch were definable. These were then used to classify calls as containing upward (u) or downward (d) changes in pitch, both directions, or no jumps. The fraction of calls per animal containing u or d-type changes is shown in (Figure 2H). Across all four classes, I did not detect any significant changes to calls containing pitch jumps of different kinds. Taken together, these results imply that the primary change in *Celf6*^{-/-} mice is reduction to the overall amount of response to maternal separation, measured as total calls or total bouts of vocalization, but that otherwise these animals are capable of producing the same kinds of calls as their wild-type and heterozygote littermates.

Mice with CELF6 loss in 5HT and DA neurons fail to show reductions to USV

Our group first identified CELF6 as exhibiting enriched expression in serotonergic neurons, but with follow up immunohistochemical work, members of our group found that it was enriched in several other populations including populations of the hypothalamus, the 5HT, DA, and noradrenergic (NE) monoaminergic neurotransmitter populations, as well as a subpopulation of neurons in the lateral habenula (likely Substance P+ cells), with weaker expression elsewhere[24]. Known markers for these populations can be used to drive expression of Cre recombinase for Cre-Lox mediated excision [25] in specific cell types. Thus we used this approach with a *Celf6*^{loxP/loxP} mouse which has loxP sites flanking a constitutive exon [5].

5HT neurons can be reliably marked by specific genes, such as those in the serotonin biosynthesis pathway (e.g., *Tph2*, *Ddc*) or the serotonin transporter (*Slc6a4*). However, genes such as *Slc6a4* are not unique to serotonergic neurons during early post-natal

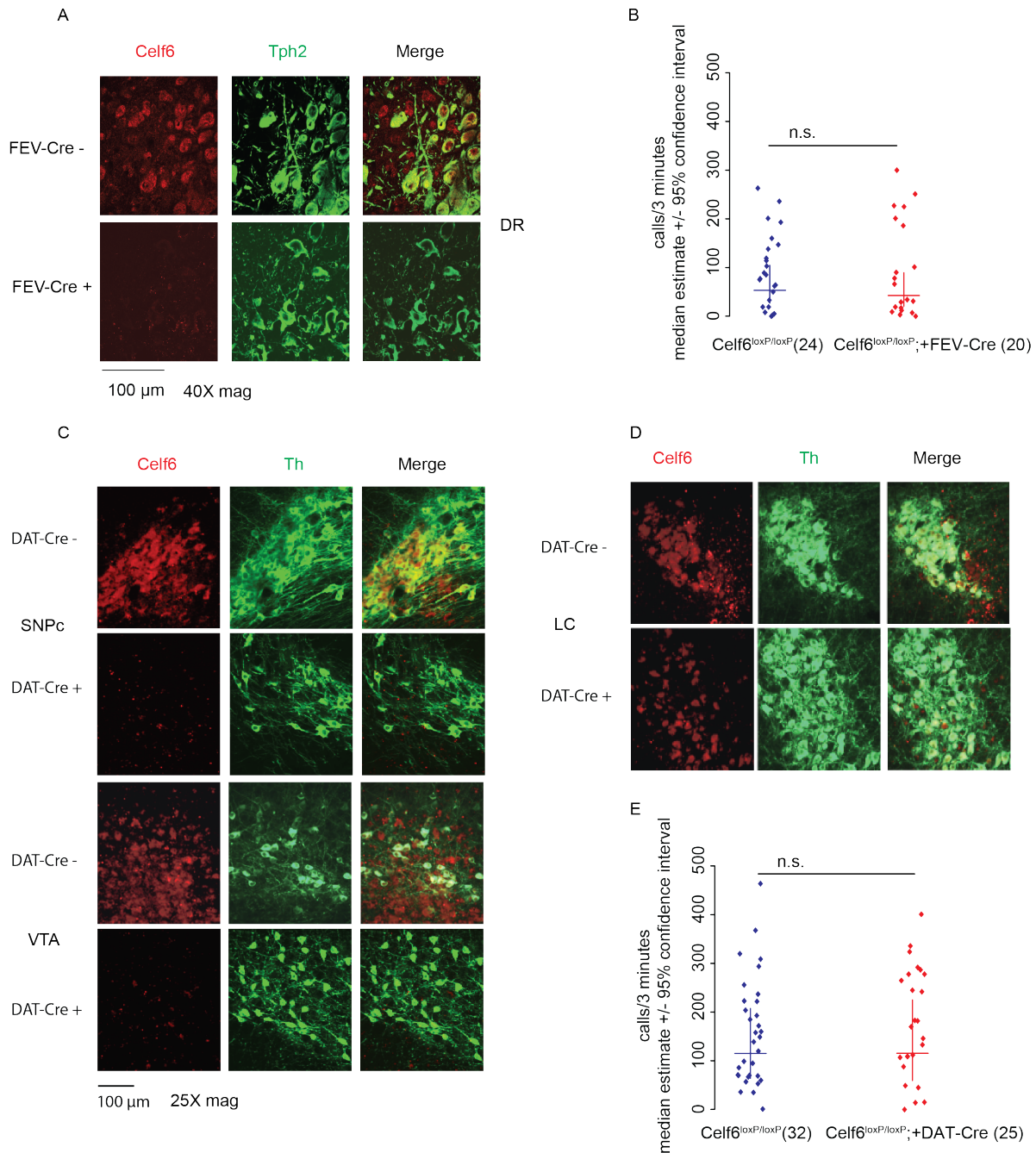


Figure 3: Mice with CELF6 loss in 5HT and DA neurons fail to show reductions to USV. (A) Conditional loss of CELF6 in *Celf6*^{loxP/loxP}; FEV-Cre+ animals. 40X confocal images are shown from 40 μ m slices of *Celf6*^{loxP/loxP}; FEV-Cre- or FEV-Cre+ animals stained with anti-CELF6 (red) and anti-TPH2 (green) antibodies in the dorsal raphe (DR) nucleus. (B) Post-natal USV measured from FEV-Cre- (24) and FEV-Cre+ (20) *Celf6*^{loxP/loxP} animals. (C) Conditional loss of CELF6 in *Celf6*^{loxP/loxP}; DAT-Cre+ animals. 20X confocal images are shown from 40 μ m slices of *Celf6*^{loxP/loxP}; DAT-Cre- or DAT-Cre+ stained with anti-CELF6 (red) or anti-TH (green) antibodies in the substantia nigra pars compacta (SNPc) and the ventral tegmental area (VTA). (D) Staining in locus coeruleus (LC) as in (C) where no conditional loss observed. (E) Post-natal USV measured from DAT-Cre- (32) and DAT-Cre+ (25) *Celf6*^{loxP/loxP} animals. Staining experiments in (A),(C),(D) are representative of 3 animals per group (Cre+ or Cre-). No significant changes to calls produced in 3 minutes were detected in (C),(E).

development when we are measuring USV but are observed elsewhere[26]. The FEV (also Pet1, ePet) transcription factor specifies the development of the raphe nuclei of the hindbrain during the late embryonic stage and maintains expression after birth [27]. Thus I employed transgenic mice expressing Cre recombinase under the control of the *FEV* promoter to drive *Celf6* knockout in 5HT neurons. Conditional loss of *Celf6* in raphe neurons was confirmed by immunofluorescent staining on 40 μm sectioned brain tissue with anti-*Celf6* and anti-Tph2 antibodies followed by confocal fluorescent microscopy. These results are shown in Figure 3A.

I recorded USV from Cre- and Cre+ *Celf6*^{loxP/loxP} pups at post-natal day 8. I did not detect an effect of genotype (Figure 3B, $p = 0.9237$) and USV call rate in both genotypes was comparable (*Celf6*^{loxP/loxP}; FEV-Cre+ 42 calls/3minutes, *Celf6*^{loxP/loxP};Cre- 53). It is worth noting that these population estimates are nearly 3-fold lower than the wild-type controls for the global knockout animal (average calls/3minutes = 182), which is likely due to intercross with the background of the FEV-Cre animals. Thus, despite our original description of CELF6 as demonstrating enriched expression in serotonergic neurons, conditional knockout in serotonergic neurons alone is insufficient to recapitulate the USV phenotype observed in the global null.

I next looked to DA neurons and used DAT-Cre transgenic animals, expressing Cre recombinase under the control of the dopamine transporter gene promoter (*DAT*) to drive recombination in dopaminergic neurons [28]. Confirmation of conditional knockout using immunofluorescent staining with anti-CELF6 and anti-Th (marking DA neurons) antibodies is shown in Figure 3C. Loss of CELF6 is shown for both DA neurons of the substantia nigra pars compacta (SNPc, upper panels) and the ventral tegmental area (VTA, lower panels) for Cre- and Cre+ animals demonstrating loss of CELF6 expression in these cells. *Th*, or tyrosine hydroxylase, is the first enzyme in the biosynthesis pathways for both DA and NE, and *Celf6* is expressed in both DA (e.g., VTA, SNPc, and smaller populations

of the hypothalamus) and NE neurons (in the locus coeruleus (LC)). Thus in Figure 3D using the same anti-Th stain to mark the noradrenergic population of the LC in the dorsal hindbrain, we show that loss of CELF6 mediated by DAT-Cre recombination is specific to the DA neurons and that the LC retains CELF6 expression.

After measuring USV in these animals I again did not detect an effect of genotype ($p \approx 1$) with comparable estimates for calls/3minutes in both *Celf6*^{loxP/loxP};DAT-Cre+ (115 calls/3minutes) and *Celf6*^{loxP/loxP};Cre- animals (116 calls/3minutes) indicating that both genotypes behaved similarly. We also measured USV in these animals on additional post natal days (5 & 9) and did not find significant differences on any days (See Chapter 3, Supplemental Figure S1). The negative data from this study were used as part of the meta-analytic study on variability in USV across early development in the following chapter. Taken together, we were unable to isolate the USV phenotype observed in *Celf6*^{-/-} to either 5HT or DA neurons.

Discussion

The ultrasonic vocalization of neonatal mouse pups has become a standard task in behavioral characterization of mouse models of ASD and other neurodevelopmental disorders [9]. Our group's publication in 2013 of reductions to vocalization in *Celf6*^{-/-} mice is expanded here to include my analysis of other aspects of USV in these mice. USV is a data rich behavior, containing information on how the animal makes use of the frequency spectrum, how the temporal sequences of calls are composed, and how much behavior overall is elicited by the stimulus (e.g., maternal separation). We have shown that *Celf6*^{-/-} mice have reduced levels vocalization produced in response to maternal separation. Because emission of USV is correlated to maternal search and retrieval behavior [29], deficits to USV may impact the communication between pups and their mothers leading to altered

levels of maternal care. We have not explicitly tested this in *Celf6*^{-/-} mice, however it will be interesting to see whether pups producing less vocalization in this model indeed receive less attention from their caregivers. In addition to USV, throughout the post-natal period, we routinely test mice for other developmental milestones such as development of the righting reflex [30] as well as pinna detachment of the ears, eye opening, and weight gain. We have not found any differences during the post-natal period, suggesting that other than the deficit to USV, *Celf6*^{-/-} are developing similarly to their wild-type littermates.

My strategy to localize the USV behavior of *Celf6*^{-/-} mice has been to use Cre-Lox recombination to mediate cell-type specific knockout. This has not yet yielded further narrowing of the phenotype to a specific population of cells. Current research in our group is employing multiple strategies to remove CELF6 across all the monoaminergic populations of cells simultaneously. When I began my knockout studies, the full extent of CELF6 protein expression was not yet known. Given a more current understanding of where in the brain CELF6 is expressed, an alternate strategy may be to use Cre-Lox to remove CELF6 from whole regions of the developing brain, such as the diencephalon or the hindbrain . Such strategies may help better define the subregions of the brain responsible for driving the phenotype in *Celf6* knockout mice.

Methods

Animals

All protocols involving animals were approved by the Animal Studies Committee of Washington University in St. Louis. Original cohort published in Dougherty et al. 2013 study was generated by breeding *Celf6*^{+/-} X *Celf6*^{+/-} crosses. For conditional knockout studies, *Celf6*^{loxP/loxP} animals were first crossed to either FEV-Cre (Jacksons Laboratory Strain

B6.Cg-Tg (Fev-cre)1Esd/J) or DAT-Cre (Jackson Laboratory strain B6.SJL -Slc6a3^{tm1.1(cre)} Bkmn/J) animals to generate *Celf6*^{loxP/loxP} to generate *Celf6*^{loxP/loxP}; Cre⁺ and Cre⁻ littermates. Cages were maintained by our facility on a 12 hr : 12 hr light:dark schedule with food and water supplied ad libidum. Animals were genotyped from toe clip tissue lysed by incubation at 50°C in Tail Lysis Buffer (0.5 M Tris-HCl pH 8.8, 0.25 M EDTA, 0.5% Tween-20) containing 4 μ L/mL 600 U/mL Proteinase K enzyme (EZ BioResearch) for 1 hour to overnight, followed by heat denaturation at 99°C for 10 minutes. Crude lysate containing DNA was genotyped using Quickload Taq Mastermix (New England Biolabs) with the following cycling conditions: 94°C 1 min, (94°C 30 s, 60 °C 30 s, 68°C 30 sec) x 30 cycles, 68°C 5 minutes, 10°hold. *Celf6*^{-/-} animals were genotyped with:

- B6_J_geno_3U: 5'-CCCTGCCACCTAGCTCTTCAGGTT-3',
- BRUNOL KO 3': 5' ATGGCTGAGCTCTTTCTTGAGAAGTAC-3'

primers. The wild-type allele is detected as a 415 bp PCR product and the knockout as a 188 bp product. *Celf6*^{loxP/loxP}; Cre⁺ animals were genotyped using: B6_J_geno_3U and BRUNOL KO 3' primers above which yield a 500 bp product for the flanked-by-loxP allele. Cre recombinase was genotyped using:

- Cre-F: 5'-CCGGTCGATGCAACGAGTGATGAGGTTC-3' and
- Cre-R 5'-GCCAGATTACGTATATCCTGGCAGCG-3'

primers yielding a 450 bp product and using primers amplifying *Actb* as an internal control for PCR failure.

USV recording

USVs were recorded on post-natal day 8 or 7 (DAT-Cre cohort) using a CM16 microphone and RECORDER software (Avisoft Bioacoustics). Full description of USV recording protocol is provided in **Chapter 3 - Methods**. Mice in original global *Celf6*^{-/-} cohort and

crosses to FEV-Cre were measured on post-natal day 8. Mice in crosses to DAT-Cre were measured on post-natal day 7 and not post-natal day 8 due to changing procedures in our laboratory, where we have standardized USV measurements across multiple time points in development. The USV deficit in *Celf6*^{-/-} mice has been confirmed as well for post-natal day 7 (Maloney et al., manuscript in preparation).

USV processing

Acquired USV recordings were processed using a custom pipeline in MATLAB based on code authored by Timothy Holy[12]. Description of this pipeline is given in **Chapter 3 - Methods** and a tutorial in its usage is given in **Appendix I - Using the MATLAB VocalizationFunctions Package**.

Immunofluorescent staining

Mice on post-natal day 21 were euthanized by CO₂ according to protocols approved by the Animal Studies Committee of Washington University in St. Louis and in accordance with NIH guidelines. Mice were subjected to transcardial perfusion with 1X PBS for 5 minutes, followed by 4% paraformaldehyde in 1X PBS for 7 minutes. After perfusion, brains were dissected and post-fixed in 4% paraformaldehyde overnight, followed by progressive sucrose protection in 5% sucrose/1X PBS for 1 hour, 15% sucrose/1X PBS overnight, and 30% sucrose/1XPBS overnight. Brains were frozen in Neg-50 medium (Thermo) and sectioned on a Leica cryostat as 40 μm floating sections in 1X PBS. Free floating sections were blocked in 0.25% Triton-X 100/1X PBS (PBST) supplemented with 10% donkey serum (Jackson ImmunoResearch), and then incubated overnight in PBST/10% donkey serum containing 1:1000 diluted mouse anti-Th (Millipore MAB318) or anti-Tph2 (Thermo) antibodies and 1:250 diluted custom rabbit anti-CELF6 antibody (#176, as described in [5]) overnight at room temperature. Sections were washed 3x10 minutes in

PBST, and then incubated for 1.5 hours at room temperature in PBST/10% donkey serum containing secondary antibodies. 5HT conditional knockout used 1:500 Alexa Fluor 488 anti-Rabbit secondary for CELF6 staining and Alexa Fluor 546 anti-Mouse secondaries for anti-Tph2 costain. We subsequently observed better resolution of CELF6 staining using more spectrally distant fluorophores. For DA conditional knockout we used 1:500 Alexa Fluor 647 (far red) anti-Rabbit antibody to mark CELF6 staining and 1:500 Alexa Fluor 488 anti-Mouse for anti-Th costain (Thermo). After incubation with secondary antibodies, slices were washed 3x10 minutes in PBST and then incubated for 10 minutes with 1XPBS, 1:30,000 DAPI, followed by 3x10 minute washes in PBS. Slices were mounted to slides and coverslips were applied using ProLong Gold antifade mounting medium (Thermo). Sections were imaged with a Perkin Elmer UltraView Vox spinning-disk confocal on a Zeiss Axiovert. Monochromatic images were adjusted for background signal in MATLAB and colored images were prepared as RGB TIFF format storing CELF6 signal in the red, and Th/Tph2 signal in green, and DAPI in blue.

Statistical Analysis

Main effects of genotype across USV features described were assessed using univariate linear models and analysis of variance (ANOVA) in MATLAB. Pairwise post hoc comparisons were performed with simple Student's t-test and adjusted for multiple comparisons by Bonferroni correction. Main effect of genotype and interaction of genotype with subclass of USVs containing upward and downward jumps in pitch were assessed by repeated measures ANOVA in MATLAB.

Chapter 2 References

1. DeLorey, T. M. & Olsen, R. W. GABA and epileptogenesis: comparing Gabrb3 gene-deficient mice with Angelman syndrome in man. *Epilepsy research* **36**, 123–132 (1999).
2. Cook Jr, E. H. *et al.* Linkage-disequilibrium mapping of autistic disorder, with 15q11-13 markers. *The American Journal of Human Genetics* **62**, 1077–1083 (1998).
3. Schain, R. J. & Freedman, D. X. Studies on 5-hydroxyindole metabolism in autistic and other mentally retarded children. *The Journal of pediatrics* **58**, 315–320 (1961).
4. Piven, J. *et al.* Platelet serotonin, a possible marker for familial autism. *Journal of autism and developmental disorders* **21**, 51–59 (1991).
5. Dougherty, J. D. *et al.* The disruption of Celf6, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors. *Journal of Neuroscience* **33**, 2732–2753 (2013).
6. Dasgupta, T. & Ladd, A. N. The importance of CELF control: molecular and biological roles of the CUG-BP, Elav-like family of RNA-binding proteins. *Wiley Interdisciplinary Reviews: RNA* **3**, 104–121 (2012).
7. Wagnon, J. L. *et al.* CELF4 regulates translation and local abundance of a vast set of mRNAs, including genes associated with regulation of synaptic function. *PLoS Genet* **8**, e1003067 (2012).

8. Wagnon, J. L. *et al.* Etiology of a genetically complex seizure disorder in *Celf4* mutant mice. *Genes, Brain and Behavior* **10**, 765–777 (Aug. 2011).
9. Branchi, I., Santucci, D. & Alleva, E. Ultrasonic vocalisation emitted by infant rodents: a tool for assessment of neurobehavioural development. *Behavioural Brain Research* **125**, 49–56 (Nov. 2001).
10. Scattoni, M. L., Ricceri, L. & Crawley, J. N. Unusual repertoire of vocalizations in adult BTBR T+ *tf/J* mice during three types of social encounters. *Genes, Brain and Behavior* **10**, 44–56 (2011).
11. Portfors, C. V. Types and functions of ultrasonic vocalizations in laboratory rats and mice. *Journal of the American Association for Laboratory Animal Science* **46**, 28–34 (2007).
12. Holy, T. E. & Guo, Z. Ultrasonic Songs of Male Mice. *PLoS Biol* **3**, e386 (Nov. 2005).
13. Burkett, Z. D., Day, N. F., Peñagarikano, O., Geschwind, D. H. & White, S. A. VoICE: A semi-automated pipeline for standardizing vocal analysis across models. *en. Scientific Reports* **5**. (2015) (May 2015).
14. Tupal, S. *et al.* Testing the role of preBötzing Complex somatostatin neurons in respiratory and vocal behaviors. *European Journal of Neuroscience* **40**, 3067–3077 (2014).
15. Schmidt, M. F. & Wild, J. M. The respiratory-vocal system of songbirds: anatomy, physiology, and neural control. *Progress in brain research* **212**, 297–335 (2014).
16. Ehret, G. Categorical perception of mouse-pup ultrasounds in the temporal domain. *Animal Behaviour* **43**, 409–416 (1992).
17. Neilans, E. G., Holfoth, D. P., Radziwon, K. E., Portfors, C. V. & Dent, M. L. Discrimination of ultrasonic vocalizations by CBA/CaJ mice (*Mus musculus*) is related to spectrotemporal dissimilarity of vocalizations. *PLoS one* **9**, e85405 (2014).

18. Shepard, K. N., Lin, F. G., Zhao, C. L., Chong, K. K. & Liu, R. C. Behavioral relevance helps untangle natural vocal categories in a specific subset of core auditory cortical pyramidal neurons. *Journal of Neuroscience* **35**, 2636–2645 (2015).
19. Liu, R. C., Miller, K. D., Merzenich, M. M. & Schreiner, C. E. Acoustic variability and distinguishability among mouse ultrasound vocalizations. *J Acoust Soc Am* **114**, 3412–3422 (Dec. 2003).
20. Liu, R. C. & Schreiner, C. E. Auditory Cortical Detection and Discrimination Correlates with Communicative Significance. *PLOS Biol* **5**, e173 (June 2007).
21. Nunez, A. A., Pomerantz, S. M., Bean, N. J. & Youngstrom, T. G. Effects of laryngeal denervation on ultrasound production and male sexual behavior in rodents. *Physiology & behavior* **34**, 901–905 (1985).
22. Arriaga, G., Zhou, E. P. & Jarvis, E. D. Of Mice, Birds, and Men: The Mouse Ultrasonic Song System Has Some Features Similar to Humans and Song-Learning Birds. *PLoS ONE* **7**, e46610 (Oct. 2012).
23. Scattoni, M. L., Gandhi, S. U., Ricceri, L. & Crawley, J. N. Unusual Repertoire of Vocalizations in the BTBR T+tf/J Mouse Model of Autism. *PLOS ONE* **3**, e3067 (Aug. 2008).
24. Maloney, S. E., Khangura, E. & Dougherty, J. D. The RNA-binding protein Celf6 is highly expressed in diencephalic nuclei and neuromodulatory cell populations of the mouse brain. *Brain Structure and Function* **221**, 1809–1831 (2016).
25. Nagy, A. Cre recombinase: the universal reagent for genome tailoring. *genesis* **26**, 99–109 (2000).
26. Lebrand, C. *et al.* Transient developmental expression of monoamine transporters in the rodent forebrain. *Journal of Comparative Neurology* **401**, 506–524 (1998).

27. Scott, M. M. *et al.* A genetic approach to access serotonin neurons for in vivo and in vitro studies. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 16472–16477 (2005).
28. Bäckman, C. M. *et al.* Characterization of a mouse strain expressing Cre recombinase from the 3' untranslated region of the dopamine transporter locus. *Genesis* **44**, 383–390 (2006).
29. Wöhr, M., Oddi, D. & D'Amato, F. R. Effect of altricial pup ultrasonic vocalization on maternal behavior. *Handbook of behavioral neuroscience* **19**, 159–166 (2010).
30. Fox, W. Reflex-ontogeny and behavioural development of the mouse. *Animal behaviour* **13**, 234–IN5 (1965).

Chapter 3

Analysis of within subjects variability across Celf6 conditional knockout studies: pups exhibit inconsistent, state-like patterns of call production

Adapted from original work in:

Rieger, M.A. and Dougherty, J.D., 2016. Analysis of within subjects variability in mouse ultrasonic vocalization: Pups exhibit inconsistent, state-like patterns of call production. *Frontiers in behavioral neuroscience*, 10, p.182.

Abstract

Mice produce ultrasonic vocalizations (USV) in multiple communicative contexts, including adult social interaction (e.g., male to female courtship), as well as pup calls when separated from the dam. Assessment of pup USV has been

widely applied in models of social and communicative disorders, dozens of which have shown alterations to this conserved behavior. However, features such as call production rate can vary substantially even within experimental groups and it is unclear to what extent aspects of USV represent stable trait-like influences or are vulnerable to an animal's state. To address this question, we have employed a mixed modeling approach to describe consistency in USV features across time, leveraging multiple large cohorts recorded from two strains, and across ages/times. We find that most features of pup USV show consistent patterns within a recording session, but inconsistent patterns across postnatal development. This supports the conclusion that pup USV is most strongly influenced by "state"-like variables. In contrast, adult USV call rate and call duration show higher consistency across sessions and may reflect a stable "trait". However, spectral features of adult song such as the presence of pitch jumps do not show this level of consistency, suggesting that pitch modulation is more susceptible to factors affecting the animal's state at the time of recording. Overall, the utility of this work is threefold. First, as variability necessarily affects the sensitivity of the assay to detect experimental perturbation, we hope the information provided here will be used to help researchers plan sufficiently powered experiments, as well as prioritize specific ages to study USV behavior and to decide which features to consider most strongly in analysis. Second, via the mouseTube platform, we have provided these hundreds of recordings and associated data to serve as a shared resource for other researchers interested in either benchmark data for these strains or in developing algorithms for studying features of mouse song. Finally, we hope that this work informs both interpretation of USV studies in models of developmental disorder, and helps to further research into understanding the

neural processes that contribute to the production and predictability of USV behavior.

Introduction

The ultrasonic vocalizations (USV) of young mouse pups in response to maternal isolation has been studied for over five decades [1–6]. The ability of isolation to elicit pup USV begins within days of birth and shows a peak in early postnatal development followed by a steady decline until two weeks of age [5]. These vocalizations function as a simple form of communication as they stimulate search and retrieval behavior from dams [2, 7, 8]. Because pup USV is easily elicited in the laboratory [6], and amenable to automated analysis [9, 10], it has been assessed routinely as an anxiety- and communication-related phenotype in models of neurodevelopmental disorder [11, 12]. A number of knockout mouse lines for ASD spectrum disorder [13–15], as well as for speech and language disorder risk genes [16] and stuttering [17], show changes to pup USV. These include either changes in the rate of USV production, or other spectral or temporal features of vocalization. Although this behavior is not human language, pup USV is a robust milestone of early postnatal development, and isolation-induced infant vocalization is a conserved behavior across mammals [3, 18–21]. Thus, understanding the neurobiological mechanisms mediating deficits of pup USV in disease models may help elucidate some conserved biology underlying these disorders of neurodevelopment.

Though production of USV is typically a robust behavior across a litter of animals, individual mouse pups show substantial variability, ranging from 0 to several hundred calls in a typical recording of wildtype C57BL/6J animals during the first week of life. Although most studies of USV in neurodevelopmental disorder models focus on mean differences between experimental and control groups, it is not often reported how variable this be-

havior is between and within subjects. While two mice of an inbred line are assumed to possess identical genetic backgrounds, this does not preclude a large degree of individual difference in behavioral expression [22–24]. The relative degree of inter- and intra-individual variation provides an estimate of the consistency or predictability of USV. The utility of estimation of the consistency of behavior and modeling intra-individual variation has been recognized in human clinical studies [25], human psychology [26, 27], and ecology, but such variability is not typically reported in studies of mouse USV, though it has been explored in the vocalizations of other species [28, 29]. In human personality theory, it has been useful to consider the differences between “trait” versus “state” influences on behavior: a state is a transient condition that influences behavior (e.g., feeling fear when seeing a snake), while a “trait” is a more stable aspect of personality that has a durable influence on behavior across time and situations (e.g., being a generally anxious person) [30, 31]. Borrowing these terms, individual-level behavioral expression patterns in USV might be due to any number of uncontrolled covariates that could mediate either state-like or trait-like differences in behavior. These include differences in intra-uterine environments and maternal health during pregnancy [32–34] or maternal experience and quality of care (feeding, licking, etc.) [35], which might have stable, trait-like impacts. Additionally, extrinsic factors such as degree of handling during the assay and temperature of the assay chamber [6], maternal behavior just prior to the assay, or physiological variables (hunger/satiety, heart rate, breathing, etc.) may have a more immediate impact. Only a subset of these external factors can be reasonably measured during the course of an experiment. For example, typical USV protocols call for controlling temperature using an incubator or a heating pad before recording, as well as minimizing handling [6]. However, even if all such factors could be controlled, some aspects of USV may yet exhibit stochasticity. Such “randomness” in behavior is demonstrable even in simpler organisms. In *C. elegans*, although the average response of worms is to move towards an attractive

olfactory stimulus, individual worms deviate from the expected pattern. In this organism, this has been shown to be controlled by neural states, where specific neurons control apparent randomization of the output behavior [36]. In mice, integration of environmental covariates and intrinsic neuronal states may differ between time points and individuals, generating a variable amount of produced USV.

Furthermore, USV is a data-rich behavioral response with numerous features in the spectral and temporal domains of audio. In particular, some features of USV may be highly consistent within an animal relative to the population across days, showing a strong “trait”-like influence on variability. Other features may be more consistent within a recording session, but display high levels of intra-individual variability across days, perhaps reflecting an individual mouse’s acute “state” on a given day. Finally some features may yet remain unpredictable even within a recording session. These degrees of consistency within and between individuals may reflect differential susceptibilities among features of USV to genetic, environmental, and intrinsic neuronal factors, leading some behaviors to show more stable “trait”-like influences (high consistency across days) while others might show patterns of variation more consistent with “state”-like responses (low consistency across days). Importantly, prior studies of features of pup USV have not considered the consistency of individuals, and determining whether a feature is more state- or trait-like may alter both interpretation of findings in disease models and the search for neurobiological mediators of pup USV.

Thus, to address the concept of consistency in USV behavior, we have used mixed modeling statistical approaches. Linear mixed models (LMMs) have proven a powerful way to estimate behavioral consistency patterns by partitioning random variance terms which describe the degree of inter- and intra-individual variability. In this study, we have employed the mixed model intra-class correlation (ICC) coefficient ([25], also referred to as “repeatability” [37, 38]) in order to understand consistency in features of USV

across three independent discovery cohorts, totaling 285 subjects, and across two strains: FVB/Ant & C57BL/6J (“Pooled Cohort Study”, PCS). We analyzed call rate (calls per minute), spectral, and temporal features of USV across three time points during postnatal development after controlling for effects of animal strain, age, and relative size. We also analyzed these features binned within recording session at each postnatal time point in order to understand consistency within a session. In order to validate our findings, we recorded additional litters of each strain at high temporal density (postnatal days 3-14, “Time Course Study”, TCS) as a replication study and to further probe the temporal dynamics of consistency. We found that despite clear group-level changes (due to age or strain) in both discovery and replication cohorts, features nevertheless varied in consistency across development, with some features, such as call rate, being largely unpredictable from day to day for a given animal. Within session however, we found that most features of USV exhibited significantly higher consistency on any given postnatal day. Furthermore, some features that showed low consistency over postnatal days, such as USV call rate, demonstrated a narrow window of high consistency near the peak of USV behavior. Early postnatal development is a highly dynamic time period for pups. To explore whether features of USV exhibit more stable behavior across measurements after animals have fully developed, we additionally looked at consistency in features of USV exhibited during adult male-female encounters. In contrast to pup USV, some features of adult USV showed dramatically higher consistency across test days, including the rate of ultrasonic calling and average call duration. Remaining features, such as the fraction of calls containing instantaneous jumps in pitch, did not show increased consistency. Thus while the amount of USV produced by an animal may acquire trait-like stability later in life, other features remain dependent on the state of the animal, environmental context, or other influences.

Results

Assessment of consistency of USV features across early postnatal development

In order to examine consistency, we have employed the intra-class correlation coefficient defined from the LMM. For a LMM of a response y (e.g., a feature of USV such as call duration), modeling fixed effects and a random intercept, we have a model of the form:

$$y^{(i)} = \mathbf{X}^{(i)} * \beta + \alpha^{(i)} + \epsilon^{(i)}$$

where $y^{(i)}$ is the i th measurement, $\mathbf{X}^{(i)}$ is the i th row of the design matrix of fixed effect covariates \mathbf{X} , β is the vector of fitted coefficients (e.g. slopes or contrasts between group means), $\alpha^{(i)}$ is the i th random intercept (a function of subject identity), and $\epsilon^{(i)}$ is the i th error. Both α and ϵ are assumed to be normally distributed random variables, which have means of 0 and variances described by σ_α^2 and σ_ϵ^2 , and these variance terms are fitted as part of the likelihood-based modeling procedure. The intraclass correlation coefficient is defined as:

$$ICC = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}$$

and ranges between 0 and 1. Figure 1 illustrates how the ICC measures the degree of consistency between subject measurements. If the response variable y is adjusted for its expected value based on fixed effects as $y - E(y)$, where $E(y) = \mathbf{X} * \beta$ (e.g., a group mean), the resulting data will be centered around 0. In the simplest scenario, the random intercept will represent the average of subject values after accounting for $E(y)$. If measurements are consistent, then respective individuals will vary tightly around this intercept (Figure 1A) after adjustment. If this is the case, very little variance between subjects will remain after adjusting for these intercepts, $\sigma_\alpha^2 \gg \sigma_\epsilon^2$, and the ICC will approach 1.0. However,

if individuals vary inconsistently, then their intercepts after adjusting will be close to 0. In other words, it will be difficult to predict where, with respect to the group estimate, an individual will be encountered from measurement to measurement (Figure 1B) and the fitted intercepts will do little to account for the remaining variance. In this scenario, $\sigma_{\alpha}^2 \ll \sigma_{\epsilon}^2$ and the ICC will approach 0. Although the fitted values of σ_{α}^2 and σ_{ϵ}^2 derived from the mixed model are point estimates, using a bootstrap approach, we are able to assign confidence intervals to these values, and thus the value of the ICC.

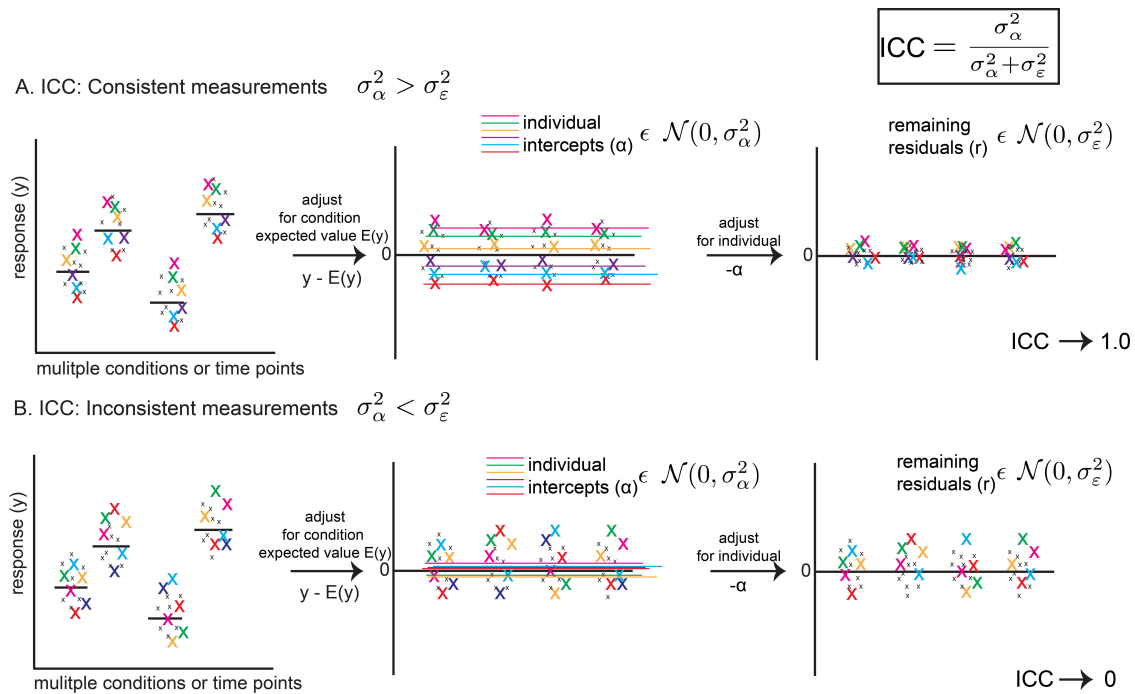


Figure 1: The Intra-Class Correlation (ICC) defined from a linear mixed model (LMM) reflects the level of behavioral consistency of individual animals across multiple measurements.

The ICC (upper right) is defined as $\sigma_{\alpha}^2 / (\sigma_{\alpha}^2 + \sigma_{\epsilon}^2)$, using mixed model random variance terms, and represents consistency of behavior across multiple measurements, where σ_{α}^2 is the random effect variance term and σ_{ϵ}^2 is the error variance term from the LMM (fitting a random intercept only, as a function of animal identity). (A) **Hypothetical scenario showing how the ICC reflects a consistent pattern.** Left Panel: a response variable (e.g., rate of ultrasonic calls per minute) is measured for the same six animals (color coded x's) across conditions and time points. Middle Panel: After adjusting for expected values $E(y)$ (e.g., group means or regression predictions for variables such as age or strain), random intercepts reflect an average expectation for a particular animal's position in the distribution of residuals. If measurements are consistent, the variance of these intercepts (σ_{α}^2) reflects most of the remaining variance in the data. Right Panel: After adjusting for intercepts α , residuals are squeezed towards zero. Thus $\sigma_{\alpha}^2 \gg \sigma_{\epsilon}^2$ and the ICC approaches 1.0. (B) **Hypothetical scenario showing inconsistent measurements.** After adjusting for time point or condition (Left Panel), residuals (Middle Panel) vary inconsistently from measurement to measurement for a given animal, and average values across measurements (random intercepts) are close to zero, and σ_{α}^2 is small and reflects little of the remaining variance in the data. After adjusting for random intercepts, the residuals are mostly unchanged (Right Panel). Thus $\sigma_{\epsilon}^2 \gg \sigma_{\alpha}^2$ and the ICC approaches 0. Thus, the ICC is a metric which summarizes consistency of patterns of behavior across measurements. The ICC is a point estimate, but using a bootstrap procedure we are able to assign confidence intervals to the ICC.

Using the ICC, we sought to explore consistency across some of the most commonly estimated features of USV in the time and frequency domains (Figure 2). In addition to the call production rate, we also looked at the fraction of calls with pitch jumps (≥ 10 kHz), as well as the duration, median pitch, and peak power. Because animals differ in the number of calls they produce, duration, pitch, and power estimates were computed as either an average over all calls for each recording, or the variability over all calls expressed as the coefficient of variation (standard deviation/mean). These features were selected based upon their salience in previous studies of USV. Pup calls are distinguishable from adult USV [39] and pitch and duration of these calls elicit maternal neuronal response and search behavior [40–42]. Using these call features as our dependent variables, the ICC represents a summary statistic describing how consistently an individual's place in the population varies across measurements from time point to time point. It helps to address, for example, whether an animal producing the longest or loudest calls on day 5 is also producing the longest or loudest calls on day 7, relative to the rest of the population.

Thus, we first analyzed a discovery cohort (PCS) gleaned from 3 datasets, 2 from C57BL/6J animals and 1 from FVB/Ant animals. In our statistical model, we controlled for effects of: strain (genetic effects and shared environment), age (postnatal day), and relative animal size (weight normalized by strain and postnatal day). We also considered other factors such as sex and litter size, however exploratory preliminary analysis did not determine statistically significant effects for these factors and they were excluded from the model [not shown]. Descriptive statistics for call features between groups and across days in the PCS are shown in Supplemental Table 1.

Examining the consistency of these eight features, we found that each generally showed low consistency across days for a given animal (Figure 3). Specifically, the most widely

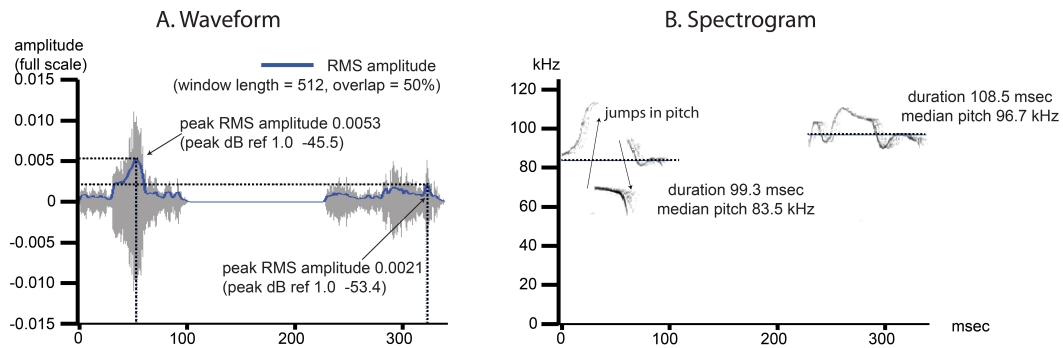


Figure 2: USV features under investigation include commonly measured features from the time and frequency domain.

Example image shows two pup isolation-induced ultrasonic calls. (A) Waveform: the time domain data for the noise- and frequency-filtered calls. A root mean squared (RMS) amplitude envelope is determined for each call, and the peak power from this envelope ($power = amplitude^2$) is determined and reported as dB ref 1.0. The average of this measurement is determined over all calls, per recording, as well as the variability in this measurement, reported as the standard deviation divided by the mean (coefficient of variation). (B) Spectrogram: the frequency domain data for the noise- and frequency-filtered calls. The presence of a pitch jump is determined by an instantaneous change in the frequency of maximum power ≥ 10 kHz, and the fraction of all calls containing at least 1 such jump was computed. The median value of the pitch (kHz) as well as the duration (msec) were determined, and both the average over all calls by recording as well as the coefficients of variation (sd/mean) representing the variability in these measurements over all calls, were computed.

assessed variable in studies of pup USV, call rate (Figure 3A), showed an ICC of 0.20 (c.i. [0.12,0.27]) indicating low consistency over postnatal days, and Studentized residuals z show rank correlations of 0.26 (Day 5 vs. Day 7) and 0.24 (Day 7 vs. Day 9) which are within the range of the ICC interval. In contrast, average call duration (Figure 3B) showed a marginally higher root ICC at 0.40 (c.i. [0.32, 0.48]) and residuals showed rank correlations of 0.48 and 0.40 for Days 5/7 and Days 7/9 respectively. However, aside from call duration and median pitch (Figure 3E), most features showed low consistencies with ICC values and rank correlations less than 0.3.

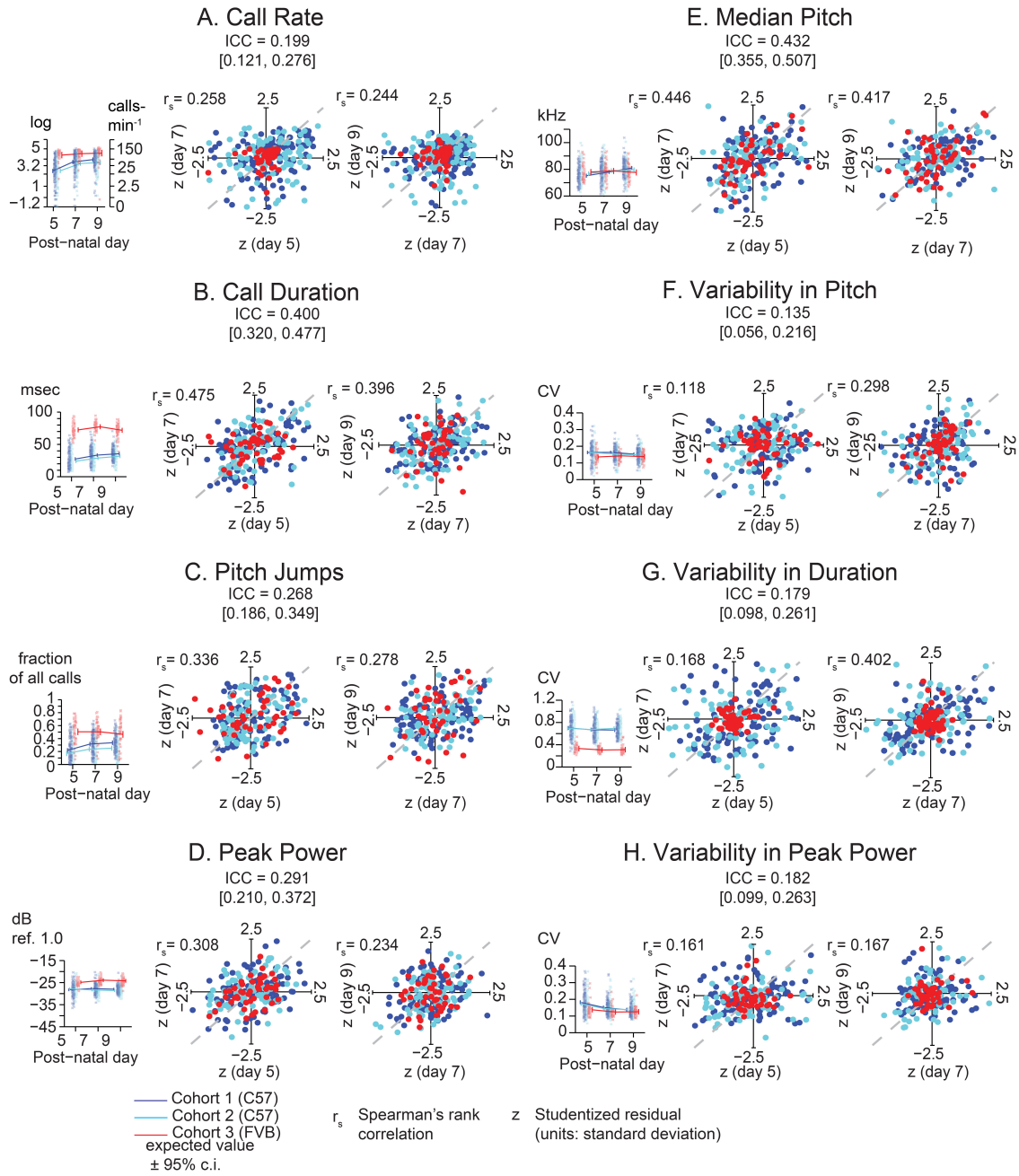


Figure 3: Lack of Strong consistency across pup USV features in Pooled Cohort Study (PCS)

. Each panel shows: (left) the value of the ICC with its bootstrapped 95% confidence interval, the data, with bee plots of individual animals, and trend lines, color-coded by cohort (blue: Cohort 1 (C57BL/6JBL/6J, N=133), cyan: Cohort 2 (C57BL/6JBL/6J, N=105), red: Cohort 3 (FVB/AntJ/AntJ N=47), showing expected values from the LMM (fixed effects only, $w=0$ (average weight)) \pm bootstrapped 95% confidence intervals on regression estimates (right) Studentized residuals (z) after adjusting for fixed effects plotting day 5 vs. day 7, and day 7 vs. day 9, and their respective Spearman rank correlation coefficients. The ICC is a summary statistic for each USV feature's consistency, but note that Spearman rank correlation coefficients are typically within or near the range of the respective ICC's confidence bounds. **(A) Call Rate** (calls-min⁻¹). LMM was fitted on $\log\left(\frac{\text{counts}+1}{\text{minutes}}\right)$ (abbreviated "log" on y-axis) with data (left panel) shown alongside linear scale values for ease of interpretation. LMMs can tolerate missing data points, and not all animals have data on all three time points due to pup death. Residual plots and associated correlation coefficients were only computed for animals with data on all three time points: N= (1) 119, (2) 101, (3) 47. **(B) Call Duration** (averaged over all calls, milliseconds). **(C) Pitch Jumps** (fraction of all calls). **(D) Peak Power** (averaged over all calls, dB ref. 1.0). **(E) Median Pitch** (averaged over all calls, kHz). **(F) Variability in Pitch. (G) Variability in Duration. (H) Variability in Peak Power** (F-H: (coefficient of variation (σ/μ) over all calls). Other than call rate (A), other features of USV (B-H) were only computed for animals possessing at least 10 calls (Day 5: N=(1) 114, (2) 90, (3) 47 | Day 7: N=(1) 122, (2) 99, (3) 47 | Day 9 N= (1) 116, (2) 98, (3) 46). LMMs fitted in R using *lme4* with models in Wilkinson notation as: feature ~cohort*w*(d+d²) + (1|id) where cohort is categorical, w is a z-score of the animals weight by cohort & day reflecting its relative size, and d is postnatal day centered around day 7, fitting both linear and quadratic terms, and (1|id) is a random intercept for each animal. Residual plots for (B)-(H) had at least 10 calls and data for all time points, N = (1) 98, (2) 80, (3) 46. Highest ICC was observed for call duration with ICC = 0.400 [0.320,0.477], with rank correlations of 0.475 on day 7 vs. day 5, and 0.396 on day 9 vs. day 7, and median pitch ICC = 0.432 [0.355, 0.507], with correlations of 0.446 on day 7 vs. day 5 and 0.417 on day 5 vs. day 9. Most features of USV have values off ICC near 0.3 or 0.2 indicating overall low levels of day-to-day consistency. Gray dotted lines show correlation of 1.0 for comparison.

To replicate these findings and improve the temporal resolution of these data, we recorded animals from two litters each of C57BL/6J and FVB/Ant every day between postnatal days 3-14 (TCS). Descriptive statistics for this group of animals is shown in Supplemental Table 2. Since most FVB/Ant animals did not call beyond postnatal day 10, only 3-10 are considered for USV features other than call rate. These data are shown in Figure 4. Features of USV in the TCS largely recapitulated the overall low consistency exhibited by animals in the PCS. Pairwise, day-by-day, rank correlations of residuals are shown as heatmaps. ICC is an aggregate measure across all time points. Although on the whole call rate shows low consistency, inspecting the heat maps (Figure 4A), one can observe an increase in pairwise correlation near the peak of vocalization behavior (just before postnatal day 5 for FVB/Ant (Spearman's rank correlation r_s days 4 & 3 = 0.54,

days 5 & 4 = 0.63) and just after postnatal day 7 (Spearman's rank correlation r_s days 8 & 7 = 0.58, days 9 & 8 = 0.75) for C57BL/6J). Thus call rate appears to show a trend towards increased stability at specific times. Interestingly, the pattern of correlation over time is different across other features of USV. Strong correlation of the median pitch (Figure 4E) for C57BL/6J appears to be restricted to an early time window (days 3-4), which degrades later in development, while FVB/Ant shows this stronger correlation for a wider time window (days 3-7). Both strains show similar increased consistency in peak power later in development (after postnatal day 7). Thus features of USV, while on the whole inconsistent across developmental time, show windows of stability which depend on the feature and the strain.

The values of ICC are tabulated for the PCS and TCS in Table 1. The point estimates of ICC for each USV feature between the PCS and TCS are replicable (Pearson's $R=0.77$, $p=0.025$, note largely overlapping confidence intervals for most variables), although some features such as median pitch did not replicate well as indicated by poorly overlapping confidence intervals. Considering results from both datasets, after predicting an animal's response using fixed effects, where in the distribution the animal will lie above or below this estimate is not strongly consistent from day to day. However, although overall consistency is low for features of USV, the actual estimates of the ICC values are reproducible across studies, describing a seemingly robust property of these features. This is remarkable, considering that the PCS and TCS differ markedly in terms of their size, composition, and number of time points.

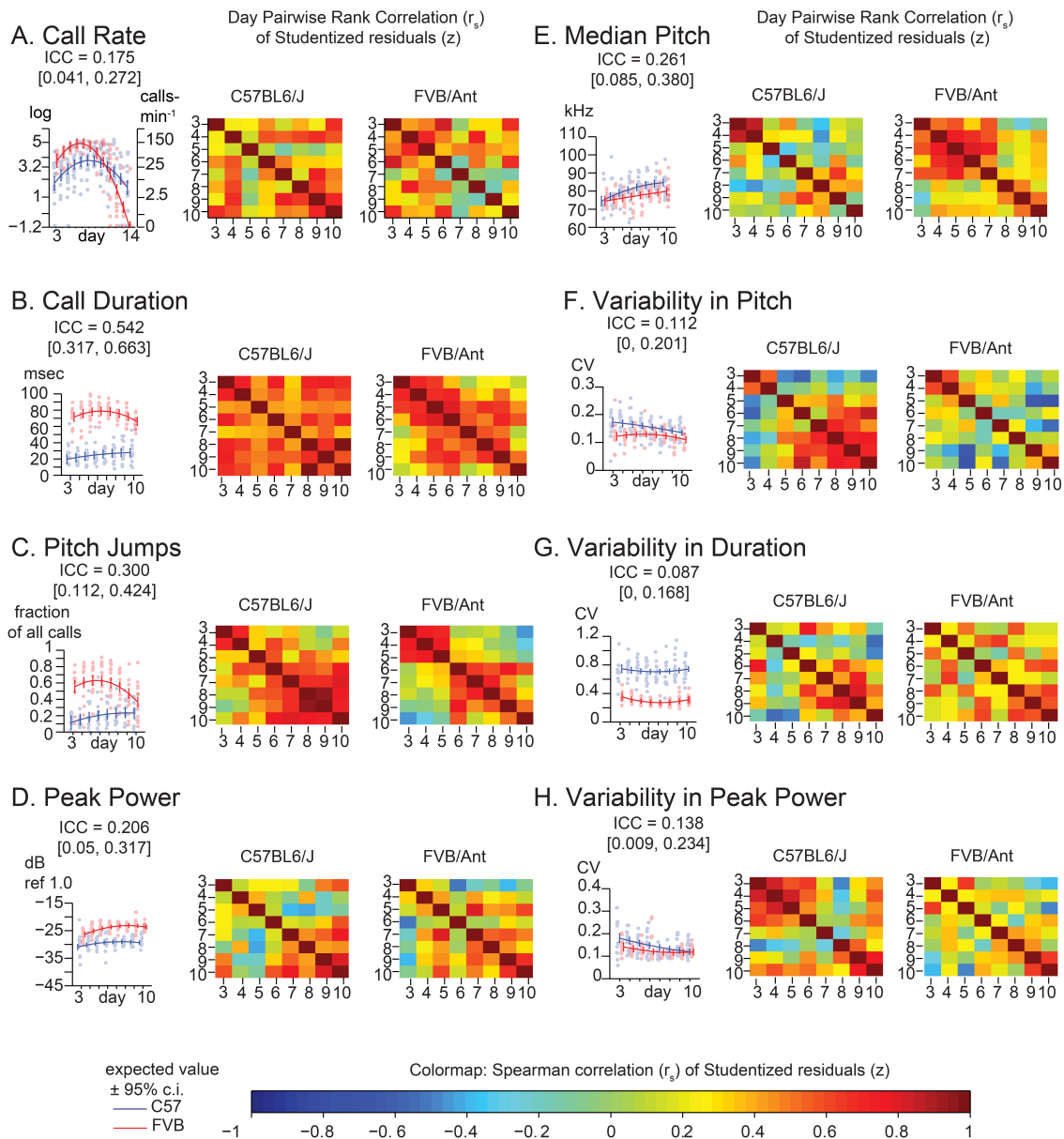


Figure 4: Lack of strong consistency across pup USV features in Time Course Study (TCS)

.Univariate linear mixed models (LMMs), ICC values, and residuals, and associated correlation coefficients were computed for data from 2 litters each of C57BL/6JBL/6J (N = 8,5) and FVB/AntJ/AntJ (N=8,5) as in Figure 2, measured each day postnatally between days 3-14. Each panel shows: (left) value of the ICC, data, with expected values ($w=0$, average weight) and 95% confidence intervals above beepplots and trendlines as in Figure 2, (right) pairwise day-by-day Spearman correlation of Studentized residuals after adjusting for fixed effects, as in Figure 2, displayed as heat maps (range: blue $r_s=-1.0$, red $r_s=1.0$). **(A) Call rate.** Data were transformed as $\log\left(\frac{\text{counts}+1}{\text{minutes}}\right)$ (abbreviated “log” on y-axis) as in Figure 3, with linear scale values shown for ease of interpretation. Note that modeling day as both linear and quadratic terms allows for prediction of the characteristic rise and fall in call rate observed through the first two weeks of life. Overall ICC is low and within range of PCS (ICC = 0.175 [0.041, 0.272]), however heatmaps reveal a density of stronger correlation near the respective peak for each strain. Beyond day 10, most FVB/AntJ animals did not exhibit >10 calls per sonogram, so graphs in B-H, and all correlation heat maps only show data between days 3-10. **(B) Call Duration.** **(C) Pitch Jumps.** **(D) Peak Power.** **(E) Median Pitch.** **(F) Variability in Pitch.** **(G) Variability in Duration.** **(H) Variability in Peak Power.** Heatmaps showing residual Spearman cross-correlation for (B)-(H) had at least 10 calls and data for all time points, N = 9 (C57BL/6JBL/6J), 13 (FVB/AntJ/AntJ). As in the PCS (Figure 2), call duration (B) shows the highest ICC 0.542 [0.317, 0.663] with higher levels of correlation day to day across all days. Median pitch did not reproduce the result in Figure 2 when all days were taken into account though slightly overlaps the confidence interval: ICC = 0.261 [0.085, 0.380]. Note both strains show an inflection in their correlations for fraction of calls with pitch jumps near day 5-6 for FVB/AntJ and day 4-5 for C57BL/6J which may indicate that something around this time is important for the development of this kind of call.

Table 1: Values of the ICC and Confidence Intervals Computed in the PCS TCS.

Feature	PCS			TCS		
	Estimate	Lower 95% ^o	Upper 95% ^o	Estimate	Lower 95% ^o	Upper 95% ^o
log call rate	0.199	0.121	0.276	0.175	0.041	0.272
duration	0.400	0.320	0.477	0.542	0.317	0.663
calls with pitch jumps	0.268	0.186	0.349	0.300	0.112	0.424
median pitch	0.432	0.355	0.507	0.261	0.085	0.380
peak power	0.291	0.210	0.372	0.206	0.050	0.317
variability in duration	0.179	0.099	0.261	0.087	0	0.168
variability in pitch	0.135	0.056	0.216	0.112	0	0.201
variability in peak power	0.182	0.099	0.263	0.138	0.009	0.234

^o 95% confidence intervals computed from parametric bootstrap ($N = 1 \times 10^5$) on linear mixed model parameters (see Methods).

Consistency of USV features within recording sessions

The relatively low consistency observed in the preceding section over developmental time could arise because USV is highly susceptible to uncontrolled intrinsic or environmental covariates, present at the time of experimentation, which perturb each individual animal's response for the duration of the recording. Alternatively, low consistency could be due to the inherent noisiness of features of USV. If the latter were the case, we hypothesized that, even within a recording session, we would find that USV features were inconsistent across the course of the session. If so, ICC computed across a recording session should be similar to ICC computed across development. If, however, consistency of USV features were higher within a recording session compared to across sessions, then we hypothesize instead that USV itself is not inherently noisy, but rather reflects perturbation of the pup's state at the time of recording by some unmeasured developmental or environmental variable.

To address this question, we computed the ICC in the PCS and TCS on each postnatal day within recordings, where repeated measures consisted of 1 minute bins through the

3 minute recording. In addition to the fixed effects of strain and size modeled previously, we also controlled for the effect of bin, as the pup's temperature may change through the course of the recording, and temperature has been shown to have an effect on aspects of USV [43, 44]. The estimates of ICC computed across bins by day in the PCS are shown in Table 2, and for the TCS in Table 3. ICC values are as much as 3-fold higher when computed across bins than when computed across developmental time, and these results are summarized in Figure 5 (TCS-Within Session vs. TCS across days Mann Whitney $p = 0.0011$, PCS-Within Session vs. PCS across days $p=1.6x10^{-4}$). ICC values computed within bins and averaged across days for each study are tabulated in Table 4. Again, the results are strongly reproducible (Pearson's $R = 0.95$, $p = 3.1x10^{-4}$) across studies.

Thus, these data support the hypothesis that most features of USV are not inherently inconsistent, but instead inconsistencies across development may arise from unknown variables affecting the animal's state at the time of recording. Examining results from both PCS and TCS indicate our estimates of ICC both between and within sessions are robust to relatively large differences in experimental design, such as the number of time points considered and sample size.

Table 2: Values of the ICC and Confidence Intervals Computed in the PCS across minute bins.

Feature	Day	Estimate	Lower 95% ^o	Upper 95% ^o	Feature	Day	Estimate	Lower 95% ^o	Upper 95% ^o
	5	0.614	0.551	0.668		5	0.619	0.544	0.688
log call rate	7	0.673	0.617	0.722	peak power	7	0.703	0.646	0.753
	9	0.675	0.617	0.724		9	0.700	0.642	0.751
duration	5	0.696	0.632	0.754	variability in duration	5	0.559	0.476	0.636
	7	0.725	0.671	0.771		7	0.501	0.422	0.574
	9	0.748	0.697	0.792		9	0.441	0.357	0.519
calls with pitch jumps	5	0.729	0.671	0.781	variability in pitch	5	0.649	0.578	0.714
	7	0.652	0.588	0.708		7	0.524	0.447	0.594
	9	0.687	0.627	0.739		9	0.595	0.524	0.659
median pitch	5	0.690	0.625	0.748	variability in peak power	5	0.427	0.331	0.517
	7	0.712	0.656	0.761		7	0.480	0.399	0.554
	9	0.777	0.731	0.816		9	0.610	0.541	0.672

^o 95% confidence intervals computed from parametric bootstrap ($N = 1 \times 10^5$) on linear mixed model parameters (see Methods).

Table 3: Values of the ICC and Confidence Intervals Computed in the TCS across minute bins.

Features	Day	Estimate	Lower 95% ^o	Upper 95% ^o	Features	Day	Estimate	Lower 95% ^o	Upper 95% ^o
log call rate	3	0.571	0.299	0.750	peak power	3	0.862	0.696	0.939
	4	0.441	0.150	0.656		4	0.738	0.492	0.874
	5	0.563	0.291	0.745		5	0.630	0.311	0.823
	6	0.452	0.163	0.666		6	0.439	0.114	0.680
	7	0.791	0.605	0.887		7	0.721	0.492	0.851
	8	0.497	0.213	0.698		8	0.804	0.623	0.902
	9	0.816	0.648	0.902		9	0.817	0.650	0.903
	10	0.783	0.594	0.883		10	0.588	0.301	0.774
duration	3	0.860	0.695	0.938	variability in duration	3	0.585	0.251	0.794
	4	0.784	0.570	0.897		4	0.000	0.000	0.313
	5	0.851	0.686	0.935		5	0.650	0.343	0.833
	6	0.863	0.725	0.934		6	0.651	0.381	0.817
	7	0.714	0.482	0.846		7	0.267	0.000	0.530
	8	0.905	0.806	0.955		8	0.887	0.772	0.946
	9	0.867	0.737	0.931		9	0.161	0.000	0.420
	10	0.776	0.574	0.886		10	0.271	0.000	0.545
calls with pitch jump	3	0.793	0.572	0.906	variability in pitch	3	0.831	0.639	0.924
	4	0.782	0.565	0.897		4	0.339	0.000	0.623
	5	0.779	0.552	0.900		5	0.379	0.010	0.667
	6	0.773	0.566	0.887		6	0.759	0.545	0.880
	7	0.795	0.607	0.893		7	0.102	0.000	0.375
	8	0.821	0.651	0.911		8	0.065	0.000	0.355
	9	0.790	0.606	0.888		9	0.257	0.000	0.509
	10	0.832	0.669	0.917		10	0.265	0.000	0.537
median pitch	3	0.696	0.406	0.857	variability in peak power	3	0.394	0.023	0.673
	4	0.736	0.490	0.873		4	0.433	0.086	0.689
	5	0.507	0.152	0.750		5	0.364	0.000	0.655
	6	0.731	0.499	0.864		6	0.402	0.077	0.652
	7	0.809	0.629	0.901		7	0.511	0.216	0.717
	8	0.671	0.412	0.828		8	0.743	0.521	0.870
	9	0.800	0.622	0.894		9	0.581	0.307	0.758
	10	0.738	0.513	0.864		10	0.652	0.387	0.815

^o 95% confidence intervals computed from paramateric bootstrap ($N = 1 \times 10^5$) on linear mixed model parameters (see Methods).

Table 4: ICC in the PCS and TCS across Minute Bins: Averages and Standard Deviations.

Feature	PCS		TCS	
	Average	S.D.	Average	S.D.
log call rate	0.654	0.035	0.614	0.158
duration	0.723	0.026	0.827	0.063
calls with pitch jumps	0.689	0.039	0.796	0.021
median pitch	0.726	0.045	0.711	0.094
peak power	0.674	0.048	0.700	0.141
variability in duration	0.501	0.059	0.434	0.302
variability in pitch	0.589	0.063	0.510	0.137
variability in peak power	0.506	0.094	0.375	0.281

Consistency of features of USV in adult male-female C57BL/6J dyads

In the preceding sections, we have shown that there is overall low consistency across the features of USV examined in mouse pups across recording sessions, yet that consistency is high within a recording session. We next examined whether adult male USV was also primarily “state” dependent or “trait” dependent. We measured USV from 47 adult male animals on two test days with a different unfamiliar female on each day, made up entirely of C57BL/6J animals. This dataset differs in a few fundamental ways: (1) the stimulus is the presentation of an adult female mouse to the male, rather than isolation of pups from the dam, (2) the recordings are dyadic. Although historically it has been suggested that in such a paradigm only the male is vocalizing[45], recently it has been shown that an appreciable number of vocalizations can be attributed to the female [46]. We make no strong claims that our data represent something unique to male behavior. Finally, the number of measurements differs importantly in that for pups each time point represents

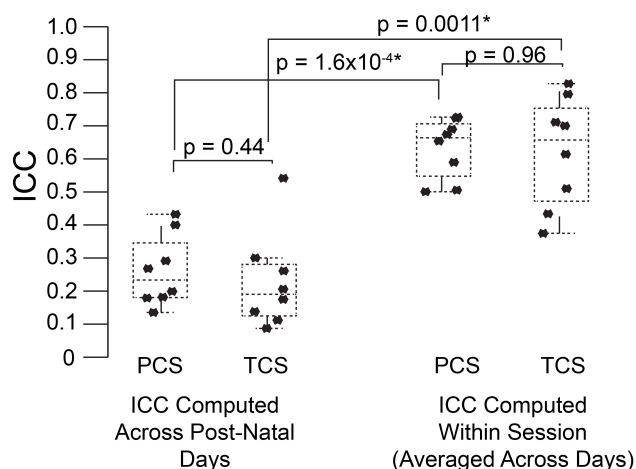


Figure 5: USV features show higher consistency within within sessions than between sessions.

ICCs were recomputed within session on each postnatal day, using 1 minute bins through the recordings (3 for each recording) as the repeated measure instead of postnatal day, in both the PCS and the TCS. Univariate LMMs were fitted using the model $feature \sim bin * strain * w$ where w is day/strain z-score of animal's weight as previously. Data are shown for (left) ICC values for all 8 USV features computed across postnatal days (see Figures 2 and 3) and (right) computed within session. Overlaid on data points are box plots. Horizontal line represents the median, and box represents lower and upper quartiles (25% (Q1) and 75% (Q3)), with whiskers extending to most extreme datapoints not exceeding 1.5 x the interquartile range. Significant differences in ICC were detected for within session vs between days (Mann Whitney $p = 1.6 \times 10^{-4}$, TCS $p = 0.0011$) with the median ICC being 2.8-fold higher within session than across days for the PCS, and 3.3-fold higher in the TCS. Fold increases in ICC in the PCS and TCS $\frac{within\ session}{across\ days}$ were largely reproducible: call rate 3.3-fold (PCS), 3.5-fold (TCS), call duration 1.8-fold (PCS), 1.5-fold (TCS), pitch jumps 2.6-fold (PCS & TCS), median pitch 1.7-fold (PCS), 2.7-fold (TCS), peak power 2.3-fold (PCS), 3.4-fold (TCS) variability in pitch 4.4-fold (PCS), 4.6-fold (TCS), variability in peak power 2.8-fold (PCS), 2.7-fold (TCS), though variability in duration was less reproducible (2.8-fold in the PCS, and 5-fold in the TCS). Linear correlation in fold change between PCS and TCS was $R=0.68$, and 0.83 if variability in duration is omitted. We also did not detect a significant difference in the magnitude of ICC values between PCS and TCS either within session or across days (ICC across postnatal days, PCS vs. TCS, Mann Whitney $p = 0.4418$; ICC within session, averaged across days, PCS vs. TCS Mann Whitney $p = 0.96$). Thus the ICC and changes to the ICC when computed within recording session vs. across development appear to be robust calculations for these USV features, despite the fact that the PCS and TCS differ widely in the number of individual animals, the number of time points.

potentially a different developmental stage, while for the adults time points are at the same developmental stage. Linear modeling in either case, however (either modeling post natal day or adult test day as a fixed effect), allows for the effect of postnatal age or test session to be regressed before assessing consistency. Consistency itself (the ICC) is thus still comparable as it resides on the same scale representing the ratio of variance amongst individuals' intercepts to the combined variance of random effects and error. For our adult recordings, as there are only two time points, the ICC values will be expected to be near the simple pairwise correlation across test days. Pups during development are changing in a rapidly dynamic fashion, which we do not discount. However, because the interpretation of the ICC is the same in either case (consistent or inconsistent), we believed the comparison between the datasets serves to identify which features of USV may stabilize later in life, and which may remain dynamic.

Dramatically the ICC for adult call rate (Figure 6, Table 5) was much higher than that observed in pups, and even higher than the value obtained within pup sessions (ICC = 0.87, c.i. [0.78, 0.93]), which is also reflected in the rank correlation ($r_s = 0.86$). Call duration showed values of ICC which were similar to that obtained within session for pups (ICC = 0.77, c.i. [0.60, 0.87], $r_s = 0.73$). However, with the exception of log call rate, other features of USV, such as the median pitch, peak power, and fraction of calls with pitch jumps, showed ICC values and rank correlations in the range of those obtained for pups. This may indicate that features such as call rate and call duration approach trait-like stability in adult animals, however other features of USV still depend on the state of the animal and its environment. Descriptive statistics for our adult data are presented in Supplemental Table 3.

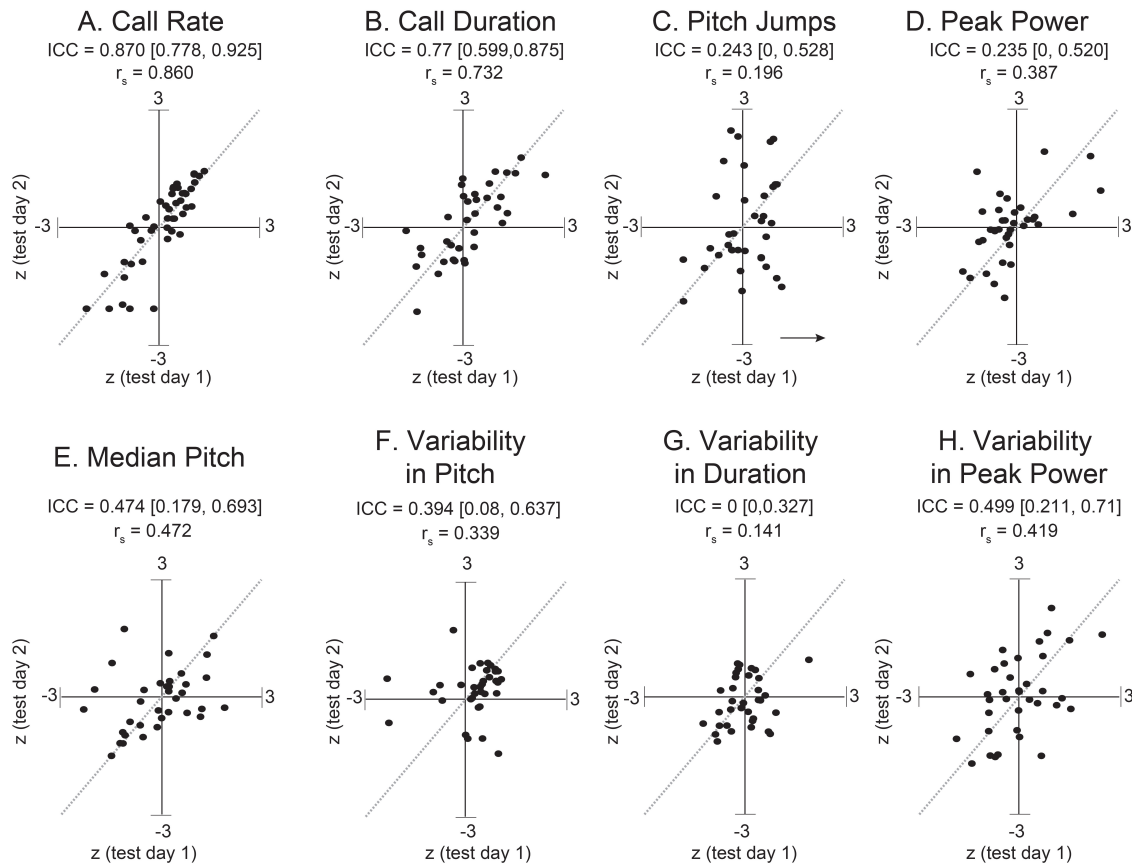


Figure 6: Stronger consistency in some adult USV features across sessions.

ICC and Spearman correlations computed for adult C57BL/6J M-F dyads across 8 features of USV. ICC and rank correlations were computed for 47 male-female pairs between 7-11 weeks of age, in which 47 group-housed males were tested two different days with a unique female each time. LMMs were fitted only using test day as a fixed effect factor: $feature \sim test\ day + (1|id)$. (Note: with only two time points, we expect the correlation coefficients to be very close to the estimates of the ICC). Studentized residuals (z) between test days are shown for (A) call rate (again, with LMM fitted for $\log(\frac{counts+1}{minutes})$), (B) call duration, (C) pitch jumps, (D) peak power, (E) median pitch, (F) variability in pitch, (G) variability in duration, and (H) variability in peak power. Call rate exhibited a much higher consistency (ICC = 0.870, [0.778, 0.925], $r_s = 0.86$) than observed for any pairwise day comparison in pup data in Figures 2 & 3. Call duration also showed higher consistency (ICC = 0.77, [0.599, 0.875], $r_s = 0.732$). However, note other features of USV showed values of ICC and corresponding correlation coefficients which are in the range of those observed for pups across early postnatal development. Thus most features of USV appear to remain relatively inconsistent from measurement to measurement, although in these data, the adult call rate & call duration appear to be stable features and exhibit trait-like behavior. Gray dotted lines show correlation of 1.0 for comparison.

Table 5: Spearman's Correlation and ICC computed for adult C57BL/6J data.

Feature	r_s	ICC	Lower 95% ^o	Upper 95% ^o
log call rate	0.86	0.87	0.78	0.93
duration	0.73	0.77	0.60	0.87
calls with pitch jumps	0.20	0.24	0	0.53
median pitch	0.39	0.24	0	0.52
peak power	0.47	0.47	0.18	0.69
variability in duration	0.34	0.39	0.08	0.64
variability in pitch	0.14	0	0	0.33
variability in peak power	0.42	0.50	0.21	0.71

Discussion

In this investigation, we have examined datasets generated in our laboratory in order to understand the extent to which features of vocalization show consistent inter- and intra-individual patterns across measurements. In young pups, we have found, in general, that across development most features of USV such as the call rate do not show consistent patterns across an individual's measurements, though some such as call duration show a larger degree of consistency. The estimates of consistency in our pup data were largely reproduced when examined in a replication cohort which increased the number of time points across development sampled. When looking within a session, pups across development show a much higher degree of consistency for most features examined. Thus we hypothesize that the expression of pup USV, although clearly under the influence of population effects such as strain or age, is highly state dependent. Therefore, we conclude that while the population average may rise or fall due to strain or age, the relative ranks of the pups in the distribution must be influenced by other unmeasured aspects of the animal's state. It could be that some of this influence derives from the litter to which the

animal pertains, however we have also estimated ICC at the level of litter and have not found any increased explanation of remaining variance upon inclusion of this hierarchy (not shown). As phenotypic expression in an individual's behavior is a complex integration of its state, and genetic and environmental factors [47], a lack of consistency is not entirely surprising. Wild species often display behavioral plasticity in the form of inconsistent individual behavior over time, yet show consistent trends at the group level [48]. The study of trait consistency over time amongst individuals has also been appreciated in the domain of human psychology [49] and ecology [50], but rarely in laboratory animals. In our mice, however, we did observe that there was an increase in intra-animal consistency near each strain's respective peak of vocalization behavior at least with respect to rate of calling. These time points may represent preferable windows to look for effects due to experimental manipulation as individual animals are performing more predictably from measurement to measurement. By contrast, in our adult dataset, consistency in call rate was dramatically higher than for pups, while pitch related features continued to show low consistency. While adults and pups are in different stages of life and react to their environment differently, there appears to be a similarity that pitch features of USV continue to show dynamic modulation even where other features such as call rate show increased consistency. However, as described in Methods and Supplemental Figure 3, our adult data were pooled from a study examining changes to USV in adults after global knockout of the *Celf6* gene, in which we did not detect significant genotype effects. Future cohorts of animals, with an increased number of test days, should be examined to discern the reproducibility of any trait stability in call rate or other features. The level of intra-individual variability and overall reaction to changes in the external environment has been shown in adult mice to be explainable to some degree by their level of subordination/dominance and aggressiveness [51] and more recently, rate of calling in adult males has been directly correlated to measures of dominance and social hierarchy in tasks such as the tube test,

and manipulation of the prefrontal cortex is able to alter the hierarchical rank order among the mice and concomitantly their rates of ultrasonic calling [52]. In our study, males were socially isolated from their cage hierarchies for 24 hours before test day #1 and up to a week before test day #2, though this may not be sufficient time to perturb the established dominance rank order in these males. For features showing poorer consistency (pitch related features) between test days, our results may be somewhat confounded by not fully knowing the animal originating the calls (male or female), and the fact that the female's estrous state was uncontrolled. It has been claimed that males can pitch modulate their song due to the presence of an alleged competitor male [53]. It is attractive to hypothesize that perhaps the state of the female or her contribution to the dyadic song somehow influences the pitch characteristics, and may explain why there are poorer correlations for these features in our study. It will be interesting to observe what other genetic or pharmacological manipulations are able to change the USV trait consistency of adult mice, which will reveal the potential neurological correlates of how these features are encoded. This very fundamental difference in the source of variability between pup USV and adult USV may explain why so few disease models show a consistent carry-over from pup to adult USV changes. Reviewing just the literature on call rate in ASD models in particular, 35 of 41 studies have shown alterations in pups behavior which typically manifests as a decrease in call rate. However, of the models where adult behavior was assayed, only 2 showed carry-over of pup USV phenotype into some kind of adult USV phenotype [54, 55]. Thus, whatever the mechanisms are that mediate the alterations in pup USV, these largely do not carry over to call rate in male-female song.

In the current study, we have not subcategorized calls into call types based upon spectral and temporal properties. We have avoided this approach as there is no standard method for call classification. Some methods, such as a method employed to study mice with a humanized *Foxp2* gene [56], classify by length of call and presence of instanta-

neous jumps in pitch, while others use jumps exclusively based upon their number and direction [9, 53]. Another commonly employed method involves manual sorting of calls into categories based upon spectral shape [57], which integrates information about pitch, the presence of jumps, harmonics, duration, and slope. Yet another method uses an unbiased classification scheme [10]. It is not clear the extent to which these different classification schemes represent biologically relevant categories. It has been well-documented that the frequency and frequency modulation of the pitch in rat USV is associated with positive and negative emotionality [58, 59] and rats will even self-administer or exhibit avoidance of the respective category of calls [60]. While mice emit USV during ostensibly rewarding circumstances such as mating or juvenile play, it is not clear that individual categories of calls based on any available scheme are associated with either reward or aversion, although it has been shown that mice can distinguish between calls of different categories [61]. However, all categorization schemes, either explicitly or implicitly, incorporate some aspect of the presence of pitch jumps in classification, and we have examined this feature, which has been shown to exhibit salience in listening animals [40, 62]. In neither our pup nor adult datasets did we see high degrees of consistency in the fraction of calls containing pitch jumps. However, it will be interesting to see whether a pup or an adult's repertoire, as categorized by one of the above schemes or some other, has the properties of a stable trait across individuals, or whether it too is highly affected by an animal's state. Some categorization schemes may turn out to be more consistent over multiple measurements than others, and this may be a useful criterion to determine which classification scheme may be measuring a stable biological feature. To enable these and other analyses that would benefit from the availability of a standardized dataset for algorithm testing and optimization, we have provided all of our recordings via the mouseTube portal (*upload pending*). We include raw audio files through this platform along with associated metadata, so that researchers may use this resource to address questions such as the

stability of categorical assemblies of call types. Future work remains to assess the relative utility of different categorization schemes and their biological relevance.

Methods

Animals

All protocols involving animals were approved by the Animal Studies Committee of Washington University in St. Louis. Animals for pooled cohort study (PCS) consisted of 133 C57BL/6J in Cohort 1 (18 litters of median size 8 animals, ranging from 4 to 11 animals per litter), 105 C57BL/6J in Cohort 2 (15 litters of median size 8 animals, ranging from 2 to 9 animals per litter), and 47 FVB/Ant in Cohort 3 (5 litters of median size 10, ranging from 6 to 12 animals). Animals in Cohorts 1 and 2 were originally planned to determine the effect of conditional knockout of the *Celf6* gene in dopaminergic or GABA-ergic neurons on USV, and were generated by crossing $Celf6^{flox/flox}$ X $Celf6^{flox/wt}$; DAT-Cre (Jackson Laboratory strain B6.SJL-Slc6a3^{tm1.1(cre)}Bkmn/J or $Celf6^{flox/flox}$ X $Celf6^{flox/wt}$; VGAT-Cre (Jackson Laboratory strain Slc32a1^{tm2(cre)Lowl}/J). No *Celf6* genotype effects were detected on any USV metric scored (See Supplemental Figures 1 and 2), and these data were pooled across genotype for the present analysis. Nonetheless, for the follow-up time course study (TCS) looking at vocalization every day postnatally between days 3 and 14, we used 13 wild-type C57BL/6J and 13 FVB/Ant (Jackson Laboratory) from two litters each, of 8 and 5 respectively. Animals were maintained in a barrier facility. Breeding cages consisted of a single male and a single female, and both parents were present during pregnancy, birth, and during the time of assay. Cages were maintained by our facility on a 12 hr : 12 hr light:dark schedule with food and water supplied ad libidum. Adult mice were composed of 47 C57BL/6J males and 41 females aged 7-11 weeks. Adult mice were originally planned to determine the effect of global knockout of the *Celf6* gene on adult

USV in male-female dyadic interactions. No *Celf6* genotype effects were detected on any USV metric scored (Supplemental Figure 3), and data were pooled across genotype for the present analysis.

USV Recording and Processing

USV recording - Pups

Ultrasonic vocalization for Cohorts 1,2, and 3 (PCS) was recorded on postnatal days 5, 7, and 9. For follow-up study (TCS), recordings were performed every day postnatally from days 3 through 14. All recordings were performed in the afternoon between 12:00 and 17:00. On first day of recording, subjects were each marked for identification immediately after recording by toe clip (PCS) or tattooing (TCS, Aramis Micro Tattoo Kit, Ketchum). On following days, subjects were recorded in random order and identifying marks were noted after recording, along with sex and weight. At the time of recording, a litter is separated from its parents by placing the parents in a temporary cage. The entire home cage with litter undisturbed is placed in an incubator and allowed to rest for 10 minutes. The pups' external temperature is regularly monitored with an infrared temperature gun digital thermometer (HDE-B01, HDE) and the incubator is maintained such that external temperature remains between 31-34°C. If the external temperature deviates below 30°C, the incubator is adjusted until external temperature returns within range, in order to minimize effects of cooling the pups on USV. For recording a pup, the pup is moved with minimal handling into an anechoic, sound attenuating chamber (Med Associates Inc.) and audio is recorded for 3 minutes using a CM16 microphone (Avisoft Bioacoustics), amplified and digitized using UltraSoundGate USG116H, using a gain of 1.4 dB, 250 kHz sampling rate, bit depth of 16, using Avisoft RECORDER software.

USV recording - Adult M-F dyads

Adult male animals were generated from group-housed weaned juveniles and were singly housed 24 hours before test time. Females were maintained group-housed, between 4-5 animals per cage. The testing chamber consisted of an empty mouse cage (no bedding) placed inside an anechoic, sound attenuating chamber (the same used for pup testing). Testing occurred during the beginning of the animals' dark cycle (between 18:00 - 20:00), and proceeded as follows: (1) Habituation phase: males were placed in the test environment for 10 minutes with concurrent recording of USV as in the case of pup recordings. No USVs were detected for males during the habituation phase. (2) Test phase: A stranger female was added to the test environment and the dyad was recorded for 10 minutes. After testing, males were returned to single housing, and the test environment was cleaned with 70% ethanol followed by 2% Nolvasan solution (Zoetis Inc) in between each animal. The number of days between tests was allowed to vary between 1-7 days, and the median number of intervening days was 4. No significant effect of the number of intervening days between test days on USV features was detected. Each male was tested on two days, with a different female each day. Pup and adult audio files were processed using the same computational pipeline.

White noise filtering in the frequency domain.

An automated method was designed to filter noise and improve automated call detection. A 10-second chunk is chosen at random from each audio file. The fast Fourier transform (FFT) is performed using 512 FFT bins corresponding to $\frac{512}{2} + 1 = 257$ audio frequencies ranging from 0 kHz to 125 kHz, and 50% temporal overlap corresponding to a temporal resolution of $0.5 \cdot \frac{250000}{512} = 1.024 \times 10^{-3}$ seconds. A histogram of $\log_{10}(\text{FFT magnitude})$ is computed for all magnitudes in FFT bins corresponding to frequencies between 20 and 120 kHz. The main bulk of this histogram corresponds to the noise level in the spectrum

which is assumed to be Gaussian in distribution. The mean of the noise distribution is estimated to be the peak of this histogram and a threshold is set at $\mu_{noise} + 2.5\sigma$ where only spectral magnitudes greater than threshold are designated as signal. This reliably separates the baseline of the FFT magnitudes from signal peaks for pup and adult calls. Such a threshold is determined for each file individually, however thresholds varied little across all files indicating a relatively constant background recording environment [not shown]. The noise distribution was estimated between 20 - 120 kHz since all sound outside of this range is band-pass filtered.

Spectrogram preparation and band-pass filtering & automated call detection

Spectrogram preparation and automated call detection were performed in MATLAB using code adapted from [9]. Briefly, after determining a threshold for white noise, the entire FFT (512 bins, 50% overlap, time resolution 1.024 ms, frequency resolution 488.2 Hz) is computed for each file, where magnitude < threshold is set to 0 and sound is band-passed filtered to reside within 20-120 kHz. All sound < 20 kHz and > 120 kHz is also set to zero. Ultrasound calls are detected using thresholds of 5 msec minimum duration, 0.15 minimum spectral purity, 1.0 maximum spectral discontinuity, with gaps < 30 ms between adjacent calls merged. In [9], 0.25 spectral purity was suggested as appropriate threshold. Empirically we have determined that 0.15 is more reliable and results in fewer instances where spectrally impure parts of longer calls lead to a call artificially scored as two calls. After automated call detection, random subsets of spectrograms (10 - 20% of all files) are inspected manually to ensure that automated scores overlap with human-distinguishable calls observed in the spectrogram.

Call feature extraction.

After calls are detected, features for each call are extracted as follows. The dominant frequency (“pitch”) is determined for each 1.024 ms time bin in the spectrogram for each call by determining the FFT bin with maximum power ($\text{Power} \propto \text{magnitude}^2$). The median pitch is determined, as well as the total duration of each call. The presence of discontinuous jumps in pitch was determined as changes over time greater than ± 10 kHz. Calls can also contain harmonic frequencies; these were not analyzed. . The inverse FFT was computed from each call’s spectrogram to yield the noise- and frequency-filtered waveform. A smoothed waveform envelope was estimated by computing a windowed RMS amplitude (512 samples, 50% overlap). The peak RMS amplitude was extracted from this envelope and power was computed as dB ref 1.0. The CM16 microphone was not calibrated, thus dB SPL were not computed, but dB are expressed with full-scale reference (max = 0 dB, $\text{dB} = 10 \cdot \log_{10}(\text{full scale amplitude}^2)$).

Statistical Analysis

Univariate linear mixed models (LMM) for each feature of USV were computed using the *lme4* package [63] in R [64] fitting a random intercept model grouped by subject id. Models were fitted using strain, postnatal day, and animal size as fixed effect factors. Postnatal day and animal size both entered models as continuous variables. Postnatal day was recentered at day 7 and fitted for both linear and quadratic effects in order to account for the “inverted U” pattern in development with a rise, peak, and fall in behavior. Animal size was z-score normalized weight with respect to day and strain as raw weight itself varies with both. For adults, only test day was used as a fixed effect factor. Significance of main effects and interactions in the data were computed by likelihood ratio tests (analysis of deviance) on nested models of increasing complexity using the *anova()* function in R.

Call rate (calls per minute) was transformed as the natural logarithm before modeling

$(\log \frac{calls+1}{minutes})$. Other USV features were not transformed. We also fitted call rate using a negative binomial generalized linear mixed model (NB-GLMM). Fixed effect coefficients between the LMM on log-transformed call rate and the NB-GLMM on untransformed count data were highly similar (Pearson's $R = 0.99$). The log-transformed model was used in order to compare mixed model parameters across all features of USV fitted with the same algorithm. The intra-class correlation coefficient (ICC) was determined using the fitted point estimates of random intercept variance (σ_{α}^2) and residual error variance (σ_{ε}^2) from the the LMMs as described in Results.

In order to determine confidence bounds for model parameters, we employed a parametric bootstrap procedure. Using the point estimates of σ_{α}^2 and σ_{ε}^2 as starting points, the i th bootstrap sample $y_{*}^{(i)}$ were computed as:

$$y_{*}^{(i)} = X^{(i)} \cdot \beta + rnorm\left(mean = 0, sd = \sqrt{\sigma_{\alpha}^2}\right) + rnorm\left(mean = 0, sd = \sqrt{\sigma_{\varepsilon}^2}\right)$$

where $X^{(i)}$ is the i th row of the fixed effects design matrix and β is the vector of fixed effects coefficients. Thus $X^{(i)}\beta$ represents the expected value $E(y)$ for the i th observation, which is then perturbed by drawing a random intercept and error from normal distributions (the R `rnorm()` function) with means of 0 and standard deviation as the square-root of the fitted LMM variance estimates. Each vector y_{*} represents a bootstrap sample dataset. The LMM was re-fitted using each y_{*} sample dataset for 100,000 iterations. The 95% confidence bounds for fixed effect coefficients, σ_{α}^2 , σ_{ε}^2 , and the ICC were determined as the lower 2.5% and upper 97.5% quantiles of the bootstrap distribution. This procedure is preferable to a strict resampling with replacement of the original values of y , as it does not result in bootstrap sample datasets lacking factor levels and leaves the fixed effect correlation structure intact.

To compute values of ICC within session, recordings were binned into 3x1-minute bins and USV aggregate features (e.g. average duration) were recomputed for each bin. LMMs

were fitted on each postnatal day using strain and bin number as categorical variables, and z-score normalized weight as previously. ICC values obtained from within session calculations were compared to ICC values compared from calculations across postnatal days using a non-parametric Wilcoxon rank-sum test.

In order to explore consistency graphically across all USV features (regardless of scale, Figures 2,3, & 5), we computed Studentized residuals. Residuals from the full model take into account both fixed and random effects, and as such are not useful for looking directly at consistency as any consistent patterns expressed in the random intercepts have been removed. Thus, we computed a first-level residual where a residual $\tilde{\varepsilon}$ is the result of a data point $y^{(i)}$ adjusting for the model's expected value $E(y)^{(i)}$ (not taking into account random effects) as:

$$\tilde{\varepsilon}^{(i)} = y^{(i)} - E(y)^{(i)}$$

Such a residual is represented in the middle panel of Figure 1A and 1B and has units that are the same as the units of y . To normalize for units, a Studentized residual was computed as:

$$z^{(i)} = \frac{\tilde{\varepsilon}^{(i)}}{\tilde{\sigma}(1 - h^{(i)})} = \frac{y^{(i)} - E(y)^{(i)}}{\tilde{\sigma}(1 - h^{(i)})}$$

where $\tilde{\sigma}(1 - h^{(i)})$ is the estimate of the standard deviation at $\tilde{\varepsilon}^{(i)}$. We took $\tilde{\sigma}$ as the estimate of the model standard error before partitioning variance:

$$\tilde{\sigma} = \sqrt{\sigma_{\alpha}^2 + \sigma_{\varepsilon}^2}$$

and $h^{(i)}$ is the i th diagonal entry from the hat matrix \mathbf{H} :

$$\mathbf{H} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

and $h = \text{diag}(\mathbf{H})$. Thus $z^{(i)}$ represents the linear modeling analog to a z-score (e.g. $\frac{y^{(i)} - \bar{x}}{sd^{(i)}}$) and has units of standard deviation.

Supplemental Information

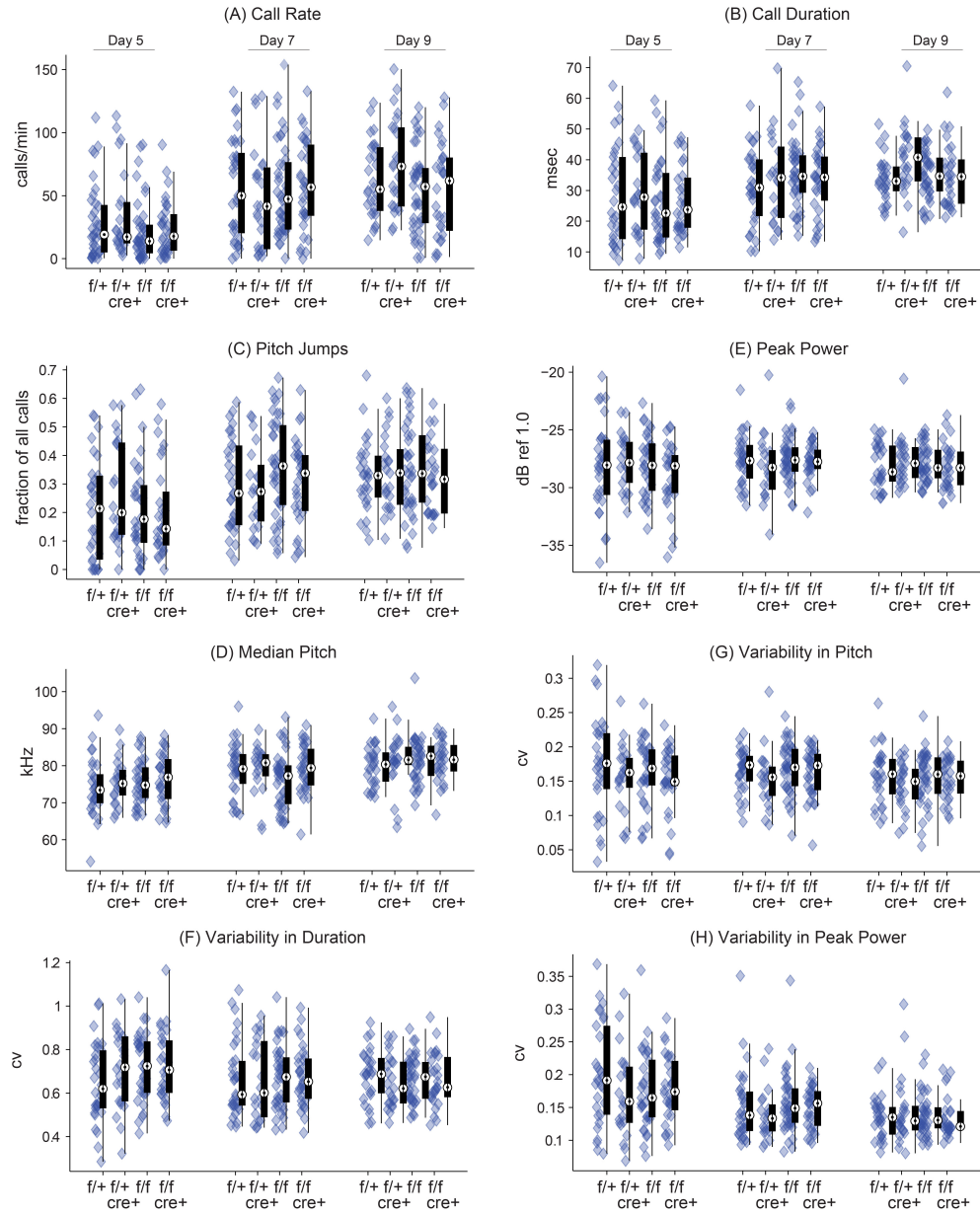


Figure S1: Conditional knockout of *Celf6* in dopaminergic neurons does not significantly alter features of USV in Cohort 1 of PCS

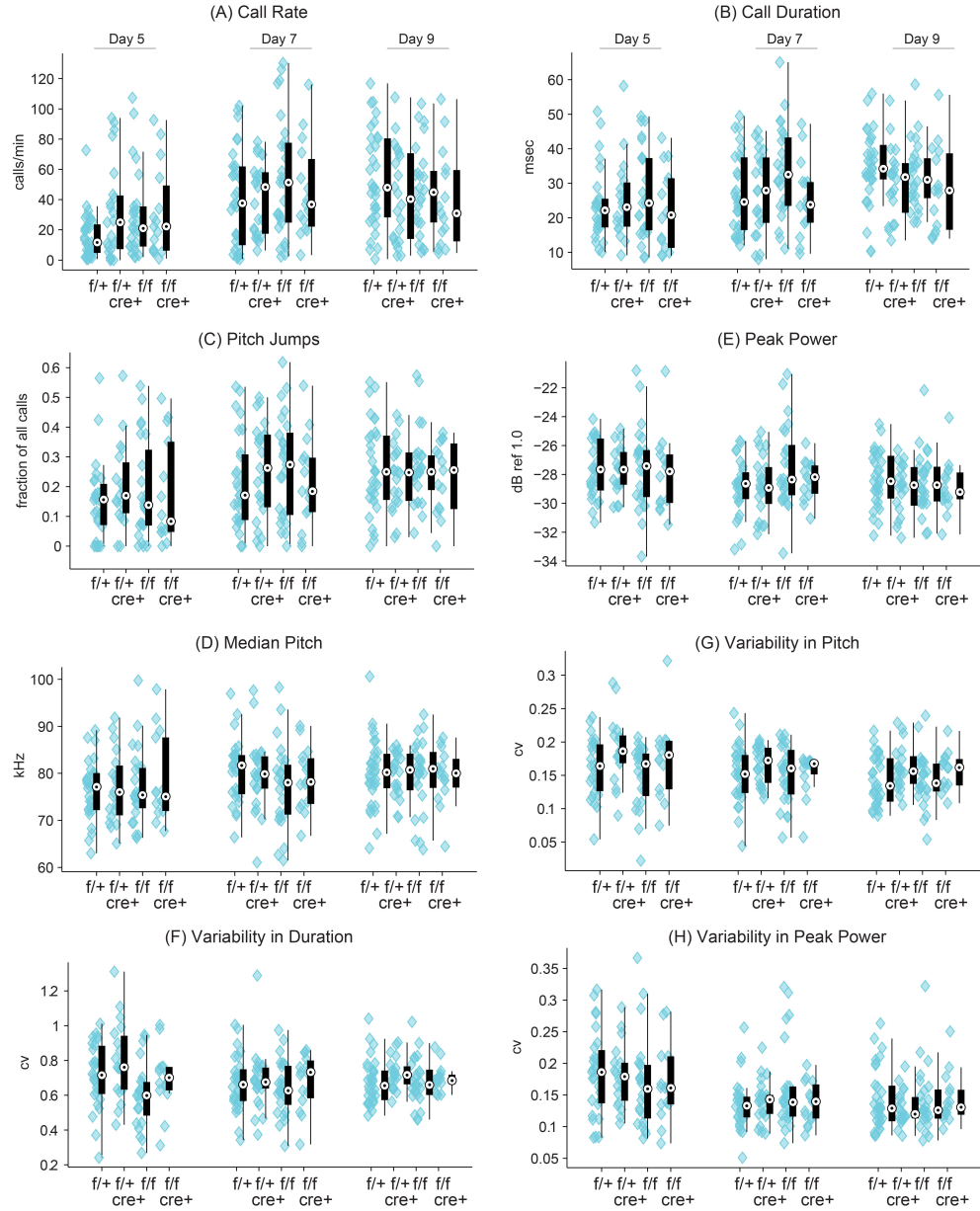


Figure S2: Conditional knockout of *Celf6* in GABA-ergic neurons does not alter features of USV in Cohort 2 of PCS.

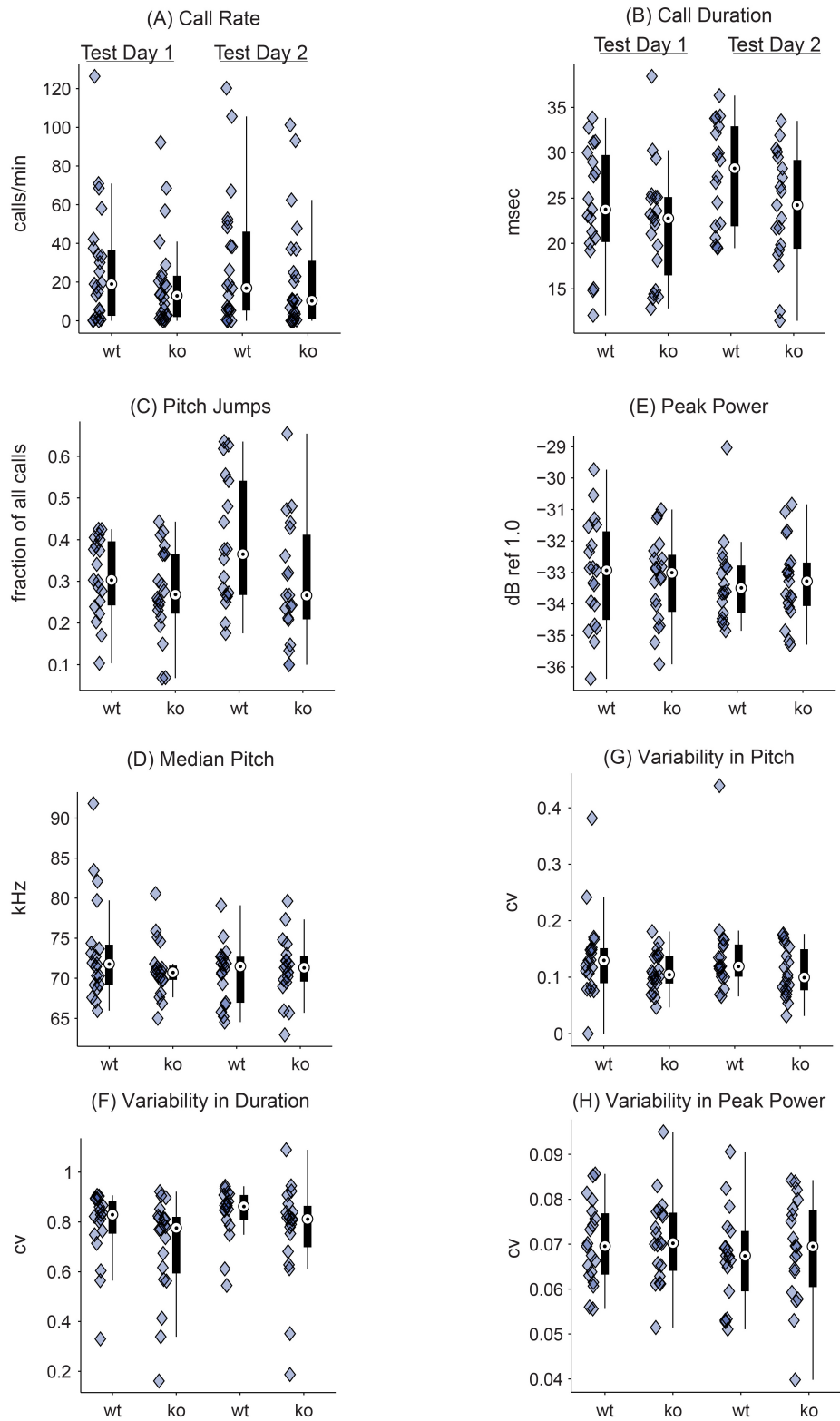


Figure S3: Global knockout of *Celf6* does not perturb USV features in adult male-female dyads.

Figure S1: Conditional knockout of *Celf6* in dopaminergic neurons does not significantly alter features of USV in Cohort 1 of PCS. A mouse model was generated possessing a conditional allele of the *Celf6* gene, in which the 4th exon was flanked by loxp sites (*Celf6^{fllox}*, abbreviated \textquotedblf\textquotedbl) in order to allow for Cre recombinase mediated excision using the DAT-Cre mouse, which expresses Cre in dopaminergic neurons under control of the dopamine transporter gene *Slc6a3*. A cross between conditional *Celf6^{fllox/+}* (\textquotedblf+\textquotedbl) heterozygotes positive for DAT-Cre and X *Celf6^{fllox/fllox}* (\textquotedblf/\textquotedblf") yielded a cohort of 133 animals constituting Cohort 1 of the PCS: 36 f/+, 23 f/+ Cre+, 42 f/f, 32 f/f Cre+, distributed across 18 litters of median size 8 animals, ranging from 4 to 11 animals per litter. Data are shown by genotype and postnatal day as bee plots for individual animals and boxplots showing the median (circle) and 25% (Q3) -75% (Q1) quartiles, with whiskers extending to most extreme datapoints not exceeding 1.5 x the interquartile range, for: (A) Call rate, (B) Call Duration, (C) Pitch Jumps, (D) Peak Power, (E) Median Pitch, (F) Variability in Pitch, (G) Variability in Duration, (H) Variability in Peak Power. Linear mixed models and likelihood ratio tests were used to detect main effects of genotype, postnatal day, or two-way interaction of genotype and postnatal day. Significant main effects of genotype were not detected for any variable, nor were any significant interactions detected. Collapsing across genotype, significant main effect of postnatal day was detected for: (A) Call rate (fitted as $\log((\text{calls}+1)/\text{minutes})$, $p = 6.5 \times 10^{-19}$), which was driven by a significant increase between days 5 and 7 (log fold change: 0.89, 95% c.i. [0.64, 1.13]) as well as a lesser but significant increase between days 7 and 9 (log fold change day 9 - 7: 0.31, 95% c.i. [0.05, 0.56]), (B) Call Duration (msec, average over all calls, $p = 6.5 \times 10^{-19}$), driven by a significant increase between days 7 and 5 (5.6 msec, 95% c.i [3.4, 7.9]) and a slightly smaller increase between days 7 and 9 (2.6 msec, 95% c.i. [0.4, 4.9]), (C) Pitch Jumps (fraction of all calls, $p = 1.2 \times 10^{-9}$), driven by a significant increase between days 7 and 5 (0.09 (9% increase), 95% c.i. [0.05, 0.12]) but no significant increase detected between days 7 and 9, (E) Median Pitch (kHz, $p = 1.9 \times 10^{-13}$), driven by a small but significant increase between days 7 and 5 (3.1 kHz, 95% c.i. [1.78, 4.42]) and a similar increase between days 9 and 7 (2.6 kHz, 95% c.i. [1.25, 3.90]), and (H) Variability in Peak Power (coefficient of variation (standard deviation over all calls/mean, per animal), $p = 5.8 \times 10^{-12}$), exhibiting a decrease in variability between days 5 and 7 (-0.032, 95% c.i. [-0.043, -0.020]), and smaller amount of decrease between days 7 and 9 (-0.012, 95% c.i. [-0.02, -4.8 $\times 10^{-4}$]).

Figure S2: Conditional knockout of Celf6 in GABA-ergic neurons does not alter features of USV in Cohort 2 of PCS. As in Supplemental Figure 1, conditional knockout of Celf6 in GABA-ergic neurons was performed by crossing Celf6^{fllox/+} (\textquotedblf/\textquotedbl) heterozygotes positive for VGAT-Cre and X Celf6^{fllox/fllox}, where the VGAT-Cre mouse expresses Cre recombinase under the control of the GABA transporter gene Slc32a1. This cross yielded a cohort of 105 animals constituting Cohort 2 of the PCS: 34 f/+, 24 f/+ Cre+, 29 f/f, 18 f/f Cre+, distributed across 15 litters of median size 8 animals, ranging from 2 to 9 animals per litter, with 12 of 15 litters possessing 7 to 9 animals. Data are shown by genotype and postnatal day as bee plots for individual animals and boxplots showing the median (circle) and 25% (Q3) -75% (Q1) quartiles, with whiskers extending to most extreme datapoints not exceeding 1.5 x the interquartile range, for: (A) Call rate, (B) Call Duration, (C) Pitch Jumps, (D) Peak Power, (E) Median Pitch, (F) Variability in Pitch, (G) Variability in Duration, (H) Variability in Peak Power. Linear mixed models and likelihood ratio tests were used to detect main effects of genotype, postnatal day, or two-way interaction of genotype and postnatal day. Significant main effects of genotype were not detected for any variable, nor were any significant interactions detected. Collapsing across genotype, significant main effect of postnatal day was detected for: (A) Call rate ($p = 5.54 \times 10^{-9}$), showing a significant increase between days 7 and 5 (log fold change in calls/min, 0.75, 95% c.i. [0.47, 1.05]), (B) Call duration ($p = 3.03 \times 10^{-7}$), showing a significant increase between days 7 and 5 comparable to Supplemental Figure 2 (4.7 msec, 95% c.i. [2.1,7.3], and a small increase between days 9 and 7 (2.5 msec, 95% c.i. [0.01, 5.03]), (C) Pitch Jumps ($p = 2.67 \times 10^{-5}$), as in Supplemental Figure 2 showing an increase between days 7 -5 in fraction of all calls (0.065, 95% c.i. [0.03, 0.10]) and no significant increase between days 9 and 7, (D) Peak Power ($p = 0.005$) showing a small decrease between days 7 and 5 (-0.55 dB, 95% c.i. [-1.08, -0.01]) (E) Median Pitch ($p = 0.002$), exhibiting a small increase between days 7 and 5 similar to Supplemental Figure 2 (2.0 kHz, 95% c.i. [0.42, 3.61] but no significant change between days 9 and 7, (F) Variability in Pitch ($p = 0.008$), showing a slight decrease in variability between days 7 and 5 (-0.013, 95% c.i. [-0.023, -1.9x10⁻³]) but no change between 9 and 7, and (H) Variability in Peak Power ($p = 4.25 \times 10^{-7}$), showing a comparable decrease in peak power variability between days 7 and 5 to Supplemental Figure 2 (-0.030, 95% c.i. [-0.043, -0.016]) but no change between days 7 and 9.

Figure S3: Global knockout of Celf6 does not perturb USV features in adult male-female

dyads. 24 Celf6 WT and 23 KO adult males (7-11 weeks) were tested on two test days with a different, stranger female each day as described in Methods. Data are shown by genotype and test day as bee plots for individual animals and boxplots showing the median (circle) and 25% (Q3) -75% (Q1) quartiles, with whiskers extending to most extreme datapoints not exceeding 1.5 x the interquartile range, for: (A) Call rate, (B) Call Duration, (C) Pitch Jumps, (D) Peak Power, (E) Median Pitch, (F) Variability in Pitch, (G) Variability in Duration, (H) Variability in Peak Power. Linear mixed models and likelihood ratio tests were used to detect main effects of genotype, postnatal day, or two-way interaction of genotype and postnatal day. No significant effects of genotype nor interactions between genotype and test day were detected. Collapsing across genotype, a significant effect of test day was detected in the case of (B) Call Duration (msec, $p = 0.004$), in which there was an increase in call duration by 2.1 msec (95% c.i. [0.76, 3.54]) on test day 2. A non-significant trend towards an increase in (C) Pitch Jumps (fraction of all calls, $p = 0.07$) was also observed between test days. USV features do not distinguish between vocalizations made by either animal in the dyadic assay and represent an aggregate of both animals, which may affect the precision of some measurements (especially if animals vocalize at the same time).

Chapter 3 References

1. Sewell, G. D. Ultrasonic Communication in Rodents. *Nature* **227**, 410 (July 1970).
2. Smith, J. C. Responses of adult mice to models of infant calls. *Journal of Comparative and Physiological Psychology* **90**, 1105–1115 (1976).
3. Ehret, G. Development of sound communication in mammals. *Advances in the Study of Behavior* **11**, 179–225 (1980).
4. Elwood, R. W. & Keeling, F. Temporal organization of ultrasonic vocalizations in infant mice. en. *Developmental Psychobiology* **15**, 221–227 (May 1982).
5. Hahn, M. E. *et al.* Genetic and Developmental Influences on Infant Mouse Ultrasonic Calling. II. Developmental Patterns in the Calls of Mice 2–12 Days of Age. en. *Behavior Genetics* **28**, 315–325 (July 1998).
6. Hofer, M. A., Shair, H. N. & Brunelli, S. A. Ultrasonic Vocalizations in Rat and Mouse Pups. en. *Current Protocols in Neuroscience*. (2014) (2001).
7. D’Amato, F. R. & Populin, R. Mother-offspring interaction and pup development in genetically deaf mice. en. *Behavior Genetics* **17**, 465–475 (Sept. 1987).
8. Hahn, M. E. & Lavooy, M. J. A Review of the Methods of Studies on Infant Ultrasound Production and Maternal Retrieval in Small Rodents. en. *Behavior Genetics* **35**, 31–52 (Jan. 2005).
9. Holy, T. E. & Guo, Z. Ultrasonic Songs of Male Mice. *PLoS Biol* **3**, e386 (Nov. 2005).

10. Burkett, Z. D., Day, N. F., Peñagarikano, O., Geschwind, D. H. & White, S. A. VoICE: A semi-automated pipeline for standardizing vocal analysis across models. en. *Scientific Reports* **5**. (2015) (May 2015).
11. Branchi, I., Santucci, D. & Alleva, E. Ultrasonic vocalisation emitted by infant rodents: a tool for assessment of neurobehavioural development. *Behavioural Brain Research* **125**, 49–56 (Nov. 2001).
12. Scattoni, M. L., Crawley, J. & Ricceri, L. Ultrasonic vocalizations: A tool for behavioural phenotyping of mouse models of neurodevelopmental disorders. *Neuroscience & Biobehavioral Reviews. Risk Factors for Mental Health: Translational Models from Behavioral Neuroscience* **33**, 508–515 (Apr. 2009).
13. Scattoni, M. L. *et al.* Reduced ultrasonic vocalizations in vasopressin 1b knockout mice. *Behavioural Brain Research* **187**, 371–378 (Mar. 2008).
14. Dougherty, J. D. *et al.* The disruption of *Celf6*, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors. *Journal of Neuroscience* **33**, 2732–2753 (2013).
15. Yang, M. *et al.* 16p11.2 Deletion Syndrome Mice Display Sensory and Ultrasonic Vocalization Deficits During Social Interactions. en. *Autism Research* **8**, 507–521 (Oct. 2015).
16. Fujita, E. *et al.* Ultrasonic vocalization impairment of *Foxp2* (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. en. *Proceedings of the National Academy of Sciences* **105**, 3117–3122 (Feb. 2008).
17. Barnes, T. *et al.* A Mutation Associated with Stuttering Alters Mouse Pup Ultrasonic Vocalizations. *Current Biology* **26**, 1009–1018 (Apr. 2016).
18. Motomura, N. *et al.* A Comparative Study of Isolation-Induced Ultrasonic Vocalization in Rodent Pups. *Experimental Animals* **51**, 187–190 (2002).

19. Stoeger-Horwath, A. S., Stoeger, S., Schwammer, H. M. & Kratochvil, H. Call repertoire of infant African elephants: First insights into the early vocal ontogeny. *The Journal of the Acoustical Society of America* **121**, 3922–3931 (June 2007).
20. Elliot, O. & Scott, J. P. The Development of Emotional Distress Reactions to Separation, in Puppies. *The Journal of Genetic Psychology* **99**, 3–22 (1961).
21. Shair, H. N. Acquisition and expression of a socially mediated separation response. *Behavioural Brain Research. Mammalian Vocalization: Neural, Behavioural, and Environmental Determinants* **182**, 180–192 (Sept. 2007).
22. Chesler, E. J., Wilson, S. G., Lariviere, W. R., Rodriguez-Zas, S. L. & Mogil, J. S. Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive. *Neuroscience & Biobehavioral Reviews* **26**, 907–923 (Dec. 2002).
23. McClearn, G. E. Contextual genetics. *Trends in Genetics* **22**, 314–319 (June 2006).
24. Ramos, A. Animal models of anxiety: do I need multiple tests? *Trends in Pharmacological Sciences* **29**, 493–498 (Oct. 2008).
25. Vangeneugden, T., Laenen, A., Geys, H., Renard, D. & Molenberghs, G. Applying linear mixed models to estimate reliability in clinical trial data with repeated measurements. *Controlled Clinical Trials* **25**, 13–30 (Feb. 2004).
26. Mroczek, D. K. & Spiro, A. Modeling Intraindividual Change in Personality Traits: Findings From the Normative Aging Study. en. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* **58**, P153–P165 (May 2003).
27. Hoffman, L. Multilevel Models for Examining Individual Differences in Within-Person Variation and Covariation Over Time. *Multivariate Behavioral Research* **42**, 609–629 (Dec. 2007).

28. Boncoraglio, G. & Saino, N. Barn swallow chicks beg more loudly when broodmates are unrelated. en. *Journal of Evolutionary Biology* **21**, 256–262 (Jan. 2008).
29. Roulin, A., Dreiss, A., Fioravanti, C. & Bize, P. Vocal sib–sib interactions: how siblings adjust signalling level to each other. *Animal Behaviour* **77**, 717–725 (Mar. 2009).
30. Dall, S. R. X., Houston, A. I. & McNamara, J. M. The behavioural ecology of personality: consistent individual differences from an adaptive perspective. en. *Ecology Letters* **7**, 734–739 (Aug. 2004).
31. Spielberger, C. D. State-Trait Anxiety Inventory. en. *The Corsini Encyclopedia of Psychology*. (2016) (2010).
32. Malkova, N. V., Yu, C. Z., Hsiao, E. Y., Moore, M. J. & Patterson, P. H. Maternal immune activation yields offspring displaying mouse versions of the three core symptoms of autism. *Brain, Behavior, and Immunity* **26**, 607–616 (May 2012).
33. Venerosi, A., Ricceri, L., Scattoni, M. L. & Calamandrei, G. Prenatal chlorpyrifos exposure alters motor behavior and ultrasonic vocalization in cd-1 mouse pups. *Environmental Health* **8**, 12 (2009).
34. Golub, Y. *et al.* Effects of In utero environment and maternal behavior on neuroendocrine and behavioral alterations in a mouse model of prenatal trauma. en. *Developmental Neurobiology*, n/a–n/a (Mar. 2016).
35. Thornton, L. M., Hahn, M. E. & Schanz, N. Genetic and Developmental Influences on Infant Mouse Ultrasonic Calling. III. Patterns of Inheritance in the Calls of Mice 3–9 Days of Age. en. *Behavior Genetics* **35**, 73–83 (Jan. 2005).
36. Gordus, A., Pokala, N., Levy, S., Flavell, S. & Bargmann, C. Feedback from Network States Generates Variability in a Probabilistic Olfactory Circuit. *Cell* **161**, 215–227 (Apr. 2015).

37. Nakagawa, S. & Schielzeth, H. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. en. *Biological Reviews* **85**, 935–956 (Nov. 2010).
38. Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**, 133–142 (2013).
39. Liu, R. C., Miller, K. D., Merzenich, M. M. & Schreiner, C. E. Acoustic variability and distinguishability among mouse ultrasound vocalizations. eng. *J Acoust Soc Am* **114**, 3412–3422 (Dec. 2003).
40. Liu, R. C. & Schreiner, C. E. Auditory Cortical Detection and Discrimination Correlates with Communicative Significance. *PLoS Biol* **5**, e173 (June 2007).
41. Ehret, G. Adaptations in the Mouse Auditory System for Perception of Ultrasonic Communication Calls. *Journal of Evolutionary Biochemistry and Physiology* **37**, 562 (Sept. 1, 2001).
42. Ehret, G. & Haack, B. Ultrasound recognition in house mice: Key-Stimulus configuration and recognition mechanism. *Journal of Comparative Physiology â A* **148**, 245 (June 1, 1982).
43. Okon, E. E. The effect of environmental temperature on the production of ultrasounds by isolated non-handled albino mouse pups. en. *Journal of Zoology* **162**, 71–83 (Sept. 1970).
44. Branchi, I., Santucci, D. & Alleva, E. Ultrasonic vocalisation emitted by infant rodents: a tool for assessment of neurobehavioural development. *Behavioural Brain Research* **125**, 49–56 (Nov. 2001).
45. Warburton, V. L., Sales, G. D. & Milligan, S. R. The emission and elicitation of mouse ultrasonic vocalizations: The effects of age, sex and gonadal status. *Physiology & Behavior* **45**, 41–47 (Jan. 1989).

46. Neunuebel, J. P., Taylor, A. L., Arthur, B. J. & Egnor, S. R. Female mice ultrasonically interact with males during courtship displays. en. *eLife* **4**, e06203 (May 2015).
47. Hofmann, H. A. Functional genomics of neural and behavioral plasticity. en. *Journal of Neurobiology* **54**, 272–282 (Jan. 2003).
48. Lee, J. S. F. & Berejikian, B. A. Stability of behavioral syndromes but plasticity in individual behavior: consequences for rockfish stock enhancement. en. *Environmental Biology of Fishes* **82**, 179–186 (Oct. 2007).
49. Roberts, B. W. & DelVecchio, W. F. The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin* **126**, 3–25 (2000).
50. Bell, A. M., Hankison, S. J. & Laskowski, K. L. The repeatability of behaviour: a meta-analysis. *Animal Behaviour* **77**, 771–783 (Apr. 2009).
51. Benus, R. F., Koolhaas, J. M. & Oortmerssen, G. A. v. Individual Differences in Behavioural Reaction to a Changing Environment in Mice and Rats. *Behaviour* **100**, 105–122 (1987).
52. Wang, F. *et al.* Bidirectional control of social hierarchy by synaptic efficacy in medial prefrontal cortex. eng. *Science (New York, N.Y.)* **334**, 693–697 (Nov. 2011).
53. Arriaga, G., Zhou, E. P. & Jarvis, E. D. Of Mice, Birds, and Men: The Mouse Ultrasonic Song System Has Some Features Similar to Humans and Song-Learning Birds. *PLoS ONE* **7**, e46610 (Oct. 2012).
54. Michetti, C. Modeling Social Communication Deficits in Mouse Models of Autism. *Autism- Open Access* **01**. (2016) (2012).
55. Roulet, F. I., Lai, J. K. Y. & Foster, J. A. In utero exposure to valproic acid and autism — A current review of clinical and animal studies. *Neurotoxicology and Teratology*.

- Special Issue: Environmental Influences and Emerging Mechanisms in the Etiology of Autism* **36**, 47–56 (Mar. 2013).
56. Enard, W. *et al.* A Humanized Version of Foxp2 Affects Cortico-Basal Ganglia Circuits in Mice. *Cell* **137**, 961–971 (May 2009).
 57. Scattoni, M. L., Gandhi, S. U., Ricceri, L. & Crawley, J. N. Unusual Repertoire of Vocalizations in the BTBR T+tf/J Mouse Model of Autism. *PLOS ONE* **3**, e3067 (Aug. 2008).
 58. Knutson, B., Burgdorf, J. & Panksepp, J. Ultrasonic vocalizations as indices of affective states in rats. *Psychological Bulletin* **128**, 961–977 (2002).
 59. Knutson, B., Burgdorf, J. & Panksepp, J. Anticipation of play elicits high-frequency ultrasonic vocalizations in young rats. *Journal of Comparative Psychology* **112**, 65–73 (1998).
 60. Burgdorf, J. *et al.* Ultrasonic vocalizations of rats (*Rattus norvegicus*) during mating, play, and aggression: Behavioral concomitants, relationship to reward, and self-administration of playback. *Journal of Comparative Psychology* **122**, 357–367 (2008).
 61. Neilans, E. G., Holfoth, D. P., Radziwon, K. E., Portfors, C. V. & Dent, M. L. Discrimination of Ultrasonic Vocalizations by CBA/CaJ Mice (*Mus musculus*) Is Related to Spectrotemporal Dissimilarity of Vocalizations. *PLOS ONE* **9**, e85405 (Jan. 2014).
 62. Portfors, C. V., Roberts, P. D. & Jonson, K. Over-representation of species-specific vocalizations in the awake mouse inferior colliculus. *Neuroscience* **162**, 486–500 (Aug. 2009).
 63. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models using lme4. *arXiv:1406.5823 [stat]*. arXiv: 1406.5823. (2016) (June 2014).
 64. R Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing. Vienna, Austria*, <http://www.R-project.org> (2013).

Chapter 4

The Functions of CELF RNA binding proteins

Abstract

RNA binding proteins (RBPs) are important regulators of RNA molecules involved throughout their life-cycle from pre-mRNA synthesis and processing, to mature RNA stability, translation, and localization. RBPs can be defined by whether they bind their targets in sequence non-specific or a sequence specific manner, and these RNA-protein interactions are mediated via distinct families of RNA binding domains (RBDs). A single RBD does not confer much binding specificity and thus most RBPs possess more than one RBD. CUGBP and ELAV-like family member 6 (CELF6), which we identified as associated with behavioral changes in knockout mice and enriched in monoaminergic neurons, is a member of the CELF family of RBPs which contain three RBDs of the RNA Recognition Motif (RRM) family. The binding targets of CELF6 are currently unidentified. Using modern techniques such

as crosslinking immunoprecipitation (CLIP) followed by next generation sequencing, we can determine the identity of RNA targets of an RBP and where on each RNA the RBP is bound. As the CELF family is diverse in its functions, with other members notably regulating alternative splicing, mRNA stability, and translation, the knowledge of CELF6's targets, and its molecular function, is essential in understanding its biological role in the brain. In this section, I review the literature related to interactions between RBPs and their RNA targets, how these interactions can be defined, and the known functions of other CELF family members.

Introduction

RNA binding proteins (RBPs) are involved in every step of a messenger RNA (mRNA) molecule's life cycle: from pre-mRNA formation and splicing into mature mRNA, capping and polyadenylation, trafficking from the nucleus into the cytoplasm, localization in the cell, maintenance and degradation, sequestration under stress, and ribosomal machinery recruitment and translation into polypeptide sequence [1–7]. Because RBPs are involved throughout the life of an mRNA they have the opportunity to act as key regulators of most cellular functions.

Neurons are highly polarized cells, and it has long been appreciated that RNA molecules must be trafficked appropriately in order to generate patterns of localized translation required for synaptic function [7, 8]. Thus, unsurprisingly, aberrations to a number of RNA binding proteins are implicated in neurological disorder, such as Fragile X Mental Retardation Protein (FMRP, gene *FMR1*) which induces ribosomal stalling and is involved in mRNA localization and translation regulation in neurons [9, 10]. Mutations to *FMR1* are associated with ASD and intellectual disability in humans [11, 12]. Members of the

CUGBP and ELAV-like Factor (CELF) family of RNA binding proteins, which have multiple functions including alternative splicing and regulation of mRNA decay [3], are known to be enriched for expression in the central nervous system, and members have been associated with neurodevelopmental abnormalities such as seizures and ASD [13, 14]. Our group identified mutations to *CELF6* as associated with ASD in humans. Mice with loss of this protein exhibit reductions to reward-seeking behavior and neonatal ultrasonic vocalizations. As described in preceding sections, *CELF6* expression is enriched in monoaminergic neurotransmitter populations such as dopaminergic and serotonergic cells, however single cell type knockouts have thus far been unable to replicate the associated behavioral phenotypes. Thus either either the specific cell type underlying *CELF6*'s association with behavioral phenotype has not yet been identified or its effect is mediated by multiple redundant systems.

In order to better understand how *CELF6* may play a role in behavior, its biological function must be understood. Therefore, in this chapter, I review the foundational concepts and techniques upon which my experiments to understand *CELF6* protein function are based. I review the basic structure and binding properties of RNA binding proteins such as *CELF6*. I go on to briefly summarize how target-protein interactions are defined experimentally and how binding sites may enable us to hypothesize about function. Finally, I review what has been previously determined for CELF family RBPs with respect to their regulation of target mRNAs. It is our hope that defining *CELF6* and its role in regulating its targets will enable future experimentation to understand the plausible mechanisms by which it regulates behavior.

Interactions with RNA through RNA Binding Domains

RBPs include proteins which bind RNA both specifically and non-specifically with respect to the RNA sequence. Common RBPs that bind RNA non-specifically include RNA polymerase which synthesizes RNA from DNA template and associates with RNA:DNA hybrids via positively charged amino acid side chains and the negatively charged phosphate backbone of the nucleic acid [15]. RBPs that bind RNA in sequence-specific ways do so through the presence of RNA Binding Domains (RBDs) in the protein's sequence. RNAs can be ontologically organized via which families of RBDs they possess. Individual RBDs typically associate with a small number of nucleotides (as few as two nucleotides), and thus higher specificity is achieved in RBPs by possession of multiple RBDs separated by linker sequences [16].

Different RBDs are characteristic of different families of RBPs. Pumilio homology domains (PUM-HDs) have repeated 37 amino acids segments which stack alternating with the RNA bases, leaving the phosphate backbone exposed to solvent [17]. K Homology (KH) domains (found in the RBP NOVA [18], members of the heterogeneous ribonucleoprotein particle proteins (HNRNP) family of proteins [19], and FMRP [20]) contain 70 amino acids with a central (I/L/V)-IGxxGxx-(I/L/V) sequence, and bind at least 4 nucleotides [16]. Among these proteins, NOVA is a nuclear localized neuronal RBP involved in splicing regulation [21]. HNRNPK exhibits diverse functions including regulating RNA splicing and stability, both activation and repression of translation, and even transcriptional activation and repression [1]. FMRP is a translation regulator as described in preceding sections. Thus the specific RBDs that a RBP possesses are not necessarily predictive of the RBPs molecular function. It is likely that the binding location on the RNA, localization in the cell, and protein-protein interactions confer added specificity to the functional consequence of a RBP's binding. CELF proteins contain the RNA recognition motif (RRM) family of RBDs. RRM are approximately 90 amino acids in size and

contain 2 conserved stretches: (K/R)-G-(F/Y)-(G/A)-(F/Y)-(V/I/L)-x-(F/Y) and (V/I/L-F/Y-V/I/L-x-N/L). Each domain has sites which bind 2-4 nucleotides, and thus a full length CELF protein with 3 RRMs can potentially bind up to 12 nucleotides [16], although recent work determining RNA binding preferences *in vitro* have determined binding sites of 7 nucleotides for CELF proteins 3,4, and 6 [22]. The domain structure of human CELF6 is shown below in Figure 1A. CELF6 like other CELF proteins has two N-terminal RRM domains adjacent to one another and a third C-terminal RRM domain, separated by a "divergent" domain (DD) less conserved throughout the CELF family with little predicted secondary structural content. It is believed the DD may function as a flexible linker region or facilitate protein:protein interactions [5].

I inspected the CELF6 protein sequence and its mammalian homologues across 4 species (*H. sapiens*, *P. troglodytes*, *M. musculus*, *R. norvegicus*) as well as its homologues in chicken (*G. gallus*) and zebrafish (*D. rerio*) and the homologue of the CELF family in *D. melanogaster*, the translation regulator Bruno[23]. The amino acid sequences were subjected to multiple sequence alignment using CLUSTAL Omega [24] and to look at the rise and fall in sequence variation, I computed the Wu and Kabat variability score across the multiple alignments [25]. Peaks indicate high variability and valleys indicate low variability on this scale, and these results are shown in relation to the human domain structure in Figure 1B. The N-terminal region and the DD both exhibit the highest variability across taxa, with RRM1 and RRM3 having low variability and thus increased conservation across all species considered. Human CELF6 is a 481 amino acid protein, so for visual ease, rather than presenting the entire multiple sequence alignment, I have visualized the output of CLUSTAL Omega using a novel amino acid color coded scheme with black indicating alignment gaps. Higher variability scores for RRM2 are observed, but upon inspection of the alignment it appears this is driven by the presence of the chicken, zebrafish, and fruitfly sequences, and the RRM2 shows very similar sequence across all

Figure 1

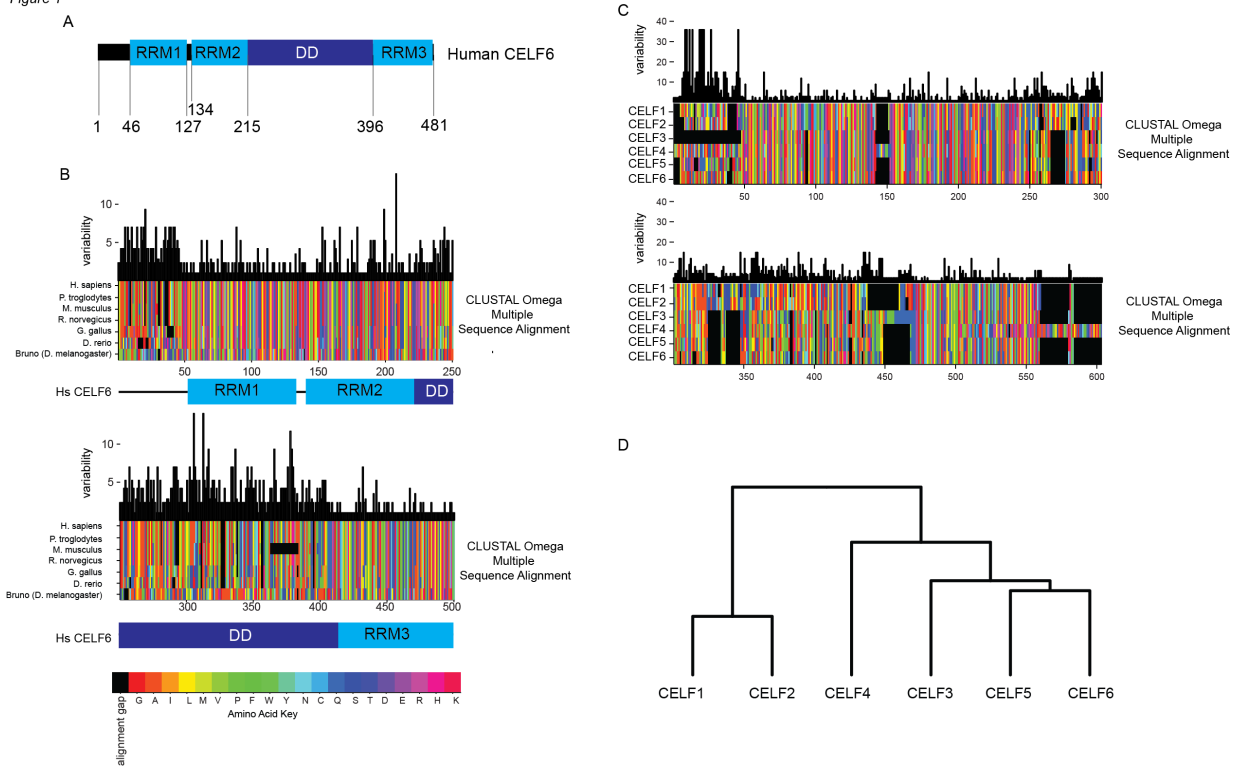


Figure 1: The CELF family of RNA binding proteins shows homologous structure with 3 RRM and two distinct subclasses.

(A) A linear schematic of the structure of human CELF6 showing all 3 RRM and the DD. (B) CLUSTAL multiple sequence alignment for CELF6 across 4 mammalian species (H. sapiens, P. troglodytes, M. musculus, R. norvegicus) and its homologues in G. gallus, D. rerio and D. melanogaster. Multiple sequence alignments are color coded according to the amino acid key given, with gaps in alignment represented in black. Above the alignment is shown the variability in the amino acid position across species using the Wu and Kabat variability score. Below the alignment is the structure of human CELF6 according to its position in the alignment. (C) Multiple sequence alignments as in (B) across human CELF proteins. (D) Clustering dendrogram based on the Hamming distance between multiple alignments in (C).

four mammalian species.

The mammalian CELF family consists of 6 proteins, and their multiple sequence alignments are shown in Figure 1C as color-coded visualization with variability scores above. All possess high variability at the N-terminal region of the protein and an increase in variability near the DD. Inspecting the alignment, it appears that CELF1 and CELF2 are more highly similar than CELF3-6. Indeed, using the multiple sequence alignment to hierarchically cluster the CELFs, CELF1 and CELF2 form a distinct cluster, with CELF3-6 in a second cluster, and CELF5 and CELF6 the most similar. This is shown as a dendrogram in Figure 1D.

The mammalian CELF family consists of 6 proteins, and the human multiple sequence alignments are shown in Figure 1C as color-coded visualization with variability scores above. All possess high variability at the N-terminal region of the protein and an increase in variability near the DD. Inspecting the alignment, it appears that CELF1 and CELF2 are more highly similar than CELFs 3-6. Indeed, using the multiple sequence alignment to hierarchically cluster the CELFs, CELF1 and CELF2 form a distinct cluster, with CELFs3-6 in a second cluster, and CELF5 and CELF6 the most similar. This is shown as a dendrogram constructed from in Figure 1D. Thus, all CELF proteins are all highly similar, particularly in their RRM. This suggests they may share similar targets as well. Thus it is necessary to understand the specific RBP-RNA interactions and I review methods to do so in the following section.

Identifying and defining RBP:RNA interactions

In order to study the function of RBPs, it is desirable to identify its RNA targets and the location on these targets where the RBP binds. RBP target RNAs are traditionally identified by some manner of affinity purification: either purification of RNA followed by

assay of associated proteins, or by purification of a specific RBP and assay of the associated RNA. In the 1990s it was learned that ultraviolet radiation (e.g. 254 nm) could be used to covalently crosslink amino acids which were positioned near nucleotides [26]. Methods to purify RNA may make use of this crosslinking technique to purify proteins associated with, for example, all polyadenylated mRNA using oligo-dT capture and followed by mass spectrometry to identify proteins [27], or by using anti-sense probes for specific RNA molecules. Probe mRNAs can be biotinylated and immobilized on a streptavidin-coated solid support [28], or using other high affinity tagging methods such as the MS2- and MS2 coat protein protein system [29]. With RNA-protein UV crosslinking, stringent washes (e.g. high salt concentration, detergent concentrations) can be used to remove non-specific binding events. In all cases, after purification, the complexes must be eluted, and digested polypeptides can be analyzed by mass spectrometry to identify which proteins are associated [30–33].

Using the RBP as prey for affinity purification requires the availability of robust antibodies or use of epitope tagging. Purifying a specific RBP and its associated RNA was originally devised without crosslinking ("RNA Immunoprecipitation", RIP) [34] but including UV crosslinking allows for more stringent washing and this is termed Crosslinking Immunoprecipitation (CLIP) [35]. CLIP'd RNA sequences were originally subcloned and subjected to Sanger sequencing for identification [36] but modern approaches employ next-generation sequencing [9, 37–41]. Statistical determination of significant peaks in sequencing read density over background can be used to identify which RNA molecules a specific RBP binds, and also where on the body of the RNA this occurs. Once the location on the RNA where a RBP binds is known, the sequence can be scanned for evidence of specific binding motifs. These sequence motifs should show enrichment in bound versus input or unbound fraction of RNAs. Some functionality may be able to be hypothesized given the location of binding. For example, if CLIP peaks occur at anno-

tated splice sites on pre-mRNAs or at exon-junctions, it may be that the RBP of interest regulates splicing at those locations.

In addition to annotated splice sites and exon junctions, binding in the untranslated regions of mRNA such as the 5' untranslated region (5'UTR) or 3' untranslated region (3'UTR) may indicate regulation of mRNA stability or translation. Certain motifs, such as the TOP motif (Terminal oligopyrimidine) C[CU]_{x4-15} [42] exist in the 5' UTR and are known to be correlated with the translation of mRNAs in response to growth stimuli [43, 44]. Although primarily known from viruses [45], internal ribosome entry segments (IRES) are also found endogenously in about 10% of eukaryotic 5'UTRs [46] where they are bound by IRES Trans activating factors (ITAFs) [47] and promote cap-independent translation, such as during apoptosis when normal cap-dependent translation is suppressed [48].

In the 3'UTR, the commonest example of a regulatory sequence bound by RBPs is the polyA tail itself, which is bound by polyA binding proteins, with increased translational repression and lowered stability of polyA tails under 50 nucleotides in length [49]. Polyadenylation is itself regulated by other interactions between the 3'UTR elements and RBPs, such as by Cytoplasmic Polyadenylation Elements (CPEBs) which associate with U-rich sequences 5'-UUUU[A]₁₋₂-U-3' and serve to activate or repress polyadenylation [50, 51]. AU rich elements (AREs) are bound by several classes of RBPs termed ARE-binding proteins (AREBPs) including the ELAV family which stabilizes transcripts and promotes translation, and the TIA family which promotes degradation and translational repression [52, 53]. Binding in the 3'UTR may also indicate that the RBP facilitates subcellular localization of the mRNA such as that of CaMKIIa [54] or calreticulin [55] in neurons. Thus taken together, once the targets of a RBP are known, the location and sequence at which the RBP binds can help to formulate hypotheses about its downstream molecular function, guiding further experimentation. Additionally, knowledge of the specific targets of a RBP can help generate hypotheses about its overall biological significance where the

downstream functions of targets are known.

Regulation of mRNAs by CELF proteins

The preceding sections have briefly surveyed the RNA:RBP binding interactions both in terms of protein domains and sequence elements, and how experimental techniques such as CLIP are used to identify and define these interactions. In this section, I would like to proceed to discuss what is currently known about the CELF family of RBPs setting the foundation for my work on CELF6. As the CELF family members all possess similar sequence and domain structure (Figure 1), understanding the prior literature on CELF proteins may lend insight into CELF6's possible functions.

In general, RBPs can exhibit a variety of functions uncoupled to their specific RBD context. For example, in addition to the CELFs, several heterogeneous ribonucleoprotein particle proteins (HNRNPs), ELAV-like protein human antigen R (HuR), and SRSF1/2 all contain RRM domains [56], and these proteins exhibit diverse functions. For example, HNRNPA and HNRNPB are known splicing repressors which can act to suppress splicing at specific splice acceptors to promote use of more distal sites by blocking exon recognition [57], and HuR binds AU-rich elements and promotes mRNA stabilization [58] and can promote dissociation of the RNA-induced silencing complex [59]. All of these proteins use RRM domains to facilitate association with their target mRNAs, but their specific functions differ.

For the CELF family of RRM containing proteins, most of the literature on molecular functions has focused on CELF1, also known as CUG-binding protein 1 (CUGBP1). As its name suggests, CELF1 was identified via its binding to CUG RNA sequence elements and is ubiquitously expressed [60]. CUG-containing sequences are present in the 3' UTR of myotonin protein kinase gene (*DMPK*), expansions of which cause of myotonic muscu-

lar dystrophy (DM) [61], and CELF1 was primarily studied in the context of this disease. In DM, muscle cells show upregulation of CELF1 protein expression and this upregulation is capable of inducing changes to splicing in other genes containing CUG repeats such as skeletal muscle specific chloride channel 1 (*CLC1*) and insulin receptor (*IR*) [62]. DM skeletal muscle tissue also shows elevated levels of MEF2A transcription factor associated with delayed myogenesis, and CELF1 is capable of binding MEF2A mRNA and enhancing its translation [63]. In addition to changes to splicing and translation, CELF1 can bind AU-rich elements in *TNF α* and *c-fos*, a process which is associated with subsequent recruitment of polyA-specific ribonuclease leading to poly-deadenylation and degradation [64]. The *Xenopus* homologue of CELF1, Embryo Deadenylation Element Binding Protein (EDEN-BP) also stimulates deadenylation and destabilization of target mRNAs, and furthermore requires oligomerization of the protein to do so [65, 66]. Thus CELF1 is a multi-functional protein, originally characterized in skeletal muscle in the context of DM, and is capable of promoting alternative splicing, enhancing translation, and promoting mRNA degradation.

CELF2, also referred to as neuroblastoma apoptosis-related RNA binding protein (NA-POR) and ETR3, is expressed across a variety of tissues including brain, heart, lung, and skeletal muscle, as well as late embryonic day development central nervous system [67]. CELF2, like CELF1 is also associated with alternative splicing activity in skeletal muscle using minigene reporter assays [68, 69]. Interestingly, CELF2 has also been associated with RNA editing - binding to AU rich sequences ahead of edited cytidine and repressing editing of the apolipoprotein B (apoB) mRNA [70] involved in atherosclerosis and cardiovascular disease. Like its relative CELF1, CELF2 is also found to regulate mRNA stability and translation in the cytoplasm. CELF2 is able to associate to AU rich sequences in the 3'UTR of cyclooxygenase-2 (COX2), and both stabilizes the COX2 mRNA while also inhibiting its translation [71].

CELF3-6 are all enriched for expression in the brain with CELF6 additionally found in the kidney and testes. There are conflicting reports indicating CELF4 is also expressed broadly or restricted to the nervous system [3, 72]. *Celf4* knockout mice experience convulsive and non-convulsive seizures when electrically stimulated in the cortex (where CELF4 has highest expression) [73]. Upon identifying CELF4 targets by CLIP, it was found that CELF4 predominantly associates in the 3'UTR region of mRNAs enriched for synaptic function, and sequences under CLIP peaks show UGU-containing motifs [74]. There is little literature on the roles of CELF3, CELF5, or CELF6 in the brain. One study shows an association with a polymorphism in CELF5 and schizophrenia [75]. Aside from our own group's work defining the behavior of *Celf6*^{-/-} mice and describing CELF6's pattern of expression in the brain [14, 76], there is one report that like CELF2 above, CELF6 is able to activate exon inclusion of a minigene *in vitro* and the same group also demonstrated these results for CELFs 3,4, and 5 [77]. However the binding targets of CELF6 have not yet been defined nor has the function this protein has on these targets.

Thus the emerging story of the CELF family of RBPs is that they may all act multifunctionally on different aspects of mRNA regulation, and this is largely supported by a history of work on CELFs 1 and 2. The *CELF4* and *CELF6* genes have been linked to neurological phenotypes in mice. CELF4 is enriched for expression in the cortex [74], whereas CELF6 shows primary expression in monoaminergic neurotransmitter populations, and limbic structures [76]. In Chapter 5 of this dissertation, I go on to show that CELF4 and CELF6 are both able to induce transcript repression *in vitro*. If it is the case that these proteins *in vivo* exhibit redundant functions, then perhaps the cell types in which they are expressed determine to some extent the phenotypic outcome when these proteins are absent.

Summary

In summary, CELF proteins are a versatile family of RBPs, and members exhibit the ability to regulate mRNA splicing, stability, and translation. Thus, one may hypothesize that CELF6, the specific object of investigation in this work, may take on any of these functions in the cell. Using the modern molecular techniques of CLIP followed by next generation sequencing, I will be able to define the targets of CELF6 and use the knowledge of where CELF6 CLIP peaks occur to hypothesize how it may act on the mRNA molecules that it regulates. In the subsequent chapter, I present the current unpublished work regarding identification of CELF6 targets, as well as the results from a massively parallel reporter assay I have employed to study the downstream consequence of CELF6 interaction with target mRNAs.

Chapter 4 References

1. Bomsztyk, K., Denisenko, O. & Ostrowski, J. hnRNP K: one protein multiple processes. *BioEssays : news and reviews in molecular, cellular and developmental biology* **26**, 629–638 (6 June 2004).
2. Busch, A. & Hertel, K. J. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdisciplinary Reviews: RNA* **3**, 1–12 (2012).
3. Dasgupta, T. & Ladd, A. N. The importance of CELF control: molecular and biological roles of the CUG-BP, Elav-like family of RNA-binding proteins. *Wiley Interdisciplinary Reviews: RNA* **3**, 104–121 (2012).
4. Fleming, V. A., Geng, C., Ladd, A. N. & Lou, H. Alternative splicing of the neurofibromatosis type 1 pre-mRNA is regulated by the muscleblind-like proteins and the CUG-BP and ELAV-like factors. *eng. BMC molecular biology* **13**, 35 (2012).
5. Barreau, C., Paillard, L., Méreau, A. & Osborne, H. B. Mammalian CELF/Bruno-like RNA-binding proteins: molecular characteristics and biological functions. *Biochimie* **88**, 515–525 (2006).
6. Louis, I. V.-S., Dickson, A. M., Bohjanen, P. R. & Wilusz, C. J. CELFish ways to modulate mRNA decay. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1829**, 695–707 (2013).
7. Schuman, E. M., Dynes, J. L. & Steward, O. Synaptic Regulation of Translation of Dendritic mRNAs. *en. The Journal of Neuroscience* **26**, 7143–7146 (July 2006).

8. Holt, C. E. & Schuman, E. M. The Central Dogma Decentralized: New Perspectives on RNA Function and Local Translation in Neurons. *Neuron* **80**, 648–657 (Oct. 2013).
9. Darnell, J. *et al.* FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. *Cell* **146**, 247–261 (July 2011).
10. Wang, D. O., Martin, K. C. & Zukin, R. S. Spatially restricting gene expression by local translation at synapses. *Trends in neurosciences* **33**, 173–182 (2010).
11. Turk, J. & Graham, P. Fragile X syndrome, autism and autistic features. *Autism* **1**, 175–197 (1997).
12. Cohen, I. L. Behavioral profiles of autistic and nonautistic fragile X males. *Developmental Brain Dysfunction* (1995).
13. Wagnon, J. L. *et al.* Etiology of a genetically complex seizure disorder in Celf4 mutant mice. *Genes, Brain and Behavior* **10**, 765–777 (Aug. 2011).
14. Dougherty, J. D. *et al.* The disruption of Celf6, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors. *Journal of Neuroscience* **33**, 2732–2753 (2013).
15. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* **303**, 1014–1016 (2004).
16. Auweter, S. D., Oberstrass, F. C. & Allain, F. H.-T. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic acids research* **34**. reviewed, 4943–4959 (2006).
17. Wang, X., McLachlan, J., Zamore, P. D. & Hall, T. M. T. Modular recognition of RNA by a human pumilio-homology domain. *Cell* **110**, 501–512 (2002).

18. Lewis, H. A. *et al.* Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100**, 323–332 (2000).
19. Ostareck-Lederer, A., Ostareck, D. H. & Hentze, M. W. Cytoplasmic regulatory functions of the KH-domain proteins hnRNPs K and E1/E2. *Trends in biochemical sciences* **23**, 409–411 (1998).
20. Ashley, C. T., Wilkinson, K. D., Reines, D. & Warren, S. T. FMR1 protein: conserved RNP family domains and selective RNA binding. *Science* **262**, 563–566 (1993).
21. Buckanovich, R. J. & Darnell, R. B. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Molecular and cellular biology* **17**, 3194–3201 (1997).
22. Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172 (2013).
23. Chekulaeva, M., Hentze, M. W. & Ephrussi, A. Bruno Acts as a Dual Repressor of oskar Translation, Promoting mRNA Oligomerization and Formation of Silencing Particles. *Cell* **124**, 521–533 (Feb. 2006).
24. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**, 539 (2011).
25. Kabat, E., Wu, T. & Bilofsky, H. Unusual distributions of amino acids in complementarity determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites. *Journal of Biological Chemistry* **252**, 6609–6616 (1977).
26. Pashev, I. G., Dimitrov, S. I. & Angelov, D. Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends in Biochemical Sciences* **16**, 323–326 (1991).

27. Castello, A. *et al.* System-wide identification of RNA-binding proteins by interactome capture. *Nature Protocols* **8**, 491– (Feb. 2013).
28. Sharma, S. Isolation of a Sequence-Specific RNA Binding Protein, Polypyrimidine Tract Binding Protein, Using RNA Affinity Chromatography. *RNA-Protein Interaction Protocols* (ed Lin, R.-J.) 1–8 (2008).
29. Bardwell, V. & Wickens, M. Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. *Nucleic Acids Research* **19**, 1980 (1991).
30. Chu, C., Quinn, J. & Chang, H. Y. Chromatin Isolation by RNA Purification (ChIRP). *Journal of Visualized Experiments : JoVE*, 3912– (Mar. 2012).
31. Simon, M. D. Capture Hybridization Analysis of RNA Targets (CHART). *Current Protocols in Molecular Biology* (2001).
32. Minajigi, A. *et al.* A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349** (2015).
33. Soeno, Y. *et al.* Identification of novel ribonucleo-protein complexes from the brain-specific snoRNA MBII-52. *RNA* **16**, 1293–1300 (Apr. 2010).
34. Penalva, L. O. F., Tenenbaum, S. A. & Keene, J. D. Gene Expression Analysis of Messenger RNP Complexes. *mRNA Processing and Metabolism: Methods and Protocols* (ed Schoenberg, D. R.) 125–134 (2004).
35. Darnell, R. B. HITS-CLIP: panoramic views of protein - RNA regulation in living cells. *Wiley Interdisciplinary Reviews: RNA* **1**, 266–286 (2010).
36. Ule, J. CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science* **302**, 1212–1215 (Nov. 2003).
37. Clark, P. M. *et al.* Argonaute CLIP-Seq reveals miRNA targetome diversity across tissue types. *Scientific Reports* **4** (Aug. 2014).

38. Huppertz, I. *et al.* iCLIP: Protein–RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 (Feb. 2014).
39. Kishore, S. *et al.* A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nature Methods* **8**, 559–564 (May 2011).
40. Lebedeva, S. *et al.* Transcriptome-wide Analysis of Regulatory Interactions of the RNA-Binding Protein HuR. *Molecular Cell* **43**, 340–352 (Aug. 2011).
41. Lee, F. C. & Ule, J. Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Molecular Cell* **69**, 354–369 (Feb. 2018).
42. Meyuhas, O. & Kahan, T. The race to decipher the top secrets of TOP mRNAs. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1849**, 801–811 (2015).
43. Patursky-Polischuk, I. *et al.* Reassessment of the role of TSC, mTORC1 and microRNAs in amino acids-mediated translational control of TOP mRNAs. *PloS one* **9**, e109410 (2014).
44. Hornstein, E., Tang, H. & Meyuhas, O. Mitogenic and nutritional signals are transduced into translational efficiency of TOP mRNAs. *Cold Spring Harbor symposia on quantitative biology* **66**, 477–484 (2001).
45. Pelletier, J., Kaplan, G., Racaniello, V. & Sonenberg, N. Cap-independent translation of poliovirus mRNA is conferred by sequence elements within the 5'noncoding region. *Molecular and Cellular Biology* **8**, 1103–1112 (1988).
46. Weingarten-Gabbay, S. *et al.* Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **351**, aad4939 (2016).
47. King, H. A., Cobbold, L. C. & Willis, A. E. The role of IRES trans-acting factors in regulating translation initiation. *Biochemical Society Transactions* (2010).

48. Bushell, M. *et al.* Polypyrimidine tract binding protein regulates IRES-mediated gene expression during apoptosis. *Molecular cell* **23**, 401–412 (2006).
49. Ivanov, A. *et al.* PABP enhances release factor recruitment and stop codon recognition during translation termination. *Nucleic acids research* **44**, 7766–7776 (2016).
50. Piqué, M., López, J. M., Foissac, S., Guigó, R. & Méndez, R. A combinatorial code for CPE-mediated translational control. *Cell* **132**, 434–448 (2008).
51. Ivshina, M., Lasko, P. & Richter, J. D. Cytoplasmic polyadenylation element binding proteins in development, health, and disease. *Annual review of cell and developmental biology* **30**, 393–415 (2014).
52. Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nature reviews Molecular cell biology* **8**, 113 (2007).
53. Harvey, R. F. *et al.* Trans-acting translational regulatory RNA binding proteins. *Wiley Interdisciplinary Reviews: RNA*. reviewed (2018).
54. Mori, Y., Imaizumi, K., Katayama, T., Yoneda, T. & Tohyama, M. Two cis-acting elements in the 3' untranslated region of CaMKII α regulate its dendritic targeting. *Nature Neuroscience* **3**, 1079–1084 (Nov. 2000).
55. Vuppalanchi, D. *et al.* Conserved 3'-Untranslated Region Sequences Direct Subcellular Localization of Chaperone Protein mRNAs in Neurons. *Journal of Biological Chemistry* **285**, 18025–18038 (Mar. 2010).
56. Tang, Y. H. *et al.* Complex Evolutionary Relationships Among Four Classes of Modular RNA-Binding Splicing Regulators in Eukaryotes: The hnRNP, SR, ELAV-Like and CELF Proteins. *Journal of Molecular Evolution* **75**, 214–228 (Nov. 2012).
57. He, Y. & Smith, R. Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cellular and Molecular Life Sciences* **66**, 1239–1256 (Dec. 2008).

58. Fan, X. C. Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the *in vivo* stability of ARE-containing mRNAs. *The EMBO Journal* **17**, 3448–3460 (June 1998).
59. Kundu, P., Fabian, M. R., Sonenberg, N., Bhattacharyya, S. N. & Filipowicz, W. HuR protein attenuates miRNA-mediated repression by promoting miRISC dissociation from the target RNA. *Nucleic Acids Research* **40**, 5088–5100 (Feb. 2012).
60. Timchenko, L., Timchenko, N., Caskey, C. & Roberts, R. Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy. *Human molecular genetics* **5**, 115–121 (1996).
61. Lee, J. & Cooper, T. Pathogenic mechanisms of myotonic dystrophy. *Biochemical Society Transactions* **37**, 1281–1286 (Dec. 2009).
62. Philips, A. V., Timchenko, L. T. & Cooper, T. A. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280**, 737–741 (1998).
63. Timchenko, N. A. *et al.* Overexpression of CUG triplet repeat-binding protein, CUGBP1, in mice inhibits myogenesis. *Journal of Biological Chemistry* **279**, 13129–13139 (2004).
64. Moraes, K. C., Wilusz, C. J. & Wilusz, J. CUG-BP binds to RNA substrates and recruits PARN deadenylase. *RNA* **12**, 1084–1091 (June 2006).
65. Cosson, B. *et al.* Oligomerization of EDEN-BP is required for specific mRNA deadenylation and binding. *Biology of the Cell* **98**, 653–665 (2006).
66. Gautier-Courteille, C. *et al.* EDEN-BP-dependent post-transcriptional regulation of gene expression in *Xenopus* somitic segmentation. *Development (Cambridge, England)* **131**, 6107–6117 (24 Dec. 2004).

67. Choi, D.-K., Ito, T., Tsukahara, F., Hirai, M. & Sakaki, Y. Developmentally-regulated expression of mNapor encoding an apoptosis-induced ELAV-type RNA binding protein. *Gene* **237**, 135–142 (1999).
68. Ladd, A. N., Charlet-B., N. & Cooper, T. A. The CELF Family of RNA Binding Proteins Is Implicated in Cell-Specific and Developmentally Regulated Alternative Splicing. *Molecular and Cellular Biology* **21**, 1285–1296 (Feb. 2001).
69. Ladd, A. N. Multiple domains control the subcellular localization and activity of ETR-3, a regulator of nuclear and cytoplasmic RNA processing events. *Journal of Cell Science* **117**, 3519–3529 (July 2004).
70. Anant, S. *et al.* Novel Role for RNA-binding Protein CUGBP2 in Mammalian RNA Editing CUGBP2 MODULATES C TO U EDITING OF APOLIPOPROTEIN B mRNA BY INTERACTING WITH APOBEC-1 AND ACF, THE APOBEC-1 COMPLEMENTATION FACTOR. *Journal of Biological Chemistry* **276**, 47338–47351 (2001).
71. Mukhopadhyay, D., Houchen, C. W., Kennedy, S., Dieckgraefe, B. K. & Anant, S. Coupled mRNA Stabilization and Translational Silencing of Cyclooxygenase-2 by a Novel RNA Binding Protein, CUGBP2. *Molecular Cell* **11**, 113–126 (Jan. 2003).
72. Brimacombe, K. R. & Ladd, A. N. Cloning and embryonic expression patterns of the chicken CELF family. *Developmental Dynamics* **236**, 2216–2224 (2007).
73. Sun, W. *et al.* Aberrant sodium channel activity in the complex seizure disorder of Celf4 mutant mice. *The Journal of Physiology* **591**, 241–255 (Dec. 2012).
74. Wagnon, J. L. *et al.* CELF4 regulates translation and local abundance of a vast set of mRNAs, including genes associated with regulation of synaptic function. *PLoS Genet* **8**, e1003067 (2012).

75. Wang, K.-S., Liu, X.-F. & Aragam, N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophrenia Research* **124**, 192–199 (Dec. 2010).
76. Maloney, S. E., Khangura, E. & Dougherty, J. D. The RNA-binding protein Celf6 is highly expressed in diencephalic nuclei and neuromodulatory cell populations of the mouse brain. *Brain Structure and Function* **221**, 1809–1831 (2016).
77. Ladd, A. N., Nguyen, N. H., Malhotra, K. & Cooper, T. A. CELF6, a member of the CELF family of RNA-binding proteins, regulates muscle-specific splicing enhancer-dependent alternative splicing. *Journal of Biological Chemistry* **279**, 17756–17764 (2004).

Chapter 5

CELF6 preferentially binds 3'UTR sequences in the brain and is associated with decreased mRNA abundance

Abstract

CUGBP and ELAV-like Factor (CELF) proteins have been shown to be involved in the splicing, degradation, and translational regulation of mRNAs. The CELF protein subfamily, CELF3-6, are enriched for expression in the brain, and CELF6 in particular is enriched for expression in the diencephalon and the monoaminergic dopamine, serotonin, and norepinephrine-producing cell populations. Recently our laboratory has identified polymorphisms in the *CELF6* are associated with ASD in humans, and mice lacking expression of *Celf6* show reduced ultrasonic vocalizations and reward-seeking behavior, however

the molecular function of CELF6 in the brain has not yet been described. We therefore used cross-linking immunoprecipitation followed by next generation sequencing (CLIP-Seq) to identify the binding targets of CELF6 *in vivo* in the mouse brain. The vast majority (>85%) of all targets showed preference for binding in the 3' untranslated region (3'UTR). 3'UTR sequence elements under CLIP-Seq peaks were subjected to analysis for differential enrichment of RNA binding motifs, and elements showed enriched presence of previously identified CELF protein UGU-containing binding motifs. In order to determine the functional consequence of CELF6 interaction with these 3'UTR elements, we subcloned a library of 436 3'UTR elements under CLIP-Seq peaks as well as mutations to motifs within these elements into a reporter construct. We found that when transiently transfected *in vitro*, reporters in general showed decreased levels of abundance with CELF6 overexpression which was abolished by mutation, but few showed changes to translation efficiency. Additionally, we tested whether CELF3, CELF4, and CELF5 could exhibit this activity, and overexpression of all CELFs 3-6 resulted in repression of reporter levels, with CELF3 and CELF4 exerting the strongest effects, and CELF5 and CELF6 exerting more intermediate effects on reporter levels. We also noted an association between strength of motif similarity to consensus motifs, and the number of motifs present in 3'UTR elements, on the magnitude of repression of reporter levels. Thus taken together, these data suggest that CELF6 and other members of its subclass form a group of RNA binding proteins that can mediate repression of mRNA via the 3'UTR, as has been previously described for other family members such as CELF1. Future research will help broaden our understanding of the exact mechanism by which CELF6 down-regulates its targets and how this may impact neuronal function.

Introduction

Messenger RNAs (mRNA) are regulated by RNA binding proteins (RBPs) in every aspect of their life cycle from early steps including transcription, splicing, nuclear export, localization, to their maintenance, translation into protein, and finally degradation in the cell [1–6]. CUGBP and ELAV-like Factor (CELF) proteins, originally characterized in relation to the pathogenesis of myotonic muscular dystrophy [7], are multi-functional RBPs which can act to regulate splicing in the nucleus, as well as mRNA half-life and translation efficiency in the cytoplasm [8]. The mammalian family contains 6 proteins (CELF1-6) which can be divided into two subgroups based on their amino acid compositional similarity: CELF1 and CELF2 which are found ubiquitously, and CELF3-6 which show enriched expression in the central nervous system [8, 9]. CELF1 binds CUG-repeat containing sequences [10] and (U)GU-rich motifs [11]. CELF1 has been shown to promote exon skipping [11], as well as promote degradation via recruitment of deadenylation machinery [12]. CELFs 3-6 however, have not been as well characterized.

Our laboratory recently identified polymorphisms in CELF6 associated with ASD in humans and reductions to ultrasonic vocalization and exploratory behavior in mice [13]. Upon further study, we found that in the brain, it showed enriched expression in the hypothalamus and in monoaminergic neurotransmitter cell populations (dopamine, serotonin, norepinephrine) [14]. Functionally, CELF6 has been shown in one report to be capable of regulating splicing in vitro [15], however its targets and function in the brain have not yet been described. The related RBP CELF4 in the brain shows preference for binding in the 3' untranslated region (3'UTR) of its targets affecting mRNA abundance and translation [16] but has also been shown to regulate alternative splicing in skeletal muscle [17], thus the function of a CELF RBP may be dependent on the tissue in which it is expressed.

In order to better understand the function of CELF6 in the brain, we performed cross-

linking immunoprecipitation followed by next generation sequencing (CLIP-Seq) on the brains of mice expressing an epitope-tagged CELF6-YFP/HA transgenic construct which recapitulates the endogenous pattern of expression [14]. Targets show CELF6 primarily associated with 3'UTRs of mRNAs. These sequences showed increased presence of UGU-containing motifs, consistent with previous research on other CELFs, and validating recently described *in vitro* binding preferences for CELF6 (Ray et al., 2013) *in vivo*. To comprehensively define the function of these motifs, we cloned over 400 independent sequences found under CLIP-Seq peaks into the 3'UTR of a reporter construct and measured reporter library mRNA abundance and translation with and without CELF6 expression, and with or without mutation of binding motif sequences. We found that CELF6 functioned generally as a repressor by decreasing the abundance of mRNAs containing the wild-type sequences, and this was abolished by mutating the motifs within the element. We also found that this behavior was redundant across CELF3-6, with CELF3 and CELF4 associated with the largest magnitudes of repression, and CELF5 and CELF6 showing intermediate levels of repression. We also assessed the levels of the reporter library on translating ribosomes using Translating Ribosome Affinity Purification (TRAP) [18], and found few changes to translation efficiency. The match strength of motifs to a UGU-containing consensus sequence [19] and the number of such motifs per element showed some ability to predict the fold repression observed in the reporter after overexpression of CELF protein. Thus we infer that CELF6 largely targets 3'UTRs in the brain, that this is at least in part mediated by UGU-containing sequences, and that this association is likely to result in down-regulation of target protein by decreasing mRNA abundance.

Results

Celf6 primarily associates with 3'UTRs of target mRNAs *in vivo*.

In order to define the *in vivo* binding locations of CELF6, we performed CLIP on brains from BAC transgenic mice expressing an epitope tagged CELF6-YFP/HA (78 kDa) with the endogenous CELF6 pattern [14]. We harvested tissue on post-natal day 9, near peak CELF6 expression and at time when Celf6 null mice exhibit a robust behavioral phenotype (decreased ultrasonic vocalizations) [13, 14]. First, we confirmed CELF6 binds RNA *in vivo*. We performed CLIP with anti-EGFP antibodies on CELF6-YFP/HA mice followed by radiolabeling of nucleic acid (Figure 1A), as compared to immunoprecipitated sample from wild-type tissue and an uncrosslinked sample, across several RNase concentrations. As expected, there was a lack of detectable RNA in immunoprecipitated(IP) from WT tissue and uncrosslinked YFP+ tissue. Next, to capture the targets of CELF6 *in vivo*, we chose a region approximately 60-200 nucleotides in size (80-150 kDa region) and purified this region from PAGE-separated RNase-digested (0.05 U/uL) lysates from 4 pools of 4 CELF6-YFP/HA+ brains. Similar to Vidaki et al.'s [20] study of Mena, we also found that the stringent lysis and wash conditions of typical CLIP protocols was incompatible with our CELF6 IP (not shown). Therefore we also collected control samples to be able to enable quantification and computational adjustments for any remaining background signal. These included 2% input samples as a measure of starting transcript abundance in the tissue, as well as IP from 3 pools of 4 WT littermate brains to identify those RNAs that interact non-specifically with the capture reagents.

To identify CLIP targets by next generation sequencing, we prepared libraries using an adaptation of the eCLIP/iCLIP workflow [21, 22](Protocol in Appendix 2, Adapter oligonucleotides in Supplemental Table 1). Currently there is no standard statistical approach for identification of targets from CLIP data. Methods typically include clustering aligned se-

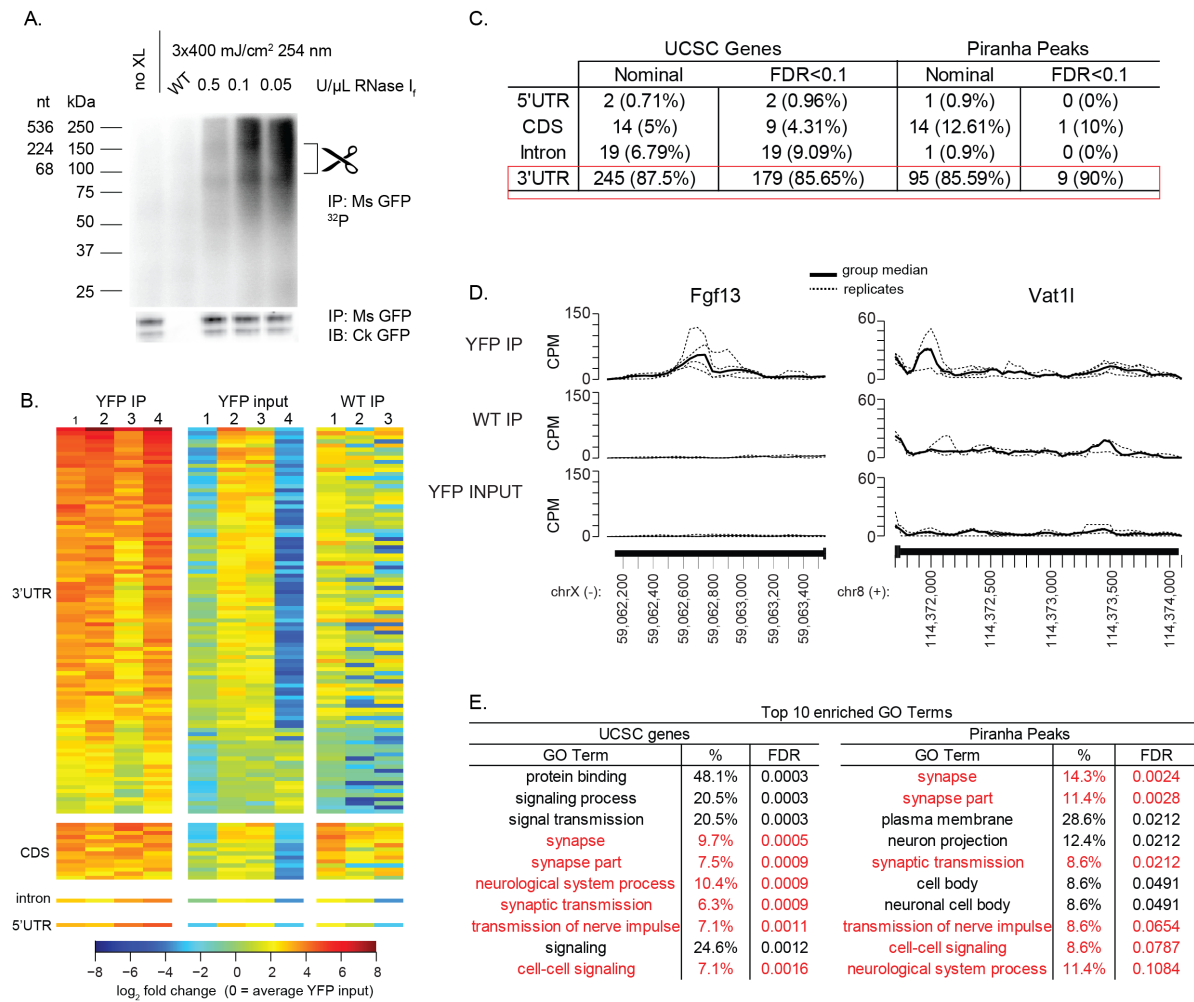


Figure 1: Celf6 primarily associates with 3'UTRs of target mRNAs *in vivo*

(A) Crosslinking followed immunoprecipitation and ^{32}P labeling of bound RNA. Brain tissue was dissected and powdered according to Methods, and tissue from single brains were used for (1) no crosslink, and (2) WT controls, and 1 pool of 3 brains was used for RNase I digestion. Upper panel shows autoradiogram of radioactively labeled RNA signal, and bottom panel shows immunoblot with chicken anti-GFP antibodies showing expression of CELF6-HA/YFP (78 kDa) which was isolated for RNA in sequencing experiments. Note immunoblot with chicken anti-GFP detects two bands. The top band is approximately 78 kDa. A second isoform lacking the first RRM of CELF6 is annotated (ENSMUST00000121266), and because the transgenic animal was generated by using a bacterial artificial chromosome containing the endogenous structure of the *Celf6* gene, both isoforms can be generated and targeted in this animal. ((B) \log_2 counts-per-million (CPM) RNA in 4 CELF6-HA/YFP+ replicates and 3 WT replicate samples across nominally significant differentially abundant regions identified by Piranha peak calling followed by edgeR differential enrichment analysis, showing enrichment in HA/YFP+ immunoprecipitate (IP) samples relative to input and WT controls. \log_2 CPM are normalized to the mean of the YFP input samples for each row. (C) Summary of differential enrichment analysis in edgeR for genes showing significant ($p < 0.05$) enrichment in HA/YFP+ IP samples relative to both HA/YFP+ input and WT IP controls, for both the UCSC Gene Annotation method (summing counts for UTR, coding sequence, and intronic sequence for all annotated genes) as well as the Piranha peak calling method (Calling peaks across the entire genome, and summing reads under peaks for differential analysis). No Piranha peaks were found in intergenic regions. Shown for each method is the number of identified unique genes at nominal significance level ($p < 0.05$) and Benjamini-Hochberg adjusted false discovery rate (FDR). (D) Example traces for two identified CLIP target 3'UTRs, showing CPM in YFP IP samples compared to controls. (E) BINGO analysis for gene ontology (GO) terms enriched in CLIP target genes determined according to both differential analysis methods.

quences in individual CLIP RNA samples or replicate averages, with varying probabilistic modeling approaches to assessing signal-to-noise in read density [11, 23–32] but rarely take into account variance across replicates or differential abundance compared to control samples. It is also worth noting that despite lack of apparent detectable RNA bound to WT samples by P32 end labeling (Figure 1A), we still found it possible to generate next generation sequencing libraries readily from WT IP samples, all of which generated many unique alignable reads even after removing PCR duplicates (Supplemental Table 2). All samples had similar read depth, and on average 94% of all uniquely aligning reads mapped to exons (%) or introns (%) in genic regions with very few reads mapping to intergenic regions, consistent with RNA expression. Next, to define specific sites of CELF6 binding, we called peaks throughout the genome using Piranha [33] and summed reads overlapping these peaks. We then performed differential enrichment analysis comparing CLIP to controls using edgeR [34, 35], requiring significant enrichment in both CLIP vs. input and CLIP vs. WT IP comparisons to define a CLIP target ("Piranha Peak Calling method". We found CLIP peaks called by Piranha to often have very wide bases (sometimes in excess of 1 kb or greater), or to regularly miss peaks manually observable in a genome browser display. Because of this, we took two approaches. First, we analyzed read counts summed narrowly under a 100-bp region below the called Piranha peak maxima. As an alternate approach, we also summed all reads mapping to genic regions corresponding to 5' untranslated (5'UTR), coding sequence (CDS), introns, or 3' untranslated (3'UTR) of UCSC annotated genes and performed differential enrichment analysis to define target regions, and then later called peaks based only on the read density in those regions ("UCSC Gene Annotation method"). Full differential enrichment results from both of these analysis methods are shown Supplemental Tables 3 & 4.

CELF proteins have previously been identified to function in alternative splicing of mRNAs as well as post-transcriptional regulation, thus to gain insight into potential molecular

functions of CELF6 we first examined where enriched CLIP targets were identified in relation to subgenic regions (5'UTR, 3'UTR, introns, coding sequence). We hypothesized that Celf6 splicing-related functions would correspond to increased density of reads in internal coding exons and alternatively spliced introns, while post-transcriptional regulatory functions would manifest as an increased density at untranslated regions such as the 3'UTR known to be involved in mRNA stability and translation regulation. Figure 1B shows a heatmap of abundance in all samples relative to input samples, for differentially enriched peaks and the vast majority of differentially enriched peaks are in 3'UTR regions. These results are further summarized in Figure 1C for both Piranha peak calling and UCSC Gene Annotation method. 37 genes overlap between the two analyses and these likely represent the highest confidence targets. In all analyses, the strongest signal came from 3' UTR regions (>85% of all CELF6 bound regions). The read distribution in CPM across the 3'UTRs for two high confidence targets, *Fgf13* and *Vat1l*, are shown across all samples in Figure 1D. The *Fgf13* gene is part of the FGF-like family of genes, and controls localization of voltage-gated Na⁺ channels in axons [36] and *Vat1l* (Vesicle Amine Transport 1-Like) is a paralog of the *Vat1* gene which regulates monoaminergic neurotransmitter storage and release [37]. The regulation of such genes may thus directly impact neuronal cell function. An enrichment for binding in the 3'UTR may indicate that regulation occurs via altered stability or translation of targets.

In order to examine whether CELF6 CLIP targets represented a specific subclass of mRNAs we performed gene ontology (GO) analysis using BiNGO [38] (Maere et al., 2005), on enriched CLIP targets. Although a few terms were found enriched and in common between both analysis methods, such as synaptic transmission, the gene subsets driving those terms were not the majority of differentially enriched genes (6.3 - 8.6%) and the largest subsets were for less specific GO terms such as "protein binding". Thus we do not believe that the majority of identified targets represent a specific functional class of

mRNAs. These results are summarized in Figure 1E.

CELF6-associated 3' UTRs are enriched for U-rich and UG-, CU- containing motifs

RBPs can bind RNAs via specific sequences, or non-specifically via interaction with the phosphate backbone. Enriched binding in specific peaks in 3' UTRs suggests CELF6 has affinity for specific nucleotide sequences *in vivo*, as previously reported in biochemical assays on recombinant protein (RNACompete, [19]). Thus we next desired to define the sequence specificity of CELF6 binding in these CNS targets. Manual inspection of the sequence underneath peaks in target 3'UTRs found many contain sequences matching UGU-rich *in vitro* binding preferences for CELF6 identified via RNACompete (Figure 2A) [19, 39]. To systematically identify sequence motifs enriched in CELF6 3'UTR targets we used the MEME suite tools [40] to identify both *de novo* motifs [41, 42] and previously cataloged motifs via Analysis of Motif Enrichment, AME [43], which is based on the Catalog of Inferred Sequence Binding Preferences of RNA binding proteins (CISBP-RNA) database [19]. The CISBP-RNA database includes binding preferences determined from RNACompete data as well as experimental data from other sources. For our MEME analysis, we used a set of sequences 50 nucleotides in length centered on peaks identified in CELF6 CLIP enriched 3'UTR regions. We analyzed 174 unique UTR in total, from under 491 identified peak maxima, 50 nucleotides in length around the peak center. Most UTRs possessed between 1-4 peaks. As a control set of sequences, we chose 491 sequences 50 nucleotides in length, sampled randomly from the 3'UTRs of brain expressed genes which showed a log fold changes ≤ 0 (no change or depletion) in CELF6-YFP/HA IP vs. input comparisons.

We did not find significantly enriched *de novo* motifs, but we found 36 enriched motifs from the CISBP-RNA database at the FDR<0.1 ($p<0.001$) level compared to control

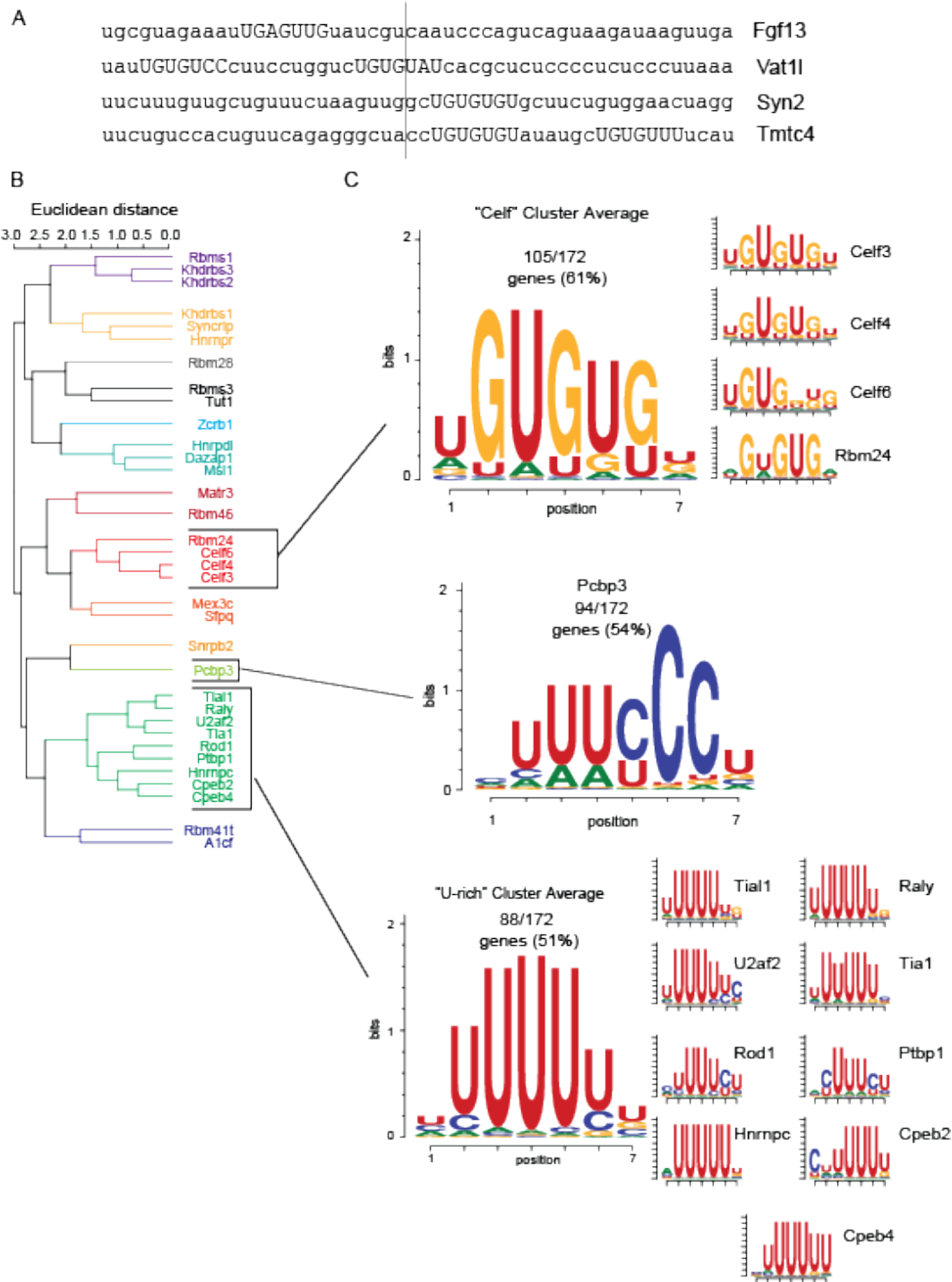


Figure 2: CELF6-associated 3'UTRs are enriched for U-rich and UG-, CU- containing motifs
(A) Example 50-nucleotide regions under CLIP target peaks in the 3'UTRs of Fgf13, Vat11, Syn2, and Tmtc for, all showing evidence for UG-rich sequence. (B) Clustering of Analysis of Motif Enrichment (AME) tool identified motifs enriched in CELF6 CLIP peak regions compared to randomly sampled control genes from background samples (HA-YFP input). Color coding shows cluster membership. (C) Motif logos for the most common clusters found in CLIP target peak regions. Logos represent the average of position weight matrices (PWMs) for each motif cluster, as well as the individual PWMs making up each cluster's membership.

sequences. These included binding motifs for CELF3, CELF4, and CELF6 as previously defined via RNACompete, as well as the U-rich motifs of the TIA and CPEB families of RNA binding proteins. RBPs that interact with specific sequence elements exhibit degenerate preferences and RNA binding domains themselves can associate with as few as 2 nucleotides and the binding motifs across RBPs are often similar or overlapping [44, 45]. To generate a more holistic understanding of binding elements enriched in our CELF6 CLIP targets, we clustered enriched motifs by the Euclidean distance between their position weight matrices (PWMs). In doing so, we unsurprisingly found that motifs clustered to some degree by RNA binding protein family (Figure 2B), however matches to preferences of some very diverse RBPs, such as the HNRNP family, do not necessarily cluster together. Motifs within a cluster are highly similar, such as the CELF motifs which all show RNACompete preferences for UGU-containing sequences. To determine how many sequences in our data were represented by each cluster, we looked for significant matches within our sequences using the MEME Suite FIMO tool to the averaged PWM representing each cluster. With this approach, 473/491 peak sequences showed significant matches across 172/174 total genes.

The top ranking clusters are depicted in Figure 2C. 105/172 unique genes' 3UTRs (61%) showed at least one match to the "CELF"- cluster, the average motif of which is [U/A]GUGU[G/U][UGA]. In addition to the CELF cluster, 94/172 genes' 3UTRs (54%) showed at least one match to the PCBP3 motif which forms its own cluster possesses a central UUU[C/U]CC sequence. PCBP3 can bind both double stranded and single stranded nucleic acid and is known primarily as a transcription factor [46] though related protein PCBP4 has also been shown to regulate mRNA stability [47]. 88/172 (51%) genes showed at least 1 match to the "U-rich" cluster, whose members all possess a central stretch of 4-5 Us, and include TIA1 which is involved in stress granule localization [48] as well as CPEB proteins which are known to be involved in polyadenylation [49], and

HNRNPC which is involved in both mRNA stability and localization/nuclear trafficking[50, 51]. 59/172 genes (34%) had at least one match to both the CELF cluster & the U-rich cluster, and 72/172 genes (42%) showed at least one match to both the CELF cluster and PCBP3. (Matches of these clusters in individual sequences is shown in Supplemental Table 5)

Massively Parallel Reporter Assay to define the function of CELF6 CLIP enriched motif sequences

Our motif analysis indicates sequence specificity mediates the interactions between CELF6 and RNA *in vivo*, but the downstream consequence of CELF6 association to elements in the 3'UTR of mRNAs has not been defined. Because 3' UTR elements often regulate mRNA stability or translation efficiency [3, 52–54], we first decided to interrogate the effect of CLIP-identified UTR sequence elements, and the identified motifs, on transcript abundance and ribosome occupancy using a recently described massively parallel reporter assay: post-transcriptional regulatory element sequencing (PTRE-Seq) [55]. For PTRE-Seq, we sub-cloned 436 independent CLIP-defined UTR elements, each 120 bp and centered under CLIP peaks, into the 3' UTR of a tdtTomato expression plasmid, excluding elements containing potential sites which would interfere with subcloning (Figure 3A). Each UTR element was included in the library design 6 times, with a unique 9-bp barcode to provide internal replication and buffer against any barcode effects. In addition, in order to assess the effect of candidate motifs on expression, we also generated a mutant version of each element. Mutations were made by swapping highest information nucleotides in the CISBP-RNA database position weight matrices for motif matches for their lowest information counterpart. Unmutated sequences we will call "reference" sequences, with mutant sequences being their counterparts after undergoing this process. This procedure was repeated for all significant matches to motifs for each element, al-

lowing us to disrupt the motif but leave most of the flanking sequence intact. In the final library, 90% of library UTR elements had <16 nucleotides mutated.

To assay the reporter library *in vitro*, we used transient transfection in human neuroblastoma SH-SY5Y cells. From RNA-Seq data on 8 replicate cultures of these cells, we assessed the mRNA expression levels of a number of RNA binding protein genes and found that, with the exception of *CELF1*, *CELFs* 2-6 were largely undetectable (Figure 3B). Thus, we were able to use this system to control *CELF6* levels by adding human *CELF6* exogenously via co-transfection with an His/Xpress epitope tagged *CELF6* construct used in previous research of *CELF6* function [15] or empty pcDNA3.1 vector (CTL). Additionally, in order to assess abundance of reporter mRNAs on translating ribosomes, we used co-transfected a construct expressing EGFP-RPL10a and employed Translating Ribosome Affinity Purification (TRAP) [56]. The Pearson correlation between recovered total reporter RNA from CTL or *CELF6* replicates (4 CTL, 4 *CELF6* replicates) ranged between 0.92 and 0.97, and in TRAP samples between 0.93 and 0.95 indicating good within-group reproducibility. To account for any differences in starting abundance of elements in the plasmid library, for every counted barcode RNA, \log_2 CPM were normalized to \log_2 CPM in the sequenced DNA plasmid pool (henceforth referred to as “ \log_2 expression”). Expression values in TRAP samples were further normalized to input RNA and this quantity will be referred to as “ \log_2 translation efficiency (T.E.)”. After removing any barcodes which were absent either in the final plasmid pool or undetectable across all samples after recovering RNA, the final analyzed library contained 424 UTR element pairs (reference and mutant) all of which contained between 3-6 surviving barcodes represented per element, across 172 total genes.

Our analysis strategy comprised using statistical models to identify effects of *CELF6* protein presence and how that interacted with motif sequence integrity, which we fitted independently for RNA expression and T.E. Specifically to determine the relationship be-

tween CELF6 overexpression and element sequence, we analyzed \log_2 RNA expression levels or TE using a 2x2 factorial design linear mixed effects model, fitting fixed effects of element sequence (“reference” or “mutant”) and overexpression condition (CTL or CELF6 overexpression) and the interaction of condition X sequence, with a random intercept term for each element, treating individual barcodes as a repeated measure. Individual models were fitted for each UTR element pair and the summary of effects and estimates of R^2 and main effect and interaction percentages of variance explained for each element is shown in Supplemental Table 6 (expression) and 7 (T.E.) along with Benjamini-Hochberg FDR controlled across all elements. At nominal $p < 0.05$, 317/424 (75%) of elements showed any significant effect at all (main effects of either sequence or condition, or sequence X condition interaction, 313/424 (74%) at FDR < 0.1).

CELF6 CLIP enriched motif sequences represent a set of repressive elements

We first examined the role of the motifs themselves by comparing the UTR elements with and without motif mutations. From our mixed model, over 91% of the elements showing any effect (74%) had a main effect of sequence. The main effects of sequence indicate there are effects of sequence mutation even in the absence of any CELF6 expression. Therefore, we looked at the distribution of average \log_2 fold changes in expression between reference and mutant sequence in the CTL condition for the 317 elements showing any nominally significant effect (Figure 3C). Considering all elements, 83.6% had fold changes less than 0 indicating a that reference sequences are repressive when compared to their mutated counterparts (median \log_2 fold change -0.42, 1.3-fold). As UTR elements with larger numbers of MEME motif matches have higher numbers of mutated bases, this raises the concern that mutating these motifs may have altered a very large fraction of the entire element. Thus we also looked at \log_2 fold changes between ref-

erence and mutant for elements where the number of mutated bases was smaller (i.e. Hamming distance (HD) for the mutation of ≤ 7 and ≤ 12 nucleotides). These were also generally repressive, with 71.6% less than zero and 77.8% less than zero for $HD \leq 7$ and $HD \leq 12$ respectively. The median \log_2 fold change for $HD \leq 7$ was -0.21 ($p=4.2E-6$, Mann-Whitney U test compared to all elements, $N=67$ $HD \leq 7$) and the median \log_2 fold change for $HD \leq 12$ was -0.30 ($p=0.0011$ Mann-Whitney U test compared to all elements, $N=189$ $HD \leq 12$, $p=0.022$ compared to $HD \leq 7$). Both the median estimates (-0.42, -0.30, -0.21) and the proportion of \log_2 fold changes less than zero (83.6%, 77.8%, 71.6%) scale with the maximum number of mutations in each case (≤ 25 , ≤ 12 , ≤ 7 nucleotides) thus indicating that while there is some effect of the number of mutations on the level of expression, even just disrupting a relatively small number of bases in the element (7 out of 120) can abrogate the repressive effects.

CELF6 protein enhances repression in a sequence dependent manner

We next assessed interactions between CELF6 overexpression and sequence. 30 elements (29 unique genes) showed a nominally significant interaction ($p < 0.05$) between overexpression condition and sequence with 4 out of these 30 elements possessing $HD \leq 7$ nucleotides, and 17/30 with ≤ 12 nucleotides. Figure 3D plots the average \log_2 fold change between reference and mutant sequence for both the CTL and CELF6 conditions across each level of HD. In every case, the fold change between reference and mutant was more steeply negative in the CELF6 condition (All HD CELF6 median \log_2 fold change reference – mutant -0.71 (~1.6-fold decrease), CTL -0.31, $p = 0.001$ | $HD \leq 7$: CELF6 -0.78, CTL -0.22, $p = 0.028$ | $HD \leq 12$ CELF6 -0.78, CTL -0.29 $p = 4.8E-4$, Mann Whitney U-tests). This appears to be due in general to a repression of the reference sequence rather than an elevation of the mutant sequence in the CELF6 condition

(Median \log_2 fold change CELF6-CTL reference sequences = -0.38, CELF6-CTL mutant sequences -0.04, $p = 2.1E-4$ Mann Whitney U-test).

In addition to impacting mRNA stability, RBPs can also alter protein production by either enhancing or repressing recruitment of the mRNA to a ribosome, thus altering its "translation efficiency" (TE i.e. the TRAP abundance divided by total RNA). However, in contrast to the consistent direction of effect of Celf6 on abundance, when we looked at changes to TE, we saw fewer significant effects overall. 78/424 sequences showed any significant effect on TE at all, with 0/424 showing a significant effect at $FDR < 0.1$. In Figure 3E, we plot the \log_2 fold change in TE between reference and mutant sequence for sequence elements showing nominally significant interactions of overexpression and sequence (15/424). We did not detect difference in overexpression condition for \log_2 fold changes in TE at any HD considered, and most \log_2 fold changes were close to 0, thus the source of interaction was not generalizable across the sequence as for total RNA. We inspected post hoc multiple comparisons for such elements to determine sources of interaction. 4/15 elements showed detectable changes in the reference sequence between CELF6 and CTL, with 2 showing an increase (Fnbp11 $p=0.0097$, \log_2 fold change 0.43 | Reep1 $p=0.004$ \log_2 fold change = 0.90), and two showing a decrease (Peg10 $p = 0.008$ \log_2 fold change -0.37 | Lin7c $p=0.006$ \log_2 fold change = -0.54). By contrast, when looking at total RNA, 23/30 post-hoc comparisons for the change to reference sequence expression (CELF6 – CTL) were significant and 19/23 of these \log_2 fold changes were negative ranging between -1.45 (2.73 fold reduction) to -0.22 (1.16 fold reduction) with a median of -0.55 (≈ 1.5 fold reduction). Example element expression is shown in Figure 3F (\log_2 expression normalized to the CTL/reference condition) and TE in Figure 3G (\log_2 TE normalized to CTL/reference condition). These findings suggest for elements showing significant interactions of CELF6 expression and sequence, these interactions are generally driven by a repression of the reference sequence via a decrease in RNA abundance

with CELF6 expression. This is abolished after mutation of the sequence. Also, there are limited effects on TE and where these occur they are not readily generalizable in their direction of effect.

CELF3-5 show redundancy in ability to enhance repression of Celf6-CLIP enriched UTR elements

The CELF3, CELF4, and CELF6 binding preferences determined by RNAcompete are highly similar (Figure 2), and as a group, CELF3-6 are more similar in amino acid identity than CELF1 or CELF2 [8]. Therefore, we were interested in testing the hypothesis that the enhanced repression of UTR elements we observed for CELF6 would also be true of CELF3-5. Thus we transiently transfected our PTRE-Seq library of reference and mutant UTR elements along with His/Xpress-tagged human CELF3, CELF4, or CELF5 used previously to study these proteins [15, 57, 58]. We then performed the analysis described in the preceding section and these results are summarized in Figure 4. Linear mixed effect models in the preceding section considered only CTL and CELF6 conditions, but in this section were refitted to include all conditions.

There were 111/424 elements representing 89/172 genes showing nominally significant sequence by overexpression condition interactions ($p < 0.05$, 91/424 FDR < 0.1) in total reporter RNA which includes an additional 74 elements compared to the analysis in the preceding section. Looking at the \log_2 fold change in reference vs. mutant sequence, CELF3 and CELF4 showed the largest differences (Figure 4A). The median element with CELF3 expression shows a $-1.15 \log_2$ fold lower expression of the reference sequence compared to the mutant sequence. CELF4 was comparable with a median $-1.01 \log_2$ fold change (CTL -0.42 , vs. CELF3 $p = 3.5E-15$, vs. CELF4 $p = 1.5E-12$, CELF3 vs. CELF4 $p = 0.13$, Mann Whitney U tests). CELF6 and CELF5 also showed comparable \log_2 fold changes, and these were intermediate between the CTL condition and CELF3/CELF4

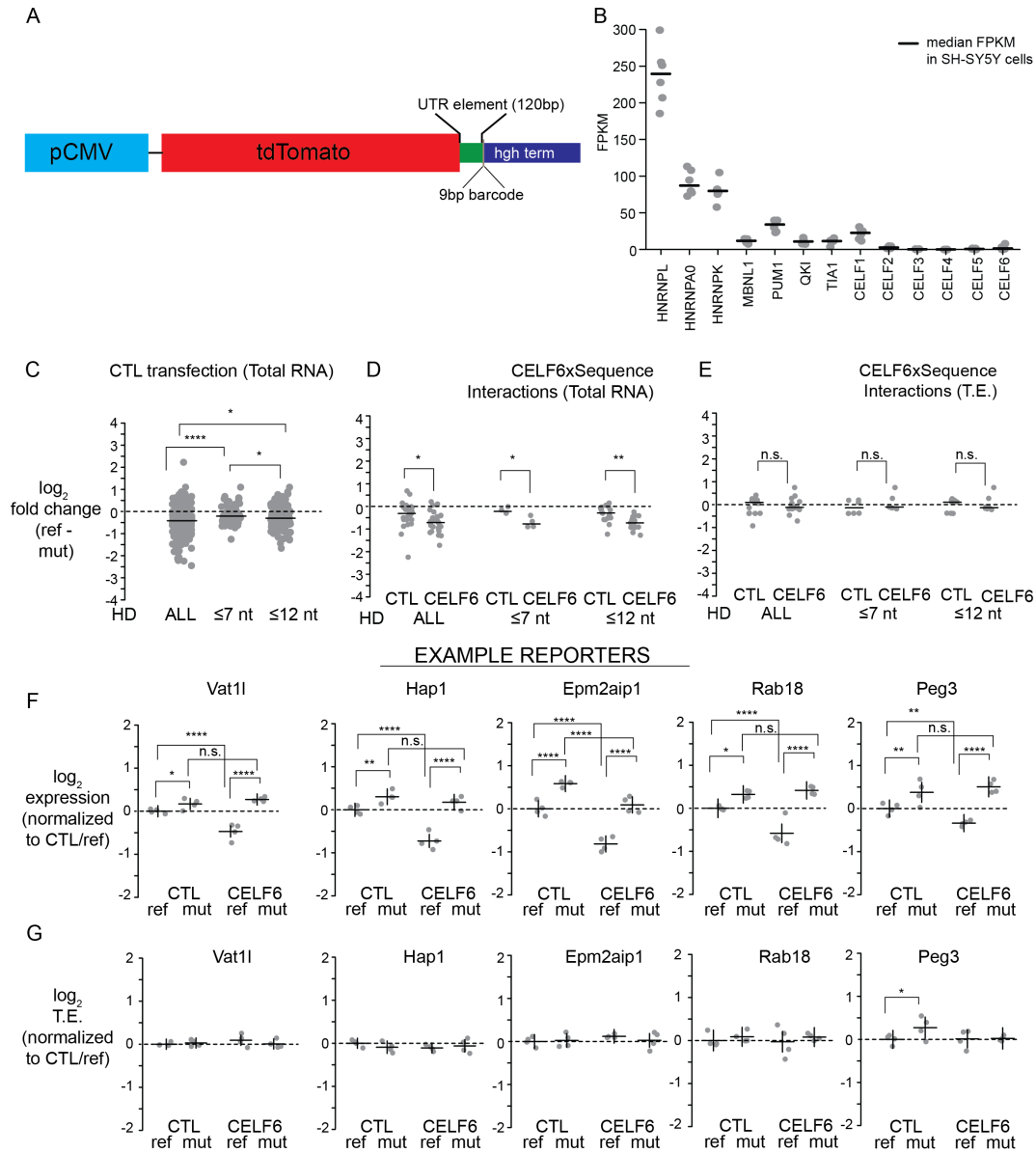


Figure 3: CELF6 CLIP enriched motif sequences represent a set of repressive elements

(A) Schematic of the tdTomato reporter used for PTRE-Seq. All reporter library elements are driven by a CMV promoter expressing mtdTomato (membrane-localizing tdTomato). The 120 nucleotide regions under CLIP peaks or mutant pairs subcloned after the tdTomato stop codon followed by a 9 nucleotide barcode sequence, and before a human growth hormone polyadenylation signal. (B) Fragments per Kilobase of transcript per Million reads (FPKM) levels are shown for several RNA binding proteins in RNA extract from 8 replicate platings of SH-SY5Y cells. (C) Recovered reporter RNA log₂fold changes in expression between reference and mutant elements in the CTL transfection conditions, considering all reference-mutant pairs, as well as only reference-mutant pairs possessing ≤7 nucleotide substitutions, or ≤12 nucleotide substitutions. (D) Recovered reporter RNA log₂ fold changes in expression between reference and mutant pairs in CTL and CELF6 overexpression conditions from elements showing significant condition X sequence interactions. (E) Recovered reporter RNA log₂fold changes in translation efficiency (T.E.) between reference and mutant pairs in CTL and CELF6 conditions for elements showing significant condition X sequence interactions. Data points in (C)-(E) are average estimates of log₂fold change reference vs. mutant per condition, averaged by replicate over barcodes, and then averaged over replicates, with lines representing medians for each distribution. Comparisons between these point estimates across conditions and total amount of sequence mutation were assessed by Mann Whitney U tests. (F) Log₂expression across 5 example reporter library elements in CTL, CELF6, reference and mutant conditions. (G) Log₂T.E. across example reporters in (F). Data points in (F)-(G) are averaged across barcodes, with each representing the estimate for each replicate. Horizontal lines represent average estimates for expression or T.E. (normalized to the CTL/reference sequence condition), and vertical lines represent 95% confidence intervals for these estimates from the linear mixed effects models. Post hoc pairwise comparisons between conditions shown were computed using the *multcomp* package in R with simultaneous multiple comparisons corrections using the multivariate normal distribution ("single step method" in *multcomp*). Significance notation: n.s. p>0.1, † p<0.1, *p<0.05, ** p < 0.001, *** p<0.0001, **** p<1E-5

(CELF5: -0.68, CELF6: -0.59, CELF6 vs. CTL $p=0.0024$, CELF5 vs. CTL $p=0.00034$, CELF5 vs. CELF6 $p = 0.26$). 12 of these elements had Hamming distances ≤ 7 nucleotides (Figure 4B). For these, CELF3 and CELF4 still showed significantly larger median \log_2 fold change compared to control ($p=0.007$, CTL -0.22, CELF3 -0.93 | $p=0.005$, CELF4 = -0.78 | CELF3 vs. CELF4 $p = 0.68$). With the added 8 elements (compared to the preceding section), CELF6 no longer showed a significantly larger median \log_2 fold change compared to control, and was comparable to CELF5 (CELF6 median \log_2 fold change -0.35, CELF5 -0.36, $p=0.85$). Considering all elements with ≤ 12 nucleotide differences (Figure 4C), the same trends were observed for all CELFs. Namely, CELF6 and CELF5 showed comparable fold change between reference and mutant sequence which were of greater magnitude than the CTL condition (CTL: -0.25, CELF6: -0.52, CELF5: -0.50, CELF6 vs. CTL $p = 0.004$, CELF5 vs. CTL $p = 0.002$, CELF6 vs. CELF5 $p = 0.56$). CELF3 and CELF4 showed the greatest magnitude changes (CELF3: -0.89 $p=1.1E-8$, CELF4: -0.88, $p=2.8E-8$) compared to CTL. As above, this appeared to be due to a general repression of the reference sequence compared to control, and CELF6 and CELF5 showed comparable intermediate effects compared to CELF3 and CELF4 which showed the strongest effects (Median \log_2 fold change CELF6-CTL reference -0.16, mutant 0.02, $p=2E-5$ | CELF5-CTL reference -0.24, mutant 0.01, $p=3.5E-8$ | CELF3-CTL reference -0.86, mutant -0.03, $p=4.5E-11$ | CELF4-CTL reference -0.62, mutant 0.04, $p=2.9E-11$). Thus taken together CELF overexpression in general showed a reduction of the reference reporter, which was abolished by mutation. Furthermore, CELF3 and CELF4 were associated with the strongest effects on reference elements, and CELF5 and CELF6 showed weaker effects. Thus within the subgroup of CELF proteins, CELFs 5 and 6 and CELFs 3 and 4 appear to form distinct subgroups with respect to their effects on this reporter library.

When looking at T.E., as in the preceding section, there were fewer significant effects

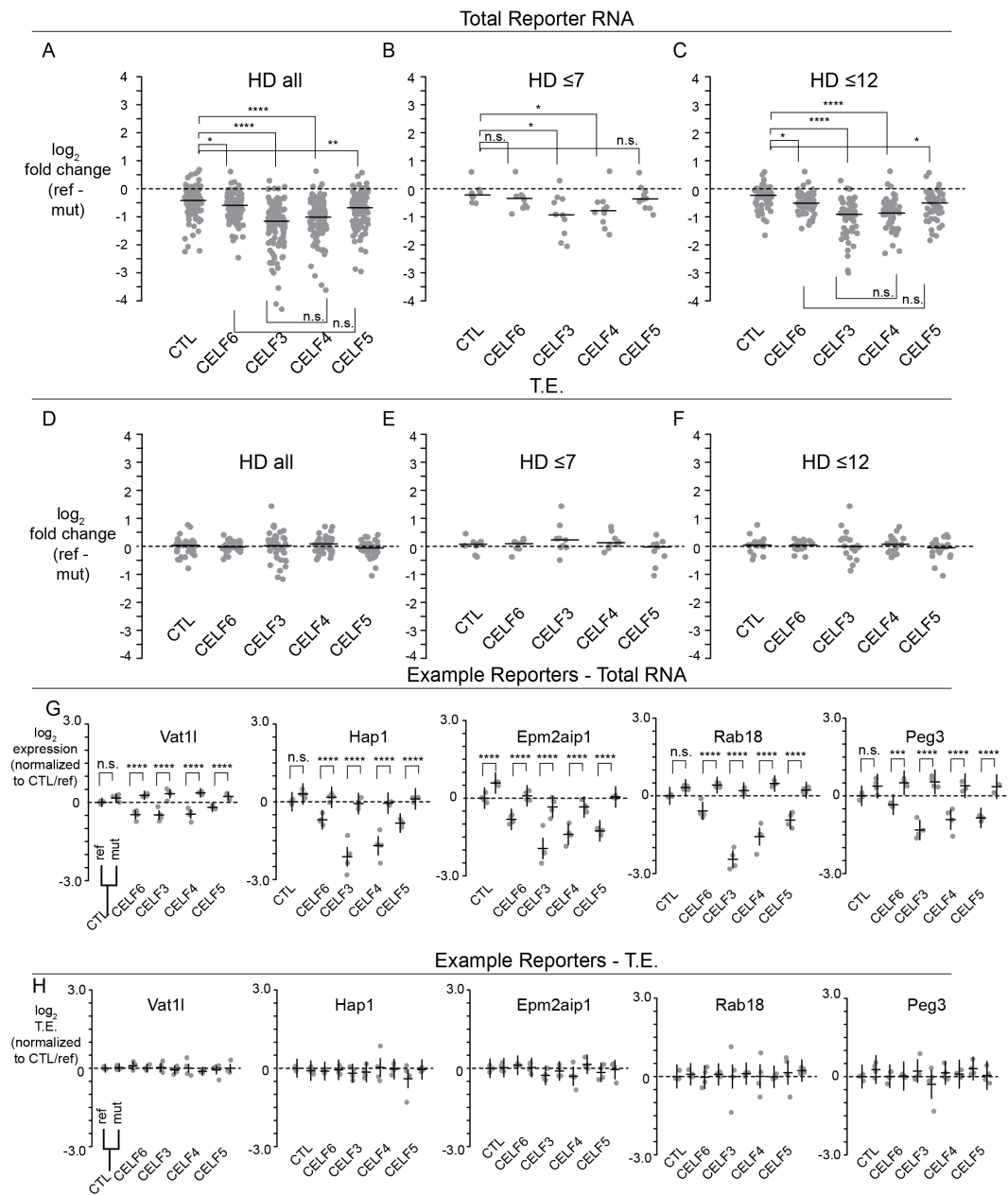


Figure 4: CELF3-5 show redundancy in ability to enhance repression of CELF6-CLIP enriched UTR elements

(A)-(C) Estimates of log₂ fold change in expression across CELF conditions considering all reporter elements, or reporter elements with a difference of ≤7 or ≤12 nucleotides between reference and mutant sequences, for all elements showing significant condition X sequence interactions. (D)-(F) Estimates of log₂ fold change in T.E. across conditions as in (A)-(C). Data points in (A)-(F) represent estimates averaged by replicate across barcode, and then across replicates, with lines representing medians. Statistical comparisons in (A)-(F) were assessed by Mann Whitney U tests. (G) Log₂ expression across 5 example reporter library elements across CELF, reference and mutant conditions. (H) Log₂ T.E. across example reporters in (G). Data points in (G)-(H) are averaged across barcodes, with each representing the estimate for each replicate. As in Figure 3, horizontal lines represent average estimates for expression or T.E. (normalized to the CTL/reference sequence condition), and vertical lines represent 95% confidence intervals for these estimates from the linear mixed effects models. Post hoc pairwise comparisons between conditions shown were computed using the *multcomp* package in R with simultaneous multiple comparisons corrections using the multivariate normal distribution ("single step method" in *multcomp*). Significance notation: n.s. p>0.1, † p<0.1, * p<0.05, ** p<0.001, *** p<0.0001, **** p<1E-5

overall. 41 elements showed significant interactions of condition and element sequence ($p < 0.05$, 6/424 at $FDR < 0.1$). However, when plotted in terms of their fold change between reference and mutant sequence, we were unable to generalize a direction of effect on TE across elements. Considering all elements median \log_2 fold changes in T.E. were near 0 with no significant differences across conditions (Figure 4D), and this was also true considering only elements with $HD \leq 7$ nucleotides (Figure 4E, 9/41 elements), and elements with $HD \leq 12$ nucleotides (Figure 4F, 22/41). In the case of reference sequences compared to CTL under CELF6 overexpression, 51% of CELF6-CTL \log_2 fold changes in T.E were negative, compared to 68% in changes to total RNA. For CELF5, these were 63% negative changes to TE, and 73% negative changes in total RNA. For CELF3 78% were < 0 for total RNA with 51% < 0 for TE, and for CELF4 these were 82% < 0 for total RNA and 37% < 0 for T.E. Thus the observable trends for decreases to reference reporter levels were apparent with CELF overexpression for total reporter levels but not for reporter translation efficiency. Total reporter expression for example reporters from Figure 3 are shown in Figure 4G with their respective T.E. shown in Figure 4H, with expression or T.E. normalized to the reference sequence/CTL condition.

In addition to these conditions, we were also curious if CELF proteins when co-expressed exerted additive or synergistic effects on reporter expression, and thus we transfected equimolar pairs of CELF6 construct with one of CELF3, CELF4, or CELF5. Among the 111/424 reporter elements (89/172 genes) showing significant sequence by condition interactions, overexpression of CELF3 and CELF6 together resulted in repression of the reference reporter which was similar in magnitude to overexpression of CELF3 by itself (Median CELF3/CELF6 \log_2 fold change compared to CTL: -0.95, CELF3 alone -0.86, $p = 0.38$). This was also true of CELF4 (Median CELF4/CELF6 \log_2 fold change compared to CTL: -0.85, CELF4 alone -0.62, $p = 0.14$). Thus the effect of CELF4 and CELF3 appears to be dominant in these cotransfections. When CELF5 and CELF6 were expressed to-

gether, the median \log_2 fold change compared to CTL was significantly greater than either overexpressed singly (Median \log_2 fold change CELF5/CELF6 vs. CTL: -0.48, CELF5 vs. CTL -0.24, CELF6 vs. CTL -0.16, p CELF5/CELF6 vs. CELF5 alone = 0.03) indicating a stronger effect can be observed by overexpressing these two proteins together, however we did not overexpress CELF6 or CELF5 alone in doubled dosage to determine whether this effect is simply due to more CELF overexpression or due to the specific combination of these proteins.

Motif strength and number alone can partially predict CELF6 impact.

Finally, to explore the relationship of enriched motifs in these reporter elements to the observed effects of CELF proteins on reporter expression, we asked whether the strength of the match to one of the 14 clusters shown in Figure 2B had any association to the magnitude of expression change with CELF overexpression compared to control. The CELF family class of UGU-rich motifs showed significant association with \log_2 fold change across all 424 reporters (172 genes) compared to CTL with respect to the match strength (measured as the $-\log_{10}$ p -value from the MEME suite FIMO tool) and this is shown in Figure 5A-D for each CELF respectively, although total effect size measured by R^2 were modest. For CELF3 and CELF4 9.9% and 11.8% of variance in reporter repression compared to CTL could be correlated to match strength (CELF3 ANOVA $p=3.57E-11$, CELF4 $p=3.96E-13$). CELF5 and CELF6 showed significant but weaker associations (CELF5 7.32% variance explained, $p=1.5E-8$, CELF6 4.2% variance explained $p=2.16E-5$). In all cases, the association was negative: stronger matches tended to be associated with more negative fold changes. All other clusters showed $\leq 3\%$ variance explained by match strength, with the exception of CELF5 condition reporter levels and the cluster containing Mex3c and Sfpq (4% variance explained, $p=2.9E-5$). This clusters near the CELF family and also has G/U-rich average motif logo ([G/A/U][G/U][A/G]GU[G/A]U) and this

is thus likely reflecting the same effect as matches to the CELF cluster. These effects are weak but nevertheless show some degree of association to previously defined CELF binding preferences and the effect of CELF overexpression. Among elements showing significant matches to the CELF motif (139/424 library elements representing 103/172 genes), we also asked whether the number of matches per element showed any relationship to the strength of reporter element repression. These results are shown in Figure 5E-H for each CELF overexpression condition. In all cases, a weak association was observed for match count to \log_2 fold change in reporter expression, with higher matches associated with greater repression. For CELF3, CELF4, and CELF5 this accounted for 8.7% ($p=4E-4$), 9.8% ($p=2E-4$), and 10% ($p=1E-4$) of total variance explained, respectively. For CELF6 this association was weaker at 4.25% variance explained by match count ($p=0.015$). Thus taken together, there is evidence for some association of CELF motif element and the number of times it is present in the UTR to reporter levels when CELF3-6 are overexpressed.

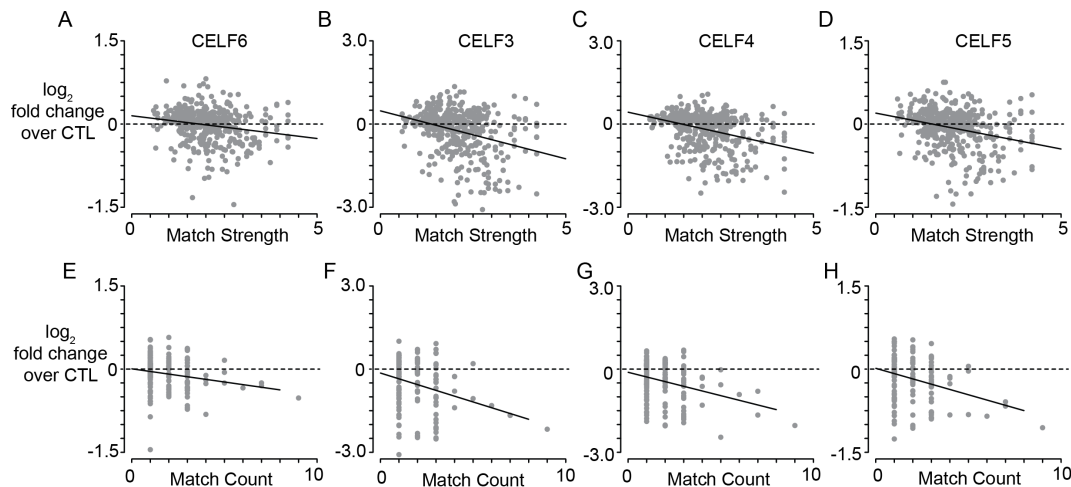


Figure 5: Motif match strength and number of matches can partially predict CELF6 impact.

(A)-(D) Association of match strength to fold change in expression, computed from the output of the Find Individual Motif Occurrences (FIMO) tool (MEME suite) for the CELF Cluster in Figure 2 across all 424 reporter elements. In all cases "Match Strength" was computed as the $-\log_{10}$ (FIMO p-value) and larger numbers indicate stronger match to the cluster's PWM. Y-axes are \log_2 fold changes between CELF and CTL conditions in the reference sequences only, collapsed by averaging by replicate across barcodes, and then across replicates. (A) CELF6, $p = 2.16E-5$, $R^2=0.042$, (B) CELF3 $p = 3.57E-11$, $R^2=0.099$, (C) CELF4 $p = 3.96E-13$, $R^2 = 0.118$, (D) CELF5 $p = 1.5E-8$, $R^2 = 0.0732$. (E)-(H) Elements showing at least 1 significant match of the CELF cluster PWM by FIMO, plotted as \log_2 fold change for reference sequences in CELF vs. CTL condition as in (A)-(D) by number of observed matches. (E)CELF6 $p = 0.015$, $R^2 = 0.0425$,(F) CELF3 $p=4E-4$, $R^2 = 0.087$, (G) CELF4 $p = 2E-4$, $R^2=0.098$, (H) $p = 1E-4$, $R^2=0.10$. Significance and coefficient of determination computed using *lm* in R.

Discussion

In this study, we have identified binding targets of CELF6 *in vivo* which primarily localize to 3'UTR regions of mRNAs, and this is consistent with what has been observed for CELF4 [16]. 3'UTR elements under CELF6 CLIP peaks are enriched for several motifs identified previously in a large biochemical study of RBP binding preferences on purified protein [19], including UGU-containing motifs, which has also been noted as binding CELF1 [11]. Additionally, using PTRE-Seq[55], we show that CELF6 and other CELFs generally down-regulate these 3'UTR elements *in vitro* which can be abolished by mutation based on their motif information content. Although CELF6 has been shown to regulate alternative splicing in skeletal muscle [15], we find very few significant binding events outside of 3'UTR regions, suggesting that CELF6 in the brain may be largely involved with post-splicing regulation of mature mRNAs. Additionally, we found few effects to the translation efficiency of our PTRE-Seq reporter library elements, suggesting that like CELF1, CELF6's repression of targets may be primarily mediated by enhancement of mRNA degradation.

Our work here raises a number of questions for future research. When transfected *in vitro*, although we are able to generally conclude that where CELF6 exerts an effect it is associated with lowered mRNA abundance, the majority of our library was not sensitive to CELF6 expression. This suggests that either CELF6 has additional functions on these transcripts not assessed in our assays (e.g. RNA localization), or that its functional impact on these sequences might depend on the cellular context with regard to the expression of other genes (e.g. other RBPs or miRNAs). We found that CELF3-6 RNA levels were largely undetectable in SH-SY5Y cells, nevertheless, the complement of other RBPs is likely to differ in the brain and likely to differ between populations of cells in which CELF6 is expressed. We did detect levels of MBNL1 in our SH-SY5Y cells, which has been shown to exist antagonistically with CELF1 [11] leading to mRNA stabilization. Additionally, it has also been shown that with regard to splicing function, CELF1, CELF2,

and CELF6 can all antagonize MBNL1 function [59]. If MBNL1 and CELF6 can act antagonistically, then some of the variability we observed in CELF6-responsiveness in cell culture may be due to competition from this or similar proteins. Antagonistic activity of RBPs on mRNA translation and stability has also been observed for CELF2 and HUR [60] and ELAVL1 and ZFP36 [61]. Antagonistic activity of RBPs may be mediated by the proximity of binding sites in the 3'UTR, with specific knockdown of RBPs freeing access to other RBP binding sites and resulting in changes to mRNA levels, as has been found for UTRs containing AU-rich elements [62]. In our analysis of CELF6 CLIP targets, we found several binding motifs showing enriched abundance in these UTRs for RBPs of different families. Although the presence of binding sites other than the UGU-rich motifs did not associate with the observed levels of reporter mRNA, we have not yet determined the extent of interaction of these binding motifs, nor whether their distance from UGU-containing motifs associates with expression levels. We did not find significant interactions between matching multiple motifs (not shown) on expression levels, but their placement with respect to one another is likely to play a role in whether they have a functional consequence of interaction. Important downstream experimentation includes knockdown of specific RBPs which may interact with CELF6 either directly or indirectly by competing for the same UTR, and determining whether this changes expression levels. Such experiments will lend important insight into the mechanism of CELF6 and other CELFs' actions at the 3'UTR.

We found that CELF6 expression was associated with down-regulation of library elements. In this study we have inferred that this is not due to changes in translation; while a subset of elements did show changes in TE in response to CELF proteins, there was not a general trend in the direction of the effect. In addition, as all reporter elements contain the same promoter, any alterations are unlikely to be due to changes to transcriptional activity. To determine whether CELF6 indeed facilitates mRNA decay, our

experiment should be repeated in the context of pulse ethynyluridine labeling to measure the decay rates of nascent RNA targets [63]. CELF1 has been shown able to recruit poly A ribonuclease (PARN) to RNA targets to facilitate mRNA decay, and Moraes and colleagues found that CELF1 could associate directly with PARN protein *in vitro* [12]. Thus it is intriguing to hypothesize that all the CELFs are able to do this, especially given our finding that CELFs3-6 can all induce repression of the same reporter elements, as well as whether knockdown of PARN is sufficient to abolish CELF-dependent down-regulation. Additionally, the *Xenopus* homologue of the CELF proteins, Embryo Deadenylation Element Binding Protein (EDEN-BP), has been shown to regulate deadenylation of target mRNA, and that oligomerization of the protein is required for this activity [64]. It is currently unknown whether any of the mammalian CELF proteins are able to oligomerize and the extent to which this may affect the functional activity of these proteins. Finally, we do not yet know the extent to which these findings generalize to CELF6 activity *in vivo*. Our reporter library is in a plasmid backbone that facilitates packaging into adeno-associated virus. Using this tool to transduce brain regions such as monoaminergic neuronal populations may help determine whether the effects we observe *in vitro* are also observed in the context of the brain.

In summary, we have presented the first evidence of CELF6 binding targets *in vivo* and our data support the hypothesis that like other CELF family members, CELF6 may function to regulate mRNA stability in addition to splicing. Future work will help further define the mechanism behind CELF6's actions *in vivo* and the downstream cellular consequences this may have on neuronal function.

Methods

Animals

All protocols involving animals were approved by the Animal Studies Committee of Washington University in St. Louis. Cages were maintained by our facility on a 12 hr : 12 hr light:dark schedule with food and water supplied ad libidum. Mice used in CLIP experiments comprised 4 pools of 4 CELF6-HA/YFP animals and 3 pools of 4 WT animals (30 animals). Animals were genotyped from toe clip tissue lysed by incubation at 50°C in Tail Lysis Buffer for 1 hours to overnight (0.5 M Tris-HCl pH 8.8, 0.25 M EDTA, 0.5% Tween-20) containing 4 μ L/mL 600 U/mL Proteinase K enzyme (EZ BioResearch), followed by heat denaturation at 99°C for 10 minutes. Crude lysis buffer was used as template for PCR with 500 nM HA/YFP genotyping primers: HA-F: 5' TTAAGCGTAGTCTGGGACGTCTG-TATGGGT 3', YFP R: 5' CTACGTCCAGGAGCGCACCATCTTCTT 3' and Actb control primers Actb F: 5'AGAGGGAAATCGTGCGTGAC 3' , Actb R: 5' CAATAGTGATGACCTG-GCCGT 3': using Quickload Taq Mastermix (New England Biolabs) with the following cycling conditions: 94°C 1 min, (94°C 30 s, 60 °C 30 s, 68°C 30 sec) x 30 cycles, 68°C 5 minutes, 10°hold.

Cell culture

SH-SY5Y neuroblastoma cells (ATCC CRL-2266) were maintained at 5% CO₂, 37°C, 95% relative humidity in 1:1 Dulbecco Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F12 Gibco) supplemented with 10% fetal bovine serum (FBS Sigma). Under maintenance conditions, cells were also incubated with 1% Penicillin-streptomycin (Thermo), but antibiotics were not used during transient transfections. Cell passage was performed with 0.25% Trypsin-EDTA (Thermo).

CLIP

Our CLIP procedure is modeled after the procedure of Wang et al. [11] and with private consultation with Eric Wang. Post-natal day 9 mice were euthanized by rapid decapitation, and brains were dissected. Cortices and cerebella were removed, retaining basal forebrain, striatum, diencephalon, colliculi, and hindbrain regions. Pools of four brains were flash frozen in liquid nitrogen and then powdered with a mortar and pestle cooled with liquid nitrogen and kept on dry ice in 10 cm Petri dishes until use. Crosslinking was performed using 3 rounds of 400 mJ/cm² dosage of 254 nm ultraviolet radiation, with petri dishes on dry ice, in a Stratalinker UV crosslinker. After each round of crosslinking, powder in the dishes was tapped and redistributed to allow for even crosslinking. After crosslinking, powders were kept on wet ice and incubated with 1 mL lysis buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 1X complete EDTA-free protease inhibitor (Sigma), 0.04 U/ μ L recombinant RNasin (Promega) 10 mM activated sodium orthovanadate, 10 mM NaF. Recombinant RNasin does not inhibit RNase I which was used for RNase digestion in CLIP and was added to prevent other environmental RNase activity. To obtain both cytoplasmic and nuclear fractions in the lysate, lysis buffer was supplemented with 1% NP40 (Sigma CA630) detergent and subjected to mechanical homogenization/lysis in a teflon homogenizer 10 times, and lysates were allowed to incubate on ice for 5 minutes. For RNase digestion, RNase I_f (New England Biolabs) was diluted to final concentration 0.5, 0.1, or 0.05 U/mL per lysate for radiolabeling. For control radiolabeled samples (no crosslink and WT tissue immunoprecipitates), the highest (0.5 U/mL) concentration of RNase was used. For samples used for sequencing, 0.05 U/mL final concentration was used. RNase-containing lysates were incubated in a thermomixer set to 1200 RPM at 37°C for 3 minutes and then clarified at 20,000xg for 20 minutes. 2 % input lysate was saved for input samples for sequencing. Per immunoprecipitation, 120 μ L of Dynabeads M280 streptavidin coated beads (Thermo) were incubated with 17 μ L 1 mg/mL biotiny-

lated Protein L (Thermo), and 36 μg each of mouse anti-EGFP clones 19F7 and 19C8 antibodies (MSKCC) for 1 hour. Beads were prepared in batch for all immunoprecipitations and then washed five times with 0.5% IgG-free bovine serum albumin (Jackson ImmunoResearch) in 1X PBS, followed by three washes in lysis buffer. Clarified lysates were incubated with coated, washed beads for 2 hours at 4°C with end-over-end rotation and then washed in 1 mL of wash buffer (50 mM Tris-HCl pH 7.4, 350 mM NaCl, 1% NP-40, 0.04 U/ μL RNasin) four times, for 5 minutes with end-over-end rotation at 4°C. For radiolabeling experiments, 60% of washed bead volume was reserved for immunoblotting and added to 20 μL of 1X Bolt-LDS non-reducing sample buffer (Thermo), and 40% proceeded to radioactive labeling. Beads for radioactive labeling were subsequently washed 3x200 μL in PNK wash buffer (20 mM Tris-HCl pH 7.4, 10 mM MgCl_2 , 0.2% Tween-20, 2.5 U/ μL RNasin) and then incubated with 10 μL PNK reaction mixture (1X PNK reaction buffer (New England Biolabs), 4 μCi ^{32}P -ATP (Perkin Elmer), 10 U T4 PNK (New England Biolabs)) for 5 minutes, at 37°C. After labeling, samples were washed in 3x200 μL PNK wash buffer to remove unincorporated label, and then added to 10 μL of 1X Bolt LDS non-reducing sample buffer. All samples in sample buffer were heated for 10 minutes at 70°C and then separated on a 4-12% gradient NuPAGE Bis/Tris gels (Thermo) and then transferred to PVDF membranes with 10% methanol for 6 hours at constant 150 mA. Samples for immunoblot were blocked for 1 hour in block solution (5% nonfat dried milk in 0.5% Tween-20/1X TBS), and then overnight with 1:1000 chicken anti-GFP antibodies (AVES) with rocking at 4°C. Blots were washed 3x5 minutes in 0.5% Tween20/1X TBS and then incubated with 1:5000 anti-chicken HRP secondary antibodies (AVES) for 1 hour at room temperature and treated with Biorad Clarity enhanced chemiluminescence reagents for 5 minutes and chemiluminescent data acquired with a Thermo MyECL instrument. Radioactive signal was acquired using an Amersham Typhoon Imaging System and a BAS Storage Phosphor screen (GE Healthcare Life Sciences).

CLIP-Seq Sequencing Library Preparation

For CLIP-Seq, EGFP immunoprecipitated WT and HA-YFP+ tissue and 2% input samples were purified from PVDF membranes as follows. Membrane slices were cut with a clean razor according to the diagram in Figure 1A, from unlabeled samples as has been performed in eCLIP [21]. In order to minimize sample processing, slices were incubated in 1.7 mL microcentrifuge tubes with 200 μ L Proteinase K buffer (100 mM Tris-HCl pH 7.4, 50 mM NaCl, 10 mM EDTA, 1% Triton-X 100) containing 40 μ L of 800 U/mL Proteinase K (NEB) and incubated in a horizontal shaker at 250 RPM, 37°C for 1 hour. Horizontal shaking reduces the need to cut the membrane into small pieces per sample which is seen in many protocols, and 1% Triton-X 100 in the Proteinase K buffer facilitates increased yield from the membrane. 200 μ L of fresh 7M Urea/Proteinase K buffer is then added to slices and tubes are incubated an additional 20 minutes with horizontal shaking at 250 RPM, 37°C. RNA is purified by addition of 400 μ L of acid phenol/chloroform/isoamyl alcohol and shaken vigorously for 15 s and allowed to incubate 5 minutes on the bench. RNA samples are centrifuged at 20,000xg for 10 minutes. Aqueous layers are purified using a Zymo-5 RNA Clean & Concentrator column. Output from CLIP'd RNA samples was estimated for total concentration using an Agilent Bioanalyzer and approximately 0.5 ng of RNA was used to prepare next generation sequencing libraries. The full protocol for sequencing library preparation is given in **Appendix 2**. This is a unified protocol based on eCLIP but which we also use for general RNA-Seq library preparation.

CLIP-Seq Sequencing Data Processing

CLIP-Seq samples were sequenced paired-end 2x40 on an Illumina Next-Seq. Unique Molecular Identifier (UMI) sequences were extracted from Illumina Read 2, and reads were trimmed from quality using Trimmomatic [65]. Reads aligning to ribosomal RNA were depleted and remaining RNA reads were aligned to the mm10 mouse reference

genome using STAR [66] and assembled into BAM formatted alignment files. BAM files were annotated with UMI information using the FGBio Java package (<https://github.com/fulcrumgenomics/fgbio>). PCR duplicates assessed by their UMIs were removed from the BAM files using picard-tools (<https://broadinstitute.github.io/picard/>). For Piranha peak calling analysis, the genome was windowed into 100 bp contiguous windows with 50% overlap. A merged BAM file pooling reads from all YFP-HA+ immunoprecipitated CLIP samples was used to count reads per window. Zero count windows were truncated, and Piranha [33] was used to call peaks on the remaining read density throughout the genome. Piranha p-values for significant peaks were adjusted for multiple testing using Benjamini Hochberg, and all peaks called with False Discovery Rate < 0.1 were kept for further analysis and stored as a Gene Transfer Format (GTF) file. For the UCSC annotated gene based analysis, UCSC table browser was used to generate GTF files containing all annotated: 3'UTR, coding sequence (CDS), 5'UTR, and intron regions. In order to ensure mapping of reads to splice sites, the table of intron annotations was allowed to overlap the surrounding exons by 10 bases. All GTFs (Piranha peaks, 3'UTR, CDS, 5'UTR, introns) were subsequently as feature sets for read counting with Subread[67, 68], summing reads for each annotated subgenic region. Samples were imported in R using the edgeR package [35] and normalized for both total library size per sample and feature length as Fragments per Kilobase per Million reads (FPKM). To identify a minimum detection level for subsequent analysis, the relationship between standard deviation across CLIP samples and mean \log_2 FPKM was computed and fitted to a spline. Variability across replicates increases dramatically at poor detection level, and the minimum FPKM required for a feature was set to where the standard deviation decayed to half maximal as a threshold (here determined at 2.5 FPKM). We required that all YFP-HA+ CLIP samples have FPKM > threshold across all YFP-HA+ CLIP samples to be included in analysis. Differential testing was then performed in edgeR against read counts deriving

from WT samples or YFP-HA+ input samples. We defined CLIP targets as having positive fold change enrichment in YFP-HA+ CLIP samples compared to both WT CLIP and input samples, with nominal edgeR p-values <0.05. We have proceeded throughout our analysis and PTRE-Seq library generation using targets defined this way in order to maximize exploration of the data. However, we also computed Benjamini-Hochberg adjusted False Discovery Rate for each feature/gene and these are summarized in Supplemental Tables 3 and 4.

CLIP Motif Enrichment Analysis

Because the Piranha peak-based analysis initially yielded often very broad peaks when called genome-wide, we repeated peak finding considering only read density across 3'UTRs of CLIP targets. It is likely that estimating the background across all transcripts even after removing zeros across the genome underestimated the background level for many genes. 50 bp regions under the maxima of peaks in CLIP target 3'UTRs were used in MEME Suite and compared to 50 bp regions sampled randomly from the 3'UTRs of "non-targets" - genes which exhibited 0 or negative fold enrichment in YFP-HA+ CLIP samples compared to input or WT controls. These sample sets were then submitted to both DREME web tool for *de novo* motif finding, and AME web tool for searching against the CISBP-RNA database (<http://meme-suite.org/>).

PTRE-Seq Reporter Library Preparation

All cloning oligos are as shown in Supplemental Table 8. We generated the pmrPTRE-AAV backbone from an existing mtdTomato construct by PCR amplification and sub-cloning of the following elements: CMV promoter and a T7 promoter, PCR amplified and subcloned into the MluI restriction site of a AAV-GFAP-mtdTomato construct and amplified from pcDNA3.1. Then, in order to add a NheI-KpnI restriction enzyme cassette into

the 3'UTR, the entire pmrPTRE-AAV plasmid was amplified and recircularized using In-fusion HD (Clontech). The correct backbone sequence of pmrPTRE-AAV was confirmed by Sanger sequencing.

Originally 473 120 bp sequences under CLIP Peaks were considered for cloning into the library across significant genes. Mutations to motifs found by AME were made as follows. Significant matches to motifs were determined using FIMO ($p < 0.005$). Next, for each matched motif, the position weight matrix (PWM) representing this motif in CISBP-RNA was used to determine the choice of mutation at each base. Only bases showing a probability of 0.8 or greater were mutated at any position. Bases showing PWM probability > 0.8 were mutated to the base showing the minimum value of PWM at that position. If all other three bases showed equal probability, a base was randomly selected from the three. The procedure was repeated at each position for each motif match, before moving to the next matching motif, ranked by FIMO p-value. Where motifs overlapped, lower ranking motifs (based on match strength) did not override mutations already made based on higher ranking motifs. After completion this generated a set of 473 mutant elements which ranged between 1 - 25 mutated nucleotides depending on the number of motif matches and number of highly conserved positions in each motif. These sequences were scanned for poly A signals and restriction enzyme sites likely to interfere with cloning, which were removed. A final set of 436 120 bp sequences, attached to 6 unique 9 base pair barcodes, as well as priming sites for amplification and cloning (final length 210 bp), and a paired set of 436 mutant sequences were synthesized by Agilent Technologies.

Obtained synthesized sequences were amplified with 4 cycles of PCR using Phusion polymerase (Thermo) using primers GFP-F and GFP-R, which are standard PCR primers in our laboratory that result in robust amplification. The library was PAGE purified and concentration of recovered library was estimated by Agilent TapeStation. The library was digested with NheI and KpnI enzymes and ligated into pmrPTRE-AAV with T4 Ligase

(Enzymatics). In order to ensure high likelihood of obtaining all library elements, we prepared our plasmid pool from approximately 40,000 colonies. .

PTRE-Seq Reporter Library Transfection

His/Xpress-tagged CELF3,4,5,&6 were obtained from the laboratory of Tom Cooper [58]. For four plasmid experiments, 2500 ng containing equimolar: 2xHis/Xpress-CELF constructs, 1 EGFP-RPL10a, and CELF6 PTRE-Seq library were prepared with Lipofectamine 2000 in OptiMem-I (Gibco). For plasmids with 3 plasmids, remaining mass was made up with empty pcDNA3.1-His. SH-SY5Y cells trypsinized and incubated in with Lipofectamine/DNA complexes overnight in DMEM/F12 supplemented with 10% FBS. The following day media was replaced with fresh DMEM/F12 supplemented with 10%FBS and cells were pelleted for TRAP and RNA extraction 40 hours post-transfection. TRAP and total RNA extraction were performed according to the protocol in **Appendix 3** with total RNA quality assessed by Agilent TapeStation and all samples had RIN_e values > 8. 5 replicates per condition were generated in batches balanced for all conditions. In each case, replicates were transfected from newly thawed aliquots of cells to control for cell passage. Read counts from 1 batch were found to cluster separately from all others after sequencing and data from this batch were excluded. The final data were analyzed from 4 replicates per condition.

PTRE-Seq Sequencing Library Preparation

PTRESeq sequencing libraries were prepared by cDNA synthesis using pmrPTRE-AAV antisense oligo (Supplemental Table 8) for library specific priming, and Superscript III Reverse Transcriptase (Thermo) according to the protocol shown in **Appendix 4**. After cDNA synthesis, cDNA libraries were enriched with PCR using Phusion polymerase (Thermo), and pmrPTRE-AAV antisense and sense oligos using 18 cycles. In parallel, plasmid pool

DNA was also amplified for sequencing the original plasmid pool. Purified PCR products were digested with NheI and KpnI enzymes and ligated to 4 equimolar staggered adapters to provide sequence diversity for sequencing on the NextSeq. Ligated products were amplified with Illumina primers as in CLIP-Seq library preparation (**Appendix 2**) and subjected to 2x40 paired end next generation sequencing on an Illumina sequencer.

PTRE-Seq Sequencing Library Data Processing

Barcode counts from sequencing read FASTQ files for each element were determined using a Python script. Read counts were imported into R using edgeR and converted to counts-per-million (CPM) to normalize for differences in library size. Elements showing no counts in the DNA plasmid pool sequencing were removed. Expression was then computed as:

$$\log_2\text{expression} = \log_2 \frac{CPM_{RNA}}{CPM_{DNA}}$$

Translation efficiency was computed as:

$$\log_2\text{TE} = \log_2\text{expression}_{n_{TRAP}} - \log_2\text{expression}_{n_{TotalRNA}}$$

The mean and standard deviation relationship within condition groups were determined for $\log_2\text{expression}$ across elements as in processing for CLIP-Seq in order to filter out poorly detected elements. A minimum $\log_2\text{expression}$ value of -4.29 was determined as a lower bound cutoff which corresponded to approximately 10 counts. We required that all 4 replicates in at least 1 condition had expression levels above this threshold. Finally, after filtering on expression, we required that all elements have at minimum 3 out of the original 6 barcodes present, and present for both reference and mutant alleles.

PTRE-Seq Sequencing Library Statistical Analysis

The final set of elements amenable to testing after filtering on expression and numbers of barcodes was 424 across 172 unique gene UTRs. Individual linear mixed models were performed using *lme4* [69] fitting the following:

$$\text{expression (or TE)} \sim \text{condition} * \text{sequence} + (1|\text{sample})$$

where barcodes were used as repeated measures for each sample, per element. Fixed effect terms of condition referred to either: (a) CTL or CELF6 expression for analyses in Figure 3, or (b) CTL, CELF6, CELF3, CELF4, CELF5, CELF3/6, CELF4/6, CELF5/6 for analyses related to Figure 4. Fixed effect term of sequence was either (a) reference or (b) mutated sequence. Omnibus tests for significant effects of fixed effect terms were computed using likelihood ratio tests in R compared to models lacking a term of interest (Analysis of Deviance). Estimates of R^2 and percentage variance explained by fixed effects were determined according to the procedure of Nakagawa and Schielzeth [70]. Omnibus p-values for fixed effects are also reported in Supplemental Tables 6 & 7 for models containing all 8 conditions alongside Benjamini Hochberg adjusted False Discovery Rates. For generalized analysis of trends across all elements in Figures 3 and 4 we have considered all elements showing main effects or interaction effects with nominal $p < 0.05$.

Figures 3 and 4 show analysis of these trends across all models and for subsets of elements with less than or equal to 7 or 12 mutations to discern whether radically mutating elements has exerted a strong effect. Because these subsets are nested and pooling independently fitted models, we have assessed significance between them using non-parametric Mann Whitney U tests for differences between medians. Post hoc multiple comparisons significance and confidence intervals reported for individual elements in

Figure 3 and 4, in order to identify the sources of interaction, were computed using the *multcomp* R package [71].

Data Accessibility

Raw and processed sequencing data from CLIP-Seq and PTRE-Seq studies will be submitted to NCBI's Gene Expression Omnibus as part of publication of this manuscript.

Supplemental Information

Table S1: Oligos for CLIPSeq Library Preparation

name	sequence (5'-3')
A01m	/5Phos/rArGrArUrCrGrGrArArGrArGrCrGrUrCrGrUrGrUrArG/3SpC3/
AR17 primer	ACACGACGCTCTCCGA
Rand103tr3	/5Phos/NNNNNNNNNAGATCGGAAGAGCACACGTCTG/3SpC3/

Table S2: CLIP-Seq Aligned Read Summary

Sample	Genotype	Fraction	total pairs (millions)	duplication %	unique aligners (millions)	% aligning to subgenic features			
						3UTR	CDS	5UTR	introns
yfp-1-ip	yfp	ip	9.9	6.59	5.9	19.2	25.53	4.5	50.77
wt-1-ip	wt	ip	15.6	4.64	9.7	15.53	27.98	4.91	51.57
yfp-1-input	yfp	input	4.6	6.15	2.7	12.4	20.47	12.58	54.55
wt-1-input	wt	input	4.9	6.96	2.7	13.5	19.15	11.38	55.97
yfp-2-ip	yfp	ip	6.9	4.48	4.1	14.7	27	10.2	48.1
wt-2-ip	wt	ip	3.2	5.14	1.6	10.86	27.8	10.28	51.06
yfp-2-input	yfp	input	29.3	19.48	21.2	11.61	16.12	4.2	68.06
wt-2-input	wt	input	5.1	3.49	3.9	11.54	16.44	4.49	67.53
yfp-3-ip	yfp	ip	5.8	35.53	3.2	15.12	26.36	7.16	51.36
yfp-3-input	yfp	input	10.7	9.11	7.8	15.11	18.69	4.14	62.06
yfp-4-ip	yfp	ip	6.7	26.52	3.3	21.48	26.19	9.01	43.31
wt-3-ip	wt	ip	14.4	49.98	3.7	15.15	31.49	7.21	46.15
yfp-4-input	yfp	input	2.2	56.86	0.8	13.43	17.16	8.36	61.06
wt-3-input	wt	input	9.8	20.42	6.7	16.46	21.24	4.15	58.14

Table S3: CLIP Targets Defined from Piranha Peaks

chr	start	end	strand	gene	geneFeature	logFC _{IP/input}	PValue	FDR	logFC _{IP/WTIP}
chrX	59062650	59062749	-	Fgf13	utr3	4.239	1.83E-5	0.0109	5.397
chrX	134305100	134305199	+	Tmem35a	utr3	3.249	3.04E-5	0.0109	2.103
chr13	54381300	54381399	+	Cplx2	utr3	2.856	2e-04	0.0328	3.227
chr15	38489550	38489649	-	Azin1	utr3	3.027	2e-04	0.0352	3.71
chrX	105123550	105123649	+	Magee1	utr3	2.897	0.001	0.0922	4.86
chr7	62417850	62417949	-	Gm32061,Mkrn3	utr3	2.228	0.0013	0.0922	3.591
chr2	165112550	165112649	-	Cdh22	cds utr3	2.5	0.0013	0.0922	0.279
chr4	47469950	47470049	-	Alg2	utr3	2.643	0.0015	0.0922	4.436
chr12	72795000	72795099	+	Ppm1a	utr3	2.233	0.0017	0.0922	3.021
chr11	116359800	116359899	-	Rnf157	cds	2.408	0.0017	0.0922	0.313
chr15	98169950	98170049	+	Ccdc184	utr3	2.457	0.0023	0.1167	2.886
chr18	12991250	12991349	+	Impact	utr3	2.407	0.0027	0.1238	4.225
chr7	100273200	100273299	+	Pgm2l1	utr3	2.558	0.0028	0.1238	2.193
chr2	28230250	28230349	+	Olfm1	utr3	2.309	0.0034	0.1329	2.751
chr2	164501050	164501149	+	Pigt	cds utr3	2.161	0.0035	0.1329	0.671
chr2	134595150	134595249	-	Tmx4	utr3	2.247	0.0054	0.1329	2.892
chr7	131566150	131566249	+	Bub3	cds utr3	1.949	0.0057	0.1329	2.505
chr5	144556150	144556249	+	Nptx2	cds	2.563	0.0058	0.1329	0.963
chr4	59221850	59221949	+	Ugcg	utr3	2.097	0.0059	0.1329	2.159
chr11	75496400	75496499	+	Prpf8	cds	1.831	0.0059	0.1329	0.865
chr13	91881950	91882049	-	Rasgrf2	utr3	2.07	0.0059	0.1329	3.467
chr5	105486400	105486499	+	Lrrc8b	utr3	1.905	0.0061	0.1329	1.945
chr6	118602050	118602149	-	Cacna1c	cds	2.118	0.0061	0.1329	2.988
chr17	64281900	64281999	-	Pja2	utr3	2.316	0.0062	0.1329	3.536

CLIP Targets Defined from Piranha Peaks Cont'd

chr	start	end	strand	gene	geneFeature	logFC _{IP/input}	PValue	FDR	logFC _{IP/WTIP}
chrX	153171600	153171699	+	Rragb	utr3	2.182	0.0063	0.1329	3.684
chr2	127226000	127226099	+	Snrnp200	cds	1.816	0.0069	0.1329	0.751
chr10	117814850	117814949	-	Rap1b	utr3	2.122	0.0073	0.1329	1.384
chr13	110055800	110055899	-	Rab3c	utr3	2.068	0.0074	0.1329	3.075
chr10	108499000	108499099	-	Syt1	utr3	2.116	0.0075	0.1329	2.778
chr17	80396600	80396699	-	Sos1	utr3	1.978	0.0078	0.1329	2.826
chrX	56588450	56588549	-	Mmg1	utr3	2.023	0.0078	0.1329	3.04
chrX	100816550	100816649	+	Dlg3	utr3	1.925	0.0081	0.1329	1.673
chrX	152610300	152610399	-	Shroom2	utr3	2.166	0.0083	0.1329	2.445
chr5	21795750	21795849	+	Psmc2	cds	1.949	0.0083	0.1329	1.236
chr7	62349525	62349624	+	Ndn	utr3	2.082	0.0088	0.1329	3.153
chr1	118300500	118300599	-	Tsn	utr3	1.958	0.0088	0.1329	3.318
chr3	51251250	51251349	+	Noct	utr3	1.977	0.0089	0.1329	1.595
chr4	135412450	135412549	-	Rcan3	cds utr3	1.88	0.0094	0.1376	1.988
chr9	71479050	71479149	-	Polr2m	utr3	2.281	0.0099	0.1376	3.733
chr13	95954850	95954949	-	Sv2c	utr3	2.022	0.0106	0.1376	3.358
chr4	33945150	33945249	+	Cnr1	utr3	1.802	0.0107	0.1376	3.775
chr7	59310600	59310699	+	Ube3a	utr3	1.947	0.0108	0.1376	2.046
chr18	24580550	24580649	-	Slc39a6	utr3	2.032	0.011	0.1376	3.12
chr16	44087100	44087199	-	Atp6v1a	utr3	2.077	0.0111	0.1376	2.83
chrX	147195850	147195949	+	Htr2c	utr3	1.999	0.0124	0.1503	2.964
chr2	181549550	181549649	+	Dnajc5	utr3	2.052	0.0138	0.1647	1.768
chr8	114371950	114372049	+	Vat1l	utr3	1.965	0.0145	0.169	1.719
chr1	180144850	180144949	+	Cdc42bpa	cds	1.938	0.0148	0.169	1.002

CLIP Targets Defined from Piranha Peaks Cont'd

chr	start	end	strand	gene	geneFeature	logFC _{IP/input}	PValue	FDR	logFC _{IP/WTIP}
chr9	26755900	26755999	+	B3gat1	cds	1.682	0.0153	0.1713	0.52
chr11	77504600	77504699	+	Git1	cds	1.822	0.0158	0.1726	0.373
chr17	34612600	34612699	+	Agpat1	utr3	1.818	0.0166	0.1754	0.669
chr1	23845900	23845999	-	Smap1	utr3	1.839	0.0166	0.1754	1.374
chr6	71810400	71810499	+	Reep1	utr3	1.95	0.0169	0.1757	1.891
chr1	33798550	33798649	-	Zfp451	utr3	1.771	0.0173	0.1759	3.289
chrX	8894600	8894699	+	B630019K06Rik	utr3	1.992	0.0174	0.1759	3.328
chr11	106058850	106058949	+	Dcaf7	utr3	1.947	0.018	0.1768	2.347
chr1	63314350	63314449	+	Zdbf2	utr3	1.909	0.0182	0.1768	2.11
chr4	118235900	118235999	-	Ptprf	cds	1.619	0.0192	0.1768	0.617
chrX	133587350	133587449	-	Pcdh19	utr3	1.837	0.0193	0.1768	2.483
chr11	84984975	84985074	-	Usp32	utr3	1.739	0.0193	0.1768	0.901
chr3	10418700	10418799	-	Snx16	utr3	1.855	0.0194	0.1768	3.879
chr4	65615600	65615699	+	Trim32	utr3	2.007	0.0201	0.1768	3.071
chr1	82290450	82290549	-	Irs1	utr5 cds	1.67	0.0202	0.1768	2.591
chr8	34827800	34827899	-	Tnks	utr3	1.811	0.0202	0.1768	2.548
chrX	36198900	36198999	+	Zcchc12	utr3	1.892	0.0215	0.1854	3.072
chr9	62792400	62792499	-	Fem1b	utr3	1.737	0.0226	0.1864	2.53
chr7	6964500	6964599	+	Usp29	utr3	1.835	0.0227	0.1864	4.552
chr15	4575200	4575299	+	Plcx3	utr3	1.896	0.0228	0.1864	5.418
chr1	169741750	169741849	-	Rgs4	utr3	1.766	0.0233	0.1864	1.675
chr2	70601800	70601899	+	Gad1	utr3	1.91	0.0235	0.1864	2.717
chr11	100347450	100347549	-	Hap1	utr3	1.754	0.0236	0.1864	2.944
chr5	129984100	129984199	+	Vkorc11	utr3	1.731	0.0237	0.1864	2.222

CLIP Targets Defined from Piranha Peaks Cont'd

chr	start	end	strand	gene	geneFeature	logFC _{IP/input}	PValue	FDR	logFC _{IP/WTIP}
chr4	101642350	101642449	+	Dnajc6	utr3	1.68	0.0254	0.1958	3.198
chr6	113168150	113168249	-	Lhfp14	utr3	1.651	0.0258	0.1969	1.921
chr4	105036050	105036149	-	Prkaa2	cds utr3	1.643	0.0275	0.2058	2.536
chr11	20225100	20225199	+	Rab1a	utr3	1.888	0.0276	0.2058	3.239
chr10	52338250	52338349	-	Gopc	utr3	1.668	0.0281	0.2058	3.138
chr12	29923000	29923099	+	Myt1l	utr3	1.599	0.0291	0.2107	2.642
chr13	25252900	25252999	-	Nrsn1	utr3	1.855	0.0303	0.2114	3.102
chr3	157315150	157315249	+	Negr1	utr3	1.659	0.0303	0.2114	1.599
chr3	124322500	124322599	+	Tram111	utr3	1.617	0.0319	0.2193	4.022
chr4	40733350	40733449	+	Dnaja1	utr3	1.742	0.0321	0.2193	2.401
chr2	26921000	26921099	-	Surf4	utr3	1.445	0.0325	0.2195	0.764
chr6	57737850	57737949	+	Landl2	utr3	1.628	0.0329	0.2205	1.496
chr18	37186700	37186799	+	Pcdha1-12	utr3	1.74	0.0333	0.2206	3.928
chr8	83435450	83435549	-	Scoc	utr3	1.735	0.0335	0.2206	2.528
chr10	112925600	112925699	-	Atxn713b	utr3	1.734	0.0344	0.2213	1.291
chr11	100459650	100459749	-	Acly	intron	1.563	0.0346	0.2213	1.804
chr17	13016450	13016549	+	Sod2	utr3	1.607	0.0378	0.2249	3.007
chr14	122480100	122480199	+	Zic2	utr3	1.608	0.0379	0.2249	4.444
chr2	121194900	121194999	-	Trp53bp1	utr3	1.57	0.0383	0.2249	2.8
chr7	122080500	122080599	+	Ubfd1	utr3	1.49	0.0385	0.2249	2.195
chr2	164999300	164999399	+	Slc12a5	cds utr3	1.735	0.0386	0.2249	2.14
chr14	19874950	19875049	-	Gng2	utr3	1.868	0.039	0.2249	3.824
chr13	102732600	102732699	-	Mast4	utr3	1.531	0.0395	0.2249	2.395
chr17	24685650	24685749	-	Syngn3	utr3	1.768	0.0401	0.2249	2.449

CLIP Targets Defined from Piranha Peaks Cont'd

chr	start	end	strand	gene	geneFeature	logFC _{IP/input}	PValue	FDR	logFC _{IP/WTIP}
chr3	9427250	9427349	-	Zfp704	utr3	1.409	0.0403	0.2249	2.453
chr5	125787250	125787349	+	Tmem132b	cds	1.398	0.0404	0.2249	0.282
chr9	27010400	27010499	-	Vps26b	cds	1.395	0.0404	0.2249	0.346
chr17	27426750	27426849	-	Grm4	cds utr3	1.543	0.0405	0.2249	0.618
chr18	76941400	76941499	+	Ier3ip1,Hdhd2	utr3	1.532	0.0425	0.2296	2.278
chr13	97251300	97251399	+	Enc1	utr3	1.76	0.0426	0.2296	2.727
chr10	24188100	24188199	+	Stx7	utr3	1.673	0.044	0.2329	3.171
chr1	131177050	131177149	-	Rassf5	utr3	1.481	0.0442	0.2329	2.781
chrX	143791700	143791799	+	Pak3	utr3	1.788	0.0448	0.2332	3.025
chr11	97509750	97509849	-	Srcin1	utr3	1.686	0.0458	0.2332	3.346
chr8	79675050	79675149	+	Otud4	utr3	1.545	0.0465	0.2332	3.731
chr7	15945300	15945399	-	Nop53	cds	1.353	0.0466	0.2332	1.587
chr2	148697200	148697299	-	Napb	utr3	1.723	0.0468	0.2332	2.628
chr12	86694850	86694949	+	Vash1	utr3	1.362	0.0472	0.2332	1.359
chr6	30446600	30446699	+	Klhdc10	cds	1.464	0.0488	0.2353	0.293

Table S4: CLIP Targets Defined From Annotated Subgenic Regions

*FDR was computed within feature for pairwise tests in 5'UTR,3'UTR, CDS, or intron data.

gene	feature	logFC _{IP/input}	p _{IP/input}	FDR	logFC _{IP/WTIP}
Fgf13	utr3	3.093	1.12E-8	2.3E-5	2.995
Magee1	utr3	2.779	1.21E-7	1e-04	2.806
Pgr15l	utr3	2.497	3.47E-6	0.0018	2.157
Syn2	utr3	1.815	5.74E-6	0.002	1.973
Olfm1	utr3	1.784	1.2E-5	0.0035	1.427

CLIP Targets Defined From Annotated Subgenic Regions (Cont'd)

gene	feature	logFC _{IP/input}	p _{IP/input}	FDR	logFC _{IP/WTIP}
Snca	utr3	2.041	1.73E-5	0.0044	1.78
Arxes1	utr3	1.963	2.57E-5	0.0059	1.877
Gad1	utr3	2.006	3.06E-5	0.0063	2.02
Slc39a6	utr3	2.427	3.39E-5	0.0063	1.88
Ndn	utr3	1.911	7.39E-5	0.0117	1.674
Atp6ap2	utr3	1.891	1e-04	0.0151	2.605
Impact	utr3	1.491	1e-04	0.0175	1.561
Klhdc2	utr3	1.849	2e-04	0.0206	2.033
Morf4l2	utr3	1.931	2e-04	0.0206	2.359
Cd200	utr3	1.558	2e-04	0.023	2.062
Enc1	utr3	1.597	2e-04	0.0235	1.314
Syt1	utr3	1.336	3e-04	0.0237	1.376
Rps6ka6	utr3	2.137	3e-04	0.0252	1.249
Zcchc12	utr3	1.787	3e-04	0.0273	2.176
Mkrn3	utr3	2.163	4e-04	0.0278	2.848
Rragb	utr3	1.764	4e-04	0.0313	2.024
Ubf1	utr3	1.351	4e-04	0.0313	1.032
Commd3	utr3	2.64	5e-04	0.0313	0.803
AW551984	utr3	1.731	5e-04	0.0313	1.835
Nrsn1	utr3	1.758	5e-04	0.0313	1.448
Nap1l5	utr3	1.428	6e-04	0.0356	1.967
Cnr1	utr3	1.517	6e-04	0.0366	1.54
Epm2aip1	utr3	1.511	6e-04	0.0366	1.63
Ccrn4l	utr3	1.869	8e-04	0.0408	1.466

CLIP Targets Defined From Annotated Subgenic Regions (Cont'd)

gene	feature	logFC _{IP/input}	p _{IP/input}	FDR	logFC _{IP/WTIP}
Htr7	utr3	1.631	9e-04	0.0419	0.475
Trim32	utr3	1.79	9e-04	0.0419	2.408
B3galnt1	utr3	1.756	9e-04	0.0419	0.878
Plcx3	utr3	2.375	9e-04	0.0419	3.014
Pfn2	utr3	1.298	0.001	0.043	0.933
Zcchc18	utr3	1.718	0.001	0.0449	1.512
Shroom2	utr3	1.248	0.0011	0.0449	1.375
Gpm6a	utr3	1.471	0.0012	0.0464	1.849
Ube2e3	utr3	1.879	0.0013	0.0469	0.891
Synpr	utr3	1.662	0.0013	0.0469	1.968
Rab18	utr3	1.288	0.0014	0.0469	1.516
Penk	utr3	1.647	0.0014	0.0469	2.103
Gpr101	utr3	1.595	0.0014	0.0469	1.168
Snap25	utr3	1.302	0.0014	0.0472	1.756
Htr2c	utr3	1.255	0.0015	0.0474	2.108
Mpped1	utr3	1.68	0.0015	0.0476	0.741
Pja2	utr3	1.502	0.0016	0.0479	1.076
Zdbf2	utr3	1.654	0.0016	0.0479	2.055
Hmgn1	utr3	1.56	0.0016	0.0481	0.747
Rae1	utr3	3.251	0.0017	0.0497	0.478
Dcaf12l1	utr3	1.749	0.0017	0.0497	1.913
Tmem35	utr3	1.503	0.0019	0.0526	1.514
Scg2	utr3	1.467	0.002	0.0526	1.288
Hsp90aa1	utr3	1.736	0.002	0.0534	1.702

CLIP Targets Defined From Annotated Subgenic Regions (Cont'd)

gene	feature	$\log FC_{IP/input}$	$p_{IP/input}$	FDR	$\log FC_{IP/WTIP}$
Uhrf2	utr3	1.614	0.0021	0.0537	0.515
Vav2	utr3	2.145	0.0022	0.0556	0.336
Isca1	utr3	1.638	0.0023	0.0566	0.938
Spsb4	utr3	1.672	0.0023	0.0566	0.597
Dek	utr3	1.487	0.0024	0.0566	2.017
Zic4	utr3	1.315	0.0026	0.0566	0.781
Rgmb	utr3	1.406	0.0026	0.0566	1.259
Irs4	utr3	1.611	0.0027	0.0566	1.948
Gabrq	utr3	1.593	0.0027	0.0566	1.791
Azin1	utr3	1.271	0.0028	0.057	1.223
Rnf14	utr3	1.104	0.0029	0.0597	0.976
Srp9	utr3	1.567	0.003	0.0599	1.008
Wif1	utr3	3.005	0.003	0.0599	5.663
Phox2b	utr3	2.121	0.003	0.0599	1.284
Ran	utr3	1.381	0.0031	0.0601	0.448
Inpp4b	utr3	2.563	0.0032	0.0609	2.786
Atmin	utr3	1.349	0.0033	0.0631	0.742
Tmtc4	utr3	2.144	0.0034	0.0639	1.62
Ttc3	utr3	1.489	0.0035	0.0651	1.281
Arxes2	utr3	1.565	0.0036	0.0654	2.156
Rraga	utr3	1.31	0.0036	0.0655	2.268
Qdpr	utr3	1.694	0.0037	0.0668	1.266
Ube2b	utr3	1.514	0.0042	0.0735	1.123
Gprasp2	utr3	1.828	0.0044	0.0751	2.206

CLIP Targets Defined From Annotated Subgenic Regions (Cont'd)

gene	feature	$\log FC_{IP/input}$	$p_{IP/input}$	FDR	$\log FC_{IP/WTIP}$
Dnm3	utr3	1.164	0.0048	0.0811	1.073
Gpr149	utr3	2.436	0.0049	0.0821	1.711
Tnks2	utr3	1.53	0.0049	0.0822	1.734
Rab3c	utr3	0.959	0.0051	0.0834	1.195
Cdk5rap2	utr3	3.507	0.0051	0.0834	3.845
Kif2a	utr3	1.229	0.0052	0.0834	1.153
Fnbp1l	utr3	1.29	0.0052	0.0834	2.051
Atf1	utr3	1.613	0.0052	0.0834	2.061
Cadps	utr3	1.46	0.0056	0.0867	1.195
Rnf11	utr3	1.338	0.0058	0.0887	1.349
Atp8a1	utr3	1.145	0.0058	0.0891	1.165
Lrrc47	utr3	1.431	0.0059	0.0891	2.41
Ddc	utr3	2.352	0.006	0.0893	2.685
Ubqln2	utr3	1.28	0.006	0.0893	1.906
Slc25a1	utr3	1.235	0.0063	0.0921	1.277
Plcb1	utr3	1.36	0.0063	0.0921	1.224
Myh10	utr3	1.064	0.0064	0.0924	0.824
Wdr82	utr3	0.992	0.0064	0.0925	0.681
Plxnc1	utr3	1.322	0.0066	0.0933	1.34
Nap1l2	utr3	1.438	0.0069	0.0961	2.311
Nsf	utr3	1.144	0.007	0.0961	0.787
Peg3	utr3	1.161	0.007	0.0961	1.719
Cltc	utr3	1.204	0.007	0.0961	1.257
Nap1l3	utr3	1.694	0.0071	0.0961	3.077

CLIP Targets Defined From Annotated Subgenic Regions (Cont'd)

gene	feature	logFC _{IP/input}	p _{IP/input}	FDR	logFC _{IP/WTIP}
Atp6v1a	utr3	1.219	0.0072	0.0971	1.108
Pgap1	utr3	1.06	0.0073	0.0971	1.205
Scoc	utr3	1.227	0.0073	0.0971	1.587
Itfg1	utr3	1.229	0.0074	0.0974	1.331
Lpin2	utr3	1.038	0.0075	0.0978	0.825
Susd2	utr3	1.4	0.0076	0.0987	1.034
Plagl1	utr3	1.339	0.0077	0.0995	1.729
Syt4	utr3	1.277	0.0078	0.0995	2.004
Dnajo6	utr3	1.281	0.0079	0.1004	1.286
Lingo3	utr3	1.286	0.0081	0.1004	0.366
Mecp2	utr3	1.022	0.0083	0.1004	0.789
Dner	utr3	1.188	0.0083	0.1004	2.069
Nxph1	utr3	1.411	0.0084	0.1004	1.075
Fos	utr3	2.405	0.0084	0.1004	1.352
Rit2	utr3	1.375	0.0086	0.1015	2.105
Slc39a11	utr3	1.351	0.0087	0.1015	0.561
Glrb	utr3	1.22	0.009	0.1016	2.03
Snx16	utr3	1.37	0.009	0.1016	2.319
Map2k1	utr3	1.365	0.0091	0.1016	1.865
Ppa2	utr3	2.533	0.0091	0.1016	1.56
Frmpd4	utr3	1.225	0.0091	0.1016	0.48
Tceal1	utr3	1.766	0.0092	0.1022	2.302
Tmem30a	utr3	1.146	0.0095	0.1037	0.848
Stx7	utr3	1.169	0.0095	0.1039	1.777

CLIP Targets Defined From Annotated Subgenic Regions (Cont'd)

gene	feature	logFC _{IP/input}	p _{IP/input}	FDR	logFC _{IP/WTIP}
Ahi1	utr3	1.165	0.0097	0.1045	1.064
C1d	utr3	1.807	0.0101	0.1069	1.596
Hap1	utr3	0.941	0.0106	0.1111	0.966
Cdk17	utr3	1.333	0.0107	0.1113	0.996
Alkbh7	utr3	2.092	0.0108	0.1117	1.351
Mir3473d	utr3	5.437	0.011	0.113	5.319
Pcdh18	utr3	1.314	0.011	0.113	2.144
Slc9a6	utr3	0.908	0.0111	0.1136	1.571
Pip5kl1	utr3	2.177	0.0114	0.1154	1.623
Zfp385b	utr3	1.593	0.0116	0.1163	2.17
Lin7c	utr3	1.046	0.0119	0.1176	1.695
Gstm4	utr3	1.907	0.012	0.1182	0.749
Slc45a1	utr3	3.654	0.0121	0.1184	0.812
Hipk1	utr3	1.132	0.0121	0.1184	1.035
Nptn	utr3	1.09	0.0123	0.119	1.517
Nbea	utr3	0.975	0.0126	0.1212	1.213
Lrp12	utr3	1.549	0.0127	0.1212	0.932
Peg10	utr3	1.047	0.0128	0.1212	0.891
Pih1d1	utr3	3.504	0.0129	0.1217	1.427
Grem1	utr3	2.572	0.0129	0.1217	2.032
Dcaf7	utr3	0.993	0.0131	0.1223	0.415
Cnrip1	utr3	1.457	0.0134	0.1246	1.193
App	utr3	0.918	0.0138	0.1261	1.437
Tle4	utr3	1.245	0.014	0.1261	1.909

CLIP Targets Defined From Annotated Subgenic Regions (Cont'd)

gene	feature	logFC _{IP/input}	p _{IP/input}	FDR	logFC _{IP/WTIP}
Arf1	utr3	0.899	0.014	0.1261	0.717
Por	utr3	1.626	0.0141	0.1261	0.619
Nrep	utr3	0.912	0.0142	0.1261	1.108
Fgd4	utr3	1.734	0.0142	0.1261	2.047
Fem1b	utr3	0.916	0.0144	0.1275	1.048
Atp2c1	utr3	1.121	0.0145	0.1275	1.888
Pgs1	utr3	1.346	0.0146	0.1275	0.264
Kcnd2	utr3	1.101	0.015	0.1298	0.76
Mapk9	utr3	0.923	0.0153	0.132	0.584
Sybu	utr3	1.362	0.0158	0.1338	1.892
Ube2i	utr3	1.302	0.0158	0.1338	1.058
Ptcd3	utr3	2.495	0.0158	0.1338	1.726
Hcrr2	utr3	1.777	0.0159	0.1338	1.392
Fam160b1	utr3	1.05	0.0161	0.135	0.618
Cul4b	utr3	1.127	0.0167	0.1383	0.91
Lysmd4	utr3	1.427	0.0174	0.1417	1.034
Lancl2	utr3	1.422	0.0174	0.1417	0.802
Cdc42se2	utr3	1.298	0.0177	0.1431	0.973
Klhl13	utr3	1.116	0.0178	0.1431	1.371
Pum1	utr3	0.954	0.0178	0.1431	1.235
Slc19a1	utr3	1.608	0.0185	0.1472	0.673
Zbtb5	utr3	1.155	0.0185	0.1472	0.408
Hoxa5	utr3	1.657	0.0187	0.1473	2.196
Blcap	utr3	0.828	0.0188	0.1473	0.841

CLIP Targets Defined From Annotated Subgenic Regions (Cont'd)

gene	feature	logFC _{IP/input}	p _{IP/input}	FDR	logFC _{IP/WTIP}
Pspc1	utr3	1.233	0.0189	0.1473	1.47
Usp14	utr3	1.365	0.0195	0.1512	1.931
Fem1c	utr3	1.484	0.0197	0.1513	1.591
Crocc	utr3	1.415	0.0197	0.1513	0.717
Serinc1	utr3	0.939	0.02	0.1521	1.398
Trim37	utr3	0.922	0.0201	0.1525	1.137
Fsd1	utr3	1.982	0.0202	0.1527	1.923
Gnai1	utr3	0.967	0.0203	0.1527	1.362
Pnma2	utr3	1.009	0.0204	0.1528	0.646
Eid3	utr3	5.333	0.0205	0.1528	5.215
Tspan7	utr3	0.921	0.0207	0.1528	1.229
Pcdh17	utr3	0.982	0.0208	0.1531	1.314
Hnmt	utr3	1.478	0.0212	0.1548	1.356
Grm7	utr3	1.239	0.0212	0.1548	1.149
Tomm20	utr3	1.059	0.0212	0.1548	0.716
Ccdc74a	utr3	2.935	0.0213	0.1548	1.229
Dnaja1	utr3	1.006	0.0216	0.1565	1.097
Zdhhc13	utr3	1.393	0.0219	0.1575	0.699
8430427H17Rik	utr3	0.852	0.0221	0.1575	0.271
Ube2d2a	utr3	1.081	0.0221	0.1575	1.147
Rab11a	utr3	1.083	0.0222	0.1575	1.175
Psd3	utr3	0.848	0.0222	0.1575	0.763
Srsf3	utr3	1.239	0.0223	0.1575	1.073
Kif26b	utr3	1.101	0.0223	0.1575	0.621

CLIP Targets Defined From Annotated Subgenic Regions (Cont'd)

gene	feature	logFC _{IP/input}	p _{IP/input}	FDR	logFC _{IP/WTIP}
Reln	utr3	1.257	0.0224	0.1575	1.672
Pbx3	utr3	1.017	0.0234	0.1618	0.787
Mir7226	utr3	4.795	0.0234	0.1618	3.003
Ppp1r12a	utr3	1.189	0.0235	0.1625	1.001
Dmxl2	utr3	1.144	0.0241	0.1651	0.499
Gucy1b3	utr3	1.026	0.0247	0.1676	1.053
Srp72	utr3	1.113	0.025	0.169	0.992
Ambra1	utr3	1.096	0.0252	0.1697	0.291
Znfx1	utr3	1.37	0.0253	0.1697	0.363
Ranbp6	utr3	1.143	0.0254	0.1705	1.124
Myo5a	utr3	0.791	0.0256	0.1708	0.377
Tmeff1	utr3	1.038	0.0258	0.1708	1.984
Pkia	utr3	0.946	0.0258	0.1708	0.836
Pcdh11x	utr3	1.071	0.0258	0.1708	1.499
Slc6a5	utr3	1.837	0.0263	0.1736	0.437
Hivep2	utr3	1.1	0.0267	0.1748	1.056
Ywhaq	utr3	1.22	0.0268	0.1748	0.46
Cav2	utr3	1.092	0.0268	0.1748	1.97
Klf10	utr3	1.479	0.0269	0.1754	1.56
Sv2c	utr3	0.949	0.0273	0.1765	0.818
Spryd7	utr3	1.234	0.0273	0.1765	1.107
Atp1b1	utr3	0.878	0.0274	0.1765	1.354
Cox14	utr3	1.319	0.0278	0.1779	0.408
Dnajb4	utr3	1.117	0.0282	0.1798	1.729

CLIP Targets Defined From Annotated Subgenic Regions (Cont'd)

gene	feature	logFC _{IP/input}	p _{IP/input}	FDR	logFC _{IP/WTIP}
Nova1	utr3	0.967	0.0288	0.18	1.719
Hcfc1	utr3	0.95	0.0288	0.18	1.065
Plch1	utr3	0.955	0.0288	0.18	0.661
Dnajb6	utr3	1.428	0.0292	0.18	2.583
Lmo3	utr3	0.928	0.0292	0.18	0.627
Jazf1	utr3	1.084	0.0293	0.18	1.141
Arl8b	utr3	1.172	0.0294	0.18	0.877
Canx	utr3	0.892	0.0294	0.18	0.828
Lztf1	utr3	1.001	0.0303	0.1823	1.046
Desi2	utr3	1.023	0.0307	0.1833	0.412
Eif4g2	utr3	0.943	0.0307	0.1833	0.716
Syt13	utr3	0.899	0.0308	0.1833	0.347
Reep1	utr3	0.762	0.0309	0.1833	0.462
Arl6ip1	utr3	0.879	0.031	0.1837	0.869
Fam134b	utr3	0.986	0.0313	0.1839	1.38
Cdk20	utr3	1.586	0.0316	0.1842	1.103
Agbl5	intron	1.185	1.23E-5	9e-04	0.498
Syn2	intron	0.994	1.24E-5	9e-04	0.937
Gm3086	intron	2.086	1.82E-5	0.0012	1.781
Apba3	intron	2.224	1.87E-5	0.0012	1.431
Cxadr	intron	0.847	4.46E-5	0.0024	0.914
Etfb	intron	1.349	1e-04	0.0044	0.746
Dohh	intron	2.048	2e-04	0.0068	0.48
Nxt2	intron	2.367	3e-04	0.0075	1.367

CLIP Targets Defined From Annotated Subgenic Regions (Cont'd)

gene	feature	logFC _{IP/input}	p _{IP/input}	FDR	logFC _{IP/WTIP}
Dnajb1	intron	2.234	3e-04	0.008	1.005
Pisd-ps2	intron	1.368	3e-04	0.008	1.492
Nudt11	intron	2.043	4e-04	0.0084	1.692
Rac3	intron	1.146	5e-04	0.0104	0.547
Araf	intron	1.084	7e-04	0.0116	0.989
Pcdha4-g	intron	0.927	7e-04	0.0118	0.621
Mrps6	intron	1.144	7e-04	0.0118	0.579
Siva1	intron	1.866	8e-04	0.0123	0.642
Ufm1	intron	1.117	0.001	0.0152	0.372
Unc5b	intron	0.907	0.0015	0.0195	0.422
Pomgnt2	intron	1.016	0.0018	0.0219	0.267
Rps27l	cds	2.888	2.44E-6	0.0019	0.318
Rpl29	cds	1.472	1e-04	0.0307	0.693
Peli3	cds	1.378	3e-04	0.0521	1.17
Gtf3a	cds	1.579	4e-04	0.0521	0.851
Rnf144a	cds	1.437	4e-04	0.0521	0.454
Urm1	cds	1.514	8e-04	0.0765	0.73
Sfrp2	cds	1.829	8e-04	0.0765	1.631
Bckdha	cds	1.041	0.001	0.0765	0.356
Rnf122	cds	1.451	0.0018	0.0991	0.36
Cbfa2t2	cds	1.054	0.0021	0.1005	0.599
Hist1h2bb	cds	1.633	0.0022	0.1005	0.448
Hist1h2af	cds	2.169	0.0027	0.113	1.092
Plod1	cds	1.088	0.0033	0.113	0.587

CLIP Targets Defined From Annotated Subgenic Regions (Cont'd)

gene	feature	logFC _{IP/input}	p _{IP/input}	FDR	logFC _{IP/WTIP}
Anapc16	cds	1.202	0.0035	0.113	0.336
Zc2hc1a	utr5	3.118	6e-04	0.0298	1.533
Rrs1	utr5	3.327	0.0012	0.041	1.569

Table S5: Top Ranking Motif Cluster Matches Across UTR peaks

Gene.UTR_Peak: Gene name, alternate annotated UTR, identified peaks (if >1)

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Pgap1_p6	+	+	-	chr1	54479487	54479536	-
Zdbf2_p1	-	+	+	chr1	63312283	63312332	+
Zdbf2_p2	+	+	+	chr1	63314408	63314457	+
Dnm3	+	+	+	chr1	161987589	161987638	-
Dnm3.1_p3	-	+	-	chr1	161990860	161990909	-
Dnm3.1_p4	-	+	-	chr1	161991885	161991934	-
Srp9_p1	-	+	-	chr1	182132193	182132242	+
Plagl1_p2	+	+	+	chr10	13131042	13131091	+
Cdk17_p1	+	+	-	chr10	93240363	93240412	+
Plxnc1_p2	-	+	-	chr10	94792253	94792302	-
Plxnc1_p3	-	+	-	chr10	94792353	94792402	-
C1d	-	+	+	chr11	17266833	17266882	+
C1d_p1	-	+	+	chr11	17267008	17267057	+
Mapk9	-	+	-	chr11	49883571	49883620	+
Mapk9.1	-	+	-	chr11	49883577	49883626	+
Mapk9.1_p1	-	+	-	chr11	49883727	49883776	+
Mapk9_p1	-	+	-	chr11	49883746	49883795	+

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Ube2b	-	+	+	chr11	51985958	51986007	-
Ube2b_p1	+	+	+	chr11	51986433	51986482	-
Arf1	-	+	-	chr11	59211649	59211698	-
Arf1_p1	-	+	-	chr11	59212399	59212448	-
Myh10_p1	-	+	-	chr11	68816284	68816333	+
Cltc	+	+	+	chr11	86694840	86694889	-
Cltc_p1	-	+	+	chr11	86695315	86695364	-
Hap1.1	+	+	-	chr11	100347464	100347513	-
Nsf_p1	-	+	-	chr11	103822995	103823044	-
Dcaf7	-	+	-	chr11	106055168	106055217	+
Dcaf7_p1	+	+	-	chr11	106055418	106055467	+
Dcaf7_p3	+	+	-	chr11	106058818	106058867	+
Dcaf7_p4	+	+	-	chr11	106058993	106059042	+
Slc39a11	-	+	-	chr11	113245117	113245166	-
Slc39a11_p1	-	+	-	chr11	113245217	113245266	-
Pgs1.1	-	+	-	chr11	118023467	118023516	+
Nrsn1_p1	+	+	+	chr13	25252926	25252975	-
Dek_p1	-	+	-	chr13	47085554	47085603	-
Ankra2.1	-	+	-	chr13	98273796	98273845	+
Ankra2_p1	-	+	-	chr13	98273797	98273846	+
Kif2a	-	+	-	chr13	106961697	106961746	-
Rab3c_p1	-	+	-	chr13	110055774	110055823	-
Rab3c_p2	-	+	+	chr13	110056049	110056098	-
Rab3c_p6	-	+	+	chr13	110057674	110057723	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Synpr	-	+	+	chr14	13614319	13614368	+
Pcdh17	+	+	+	chr14	84534745	84534794	+
Pcdh17_p2	-	+	-	chr14	84535070	84535119	+
Pcdh17_p5	+	+	+	chr14	84536395	84536444	+
Tmtc4_p2	-	+	-	chr14	122919237	122919286	-
Tmtc4_p3	-	+	-	chr14	122919912	122919961	-
Azin1	+	+	-	chr15	38489117	38489166	-
Lrp12_p1	-	+	+	chr15	39870990	39871039	-
Mpped1	+	+	-	chr15	83856404	83856453	+
Slc25a1_p2	+	+	-	chr16	17925673	17925722	-
Cd200	+	+	+	chr16	45382422	45382471	-
App	-	+	-	chr16	84954948	84954997	-
Ttc3_p2	+	+	-	chr16	94468988	94469037	+
Rgmb_p1	-	+	-	chr17	15806865	15806914	-
Alkbh7	+	+	-	chr17	56999080	56999129	+
Lpin2	-	+	-	chr17	71247584	71247633	+
Lpin2_p2	-	+	-	chr17	71249159	71249208	+
Socs5	-	+	+	chr17	87135882	87135931	+
Rab18_p2	-	+	+	chr18	6788868	6788917	+
Rab18_p3	+	+	+	chr18	6788993	6789042	+
Rab18_p5	-	+	-	chr18	6789468	6789517	+
Rab18_p6	-	+	+	chr18	6789543	6789592	+
Slc39a6_p1	-	+	-	chr18	24580368	24580417	-
Syt4	+	+	-	chr18	31437845	31437894	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Syt4_p2	-	+	+	chr18	31438645	31438694	-
Nrep	+	+	+	chr18	33437231	33437280	-
Nrep_p1	-	+	+	chr18	33437956	33438005	-
Rnf14_p2	-	+	+	chr18	38317600	38317649	+
Tle4	-	+	-	chr19	14448284	14448333	-
Ranbp6	-	+	-	chr19	29808495	29808544	-
Ranbp6_p2	-	+	-	chr19	29809370	29809419	-
Htr7_p1	-	+	-	chr19	35959491	35959540	-
Tnks2	-	+	-	chr19	36891291	36891340	+
Tnks2_p1	-	+	-	chr19	36892616	36892665	+
Tnks2_p2	-	+	-	chr19	36892691	36892740	+
Tnks2_p3	+	+	-	chr19	36892816	36892865	+
Fam160b1_p3	-	+	-	chr19	57387414	57387463	+
Fam160b1_p5	-	+	-	chr19	57387889	57387938	+
Zfp385b_p2	-	+	-	chr2	77411839	77411888	-
Ube2e3	+	+	-	chr2	78920493	78920542	+
Lin7c	+	+	+	chr2	109899289	109899338	+
Lin7c_p1	+	+	-	chr2	109899764	109899813	+
Plcb1.1_p3	-	+	+	chr2	135474540	135474589	+
Plcb1.2_p4	-	+	-	chr2	135474543	135474592	+
8430427H17Rik	-	+	+	chr2	153407623	153407672	-
8430427H17Rik_p2	+	+	-	chr2	153408823	153408872	-
Nbea_p1	+	+	+	chr3	55625835	55625884	-
Gpr149	-	+	-	chr3	62529950	62529999	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
B3galt1_p1	-	+	+	chr3	69574306	69574355	-
B3galt1_p2	-	+	-	chr3	69574831	69574880	-
Serpini1	-	+	-	chr3	75640914	75640963	+
Serpini1_p2	-	+	-	chr3	75642189	75642238	+
Glrp_p2	-	+	-	chr3	80844736	80844785	-
Hipk1	+	+	+	chr3	103740077	103740126	-
Hipk1_p1	+	+	+	chr3	103741202	103741251	-
Fnbp1l	-	+	-	chr3	122539456	122539505	-
Fnbp1l_p1	+	+	-	chr3	122540306	122540355	-
Cnr1_p3	-	+	-	chr4	33946148	33946197	+
Cnr1_p5	-	+	-	chr4	33948523	33948572	+
Dnaja1_p1	-	+	-	chr4	40733359	40733408	+
Zbtb5_p1	-	+	+	chr4	44992605	44992654	-
Zbtb5_p2	-	+	-	chr4	44992680	44992729	-
Alg2	+	+	-	chr4	47469995	47470044	-
Trim32	-	+	-	chr4	65615563	65615612	+
Dnajc6	-	+	+	chr4	101642286	101642335	+
Rnf11	-	+	-	chr4	109452894	109452943	-
Rnf11_p2	-	+	-	chr4	109453994	109454043	-
2510039O18Rik_p1	+	+	+	chr4	147947006	147947055	+
Lrrc47	-	+	-	chr4	154020139	154020188	+
Lrrc47_p1	-	+	-	chr4	154020264	154020313	+
Lrrc47_p2	+	+	-	chr4	154020364	154020413	+
Qdpr	+	+	-	chr5	45434394	45434443	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Atp8a1_p3	-	+	-	chr5	67621426	67621475	-
Atp8a1_p4	-	+	-	chr5	67621526	67621575	-
Ran	-	+	-	chr5	129022831	129022880	+
Ran_p1	-	+	-	chr5	129024231	129024280	+
Peg10_p5	-	+	+	chr6	4758564	4758613	+
Kcnd2_p3	-	+	+	chr6	21729580	21729629	+
Reep1_p2	-	+	+	chr6	71808239	71808288	+
Reep1_p3	+	+	+	chr6	71809964	71810013	+
Grm7_p2	+	+	+	chr6	111566505	111566554	+
Syn2	-	+	-	chr6	115275162	115275211	+
Peg3	+	+	-	chr7	6706522	6706571	-
Peg3_p1	-	+	-	chr7	6706822	6706871	-
Ndn	-	+	-	chr7	62349523	62349572	+
Mkrn3	-	+	-	chr7	62417905	62417954	-
Ubfd1_p2	+	+	-	chr7	122079898	122079947	+
Vat1l	+	+	-	chr8	114371933	114371982	+
Atmin	-	+	+	chr8	116958247	116958296	+
Fem1b	-	+	-	chr9	62791917	62791966	-
Fem1b_p4	+	+	-	chr9	62793592	62793641	-
Map2k1	-	+	+	chr9	64185955	64186004	-
Myo5a_p4	-	+	+	chr9	75220606	75220655	+
Tmem30a_p2	+	+	+	chr9	79769953	79770002	-
Zic4_p1	+	+	-	chr9	91388498	91388547	+
Spsb4	-	+	-	chr9	96943719	96943768	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Wdr82_p1	-	+	-	chr9	106189688	106189737	+
Atp6ap2	-	+	-	chrX	12616486	12616535	+
Zcchc12	-	+	-	chrX	36198878	36198927	+
Cul4b_p1	+	+	-	chrX	38532083	38532132	-
Dcaf12l1_p1	-	+	-	chrX	44787929	44787978	-
Slc9a6	-	+	-	chrX	56663461	56663510	+
Slc9a6_p1	+	+	+	chrX	56663661	56663710	+
Gpr101_p3	-	+	-	chrX	57498055	57498104	-
Fgf13	+	+	-	chrX	59062658	59062707	-
Gabrq.1_p1	-	+	-	chrX	72841253	72841302	+
Mecp2_p1	+	+	+	chrX	74028261	74028310	-
Mecp2_p2	+	+	-	chrX	74028336	74028385	-
Mecp2_p9	+	+	+	chrX	74035211	74035260	-
Magee1	-	+	-	chrX	105123605	105123654	+
Tmem35	-	+	-	chrX	134305178	134305227	+
Gprasp2	-	+	-	chrX	135844487	135844536	+
Arxes2_p1	-	+	+	chrX	135995039	135995088	+
Arxes1_p1	-	+	-	chrX	136034461	136034510	+
Tceal1	-	+	+	chrX	136709445	136709494	+
Morf4l2	+	+	-	chrX	136733135	136733184	-
Pgap1	+	-	+	chr1	54473087	54473136	-
Pgap1_p1	-	-	-	chr1	54475187	54475236	-
Pgap1_p2	-	-	-	chr1	54476087	54476136	-
Pgap1_p3	-	-	-	chr1	54478587	54478636	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Pgap1_p4	+	-	-	chr1	54479262	54479311	-
Pgap1_p5	-	-	-	chr1	54479337	54479386	-
Pgap1_p7	-	-	-	chr1	54480487	54480536	-
Zdbf2	-	-	+	chr1	63311458	63311507	+
Scg2	+	-	-	chr1	79435031	79435080	-
Dner	+	-	+	chr1	84370251	84370300	-
Dnm3_p1	-	-	+	chr1	161988989	161989038	-
Dnm3_p2	+	-	-	chr1	161989739	161989788	-
Dnm3.1	-	-	+	chr1	161990560	161990609	-
Dnm3.1_p1	-	-	-	chr1	161990660	161990709	-
Dnm3.1_p2	-	-	-	chr1	161990785	161990834	-
Dnm3.1_p5	-	-	-	chr1	161991960	161992009	-
Srp9	+	-	+	chr1	182131793	182131842	+
Plagl1	+	-	+	chr10	13130592	13130641	+
Plagl1_p1	+	-	-	chr10	13130767	13130816	+
Ahi1	+	-	+	chr10	21079899	21079948	+
Stx7	-	-	-	chr10	24188133	24188182	+
Stx7_p1	-	-	+	chr10	24188208	24188257	+
Stx7_p2	-	-	+	chr10	24188708	24188757	+
Susd2_p1	-	-	-	chr10	75637056	75637105	-
Susd2_p2	-	-	-	chr10	75637156	75637205	-
Slc19a1_p1	-	-	-	chr10	77049944	77049993	+
Lingo3	+	-	-	chr10	80833713	80833762	-
Cdk17	-	-	+	chr10	93240188	93240237	+

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Cdk17_p2	-	-	-	chr10	93240438	93240487	+
Plxnc1	+	-	-	chr10	94791778	94791827	-
Plxnc1_p1	+	-	-	chr10	94791903	94791952	-
Syt1	-	-	+	chr10	108497687	108497736	-
Syt1_p1	-	-	-	chr10	108497787	108497836	-
Syt1_p2	-	-	-	chr10	108498987	108499036	-
Syt1_p3	-	-	+	chr10	108500312	108500361	-
Ddc	-	-	-	chr11	11814188	11814237	-
Mapk9.1_p2	-	-	-	chr11	49884002	49884051	+
Mapk9_p2	+	-	+	chr11	49884046	49884095	+
Mapk9.1_p3	-	-	-	chr11	49884102	49884151	+
Ube2b_p2	-	-	-	chr11	51986558	51986607	-
Arf1_p2	+	-	-	chr11	59212449	59212498	-
Myh10	-	-	-	chr11	68815509	68815558	+
Nsf	+	-	-	chr11	103821945	103821994	-
Dcaf7_p2	+	-	+	chr11	106058043	106058092	+
Slc39a11_p3	-	-	-	chr11	113246042	113246091	-
Slc39a11_p4	+	-	-	chr11	113246117	113246166	-
Klhdc2	-	-	-	chr12	69310639	69310688	+
Atf1	-	-	-	chr12	69963651	69963700	+
Hsp90aa1	-	-	-	chr12	110691373	110691422	-
Nrsn1	-	-	-	chr13	25252151	25252200	-
Dek_p2	-	-	+	chr13	47085829	47085878	-
Isca1	-	-	-	chr13	59755702	59755751	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Isca1_p1	-	-	-	chr13	59755902	59755951	-
Enc1.1_p1	-	-	-	chr13	97251374	97251423	+
Enc1.1_p2	+	-	-	chr13	97251999	97252048	+
Enc1.1_p3	-	-	-	chr13	97252374	97252423	+
Ankra2	-	-	-	chr13	98273397	98273446	+
Rab3c	-	-	-	chr13	110054974	110055023	-
Rab3c_p3	-	-	-	chr13	110056199	110056248	-
Rab3c_p4	-	-	-	chr13	110056474	110056523	-
Rab3c_p5	+	-	-	chr13	110056974	110057023	-
Cadps	-	-	+	chr14	12372700	12372749	-
Cadps_p1	+	-	+	chr14	12372875	12372924	-
Cadps_p2	-	-	-	chr14	12372925	12372974	-
Synpr_p1	+	-	-	chr14	13614469	13614518	+
Synpr_p2	-	-	+	chr14	13614819	13614868	+
Synpr_p3	-	-	-	chr14	13615319	13615368	+
Pspc1	-	-	-	chr14	56722636	56722685	-
Pspc1_p1	+	-	-	chr14	56722736	56722785	-
Pcdh17_p1	-	-	-	chr14	84534920	84534969	+
Pcdh17_p3	-	-	+	chr14	84535195	84535244	+
Pcdh17_p4	+	-	+	chr14	84535270	84535319	+
Tmtc4	-	-	+	chr14	122918987	122919036	-
Tmtc4_p1	-	-	-	chr14	122919162	122919211	-
Plcx3	+	-	-	chr15	4575215	4575264	+
Azin1_p1	-	-	-	chr15	38489567	38489616	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Lrp12	-	-	-	chr15	39870890	39870939	-
Lrp12_p2	-	-	-	chr15	39871265	39871314	-
Sybu	-	-	-	chr15	44672243	44672292	-
Mpped1_p2	-	-	+	chr15	83858304	83858353	+
Cox14	-	-	-	chr15	99727814	99727863	+
Atp6v1a	-	-	-	chr16	44087166	44087215	-
Cd200_p1	+	-	-	chr16	45382597	45382646	-
Cd200_p2	+	-	-	chr16	45382797	45382846	-
Ttc3	+	-	+	chr16	94468488	94468537	+
Ttc3_p1	-	-	-	chr16	94468738	94468787	+
Hmg1n1	-	-	+	chr16	96122000	96122049	-
Hmg1n1_p1	-	-	-	chr16	96122325	96122374	-
Rgmb	-	-	-	chr17	15806765	15806814	-
Syng3	-	-	-	chr17	24685679	24685728	-
Pja2_p1	-	-	-	chr17	64281793	64281842	-
Pja2_p2	+	-	+	chr17	64281943	64281992	-
Pja2_p3	-	-	-	chr17	64283368	64283417	-
Lpin2_p1	-	-	-	chr17	71248884	71248933	+
Socs5_p1	+	-	-	chr17	87136357	87136406	+
Calm2	+	-	+	chr17	87433813	87433862	-
Rab18	-	-	-	chr18	6788643	6788692	+
Rab18_p1	-	-	+	chr18	6788818	6788867	+
Rab18_p4	-	-	-	chr18	6789168	6789217	+
Rab18_p7	-	-	-	chr18	6789643	6789692	+

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Rab18_p8	-	-	-	chr18	6789818	6789867	+
Impact	+	-	-	chr18	12991246	12991295	+
Slc39a6	-	-	-	chr18	24580293	24580342	-
Slc39a6_p2	-	-	-	chr18	24580543	24580592	-
Slc39a6_p3	-	-	-	chr18	24580618	24580667	-
Rit2	+	-	+	chr18	30975151	30975200	-
Syt4_p3	+	-	+	chr18	31439245	31439294	-
Rnf14	-	-	-	chr18	38316900	38316949	+
Rnf14_p1	+	-	-	chr18	38317525	38317574	+
Fem1c	+	-	+	chr18	46504668	46504717	-
Tle4_p1	-	-	-	chr19	14448459	14448508	-
Ranbp6_p1	-	-	-	chr19	29809245	29809294	-
Uhrf2	+	-	+	chr19	30093102	30093151	+
Uhrf2_p1	-	-	-	chr19	30093177	30093226	+
Htr7	-	-	-	chr19	35959116	35959165	-
Fam160b1	+	-	-	chr19	57386564	57386613	+
Fam160b1_p1	-	-	+	chr19	57387214	57387263	+
Fam160b1_p2	-	-	+	chr19	57387289	57387338	+
Fam160b1_p6	-	-	-	chr19	57388364	57388413	+
Fam160b1_p7	+	-	+	chr19	57388464	57388513	+
Fam160b1_p8	-	-	-	chr19	57388564	57388613	+
Fam160b1_p9	+	-	+	chr19	57388814	57388863	+
Commd3	-	-	-	chr2	18675914	18675963	+
Hnmt	+	-	+	chr2	24003125	24003174	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Vav2	+	-	-	chr2	27263747	27263796	-
Vav2_p1	-	-	-	chr2	27264122	27264171	-
Olfm1	-	-	+	chr2	28214094	28214143	+
Olfm1.1	+	-	-	chr2	28230316	28230365	+
Gad1	-	-	-	chr2	70601059	70601108	+
Gad1_p1	-	-	+	chr2	70601809	70601858	+
Zfp385b	+	-	+	chr2	77411264	77411313	-
Zfp385b_p1	+	-	-	chr2	77411339	77411388	-
Lin7c_p2	-	-	-	chr2	109899964	109900013	+
Lin7c_p3	+	-	-	chr2	109900089	109900138	+
Lin7c_p4	+	-	+	chr2	109900239	109900288	+
Lin7c_p5	-	-	-	chr2	109900364	109900413	+
Plcb1.1	-	-	-	chr2	135472765	135472814	+
Plcb1.2	-	-	-	chr2	135472768	135472817	+
Plcb1.1_p1	-	-	-	chr2	135473190	135473239	+
Plcb1.2_p1	-	-	-	chr2	135473193	135473242	+
Plcb1.2_p2	-	-	-	chr2	135473343	135473392	+
Plcb1.1_p2	-	-	-	chr2	135473390	135473439	+
Plcb1.2_p3	-	-	-	chr2	135473393	135473442	+
Plcb1.1_p4	+	-	+	chr2	135474740	135474789	+
Plcb1.2_p5	+	-	-	chr2	135474743	135474792	+
Snap25	-	-	+	chr2	136781456	136781505	+
Snap25_p1	-	-	-	chr2	136782131	136782180	+
8430427H17Rik_p1	-	-	-	chr2	153408373	153408422	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
8430427H17Rik_p3	+	-	-	chr2	153409498	153409547	-
8430427H17Rik_p5	+	-	+	chr2	153410823	153410872	-
Blcap	-	-	-	chr2	157556774	157556823	-
Blcap_p1	+	-	+	chr2	157557474	157557523	-
Snx16	-	-	-	chr3	10418729	10418778	-
Pcdh18	-	-	-	chr3	49743383	49743432	-
Pcdh18_p1	-	-	+	chr3	49743533	49743582	-
Ccrn4l	+	-	+	chr3	51251245	51251294	+
Nbea	+	-	+	chr3	55625760	55625809	-
Nbea_p2	-	-	-	chr3	55626310	55626359	-
Pfn2	-	-	-	chr3	57843007	57843056	-
Gpr149_p1	-	-	-	chr3	62530125	62530174	-
B3galnt1	-	-	-	chr3	69574206	69574255	-
Serpini1_p1	-	-	-	chr3	75641864	75641913	+
Glr3	-	-	-	chr3	80844411	80844460	-
Glr3_p1	+	-	-	chr3	80844536	80844585	-
Hipk1_p2	-	-	-	chr3	103741877	103741926	-
Hipk1_p3	+	-	+	chr3	103742202	103742251	-
Hipk1_p4	-	-	-	chr3	103742377	103742426	-
Fnbp1_p2	-	-	-	chr3	122540431	122540480	-
Prkacb	-	-	-	chr3	146729941	146729990	-
Prkacb_p1	-	-	-	chr3	146730891	146730940	-
Prkacb_p2	+	-	-	chr3	146731141	146731190	-
Prkacb_p3	+	-	+	chr3	146732291	146732340	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Penk	-	-	-	chr4	4133698	4133747	-
Cnr1	-	-	-	chr4	33945148	33945197	+
Cnr1_p1	-	-	-	chr4	33945648	33945697	+
Cnr1_p2	+	-	-	chr4	33945973	33946022	+
Cnr1_p4	-	-	-	chr4	33948048	33948097	+
Cnr1_p6	-	-	+	chr4	33948598	33948647	+
Dnaja1	-	-	-	chr4	40732934	40732983	+
Dnaja1_p2	-	-	-	chr4	40733709	40733758	+
Zbtb5	-	-	-	chr4	44992130	44992179	-
Zbtb5_p3	-	-	-	chr4	44993330	44993379	-
Trim32_p1	-	-	-	chr4	65615663	65615712	+
Rraga	+	-	-	chr4	86576973	86577022	+
Rnf11_p1	-	-	-	chr4	109453219	109453268	-
Rnf11_p3	-	-	+	chr4	109454069	109454118	-
Pum1	+	-	-	chr4	130779916	130779965	+
Pum1_p1	-	-	-	chr4	130781241	130781290	+
Crocc	-	-	-	chr4	141016974	141017023	-
2510039O18Rik	+	-	+	chr4	147946806	147946855	+
2510039O18Rik_p2	-	-	-	chr4	147947056	147947105	+
Atp8a1	+	-	+	chr5	67620376	67620425	-
Atp8a1_p1	+	-	+	chr5	67620426	67620475	-
Atp8a1_p2	+	-	-	chr5	67621326	67621375	-
Atp8a1_p5	+	-	+	chr5	67621776	67621825	-
Atp8a1_p6	-	-	-	chr5	67622026	67622075	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Por	-	-	-	chr5	135734970	135735019	+
Peg10_p2	-	-	-	chr6	4756439	4756488	+
Peg10.1	-	-	-	chr6	4757478	4757527	+
Peg10_p3	-	-	-	chr6	4757489	4757538	+
Peg10.1_p1	-	-	+	chr6	4757628	4757677	+
Peg10_p4	-	-	+	chr6	4757639	4757688	+
Peg10.1_p2	-	-	+	chr6	4758628	4758677	+
Peg10_p6	+	-	+	chr6	4758689	4758738	+
Nxph1	-	-	-	chr6	9248009	9248058	+
Kcnd2	-	-	-	chr6	21727330	21727379	+
Kcnd2_p1	+	-	+	chr6	21728280	21728329	+
Kcnd2_p2	-	-	+	chr6	21728380	21728429	+
Nap115	-	-	-	chr6	58905570	58905619	-
Reep1	-	-	-	chr6	71807789	71807838	+
Reep1_p1	-	-	-	chr6	71808039	71808088	+
Reep1_p4	+	-	-	chr6	71810389	71810438	+
Grm7	-	-	-	chr6	111566230	111566279	+
Grm7_p1	-	-	+	chr6	111566330	111566379	+
Grm7_p3	-	-	+	chr6	111566905	111566954	+
Syn2_p1	+	-	+	chr6	115276137	115276186	+
Syn2_p2	-	-	+	chr6	115276437	115276486	+
Syn2.1	+	-	-	chr6	115281711	115281760	+
Zdhhc13	-	-	+	chr7	48826995	48827044	+
Mkrn3_p1	+	-	+	chr7	62418305	62418354	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Ubfd1_p1	+	-	+	chr7	122079673	122079722	+
Ubfd1_p3	+	-	-	chr7	122080048	122080097	+
Ubfd1_p4	+	-	-	chr7	122080573	122080622	+
Ubfd1_p5	-	-	+	chr7	122081048	122081097	+
Carkd	-	-	+	chr8	11513158	11513207	+
Tmem66	-	-	-	chr8	34170659	34170708	+
Gpm6a	+	-	+	chr8	55060492	55060541	+
Psd3	-	-	-	chr8	67689569	67689618	-
Psd3_p1	-	-	-	chr8	67691819	67691868	-
Psd3_p2	-	-	-	chr8	67692969	67693018	-
Psd3_p3	+	-	-	chr8	67693044	67693093	-
Psd3_p4	+	-	+	chr8	67693144	67693193	-
Psd3_p5	+	-	+	chr8	67693819	67693868	-
Psd3_p6	+	-	-	chr8	67694369	67694418	-
Psd3_p7	+	-	+	chr8	67695094	67695143	-
Psd3_p8	-	-	-	chr8	67695294	67695343	-
Psd3_p9	-	-	-	chr8	67695969	67696018	-
Psd3_p10	-	-	-	chr8	67696569	67696618	-
Scoc	-	-	-	chr8	83435479	83435528	-
Itfg1	-	-	-	chr8	85718519	85718568	-
Atmin_p2	+	-	-	chr8	116960297	116960346	+
AW551984	-	-	-	chr9	39588383	39588432	-
AW551984_p1	+	-	-	chr9	39588608	39588657	-
AW551984_p2	-	-	-	chr9	39588783	39588832	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Nptrn.1	+	-	-	chr9	58652071	58652120	+
Nptrn.1_p1	-	-	-	chr9	58652346	58652395	+
Fem1b_p1	-	-	-	chr9	62791967	62792016	-
Fem1b_p2	+	-	+	chr9	62792217	62792266	-
Fem1b_p3	-	-	-	chr9	62792467	62792516	-
Fem1b_p5	-	-	-	chr9	62793967	62794016	-
Map2k1_p1	+	-	-	chr9	64186130	64186179	-
Myo5a	-	-	-	chr9	75217981	75218030	+
Myo5a_p1	-	-	-	chr9	75218731	75218780	+
Myo5a_p2	+	-	-	chr9	75220156	75220205	+
Myo5a_p3	+	-	-	chr9	75220506	75220555	+
Myo5a_p5	+	-	-	chr9	75220731	75220780	+
Myo5a_p6	-	-	-	chr9	75220881	75220930	+
Myo5a_p7	+	-	-	chr9	75221506	75221555	+
Myo5a_p8	-	-	+	chr9	75222331	75222380	+
Myo5a_p9	-	-	-	chr9	75222781	75222830	+
Tmem30a	-	-	-	chr9	79769703	79769752	-
Tmem30a_p1	+	-	-	chr9	79769828	79769877	-
Tmem30a_p3	-	-	+	chr9	79770203	79770252	-
Tmem30a_p4	-	-	-	chr9	79770653	79770702	-
Zic4	-	-	-	chr9	91387348	91387397	+
Zic4_p2	-	-	-	chr9	91388873	91388922	+
Spsb4_p1	+	-	-	chr9	96944019	96944068	-
Spsb4_p2	+	-	-	chr9	96944144	96944193	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Spsb4_p3	-	-	-	chr9	96944494	96944543	-
Atp2c1	-	-	+	chr9	105411624	105411673	-
Wdr82	+	-	-	chr9	106189538	106189587	+
Wdr82_p2	+	-	-	chr9	106190213	106190262	+
Epm2aip1	+	-	-	chr9	111276194	111276243	+
Epm2aip1_p1	-	-	+	chr9	111276444	111276493	+
Epm2aip1_p2	+	-	-	chr9	111276594	111276643	+
Epm2aip1_p3	+	-	+	chr9	111276969	111277018	+
Epm2aip1_p4	+	-	+	chr9	111278944	111278993	+
Zcchc12_p1	-	-	-	chrX	36198953	36199002	+
Cul4b	+	-	-	chrX	38532008	38532057	-
Cul4b_p3	-	-	-	chrX	38532708	38532757	-
Dcaf12h1	+	-	+	chrX	44787579	44787628	-
Gpr101	-	-	+	chrX	57497705	57497754	-
Gpr101_p1	+	-	-	chrX	57497805	57497854	-
Gpr101_p2	-	-	+	chrX	57497955	57498004	-
Gpr101_p4	-	-	+	chrX	57498255	57498304	-
Gpr101_p5	-	-	-	chrX	57498605	57498654	-
Gpr101_p6	+	-	+	chrX	57499330	57499379	-
Gpr101_p7	+	-	-	chrX	57500105	57500154	-
Gabrq.1	-	-	-	chrX	72839928	72839977	+
Gabrq.1_p2	+	-	-	chrX	72841503	72841552	+
Mecp2	+	-	+	chrX	74028086	74028135	-
Mecp2_p3	+	-	-	chrX	74032061	74032110	-

Top Ranking Motif Cluster Matches Across UTR peaks cont'd

Gene.UTR_Peak	Pcbp3	Celf	U-rich	chr	start	stop	strand
Mecp2_p4	-	-	-	chrX	74032736	74032785	-
Mecp2_p6	-	-	-	chrX	74033186	74033235	-
Mecp2_p7	-	-	+	chrX	74033986	74034035	-
Pgr15l	-	-	+	chrX	97079888	97079937	+
Pgr15l_p1	+	-	-	chrX	97080038	97080087	+
Pgr15l_p2	+	-	+	chrX	97080263	97080312	+
Pgr15l_p3	-	-	-	chrX	97081088	97081137	+
Pgr15l_p4	-	-	-	chrX	97081663	97081712	+
Nap1l2	-	-	-	chrX	103184646	103184695	-
Nap1l3	-	-	-	chrX	122395023	122395072	-
Arxes2	-	-	+	chrX	135994989	135995038	+
Arxes1	-	-	+	chrX	136034036	136034085	+
Zcchc18	+	-	-	chrX	136996132	136996181	+
Irs4	-	-	-	chrX	141711060	141711109	-
Irs4_p1	-	-	-	chrX	141711635	141711684	-
Irs4_p2	-	-	-	chrX	141711985	141712034	-
Htr2c	-	-	-	chrX	147195133	147195182	+
Htr2c_p1	-	-	-	chrX	147195283	147195332	+
Htr2c_p2	+	-	-	chrX	147195883	147195932	+
Shroom2	-	-	-	chrX	152610321	152610370	-
Shroom2_p1	+	-	+	chrX	152610946	152610995	-
Rragb	-	-	+	chrX	153171651	153171700	+
Ubqln2	-	-	+	chrX	153500908	153500957	+

Table S6: PTRE-Seq - Effects on Expression

Terms shown: Main effect of sequence (reference or mutant), Main effect of condition (CELF overexpression or CTL), Interaction between sequence and condition.

FDR is shown for sequence x condition interactions

Gene.UTR_Peak	p			R ²	% variance explained					FDR
	sequence	condition	interaction		sequence	condition	interaction	within samples	unexplained	
Fgf13	1.41E-62	5.51E-38	2e-04	66.48%	23.33%	40.76%	2.34%	4.1%	29.47%	1.03E-61
Magee1	6.06E-188	1.42E-20	1.92E-16	75.36%	51.67%	18.27%	5.37%	3.42%	21.27%	2.14E-186
Plcx3	3.13E-38	6.22E-11	0.8759	38.45%	26.48%	11.46%	0.49%	0.85%	60.72%	1.26E-37
Ddc	1.65E-177	3.97E-8	0.6832	69.18%	64.94%	3.84%	0.39%	<0.01%	30.84%	4.37E-176
Mkx3	1.06E-100	3.02E-6	2.6E-6	57.47%	41.81%	12.11%	3.53%	7.24%	35.32%	1.21E-99
Mkx3_p1	3.77E-48	0.1221	0.5946	36.94%	33.06%	3.02%	0.86%	3.49%	59.57%	1.86E-47
Carkd	5.55E-19	0.0435	0.8034	23.41%	19.02%	3.47%	0.91%	0%	76.6%	1.41E-18
Nap13	0.4012	1.32E-10	0.092	25.26%	0.15%	22.42%	2.69%	4.88%	69.86%	0.4264
Pgr15l	0.1286	0.412	0.8656	3.2%	0.58%	1.81%	0.81%	0%	96.8%	0.1435
Pgr15l_p1	1.64E-180	1.86E-5	0.3659	65.27%	51.19%	13.57%	0.48%	10.83%	23.93%	4.95E-179
Pgr15l_p2	5.79E-10	0.0133	0.9723	14.15%	9.39%	4.33%	0.43%	<0.01%	85.85%	1.07E-9
Pgr15l_p3	7.62E-8	0.7091	0.4398	10.34%	7.31%	1.28%	1.74%	0.82%	88.84%	1.25E-7
Slc39a6	4.66E-25	3.59E-24	4.2E-5	45.29%	16.28%	24.12%	4.86%	1.23%	53.5%	1.37E-24
Slc39a6_p1	1.78E-59	6.12E-7	0.2513	47.78%	37.68%	8.8%	1.28%	2.2%	50.03%	1.15E-58
Slc39a6_p2	7.57E-43	2.18E-7	0.0351	39.76%	29.02%	8.4%	2.32%	1.19%	59.07%	3.34E-42
Slc39a6_p3	~0	2.11E-23	1.62E-35	85.34%	71.06%	7.01%	7.24%	0.54%	14.15%	~0
Trim32	8.03E-181	4e-04	1.75E-11	71.88%	61.56%	5.47%	4.83%	4.23%	23.9%	2.62E-179
Trim32_p1	6.62E-77	1.64E-21	1e-04	58.13%	41.09%	13.54%	3.47%	<0.01%	41.9%	5.84E-76
Gpr149	2.35E-17	8.41E-9	1.98E-10	32.87%	12.44%	10.14%	10.28%	0.8%	66.34%	5.64E-17
Gpr149_p1	5.77E-11	0.0558	0.8073	17.39%	12.35%	3.96%	1.08%	0%	82.62%	1.1E-10
Tceal1	3.26E-79	0.0134	0.1449	56.79%	51.92%	3.27%	1.58%	1.24%	41.98%	2.94E-78

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Gprasp2	3.59E-56	3.12E-10	0.001	48.62%	36.49%	8.54%	3.57%	0%	51.39%	2.15E-55
Gad1	0.6579	0.009	0.9142	13.56%	0.04%	12.92%	0.6%	14.8%	71.64%	0.6821
Gad1_p1	5.52E-34	0.0212	0.0479	35.86%	29.7%	3.31%	2.85%	<0.01%	64.14%	1.98E-33
Zcchc12	7.04E-26	1.32E-11	0.4131	34.28%	20.71%	12.22%	1.34%	0%	65.73%	2.1E-25
Zcchc12_p1	1.59E-22	0.0374	0.045	28.28%	20.38%	4.83%	3.07%	3.52%	68.2%	4.41E-22
Klhdc2	3e-04	0.0491	0.4923	8.86%	3.02%	4.36%	1.48%	2.47%	88.66%	4e-04
Lrrc47	4.76E-33	0.0023	0.0153	36.91%	25.72%	8.07%	3.11%	5.88%	57.23%	1.64E-32
Lrrc47_p1	0.0193	2.34E-5	0.6748	15.2%	1.57%	12.23%	1.4%	2.58%	82.22%	0.0233
Lrrc47_p2	7.7E-38	1.67E-9	4.69E-5	41.77%	27.43%	9.08%	5.25%	0%	58.25%	3.02E-37
Alg2	~0	4.41E-58	3.41E-44	86.05%	62.51%	15.83%	7.66%	0.66%	13.34%	~0
Arxes1_p1	8.53E-59	2.25E-6	0.1049	47.05%	39.42%	5.83%	1.79%	0%	52.96%	5.32E-58
Snca	5.15E-10	0.1305	0.5689	13.67%	9.5%	2.75%	1.41%	<0.01%	86.34%	9.54E-10
Syn2	5.87E-69	1.83E-19	6e-04	53.3%	37.55%	12.64%	3.09%	0%	46.72%	4.79E-68
Syn2_p2	1.18E-13	1.53E-14	0.9021	28.61%	11.14%	16.89%	0.57%	0.3%	71.1%	2.53E-13
Syn2.1	0.0143	0.0017	0.0392	11.08%	1.52%	5.82%	3.74%	0%	88.93%	0.0173
Rragb	1.2E-43	0.0046	0.5271	40.65%	35.7%	3.81%	1.14%	0%	59.35%	5.42E-43
Nap1l2	1.8E-116	4.96E-14	4.88E-11	68.15%	44.35%	18.49%	5.25%	4.95%	26.95%	2.38E-115
Tmtc4	3.55E-146	0.0361	0.3886	65.47%	62.72%	2.04%	0.7%	1.32%	33.22%	6.28E-145
Tmtc4_p1	7.95E-34	0.5793	0.8662	28.6%	26.65%	1.37%	0.58%	1.96%	69.44%	2.81E-33
Tmtc4_p2	2.17E-105	0.0633	2.42E-17	57.23%	44.31%	4.21%	8.7%	7.08%	35.7%	2.56E-104
Tmtc4_p3	3.02E-11	2.05E-9	0.8204	25.94%	10.02%	15.08%	0.83%	1.62%	72.45%	5.84E-11
Zfp385b	7.58E-14	4e-04	0.3586	24.03%	12.52%	9.78%	1.73%	4.53%	71.45%	1.63E-13
Zfp385b_p1	2.49E-19	2e-04	0.0017	34.15%	20.87%	7.34%	5.93%	0%	65.86%	6.36E-19
Zfp385b_p2	6.51E-138	7.5E-6	0.5434	65.51%	61.39%	3.52%	0.59%	<0.01%	34.5%	1.1E-136

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Zdbf2	3.13E-55	6.03E-6	0.0156	45.98%	37.71%	5.6%	2.66%	0%	54.03%	1.82E-54
Zdbf2_p1	6.55E-31	0.5821	0.9957	32.82%	31.28%	1.32%	0.22%	0%	67.18%	2.12E-30
Zdbf2_p2	1.39E-47	1.11E-24	8e-04	50.88%	29.37%	18.02%	3.46%	0%	49.15%	6.8E-47
Arxes2	1.24E-7	0.4597	0.045	13.53%	7.22%	2.6%	3.71%	4.14%	82.34%	2E-7
Arxes2_p1	5.39E-59	0.5829	0.9109	45.66%	44.14%	1.06%	0.45%	0.67%	53.68%	3.41E-58
Penk	2.1E-11	3.25E-10	3e-04	25.43%	8.74%	11.36%	5.33%	0%	74.58%	4.12E-11
At11	9.92E-27	4.19E-6	0.0476	32.98%	21.27%	9.06%	2.64%	1.84%	65.2%	3E-26
Snx16	0.0382	0.1214	0.813	5.92%	1.01%	4.04%	0.87%	3.76%	90.32%	0.0451
Dcaf12l1	7.44E-30	0.2444	0.5463	27.31%	24.45%	1.73%	1.13%	0%	72.69%	2.32E-29
Dcaf12l1_p1	0.0284	0.0098	0.2053	10.33%	1.5%	5.79%	3.03%	<0.01%	89.67%	0.0338
Synpr	5.25E-54	0.8101	0.0026	41.57%	36.51%	1.71%	3.35%	9.79%	48.64%	2.93E-53
Synpr_p1	1.67E-33	0.0198	0.8463	32.1%	27.5%	3.95%	0.64%	1.55%	66.36%	5.87E-33
Synpr_p2	0.0039	2e-04	3.71E-8	18.1%	1.78%	6.07%	10.24%	0%	81.9%	0.005
Cd200	7.48E-14	0.004	0.9559	19.81%	14.06%	5.23%	0.52%	0%	80.19%	1.62E-13
Cd200_p1	0.0069	1.07E-5	0.5173	14.49%	2.18%	10.46%	1.84%	<0.01%	85.52%	0.0086
Cd200_p2	2.38E-151	2.06E-5	0.429	67.45%	63.68%	3.11%	0.65%	0%	32.56%	4.59E-150
Ndn	6.28E-39	2.1E-8	3.63E-7	43.41%	22.75%	14.93%	5.71%	5.43%	51.18%	2.63E-38
AW551984	3.67E-44	0.0017	0.6153	43.61%	36.27%	6.32%	1%	2.8%	53.61%	1.69E-43
AW551984_p1	6.17E-7	0.233	0.8339	9.69%	6.4%	2.39%	0.9%	0%	90.31%	9.79E-7
Ankra2	0.1558	0.0696	0.9395	12.04%	0.35%	11.29%	0.4%	22%	65.96%	0.1707
Ankra2_p1	0.0058	0.0325	0.015	13.21%	2.22%	5.91%	5.08%	3.02%	83.78%	0.0073
Ankra2.1	4.05E-8	0.7013	0.2419	12.14%	7.93%	1.8%	2.41%	3.94%	83.93%	6.71E-8
Rraga	3.99E-38	5e-04	1.54E-7	45.29%	31.77%	4.99%	8.52%	<0.01%	54.72%	1.6E-37
Irs4	5.67E-109	0.4301	9e-04	60.4%	56.11%	1.5%	2.79%	3.2%	36.4%	7.28E-108

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Irs4_p1	3.94E-43	0.1778	0.9734	34.47%	32.43%	1.74%	0.3%	0%	65.53%	1.76E-42
Irs4_p2	0.0079	0.0191	0.5271	7.85%	1.85%	4.4%	1.6%	0%	92.15%	0.0098
Dek	5.33E-12	0.1305	0.2045	18.03%	11.75%	3.87%	2.4%	3.15%	78.82%	1.08E-11
Dek_p1	1.34E-9	0.0045	0.2018	19.25%	7.57%	9.66%	2.01%	8.42%	72.33%	2.42E-9
Dek_p2	5.37E-108	0.1844	0.0015	56.05%	51.41%	2.18%	2.46%	3.55%	40.41%	6.7E-107
Rit2	8.59E-46	3.97E-6	0.4564	40.8%	28.4%	11.44%	0.95%	5.28%	53.94%	4.05E-45
Pcdh18	~0	0.3564	8.55E-8	82.73%	79.87%	0.92%	1.94%	2.44%	14.84%	~0
Pcdh18_p1	1.39E-9	0.0247	0.1786	15.47%	7.65%	5.7%	2.12%	4.7%	79.83%	2.5E-9
Hsp90aa1	0.0706	0.8391	0.9818	2.79%	1.11%	1.17%	0.51%	0%	97.21%	0.0805
C1d	1.3E-17	0.7183	0.0454	19.26%	15.19%	1.09%	2.98%	1.03%	79.71%	3.17E-17
C1d_p1	1.05E-7	0.0212	0.3052	13.13%	7%	4.07%	2.06%	0%	86.87%	1.71E-7
Alkbh7	2.88E-6	0.4154	0.1414	11.25%	6%	2.26%	2.99%	1.39%	87.36%	4.43E-6
Gabrq.1	3.15E-20	6.09E-6	0.0383	26.7%	16.06%	7.83%	2.8%	0.83%	72.48%	8.14E-20
Gabrq.1_p1	5.75E-37	0.1359	0.002	33.38%	26.65%	2.99%	3.74%	3.36%	63.26%	2.12E-36
Nap1l5	0.001	0.1613	0.5184	8.73%	3.41%	3.34%	1.97%	<0.01%	91.27%	0.0014
Htr2c_p1	2.38E-7	6.76E-5	0.8587	16%	7.03%	8.11%	0.86%	0%	84%	3.82E-7
Htr2c_p2	1.8E-65	2.2E-26	6.84E-12	58.54%	34.48%	16.16%	7.87%	0%	41.49%	1.38E-64
Commd3	0.0055	0.4373	0.0611	8.3%	2.17%	2.33%	3.8%	1.72%	89.98%	0.007
Fnbp1l	3.64E-24	0.0034	0.0039	33.57%	23.8%	4.92%	4.84%	<0.01%	66.44%	1.04E-23
Fnbp1l_p1	3.21E-16	0.3059	0.4381	18.92%	15.4%	1.92%	1.6%	<0.01%	81.08%	7.4E-16
Fnbp1l_p2	3.25E-24	5.73E-8	0.0044	33.67%	16.76%	13.54%	3.35%	4.02%	62.32%	9.31E-24
Ccrn4l	0.0124	2.64E-7	0.0019	15.93%	1.37%	9.56%	5%	0%	84.07%	0.0152
Dner	6.45E-8	3.45E-11	0.5483	23.69%	6.98%	15.29%	1.41%	0.1%	76.21%	1.06E-7
Syt4	6.34E-10	0.0024	7e-04	19.56%	8.76%	5.06%	5.74%	0%	80.44%	1.16E-9

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Syt4_p3	1.66E-8	2.69E-6	0.991	16.89%	7.54%	9.06%	0.28%	0%	83.11%	2.85E-8
Sybu	4.45E-53	8.69E-12	0.6853	44.42%	34.12%	9.6%	0.69%	0%	55.59%	2.42E-52
Tnks2	4.76E-60	0.0196	1.26E-13	47.67%	34.16%	3.89%	9.61%	3.36%	48.98%	3.15E-59
Tnks2_p1	4.79E-56	0.021	2.23E-11	45.98%	34.18%	2.99%	8.8%	1.4%	52.62%	2.82E-55
Tnks2_p2	1.73E-126	6.64E-5	0.9511	65.37%	61.45%	3.67%	0.23%	0.38%	34.27%	2.44E-125
Zcchc18	0.0071	0.0584	0.8686	6.54%	1.92%	3.78%	0.84%	0.4%	93.05%	0.0088
Map2k1	6e-04	1.48E-19	0.1715	26.45%	2.47%	21.83%	2.16%	<0.01%	73.55%	8e-04
Map2k1_p1	3.83E-11	0.614	0.0622	15.12%	10.57%	1.3%	3.25%	<0.01%	84.88%	7.35E-11
Glrp_p2	0.1669	0.2188	0.7591	4.66%	0.57%	2.84%	1.25%	<0.01%	95.34%	0.1819
Olfm1.1	6.48E-18	2.76E-46	7.61E-8	55.08%	11.63%	36.18%	7.24%	0%	44.95%	1.61E-17
Nrsn1	8.09E-9	1.04E-6	0.5903	29.72%	5.4%	23.41%	0.9%	13.3%	56.99%	1.42E-8
Nrsn1_p1	4.97E-47	3.01E-8	3.31E-9	50.46%	29.3%	13.62%	7.51%	4.48%	45.08%	2.39E-46
Epm2aip1	0.1202	0.137	0.9518	3.91%	0.61%	2.77%	0.54%	0%	96.09%	0.1348
Epm2aip1_p1	1.97E-148	0.0037	0.6417	64.62%	62.19%	1.94%	0.47%	<0.01%	35.39%	3.63E-147
Epm2aip1_p2	1.01E-15	3.07E-9	0.6988	33.36%	10.82%	21.73%	0.79%	7.62%	59.03%	2.28E-15
Epm2aip1_p3	5.52E-15	0.4225	0.9785	21.47%	18.8%	2.17%	0.49%	0%	78.53%	1.23E-14
Epm2aip1_p4	1.42E-193	7.06E-26	3.15E-11	76.07%	40.78%	32.29%	2.93%	6.19%	17.81%	6.02E-192
Tle4	1.41E-37	2.13E-5	0.2779	37.02%	29.45%	6.01%	1.55%	0%	62.99%	5.5E-37
Tle4_p1	7.53E-22	0.0066	3.94E-5	27.7%	16.88%	4.96%	5.86%	2.26%	70.05%	2.03E-21
Snap25	6.59E-7	0.4014	0.0605	11.3%	5.28%	3.14%	2.88%	6.99%	81.71%	1.04E-6
Snap25_p1	0.0049	1.06E-11	0.1573	27.46%	1.37%	24.26%	1.83%	6.25%	66.29%	0.0062
Cnr1	0.0614	0.0219	0.2675	8.25%	1.01%	4.71%	2.53%	<0.01%	91.75%	0.0706
Cnr1_p4	0.1544	1e-04	0.8457	8.3%	0.49%	7%	0.81%	0%	91.7%	0.17
Cnr1_p5	2.15E-169	1.38E-11	1.02E-13	70.05%	53.26%	11.53%	5.23%	3.44%	26.54%	5.36E-168

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Cnr1_p6	0.974	1.12E-6	0.3756	14.58%	<0.01%	12.6%	1.98%	1.6%	83.83%	0.9763
Plagl1	2e-04	1.11E-5	0.0372	14.3%	3.12%	7.84%	3.34%	0%	85.71%	3e-04
Plagl1_p1	2.55E-9	0.0557	0.2314	15.51%	9.4%	3.64%	2.46%	<0.01%	84.49%	4.54E-9
Plagl1_p2	1.12E-99	5.24E-7	2e-04	57.55%	49.78%	4.65%	3.11%	0%	42.47%	1.24E-98
Impact	8.83E-22	1.83E-11	0.2155	39.02%	14.25%	23.28%	1.48%	6.56%	54.44%	2.37E-21
Tmem35	7.54E-165	1.61E-6	6.52E-15	72.27%	59.82%	5.92%	6.51%	2.25%	25.51%	1.78E-163
Atp2c1	2.11E-6	4.72E-10	0.3188	35.84%	2.98%	31.78%	1.08%	13.4%	50.76%	3.25E-6
Fem1c	5.66E-96	7.3E-6	0.3183	57.57%	52.23%	4.34%	0.99%	0%	42.44%	6E-95
Qdpr	4.02E-7	0.053	0.0107	15.36%	6.82%	3.69%	4.85%	<0.01%	84.64%	6.44E-7
Stx7	4.56E-62	8.86E-10	0.4698	46.96%	38.27%	7.77%	0.92%	0%	53.05%	3.22E-61
Stx7_p1	2.3E-268	5.07E-26	2.27E-32	79.96%	62.74%	8.73%	8.46%	0.43%	19.65%	1.63E-266
Stx7_p2	1.02E-5	0.0284	0.8128	11.92%	5.98%	4.8%	1.14%	0%	88.08%	1.55E-5
Enc1.1	5.04E-21	1.93E-7	0.3437	29.24%	17.61%	10.06%	1.57%	0.91%	69.85%	1.33E-20
Enc1.1_p1	0.0023	0.1353	0.6598	8.13%	2.58%	4.16%	1.39%	3.1%	88.77%	0.0031
Enc1.1_p2	1.01E-80	2e-04	0.1638	58.22%	52.48%	4.21%	1.51%	0.18%	41.62%	9.29E-80
Enc1.1_p3	9.01E-11	1.66E-29	0.7954	47.35%	5.79%	41.01%	0.53%	4.22%	48.45%	1.7E-10
Peg3_p1	3.37E-84	2.57E-10	1.91E-10	60.92%	45.84%	7.85%	7.2%	0.39%	38.71%	3.33E-83
Scoc	5.71E-22	7.38E-12	2.24E-15	40.93%	15.62%	11.2%	14.1%	0%	59.08%	1.55E-21
Rab18_p1	3.69E-53	1.14E-5	0.0607	44.72%	37.09%	5.5%	2.12%	0%	55.29%	2.03E-52
Rab18_p2	0.4696	0.0282	0.1328	7.89%	0.15%	4.53%	3.21%	0%	92.11%	0.4929
Rab18_p3	8.65E-58	0.0382	0.265	46.27%	41.36%	3.48%	1.42%	2.35%	51.39%	5.32E-57
Rab18_p4	2.22E-11	0.1212	0.9392	18.67%	14.28%	3.64%	0.74%	<0.01%	81.34%	4.32E-11
Rab18_p5	3.95E-208	4.04E-32	1.63E-41	79.08%	55.85%	10.92%	12.27%	0.24%	20.72%	1.86E-206
Rab18_p6	2.95E-16	0.7679	0.9043	17.36%	15.74%	0.97%	0.65%	<0.01%	82.64%	6.83E-16

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Rab18_p7	0.8295	0.0841	0.9956	4.5%	0.02%	4.17%	0.32%	0%	95.5%	0.8475
Rab18_p8	3.16E-11	0.0331	0.5144	17.04%	11.46%	3.96%	1.62%	<0.01%	82.96%	6.09E-11
Ube2e3	2.64E-13	0.137	0.3587	20.1%	14.88%	3.07%	2.15%	<0.01%	79.9%	5.57E-13
Ttc3	3.03E-35	1.97E-8	0.5604	40.19%	24.01%	15.25%	0.91%	4.88%	54.96%	1.1E-34
Ttc3_p2	1.62E-5	9.19E-21	0.3329	42.78%	2.42%	39.31%	1.04%	7.24%	49.98%	2.44E-5
Lin7c	0.0206	1.49E-10	0.6379	18.7%	1.22%	16.3%	1.18%	1.4%	79.91%	0.0247
Lin7c_p3	3.85E-8	0.0028	0.0636	17%	7.86%	5.66%	3.48%	0%	83%	6.42E-8
Lin7c_p4	2.53E-74	0.2992	0.9766	51.79%	50.27%	1.27%	0.25%	0%	48.21%	2.1E-73
Lin7c_p5	1.54E-6	0.4142	0.9026	9.39%	6.56%	2.03%	0.8%	<0.01%	90.61%	2.4E-6
Gpr101_p1	0.9667	0.0174	0.5901	8.92%	<0.01%	7.42%	1.5%	5.35%	85.74%	0.9713
Gpr101_p2	6.77E-15	1.54E-17	0.0011	44.36%	10.88%	29.13%	4.33%	4.12%	51.55%	1.5E-14
Gpr101_p3	7.7E-87	5.84E-14	7.83E-7	60.12%	41.57%	14.14%	4.37%	2.47%	37.45%	7.96E-86
Gpr101_p4	1.53E-13	0.5174	0.151	15.78%	10.74%	2.92%	2.11%	8.78%	75.44%	3.25E-13
Gpr101_p5	1.67E-11	0.0039	0.9563	20.71%	11.83%	8.34%	0.54%	4.37%	74.93%	3.3E-11
Gpr101_p6	0.0016	0.2409	0.9412	6.29%	2.92%	2.69%	0.68%	0%	93.71%	0.0021
Gpr101_p7	2.24E-5	8e-04	0.7383	10.99%	4.18%	5.81%	1.01%	0%	89.01%	3.33E-5
Hnmt	3.54E-65	0.1518	0.8069	45.33%	41.87%	2.92%	0.54%	4.11%	50.57%	2.68E-64
Syt1	0.0069	0.0113	0.1897	9.17%	1.89%	4.7%	2.58%	<0.01%	90.84%	0.0087
Syt1_p1	2.75E-31	1.17E-6	0.4486	32.79%	23.39%	8.21%	1.18%	1.01%	66.21%	9.05E-31
Syt1_p3	0.0075	0.2326	0.7969	5.47%	1.93%	2.5%	1.04%	0%	94.53%	0.0093
Pspc1	8.96E-47	9.13E-5	0.8123	38.89%	31.91%	6.39%	0.57%	1.85%	59.28%	4.27E-46
Pspc1_p1	0.0479	1.66E-10	0.5146	19.59%	1.1%	16.75%	1.74%	0%	80.41%	0.0556
Rnf11	9.97E-130	0.2523	0.0065	60.62%	57.25%	1.45%	1.91%	2.04%	37.34%	1.57E-128
Rnf11_p1	7.8E-18	0.4534	0.0955	22.56%	17.97%	1.64%	2.95%	<0.01%	77.44%	1.92E-17

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Rnf11_p2	2.16E-8	0.002	0.7659	16.84%	9.08%	6.56%	1.19%	0%	83.17%	3.69E-8
Plxnc1	0.1293	2.34E-5	0.7028	20.57%	0.44%	19.24%	0.89%	12.35%	67.08%	0.1439
Plxnc1_p1	4.66E-13	0.3894	0.6023	15.67%	12.58%	1.78%	1.31%	0%	84.33%	9.69E-13
Plxnc1_p2	2.8E-38	0.0271	0.0164	43.55%	35.27%	4.66%	3.61%	2.71%	53.75%	1.14E-37
Plxnc1_p3	3.72E-36	3.46E-7	2.96E-7	38.9%	25.15%	6.84%	6.9%	0%	61.11%	1.36E-35
Rgmb	6.78E-52	0.5061	0.001	40.24%	34.21%	2.41%	3.61%	7.49%	52.27%	3.55E-51
Rgmb_p1	1.07E-49	6e-04	0.0012	45.65%	35.01%	6.83%	3.79%	3.49%	50.87%	5.55E-49
Cadps	2.26E-18	0.4513	0.1141	21.27%	17.15%	1.52%	2.6%	<0.01%	78.73%	5.71E-18
Cadps_p1	1.08E-40	1.87E-6	5.04E-12	50.08%	29.62%	9.27%	11.17%	2.27%	47.67%	4.61E-40
Cadps_p2	0.9781	0.0501	0.7899	4.87%	<0.01%	3.81%	1.06%	0%	95.13%	0.9781
Ube2b	2.47E-60	0.001	3.09E-6	46.33%	37.62%	3.39%	5.31%	0%	53.68%	1.66E-59
Ube2b_p1	4.53E-53	2e-04	0.0067	46.95%	34.61%	9.44%	2.87%	6.08%	46.99%	2.43E-52
Shroom2_p1	3.13E-22	0.012	0.7034	23.87%	18.04%	4.94%	0.89%	2.64%	73.49%	8.56E-22
Nptn.1	4.02E-8	0.001	0.4376	13.8%	6.78%	5.46%	1.56%	0%	86.2%	6.68E-8
B3galnt1	0.0014	0.0397	0.9907	6.93%	2.7%	3.9%	0.32%	0%	93.07%	0.0019
B3galnt1_p1	7.2E-33	0.768	0.0961	28.77%	25.51%	1.08%	2.17%	2.71%	68.53%	2.46E-32
B3galnt1_p2	2.9E-18	6.95E-5	0.8009	23.32%	14.85%	7.71%	0.75%	1.77%	74.92%	7.23E-18
Calm2	1.06E-159	0.0431	0.166	64.48%	59.36%	4.26%	0.85%	6.78%	28.75%	2.25E-158
Dnajc6	1.5E-81	8.75E-39	2e-04	69.57%	27.38%	40.01%	2.13%	4.13%	26.36%	1.42E-80
Plcb1.1	3.48E-31	0.0169	0.0189	30.67%	23.96%	3.73%	2.98%	1.3%	68.03%	1.13E-30
Plcb1.1_p1	9.02E-17	3.62E-5	0.8722	24.69%	16.33%	7.62%	0.74%	0%	75.32%	2.14E-16
Plcb1.1_p2	4.55E-277	3.04E-7	8e-04	80.69%	76.57%	2.61%	1.5%	<0.01%	19.32%	3.86E-275
Plcb1.1_p3	2.51E-20	9.77E-10	0.0516	28.84%	15.85%	10.39%	2.6%	0%	71.17%	6.52E-20
Plcb1.2	6.18E-131	1.22E-12	0.3387	66.38%	44.72%	21.01%	0.6%	7.13%	26.55%	1.01E-129

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Plcb1.2_p1	8.43E-25	0.2079	0.1815	24.66%	20.49%	2.21%	1.96%	1.13%	74.21%	2.46E-24
Plcb1.2_p2	7e-04	0.1351	0.6028	8.22%	2.91%	3.93%	1.38%	3.27%	88.52%	9e-04
Plcb1.2_p3	6.03E-69	7.2E-5	0.1668	57.78%	50.98%	5.06%	1.72%	0%	42.24%	4.83E-68
Plcb1.2_p4	2.14E-11	2.04E-12	0.0181	29.66%	8.74%	17.62%	3.29%	1.89%	68.46%	4.19E-11
Plcb1.2_p5	0.0768	0.0091	0.7275	7.62%	0.91%	5.43%	1.29%	0%	92.38%	0.0871
Srp9	1.53E-59	0.0081	6e-04	44.69%	38.24%	2.75%	3.7%	0%	55.31%	9.98E-59
Srp9_p1	3.61E-8	0.185	0.0366	15.04%	7.89%	3.26%	3.88%	2.06%	82.91%	6.08E-8
Pja2	5.9E-38	0.4074	0.9183	33.36%	31.49%	1.37%	0.5%	0%	66.64%	2.34E-37
Pja2_p1	2.89E-13	1.17E-6	0.8086	20.69%	10.95%	8.96%	0.77%	0.56%	78.76%	6.07E-13
Pja2_p2	1.69E-9	0.1975	0.1769	13.84%	8.91%	2.42%	2.51%	0%	86.17%	3.03E-9
Pja2_p3	8.27E-7	0.2718	0.8618	11.22%	7.51%	2.7%	1%	<0.01%	88.78%	1.3E-6
Isca1	0.4501	0.3835	0.9469	3.82%	0.17%	3.01%	0.64%	3.63%	92.55%	0.4748
Isca1_p1	0.2462	0.387	0.9557	2.99%	0.37%	2.05%	0.57%	<0.01%	97.01%	0.2656
Isca1_p2	0.0489	0.4219	0.8781	4.2%	1.04%	2.34%	0.82%	2.03%	93.77%	0.0567
Azin1	9.5E-10	3e-04	0.7717	19.35%	10.51%	7.69%	1.14%	0%	80.65%	1.74E-9
Lrp12	1e-04	0.0185	0.9939	9.31%	4.23%	4.79%	0.3%	0%	90.69%	2e-04
Lrp12_p1	6.44E-34	2e-04	0.0993	37.12%	29.05%	5.69%	2.37%	0%	62.89%	2.29E-33
Lrp12_p2	0.0012	0.0198	0.1116	9.95%	2.68%	4.27%	3%	<0.01%	90.05%	0.0017
Nxph1	0.593	0.401	0.3967	4.46%	0.09%	2.18%	2.19%	0%	95.54%	0.6178
Nxph1_p1	0.0231	0.0069	0.4785	9.79%	1.62%	6.11%	2.05%	<0.01%	90.21%	0.0275
Slc25a1_p1	0.4628	7.44E-42	0.6957	45.58%	0.08%	44.8%	0.7%	2.03%	52.39%	0.4869
Slc25a1_p2	7.39E-44	1.88E-33	1.79E-7	62.35%	16.32%	42.24%	3.75%	5.2%	32.48%	3.37E-43
Slc9a6	4.22E-5	0.6986	0.8326	6.64%	4.46%	1.25%	0.94%	0%	93.36%	6.11E-5
Slc9a6_p1	1.57E-75	2.86E-6	1.55E-7	54.53%	41.66%	7.36%	5.5%	2.23%	43.25%	1.33E-74

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Cltc	1.13E-128	1.25E-61	9.65E-28	72.88%	41.23%	21.46%	10.14%	<0.01%	27.17%	1.71E-127
Vav2	0.0388	4.25E-7	0.1202	13.19%	0.97%	9.62%	2.59%	0%	86.81%	0.0458
Vav2_p1	0.0866	2e-04	0.2078	11.46%	0.81%	7.96%	2.68%	0%	88.54%	0.098
Kif2a	3.26E-27	1e-04	0.4757	28.49%	21.8%	5.47%	1.22%	0%	71.51%	9.93E-27
Susd2_p1	0.203	0.0723	0.5508	5.08%	0.4%	3.22%	1.46%	0%	94.92%	0.2207
Susd2_p2	0.0039	0.0025	0.9948	8.21%	2.18%	5.77%	0.26%	0%	91.79%	0.005
Grm7	1.61E-6	2e-04	0.7029	14.2%	5.58%	7.49%	1.13%	0.74%	85.06%	2.5E-6
Grm7_p1	0.383	4e-04	0.7955	11.71%	0.16%	10.72%	0.83%	6.03%	82.26%	0.408
Ubfd1	3.28E-5	0.1578	0.7465	7.74%	4.15%	2.55%	1.03%	0%	92.26%	4.77E-5
Ubfd1_p1	7e-04	0.9952	0.563	4.55%	2.86%	0.24%	1.45%	0%	95.45%	0.001
Ubfd1_p2	0.8695	0.0177	4.44E-6	12.38%	0.01%	3.88%	8.49%	0%	87.62%	0.8863
Ubfd1_p4	0.0028	0.1452	0.1208	7.54%	2.16%	2.62%	2.76%	0%	92.46%	0.0036
Ubfd1_p5	6e-04	0.0015	0.4795	9.8%	2.76%	5.5%	1.54%	0%	90.2%	8e-04
Mpped1	7.66E-15	0.1725	0.6959	16.76%	12.67%	3.09%	0.99%	2.9%	80.35%	1.68E-14
Mpped1_p1	1.68E-5	0.4921	0.9816	8.44%	5.91%	2.05%	0.48%	0%	91.56%	2.51E-5
Mpped1_p2	2.48E-5	0.0702	0.0717	11.24%	3.95%	4.4%	2.89%	3.63%	85.13%	3.65E-5
App	4.44E-14	3.54E-5	0.563	28.25%	15.5%	11.16%	1.58%	2.34%	69.42%	9.66E-14
Syng3	0.0479	1.13E-8	0.2425	14.33%	0.87%	11.41%	2.04%	0.08%	85.59%	0.0556
Cdk17	3.96E-12	0.0367	0.4355	16.65%	11.42%	3.59%	1.64%	0.09%	83.26%	8.03E-12
Cdk17_p1	0.0136	8.15E-6	0.3393	16.1%	1.75%	12.08%	2.27%	1.64%	82.27%	0.0166
Cdk17_p2	6.93E-30	1.04E-12	0.4034	35.08%	21.85%	11.99%	1.23%	0%	64.93%	2.18E-29
Pcdh17	1.53E-62	0.0468	0.2543	46.23%	42.67%	2.18%	1.37%	0%	53.77%	1.1E-61
Pcdh17_p1	0.7976	0.0084	0.0616	10.63%	0.02%	6.97%	3.64%	3.13%	86.24%	0.8188
Pcdh17_p3	2.32E-85	1.75E-6	0.0057	60.53%	51.3%	6.54%	2.67%	1.06%	38.43%	2.34E-84

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Pcdh17_p4	2.24E-37	0.7396	0.5829	33.04%	31.14%	0.83%	1.07%	<0.01%	66.96%	8.55E-37
Pcdh17_p5	1.44E-5	0.4801	0.5572	7.53%	4.54%	1.57%	1.41%	0%	92.47%	2.17E-5
Atp8a1	0.9436	0.0354	0.7259	5.31%	<0.01%	4.11%	1.2%	0.12%	94.57%	0.9549
Atp8a1_p1	0.0143	0.0272	0.8797	8.96%	1.31%	6.97%	0.67%	7.11%	83.94%	0.0173
Atp8a1_p2	0.0416	0.6018	0.9938	3.24%	1.26%	1.66%	0.32%	0%	96.76%	0.0489
Atp8a1_p3	3.06E-10	5.1E-9	0.891	23.5%	7.57%	15.36%	0.56%	3.27%	73.23%	5.69E-10
Atp8a1_p4	1.7E-21	8.31E-48	0.6356	62.28%	7.87%	53.94%	0.45%	4.45%	33.29%	4.52E-21
Atp8a1_p5	8.4E-24	1.95E-6	0.7958	31.11%	21.85%	8.42%	0.83%	0%	68.9%	2.37E-23
Atp8a1_p6	9.27E-63	5.63E-18	1.94E-6	61%	24.62%	32.9%	3.43%	8.07%	30.98%	6.9E-62
Socs5	0.1048	2e-04	0.1596	9.79%	0.62%	6.68%	2.48%	<0.01%	90.21%	0.1179
Socs5_p1	4.89E-16	5.17E-44	1.59E-7	61.14%	5.89%	51.24%	3.99%	4.56%	34.32%	1.11E-15
Hmgn1	2.11E-67	2.57E-6	6.75E-9	54.98%	34.68%	14.32%	5.95%	8.24%	36.81%	1.66E-66
Hmgn1_p1	0.3511	0.0136	0.6076	10.86%	0.2%	9.44%	1.23%	9.84%	79.3%	0.376
Pgap1	3.04E-41	4.69E-7	4e-04	50.04%	33.25%	11.84%	4.92%	3.09%	46.9%	1.31E-40
Pgap1_p1	1e-04	0.0136	0.56	12.79%	4.34%	6.74%	1.71%	2.78%	84.43%	2e-04
Pgap1_p3	7.43E-11	3e-04	0.5903	19.11%	10.75%	6.94%	1.41%	0%	80.89%	1.41E-10
Pgap1_p4	0.1552	0.0776	0.4956	6.26%	0.48%	4.27%	1.51%	3.12%	90.62%	0.1704
Pgap1_p5	1.41E-20	0.0093	0.8669	22.05%	17.6%	3.79%	0.65%	0%	77.95%	3.68E-20
Ranbp6	1.98E-8	0.6931	0.2018	12.61%	8.63%	1.3%	2.68%	0%	87.39%	3.39E-8
Ranbp6_p2	2.49E-30	0.005	0.6049	30.87%	25.8%	3.99%	1.07%	0%	69.14%	8.01E-30
Ahi1	2.19E-5	0.2968	0.2266	8.55%	4.3%	2.01%	2.24%	0%	91.45%	3.25E-5
Dnm3	1.51E-48	2.73E-9	0.0938	42.68%	31.65%	9.21%	1.8%	0.77%	56.57%	7.55E-48
Dnm3_p1	1.2E-60	8.97E-26	0.0739	58.68%	29.29%	27.94%	1.4%	3.19%	38.17%	8.22E-60
Dnm3_p2	2.07E-24	0.0799	0.6813	24.75%	17.47%	6.47%	0.81%	10.9%	64.36%	6E-24

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Dnm3.1_p3	5.38E-5	0.0284	0.6581	9.53%	4.2%	4.04%	1.29%	<0.01%	90.47%	7.76E-5
Dnm3.1_p5	3.16E-9	0.0198	0.2398	15.7%	8.12%	5.46%	2.12%	3.08%	81.22%	5.6E-9
Slc19a1	2.43E-8	4e-04	0.8781	16.44%	7.13%	8.6%	0.7%	3.12%	80.45%	4.15E-8
Slc19a1_p1	1.06E-10	0.6271	0.8161	11.74%	9.39%	1.52%	0.83%	2.03%	86.23%	1.99E-10
Spsb4	0.0069	0.6346	0.5186	5.53%	2.16%	1.54%	1.83%	<0.01%	94.47%	0.0087
Spsb4_p1	7.29E-17	8.89E-7	0.9357	31.1%	13.43%	17.19%	0.46%	7.29%	61.62%	1.74E-16
Spsb4_p2	0.0458	0.0173	0.5033	16.21%	0.71%	14.38%	1.12%	21.27%	62.52%	0.0535
Spsb4_p3	0.0229	0.1821	0.215	6.09%	1.27%	2.48%	2.34%	0%	93.91%	0.0274
Pum1	0.2316	0.5587	0.2283	4.52%	0.39%	1.59%	2.54%	0%	95.48%	0.2511
Pum1_p1	3.91E-33	0.0698	0.0927	32.52%	27.65%	2.52%	2.35%	<0.01%	67.48%	1.36E-32
Por	0.8231	0.496	0.951	2.8%	0.01%	2.26%	0.53%	3.5%	93.7%	0.843
Nbea	0.7873	7.91E-13	0.0578	18.17%	0.02%	15.24%	2.91%	0%	81.83%	0.8122
Nbea_p1	4.75E-13	0.5399	0.9613	13.6%	11.8%	1.35%	0.45%	0%	86.4%	9.82E-13
Nbea_p2	3.53E-9	0.2983	0.9383	16.98%	12.98%	3.13%	0.87%	<0.01%	83.02%	6.23E-9
Wsb2	0.9515	0.6873	0.6682	3.27%	<0.01%	1.61%	1.66%	0%	96.73%	0.9605
Rab3c	0.1634	0.1011	0.156	6.15%	0.47%	3.09%	2.59%	0.43%	93.42%	0.1785
Rab3c_p1	6.11E-83	0.2047	0.7578	52.07%	49.96%	1.55%	0.56%	0.82%	47.11%	5.89E-82
Rab3c_p2	7.75E-20	0.1101	0.0023	26.85%	17.24%	4.99%	4.61%	6.98%	66.17%	1.99E-19
Rab3c_p3	1.38E-25	8.39E-10	0.3395	33.34%	18.34%	13.66%	1.33%	2.4%	64.28%	4.1E-25
Rab3c_p4	9.57E-18	1.2E-41	0.3791	46.06%	10.18%	34.82%	1.04%	0.89%	53.06%	2.35E-17
Rab3c_p5	0.0026	0.3318	0.7269	5.77%	2.43%	2.15%	1.19%	0%	94.23%	0.0035
Rab3c_p6	6.65E-32	0.0853	0.038	32.05%	26.75%	2.42%	2.87%	0%	67.96%	2.24E-31
Hipk1	2.91E-5	7.19E-7	0.4762	20.72%	3.32%	16.14%	1.25%	6.41%	72.87%	4.23E-5
Hipk1_p2	4.22E-37	0.7322	0.0487	33.96%	30.47%	0.83%	2.66%	<0.01%	66.04%	1.57E-36

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Hipk1_p3	0.0674	0.0054	0.7579	9.47%	0.74%	7.8%	0.93%	5.3%	85.23%	0.0772
Hipk1_p4	7.47E-39	0.0642	0.4912	33.09%	29.37%	2.61%	1.11%	0.73%	66.18%	3.11E-38
Prkacb_p1	3.57E-61	1.14E-16	5.06E-5	53.18%	31.28%	18.26%	3.61%	2.8%	44.05%	2.48E-60
Prkacb_p2	3.03E-37	9.35E-6	0.6625	36.64%	29.35%	6.39%	0.9%	<0.01%	63.37%	1.15E-36
Prkacb_p3	3.12E-163	0.0026	0.2398	66.59%	63.51%	2.29%	0.79%	0.59%	32.82%	6.97E-162
Dnaja1	0.007	0.021	0.167	10.64%	2.27%	5.14%	3.24%	0%	89.36%	0.0087
Dnaja1_p1	2.59E-8	1.61E-8	1.16E-6	27.75%	6.17%	13.58%	8%	2.35%	69.9%	4.39E-8
Dnaja1_p2	1.44E-11	4e-04	0.6387	18.09%	10.64%	6.23%	1.21%	0%	81.92%	2.87E-11
Atmin	4.24E-7	0.5125	0.7231	9.62%	6.37%	2.13%	1.12%	2.93%	87.45%	6.75E-7
Atmin_p1	0.001	5.8E-6	0.9035	12.48%	2.68%	9.1%	0.7%	0%	87.52%	0.0014
Atmin_p2	2.58E-30	6.99E-14	0.0389	40.85%	21.33%	17.09%	2.41%	1.93%	57.24%	8.24E-30
Htr7	0.9255	3e-04	0.0459	10.64%	<0.01%	6.99%	3.64%	<0.01%	89.36%	0.9388
Htr7_p1	0.0028	0.7285	0.8876	4.87%	2.66%	1.32%	0.89%	<0.01%	95.13%	0.0036
Uhrf2	0.0013	0.3105	0.3158	7.11%	2.75%	2.19%	2.17%	<0.01%	92.89%	0.0017
Uhrf2_p1	0.0942	4.26E-5	0.9181	8.88%	0.67%	7.59%	0.62%	0%	91.12%	0.1063
Rnf14	0.005	7e-04	0.5141	12.07%	2.41%	7.75%	1.91%	0%	87.93%	0.0064
Zic4	0.9575	0.0022	0.9797	6.44%	<0.01%	6.06%	0.38%	0.9%	92.66%	0.9643
Zic4_p1	1.07E-41	0.0026	0.8205	39.54%	34.68%	4.16%	0.69%	<0.01%	60.47%	4.67E-41
Zic4_p2	6.48E-5	0.3764	0.0614	8.92%	3.76%	1.99%	3.17%	0.93%	90.15%	9.24E-5
Crocc	0.5337	0.7102	0.6495	2.77%	0.1%	1.41%	1.27%	1.87%	95.36%	0.5574
Nrep	4.51E-15	2.98E-10	0.0187	27.31%	11.52%	12.63%	3.15%	0.91%	71.79%	1.01E-14
Cul4b	0.7493	0.1914	0.8638	3.35%	0.03%	2.51%	0.81%	<0.01%	96.65%	0.7749
Cul4b_p1	1.23E-236	0.7077	0.3866	75.06%	74.11%	0.44%	0.51%	0.83%	24.11%	6.52E-235
Cul4b_p2	0.1371	3.13E-7	0.8434	12.2%	0.55%	10.79%	0.86%	0%	87.8%	0.1522

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Cul4b_p3	1.96E-105	3.07E-6	0.0024	57.59%	42.35%	13.24%	1.97%	8.3%	34.15%	2.37E-104
Tmem30a_p1	0.0504	0.0414	0.2058	7.42%	1.01%	3.85%	2.56%	0%	92.58%	0.0583
Tmem30a_p2	4.27E-30	0.2566	9e-04	31.78%	25.25%	1.74%	4.79%	<0.01%	68.22%	1.35E-29
Tmem30a_p3	1.69E-23	7.9E-9	0.8568	32.62%	21.07%	10.84%	0.69%	0%	67.4%	4.71E-23
Tmem30a_p4	5.83E-159	1.51E-9	2e-04	67.76%	60.74%	4.62%	2.39%	<0.01%	32.25%	1.18E-157
Fem1b	0.0015	0.0089	0.0342	10.41%	2.36%	4.52%	3.53%	0.23%	89.36%	0.002
Fem1b_p1	2.5E-10	0.0305	0.7899	15.7%	10.58%	4.09%	1.03%	0%	84.3%	4.67E-10
Fem1b_p2	1.22E-23	9e-04	0.5505	27.16%	20.84%	5.09%	1.23%	0%	72.85%	3.43E-23
Fem1b_p3	0.7916	0.8798	0.9138	1.78%	0.02%	0.94%	0.82%	<0.01%	98.22%	0.8147
Fem1b_p5	2.18E-31	0.0849	0.6544	30.41%	26.92%	2.48%	1%	0.02%	69.58%	7.23E-31
Zdhhc13	1.41E-12	0.001	0.6949	18.39%	11.66%	5.63%	1.1%	<0.01%	81.61%	2.9E-12
Peg10	1.49E-8	0.0166	0.2657	13.15%	7.27%	3.88%	2%	0%	86.85%	2.59E-8
Peg10_p2	0.0038	0.002	0.9816	11.32%	2.53%	8.33%	0.46%	2.17%	86.51%	0.0049
Peg10_p3	0.1228	3e-04	0.1402	10.45%	0.61%	7.04%	2.8%	0%	89.55%	0.1374
Peg10_p4	0.6056	0.1264	0.8284	8.05%	0.05%	7.27%	0.72%	14.06%	77.89%	0.6293
Peg10_p5	2E-323	1.83E-12	2.12E-11	81.5%	71.86%	6.5%	3.12%	1.43%	17.09%	2.1E-321
Peg10_p6	0.0023	6.06E-7	0.5259	18.29%	2.98%	13.35%	1.96%	0%	81.71%	0.003
Peg10.1	0.0182	0.0128	0.2832	10.03%	1.75%	5.59%	2.69%	0%	89.97%	0.022
Peg10.1_p1	0.2897	0.0718	0.996	4.11%	0.31%	3.55%	0.25%	0%	95.89%	0.3118
Peg10.1_p2	3.83E-16	0.4709	0.0786	19.33%	13.38%	3.37%	2.57%	9.85%	70.82%	8.78E-16
Nsf	3e-04	4.85E-39	0.8142	53%	1.56%	50.98%	0.45%	4.38%	42.63%	5e-04
Hap1	2.15E-178	7.87E-31	7.53E-34	79.25%	43.5%	26.44%	9.24%	3.63%	17.19%	6.07E-177
Tmem66	0.0044	0.0012	0.9312	8.94%	2.11%	6.2%	0.63%	<0.01%	91.06%	0.0056
Serpini1	3.59E-13	8.21E-7	0.0379	25.41%	12.36%	9.58%	3.47%	<0.01%	74.6%	7.5E-13

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Serpini1_p1	0.2433	0.1656	0.6025	5.3%	0.36%	3.5%	1.44%	2.33%	92.37%	0.2632
Serpini1_p2	1.12E-16	0.0111	0.9091	22.27%	11.78%	10.01%	0.47%	12.08%	65.66%	2.64E-16
Myh10	0.3453	2.71E-13	0.6543	21.58%	0.19%	20.29%	1.1%	1.83%	76.59%	0.3707
Myh10_p1	2e-04	0.0661	0.6705	8.27%	3.25%	3.86%	1.15%	1.8%	89.93%	3e-04
Slc39a11	3.77E-8	0.9076	0.6158	9.11%	7.18%	0.65%	1.27%	0%	90.89%	6.32E-8
Slc39a11_p1	0.1471	6.89E-8	0.9477	15.93%	0.49%	14.93%	0.51%	2.85%	81.22%	0.1624
Slc39a11_p3	4.92E-15	0.0063	0.7417	20.44%	11.63%	7.99%	0.82%	6.89%	72.68%	1.1E-14
Slc39a11_p4	5.83E-5	0.1653	0.3222	10.54%	4.35%	4.01%	2.18%	3.67%	85.79%	8.38E-5
Kcnd2	9.86E-8	0.0079	0.9683	17.92%	8.58%	8.79%	0.55%	5.05%	77.04%	1.61E-7
Kcnd2_p1	1.78E-56	2.69E-16	0.446	53.59%	29.99%	22.75%	0.82%	4.4%	42.04%	1.08E-55
Kcnd2_p2	2.77E-18	0.091	0.6829	23.31%	15.95%	6.35%	1.01%	9.77%	66.93%	6.94E-18
Kcnd2_p3	2.08E-241	1.71E-38	1.57E-38	81.08%	58%	12.81%	10.23%	0.42%	18.54%	1.26E-239
Lpin2	0.0043	0.0221	0.3901	10.78%	2.47%	6.07%	2.24%	2.17%	87.06%	0.0056
Lpin2_p1	2.3E-16	0.3159	0.921	18.47%	13.76%	4.19%	0.53%	9.79%	71.74%	5.36E-16
Lpin2_p2	1.5E-37	1.21E-17	0.007	44.6%	25.74%	15.79%	3.04%	0.31%	55.12%	5.8E-37
Mecp2	7.88E-21	0.3226	0.0501	22.28%	17.78%	1.65%	2.85%	0%	77.72%	2.08E-20
Mecp2_p1	9.71E-33	0.0019	0.5073	32.48%	21.8%	9.7%	0.96%	8.71%	58.82%	3.29E-32
Mecp2_p3	0.421	0.0026	0.8738	6.29%	0.16%	5.37%	0.76%	0%	93.71%	0.4463
Mecp2_p4	2.32E-5	0.8829	0.8871	6.63%	4.25%	1.67%	0.71%	10.07%	83.3%	3.43E-5
Mecp2_p6	1.85E-39	1.05E-9	0.3026	42.57%	26.51%	14.76%	1.28%	3.55%	53.9%	7.85E-39
Mecp2_p7	1.48E-17	0.0061	0.8508	20.02%	15.19%	4.13%	0.7%	0%	79.99%	3.59E-17
Mecp2_p8	1.13E-5	2.71E-6	0.7896	13.83%	4.34%	8.61%	0.88%	0%	86.18%	1.72E-5
Mecp2_p9	1.88E-9	0.0347	0.9604	13.16%	8.93%	3.73%	0.49%	0%	86.84%	3.36E-9
Ran	1.29E-11	0.61	0.1619	14.95%	10.89%	1.56%	2.5%	1.61%	83.44%	2.58E-11

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Ran_p1	1.42E-29	0.5233	0.5218	30.48%	27.8%	1.34%	1.34%	<0.01%	69.53%	4.38E-29
Cox14	2.76E-5	0.4229	0.9857	6.36%	4.3%	1.73%	0.34%	<0.01%	93.64%	4.04E-5
Bicap	2.54E-5	3.72E-8	0.8983	17.37%	3.73%	13.04%	0.6%	1.96%	80.67%	3.73E-5
Bicap_p1	0.0231	0.0053	0.1747	16.59%	0.9%	13.89%	1.79%	16.46%	66.95%	0.0275
Wdr82	1.5E-48	8e-04	0.4212	39.13%	34.07%	3.93%	1.12%	0%	60.88%	7.55E-48
Wdr82_p1	3.11E-14	0.8397	0.5788	15.05%	11.72%	2.18%	1.15%	13.6%	71.35%	6.8E-14
Wdr82_p2	1.77E-10	0.039	0.2872	14.32%	9.11%	3.31%	1.91%	0%	85.68%	3.32E-10
Fam160b1	0.0037	0.1987	0.1962	13.3%	1.26%	10.57%	1.47%	29.56%	57.14%	0.0047
Fam160b1_p1	0.0687	0.3232	0.5956	5.58%	1.09%	2.67%	1.82%	0%	94.42%	0.0785
Fam160b1_p2	0.0014	0.5748	0.0514	7.85%	2.68%	1.5%	3.67%	<0.01%	92.15%	0.0019
Fam160b1_p3	1.55E-11	1.74E-11	0.0549	30.55%	7.68%	20.54%	2.33%	4.74%	64.71%	3.07E-11
Fam160b1_p5	2.06E-6	2e-04	0.1157	15.2%	5.45%	6.96%	2.79%	0%	84.8%	3.18E-6
Fam160b1_p6	0.4286	0.1062	0.4071	6.24%	0.16%	4.22%	1.86%	3.14%	90.62%	0.4532
Fam160b1_p7	3.87E-12	0.0229	0.9601	16.14%	9.96%	5.76%	0.41%	4.72%	79.14%	7.9E-12
Fam160b1_p8	1.15E-11	9.18E-15	0.9528	26.88%	9.59%	16.84%	0.44%	0%	73.13%	2.3E-11
Fam160b1_p9	1.8E-52	0.1284	0.153	39.9%	36.46%	1.76%	1.67%	0%	60.1%	9.52E-52
Arf1	0.0381	0.058	0.892	5.17%	1.07%	3.38%	0.72%	0%	94.83%	0.0451
Arf1_p2	2.31E-17	0.2817	1.4E-6	25.46%	15.05%	2.08%	8.32%	1.02%	73.52%	5.57E-17
Psd3_p1	0.0012	3e-04	0.6897	11.74%	2.88%	7.54%	1.32%	0%	88.26%	0.0017
Psd3_p3	0.0565	0.0052	0.7025	7.51%	0.96%	5.32%	1.23%	<0.01%	92.49%	0.0651
Psd3_p4	8.77E-6	2.9E-9	0.5007	23.17%	4.1%	17.75%	1.32%	3.95%	72.88%	1.34E-5
Psd3_p5	1e-04	2e-04	0.4158	15.5%	3.74%	9.92%	1.82%	2.87%	81.64%	2e-04
Psd3_p7	0.0141	0.2733	0.2161	6.59%	1.59%	2.48%	2.52%	0.62%	92.79%	0.0172
Psd3_p8	1.33E-16	0.4491	0.5355	17.5%	14.74%	1.47%	1.3%	0%	82.5%	3.11E-16

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Psd3_p9	1.27E-6	0.0052	0.9908	11.33%	5.93%	5.09%	0.3%	0%	88.67%	1.98E-6
Psd3_p10	5.2E-13	0.0375	0.8371	19.72%	14.58%	4.16%	0.97%	<0.01%	80.29%	1.07E-12
Pgs1.1	1.04E-9	0.0015	0.3734	18.17%	9.39%	6.87%	1.9%	1.39%	80.45%	1.89E-9
2510039O18Rik	5.61E-9	0.0162	0.0352	15.87%	8.14%	4.12%	3.61%	0%	84.13%	9.86E-9
2510039O18Rik_p1	1.43E-5	0.2626	0.3288	9.24%	4.87%	2.29%	2.08%	0%	90.76%	2.16E-5
2510039O18Rik_p2	0.9048	0.0072	0.5442	6.2%	<0.01%	4.73%	1.46%	0%	93.8%	0.92
Lingo3	1e-04	0.004	0.6001	12.6%	3.47%	7.83%	1.3%	4.45%	82.95%	2e-04
Zbtb5_p1	1.5E-75	1.15E-71	6.43E-11	66.63%	29.33%	31.9%	5.36%	0.14%	33.27%	1.3E-74
Zbtb5_p2	1.73E-27	2.45E-8	5.15E-7	40.38%	18.28%	15.58%	6.5%	5.24%	54.4%	5.32E-27
Zbtb5_p3	0.3608	6.85E-5	0.4975	9.13%	0.2%	7.42%	1.51%	0.14%	90.74%	0.3854
Mapk9	1.54E-45	2.29E-9	9.47E-8	46.12%	30.79%	8.29%	7.03%	<0.01%	53.89%	7.16E-45
Mapk9_p1	2.02E-38	1.51E-6	0.2334	40.24%	26.05%	12.73%	1.44%	5.32%	54.47%	8.31E-38
Mapk9_p2	0.0076	0.0185	0.445	8.07%	1.86%	4.41%	1.79%	<0.01%	91.93%	0.0094
Mapk9.1	4.73E-128	5.49E-29	2.6E-20	72.77%	33.12%	33.44%	6.16%	5.32%	21.96%	6.92E-127
Mapk9.1_p1	3.88E-37	3.25E-6	0.1034	37.68%	28.73%	6.83%	2.11%	0.11%	62.22%	1.46E-36
Mapk9.1_p2	4.59E-49	0.0469	0.063	38.57%	32.9%	3.63%	2.03%	3.3%	58.14%	2.34E-48
Mapk9.1_p3	1.32E-7	0.0017	0.4655	17.47%	5.35%	10.84%	1.28%	8.93%	73.6%	2.14E-7
Dcaf7	3.56E-8	5.91E-7	9e-04	21.6%	6.78%	9.32%	5.49%	0%	78.41%	6.01E-8
Dcaf7_p1	0.0712	0.0946	0.416	6.04%	0.87%	3.26%	1.91%	0%	93.96%	0.0809
Dcaf7_p2	1.29E-31	1.93E-10	0.8523	36.26%	24.85%	10.79%	0.61%	0%	63.76%	4.29E-31
Dcaf7_p3	6.06E-5	0.0032	0.7009	10.09%	3.76%	5.24%	1.09%	0.35%	89.56%	8.68E-5
Dcaf7_p4	1.37E-54	1.15E-5	0.0316	44%	33.02%	8.87%	2.09%	3.76%	52.26%	7.73E-54
Vat1l	5.69E-189	0.0018	1.84E-27	72.38%	60.28%	2.15%	9.95%	0.77%	26.86%	2.19E-187
Reep1	1.64E-8	0.117	0.7496	11.07%	7.4%	2.68%	0.99%	0%	88.93%	2.84E-8

PTRE-Seq - Effects on Expression Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Reep1_p1	1.45E-26	7.34E-6	0.3629	29.58%	20.66%	7.52%	1.39%	0.88%	69.54%	4.35E-26
Reep1_p2	6.46E-126	1.09E-7	7e-04	62.38%	52.78%	7.25%	2.32%	2.14%	35.51%	8.83E-125
Reep1_p3	7.03E-55	0.5456	0.0396	42.54%	39%	1.18%	2.36%	1.23%	56.23%	4.03E-54
Reep1_p4	4.39E-8	1.66E-15	0.2027	26.16%	5.7%	18.6%	1.86%	0.95%	72.9%	7.24E-8
Myo5a	0.5135	9e-04	0.6632	9.46%	0.13%	7.75%	1.57%	<0.01%	90.54%	0.5376
Myo5a_p1	0.0115	4e-04	0.0597	11.69%	1.61%	6.67%	3.41%	0%	88.31%	0.0141
Myo5a_p2	9.45E-9	0.6505	0.8371	10.22%	6.89%	2.6%	0.73%	9.66%	80.13%	1.65E-8
Myo5a_p3	1.02E-97	3.15E-15	0.0054	59.47%	44.66%	12.74%	2.03%	1.66%	38.91%	1.11E-96
Myo5a_p4	7.04E-7	0.2699	0.1955	10.15%	5.77%	2.06%	2.32%	<0.01%	89.85%	1.11E-6
Myo5a_p5	2.42E-24	0.4207	0.9684	21.4%	14.9%	6.23%	0.26%	23.53%	55.07%	6.98E-24
Myo5a_p6	1.44E-13	0.7722	0.5453	15.56%	13.15%	0.98%	1.43%	<0.01%	84.44%	3.07E-13
Myo5a_p7	1.97E-5	0.0715	0.5023	8.93%	4.33%	3.1%	1.5%	0%	91.07%	2.94E-5
Myo5a_p8	9e-04	0.0902	0.0306	9.97%	2.85%	3.16%	3.96%	0%	90.03%	0.0012
Myo5a_p9	2.38E-22	1.39E-5	0.6711	28.44%	18.83%	8.62%	0.98%	1.62%	69.94%	6.56E-22
8430427H17Rik	0.002	0.1281	0.1013	7.89%	2.3%	2.71%	2.88%	0%	92.11%	0.0026
8430427H17Rik_p2	5.22E-16	0.9948	0.788	13.71%	12.13%	0.86%	0.72%	21.5%	64.78%	1.18E-15
8430427H17Rik_p3	5e-04	0.0051	0.6375	8.88%	2.84%	4.81%	1.23%	0%	91.12%	8e-04
8430427H17Rik_p5	0.1428	0.1049	0.6061	6.99%	0.48%	5.28%	1.22%	7.04%	85.98%	0.1581
8430427H17Rik_p6	0.0421	0.5847	0.451	4.5%	1.12%	1.53%	1.85%	0%	95.5%	0.0493

Table S7: PTRE-Seq - Effects on Translation Efficiency

Terms shown: Main effect of sequence (reference or mutant), Main effect of condition (CELF overexpression or CTL), Interaction between sequence and condition.

FDR is shown for sequence x condition interactions

p

% variance explained

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Fgf13	0.1654	0.1631	0.6852	7.8%	0.44%	6.27%	1.1%	11.77%	80.43%	0.9457
Magee1	0.0024	0.4062	0.8532	6.6%	2.25%	3.54%	0.81%	7.85%	85.55%	0.9907
Plcx3	0.3675	0.4976	0.3561	3.75%	0.2%	1.6%	1.94%	<0.01%	96.25%	0.8712
Ddc	3.01E-5	0.025	0.6919	9.06%	4.13%	3.8%	1.12%	0%	90.94%	0.9457
Mkx3	0.011	0.3065	0.0847	7.21%	1.51%	2.77%	2.93%	3.18%	89.61%	0.6101
Mkx3_p1	0.0554	0.9581	0.8372	2.47%	0.9%	0.71%	0.86%	3.37%	94.15%	0.9888
Carkd	0.6736	0.0395	0.3666	6.6%	0.05%	4.32%	2.23%	<0.01%	93.4%	0.8712
Nap13	0.9285	0.3001	0.3714	5.83%	<0.01%	3.71%	2.11%	5.23%	88.95%	0.8712
Pgr15l	0.0058	0.9929	0.2946	4.23%	1.77%	0.48%	1.97%	6.4%	89.37%	0.8427
Pgr15l_p1	3.5E-7	0.1308	0.4861	10.77%	5.87%	3.43%	1.46%	2.56%	86.67%	0.8917
Pgr15l_p2	0.4423	0.4867	0.4235	3.86%	0.16%	1.77%	1.93%	0%	96.14%	0.8712
Pgr15l_p3	0.2851	0.7304	0.6429	3.2%	0.31%	1.51%	1.38%	2.32%	94.47%	0.9457
Slc39a6	0.4539	0.7327	0.046	6.2%	0.13%	2.76%	3.31%	12.6%	81.2%	0.5072
Slc39a6_p1	0.6159	0.8096	0.1754	4.43%	0.06%	1.79%	2.58%	7.24%	88.33%	0.7619
Slc39a6_p2	0.5351	0.3386	0.432	3.84%	0.1%	1.99%	1.75%	<0.01%	96.16%	0.8719
Slc39a6_p3	4.58E-6	0.6608	0.1528	9.8%	4.97%	2.3%	2.53%	7.15%	83.06%	0.7361
Trim32	0.3037	0.2695	0.834	5.49%	0.29%	4.23%	0.97%	6.58%	87.93%	0.9877
Trim32_p1	0.2457	0.1575	0.0246	8.17%	0.34%	3.78%	4.05%	3.35%	88.48%	0.4019
Gpr149	0.7417	0.3131	0.6917	3.55%	0.03%	2.34%	1.18%	1.16%	95.29%	0.9457
Gpr149_p1	0.9955	0.4117	0.0242	8.52%	<0.01%	3.82%	4.7%	7.68%	83.8%	0.4019
Tceal1	0.1215	0.0566	0.0755	9.5%	0.75%	4.74%	4.01%	1.1%	89.41%	0.5766
Gprasp2	0.6685	0.7363	0.8271	2.31%	0.05%	1.27%	0.99%	0.41%	97.28%	0.9877
Gad1	0.6795	0.1318	0.7071	5.54%	0.05%	4.16%	1.33%	2.72%	91.75%	0.9457

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Gad1_p1	0.1012	0.2444	0.646	5.04%	0.8%	2.71%	1.52%	0%	94.96%	0.9457
Zcchc12	0.1121	0.0227	0.5946	6.48%	0.67%	4.34%	1.48%	0%	93.52%	0.9445
Zcchc12_p1	0.5519	0.8838	0.9252	1.98%	0.11%	1.12%	0.76%	2.27%	95.75%	0.9962
Klhdc2	0.8542	0.4877	0.4288	3.4%	0.01%	1.63%	1.77%	<0.01%	96.6%	0.8712
Lrrc47	0.0132	0.0406	0.0048	13.52%	1.53%	6.92%	5.07%	7.11%	79.37%	0.1852
Lrrc47_p1	1.09E-6	0.0742	0.4098	14.96%	6.4%	6.63%	1.93%	7.78%	77.26%	0.8712
Lrrc47_p2	0.0015	0.5216	0.2111	6.92%	2.65%	1.74%	2.53%	0.59%	92.48%	0.7651
Alg2	2.51E-7	0.0545	6e-04	17%	5.05%	7.07%	4.87%	10.28%	72.73%	0.0511
Arxes1_p1	0.8096	0.0062	0.7659	6.38%	0.02%	5.26%	1.1%	0%	93.62%	0.9788
Snca	0.5326	0.6871	0.5878	2.97%	0.11%	1.32%	1.55%	<0.01%	97.03%	0.9441
Syn2	0.5604	0.1124	0.6819	4.21%	0.08%	2.92%	1.21%	0%	95.79%	0.9457
Syn2_p2	0.2671	0.0486	0.7497	5.9%	0.32%	4.45%	1.12%	1.64%	92.46%	0.9721
Syn2.1	0.895	0.2166	0.0805	5.95%	<0.01%	2.55%	3.4%	0%	94.05%	0.5985
Rragb	5.26E-7	0.449	0.5571	10.6%	7.05%	1.91%	1.64%	0%	89.4%	0.9226
Nap1l2	0.0069	0.4243	0.3484	8.33%	1.8%	4.59%	1.93%	12.93%	78.75%	0.8712
Tmtc4	0.4688	0.0837	0.8161	4.73%	0.14%	3.6%	0.99%	0.53%	94.73%	0.9877
Tmtc4_p1	0.1508	0.1533	0.6019	5.3%	0.5%	3.49%	1.31%	2.78%	91.92%	0.9445
Tmtc4_p2	0.4345	0.0362	0.1092	7.64%	0.14%	4.74%	2.75%	2.63%	89.73%	0.6273
Tmtc4_p3	0.9114	0.1083	0.1583	6.55%	<0.01%	3.45%	3.1%	<0.01%	93.45%	0.7457
Zfp385b	0.853	0.3157	0.2242	5.48%	0.01%	2.72%	2.75%	1.26%	93.26%	0.7744
Zfp385b_p1	0.0828	0.0517	0.0532	10.8%	1.05%	4.89%	4.86%	<0.01%	89.2%	0.513
Zfp385b_p2	0.1108	0.0016	0.0553	10.12%	0.65%	5.95%	3.53%	0%	89.88%	0.5211
Zdbf2	0.0605	0.812	0.7355	3.2%	0.97%	1.02%	1.21%	<0.01%	96.8%	0.9609
Zdbf2_p1	0.0609	0.0693	0.4825	7.46%	1.13%	4.23%	2.1%	0%	92.54%	0.8917

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Zdbf2_p2	0.1148	0.6064	0.285	5.02%	0.64%	2.16%	2.21%	4.47%	90.52%	0.8333
Arxes2	0.3582	0.6449	0.6015	3.69%	0.25%	1.82%	1.62%	1.88%	94.43%	0.9445
Arxes2_p1	0.0097	0.8694	0.9604	3.58%	2.02%	0.96%	0.6%	<0.01%	96.42%	0.9962
Penk	0.0799	0.6176	0.053	7.62%	0.58%	4.44%	2.61%	20.49%	71.88%	0.513
Atf1	0.7935	0.158	0.208	5.7%	0.02%	3.11%	2.57%	0.88%	93.42%	0.7651
Snx16	0.6086	0.043	0.6767	5.05%	0.06%	3.78%	1.2%	0.45%	94.5%	0.9457
Dcaf12l1	3e-04	0.1631	6e-04	11.48%	3.05%	2.55%	5.88%	0.42%	88.1%	0.0511
Dcaf12l1_p1	0.0645	0.1423	0.2403	7.57%	1.1%	3.52%	2.95%	<0.01%	92.43%	0.8022
Synpr	0.9653	0.1933	0.3999	6.3%	<0.01%	4.28%	2.03%	4.87%	88.82%	0.8712
Synpr_p1	0.9991	0.0194	0.6035	5.94%	<0.01%	4.48%	1.46%	0%	94.06%	0.9445
Synpr_p2	0.7208	0.7343	0.0502	4.62%	0.03%	1.09%	3.5%	<0.01%	95.38%	0.5072
Cd200	0.0229	0.1476	0.7109	10.07%	1.21%	7.8%	1.07%	15.61%	74.31%	0.9457
Cd200_p1	0.9199	0.3454	0.2453	5.58%	<0.01%	2.58%	2.99%	0%	94.42%	0.8105
Cd200_p2	0.0289	0.2727	0.0155	8.07%	1.25%	2.29%	4.54%	0%	91.93%	0.3399
Ndn	0.0581	0.9288	0.0025	6.88%	0.78%	1.31%	4.79%	9.97%	83.16%	0.1426
AW551984	0.7713	0.426	0.5918	4.81%	0.03%	3%	1.78%	3.43%	91.76%	0.9445
AW551984_p1	0.5957	0.6231	0.8898	2.37%	0.08%	1.47%	0.82%	0%	97.63%	0.9959
Ankra2	3.64E-5	0.0221	0.5226	10.57%	3.84%	5.35%	1.38%	3.25%	86.19%	0.8955
Ankra2_p1	0.6155	0.5104	0.2016	5.37%	0.08%	2.06%	3.22%	0%	94.63%	0.7639
Ankra2.1	0.8554	0.8529	0.1971	4%	0.01%	1.04%	2.95%	0.36%	95.64%	0.7639
Rraga	0.0824	0.1558	0.3618	6.91%	0.98%	3.45%	2.49%	0%	93.09%	0.8712
Irs4	0.5861	0.5536	0.2535	4.94%	0.09%	2.26%	2.59%	3.08%	91.98%	0.8105
Irs4_p1	0.1477	0.4416	0.831	3.16%	0.53%	1.74%	0.89%	0%	96.84%	0.9877
Irs4_p2	0.5425	0.0228	0.3549	6.54%	0.1%	4.37%	2.06%	0.09%	93.37%	0.8712

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Dek	0.5373	0.4453	0.0496	6.26%	0.11%	2.01%	4.14%	<0.01%	93.74%	0.5072
Dek_p1	0.0265	0.384	0.4163	5.68%	1.29%	2.53%	1.86%	2.47%	91.85%	0.8712
Dek_p2	3e-04	0.2351	0.2793	8.34%	2.93%	3.44%	1.96%	4.63%	87.03%	0.8333
Rit2	0.7804	0.0404	0.0306	10.12%	0.02%	6.8%	3.3%	7.97%	81.91%	0.4415
Pcdh18	0.0475	0.4806	0.2247	5.92%	1.01%	2.49%	2.41%	4.02%	90.06%	0.7744
Pcdh18_p1	0.3161	0.3336	0.1854	4.74%	0.25%	1.99%	2.5%	0%	95.26%	0.7619
Hsp90aa1	0.7101	0.6272	0.9994	2.03%	0.05%	1.81%	0.17%	0.05%	97.92%	0.9997
C1d	0.281	0.7413	0.7317	2.52%	0.3%	1.1%	1.12%	0%	97.48%	0.9609
C1d_p1	0.6457	0.8531	0.939	1.65%	0.06%	0.93%	0.65%	0%	98.35%	0.9962
Alkbh7	0.1137	0.832	0.4634	3.83%	0.75%	1.06%	2.01%	0%	96.17%	0.8886
Gabrq.1	0.5011	0.0623	0.5738	5.29%	0.11%	3.79%	1.39%	1.2%	93.51%	0.9424
Gabrq.1_p1	0.1423	0.3405	0.0996	5.45%	0.53%	1.95%	2.97%	<0.01%	94.55%	0.6101
Nap1l5	0.1522	0.7045	0.4879	4.82%	0.65%	2.13%	2.04%	4.63%	90.56%	0.8917
Htr2c_p1	2e-04	0.2936	0.0955	9.65%	3.81%	2.4%	3.44%	0%	90.35%	0.6101
Htr2c_p2	0.4903	0.0801	0.1566	6.34%	0.13%	3.39%	2.83%	0%	93.66%	0.7457
Commd3	0.1521	0.4742	0.1646	5.64%	0.61%	1.95%	3.09%	<0.01%	94.36%	0.7619
Fnbp1l	0.0713	0.3674	0.4533	5.79%	1.07%	2.5%	2.22%	0%	94.21%	0.8796
Fnbp1l_p1	0.3274	0.019	0.0346	8.55%	0.25%	4.37%	3.94%	<0.01%	91.45%	0.4509
Fnbp1l_p2	0.5452	0.5877	0.4193	5.84%	0.07%	4.37%	1.4%	18.6%	75.56%	0.8712
Ccrn4l	0.6856	0.5658	0.9132	2.2%	0.04%	1.48%	0.68%	<0.01%	97.8%	0.9962
Dner	0.022	0.0779	0.0668	8.92%	1.5%	3.65%	3.77%	<0.01%	91.08%	0.5744
Syt4	0.8734	0.1662	0.9867	3.25%	0.01%	2.87%	0.38%	<0.01%	96.75%	0.9997
Syt4_p3	1e-04	0.735	0.909	6%	4%	1.27%	0.72%	0.8%	93.2%	0.9962
Sybu	0.2192	0.0074	0.5445	6.52%	0.37%	4.7%	1.45%	<0.01%	93.48%	0.9126

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Tnks2	0.7385	0.3702	0.1142	4.88%	0.03%	1.99%	2.87%	0.47%	94.64%	0.6372
Tnks2_p1	0.0609	0.6622	0.3116	4.77%	0.82%	2.01%	1.93%	5.37%	89.87%	0.8427
Tnks2_p2	0.1367	0.07	0.1022	7.87%	0.64%	3.78%	3.45%	0%	92.13%	0.6101
Zcchc18	0.845	0.0779	0.818	4.48%	0.01%	3.48%	1%	<0.01%	95.52%	0.9877
Map2k1	0.6406	0.805	0.8268	2.45%	0.06%	1.43%	0.96%	3.56%	94%	0.9877
Map2k1_p1	0.3144	0.4505	0.0484	5.89%	0.27%	1.82%	3.8%	<0.01%	94.11%	0.5072
Glrp_p2	0.2946	0.0203	0.552	6.88%	0.32%	4.84%	1.72%	<0.01%	93.12%	0.9215
Olfm1.1	0.0885	0.142	0.6081	7.2%	0.9%	4.61%	1.69%	3.55%	89.25%	0.9445
Nrsn1	0.6589	0.8472	0.9827	2.52%	0.05%	2.11%	0.36%	12.12%	85.36%	0.9997
Nrsn1_p1	0.259	0.0722	0.7942	7.59%	0.34%	6.21%	1.04%	6.67%	85.74%	0.9875
Epm2aip1	0.4896	0.0034	0.6733	7.13%	0.11%	5.84%	1.17%	1.14%	91.73%	0.9457
Epm2aip1_p1	0.8947	0.594	0.0702	4.64%	<0.01%	1.38%	3.26%	0%	95.36%	0.5766
Epm2aip1_p2	0.9445	0.08	0.2619	6.98%	<0.01%	4.72%	2.26%	3.75%	89.28%	0.8105
Epm2aip1_p3	0.0647	0.8404	0.7892	4.06%	1.28%	1.3%	1.47%	0%	95.94%	0.9842
Epm2aip1_p4	6e-04	0.2632	0.1843	9%	2.53%	4.3%	2.17%	8.63%	82.37%	0.7619
Tle4	0.5659	0.8261	0.6778	2.59%	0.09%	1.18%	1.32%	1.84%	95.57%	0.9457
Tle4_p1	0.3776	0.1224	0.1002	7.66%	0.17%	4.79%	2.7%	6.24%	86.1%	0.6101
Snap25	0.9964	0.718	0.6202	2.92%	<0.01%	1.62%	1.3%	3.65%	93.43%	0.9457
Snap25_p1	0.6404	0.7483	0.6696	3.46%	0.05%	2.29%	1.11%	9.93%	86.61%	0.9457
Cnr1	0.1081	0.4565	0.8324	3.87%	0.78%	2.03%	1.06%	<0.01%	96.13%	0.9877
Cnr1_p4	0.6004	0.2607	0.6471	3.59%	0.07%	2.24%	1.28%	<0.01%	96.41%	0.9457
Cnr1_p5	0.0017	0.0041	0.0928	10.07%	2.31%	4.89%	2.87%	<0.01%	89.93%	0.6101
Cnr1_p6	0.4981	0.3502	0.4539	4.65%	0.14%	2.51%	2%	0.81%	94.53%	0.8796
Plagl1	0.2929	0.307	0.1368	5.07%	0.27%	2.06%	2.74%	0%	94.93%	0.7076

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Plagl1_p1	0.794	0.7469	0.4026	3.54%	0.02%	1.33%	2.19%	0.31%	96.15%	0.8712
Plagl1_p2	3e-04	0.8854	0.9288	4.7%	3.33%	0.75%	0.62%	<0.01%	95.3%	0.9962
Impact	0.2997	0.5767	0.207	4.48%	0.29%	1.55%	2.64%	0%	95.52%	0.7651
Tmem35	0.1239	0.0184	0.6581	7.48%	0.68%	5.36%	1.44%	1.01%	91.51%	0.9457
Atp2c1	0.6931	0.9134	0.7388	2.74%	0.03%	1.76%	0.94%	14.1%	83.16%	0.9609
Fem1c	0.0302	0.7743	0.0448	6.28%	1.23%	1.29%	3.77%	1.79%	91.92%	0.5072
Qdpr	0.9618	0.0474	0.0031	10.08%	<0.01%	4.01%	6.07%	0%	89.92%	0.144
Stx7	0.2115	0.4116	0.6971	3.39%	0.39%	1.81%	1.18%	0%	96.61%	0.9457
Stx7_p1	0.0524	0.4357	0.4676	5.46%	0.86%	3.08%	1.52%	6.88%	87.66%	0.8886
Stx7_p2	0.5915	0.3225	0.8817	4.6%	0.09%	3.54%	0.97%	3.72%	91.67%	0.9959
Enc1.1	0.0538	0.9095	0.8572	2.7%	1.03%	0.75%	0.91%	0%	97.3%	0.9907
Enc1.1_p1	0.2905	0.3021	0.7674	5.92%	0.3%	4.52%	1.1%	8.72%	85.36%	0.9788
Enc1.1_p2	0.4434	0.7487	0.8243	3.19%	0.19%	1.83%	1.17%	3.29%	93.52%	0.9877
Enc1.1_p3	0.5375	0.3051	0.1441	7.34%	0.09%	4.7%	2.54%	10.57%	82.1%	0.7252
Peg3_p1	0.3487	0.2373	0.184	5.95%	0.26%	2.72%	2.97%	0%	94.05%	0.7619
Scoc	0.0481	0.066	0.0145	9.27%	1%	3.8%	4.47%	0.99%	89.74%	0.3399
Rab18_p1	0.0624	0.0943	0.7747	5.32%	0.94%	3.29%	1.09%	<0.01%	94.68%	0.9788
Rab18_p2	0.061	0.8027	0.2347	4.94%	1.05%	1.13%	2.76%	<0.01%	95.06%	0.796
Rab18_p3	0.1163	0.5161	0.0341	7.24%	0.7%	2.26%	4.28%	2.61%	90.15%	0.4509
Rab18_p4	0.5959	0.4064	0.78	4.31%	0.11%	2.71%	1.5%	<0.01%	95.69%	0.9788
Rab18_p5	0.0322	0.7587	0.4356	4.72%	1.18%	1.77%	1.78%	5.35%	89.92%	0.8719
Rab18_p6	0.1799	0.869	0.6796	2.72%	0.5%	0.88%	1.34%	<0.01%	97.28%	0.9457
Rab18_p7	0.4666	0.9461	0.8926	1.94%	0.18%	0.76%	1%	0%	98.06%	0.9959
Rab18_p8	0.2002	0.4079	0.8911	3.56%	0.5%	2.18%	0.89%	0%	96.44%	0.9959

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Ube2e3	0.0509	0.256	0.5111	6.21%	1.25%	2.93%	2.04%	0%	93.79%	0.893
Ttc3	0.1204	0.2113	0.8541	6.45%	0.58%	5.07%	0.8%	9.17%	84.38%	0.9907
Ttc3_p2	0.0988	0.3589	0.4107	5.94%	0.61%	3.73%	1.61%	8.28%	85.77%	0.8712
Lin7c	0.1473	0.4796	0.5086	4.07%	0.57%	1.78%	1.71%	0%	95.93%	0.893
Lin7c_p3	0.7888	0.0162	0.0124	9.93%	0.02%	4.86%	5.06%	0%	90.07%	0.3086
Lin7c_p4	0.0817	0.0204	0.5387	7.43%	0.88%	4.81%	1.74%	<0.01%	92.57%	0.9063
Lin7c_p5	0.0109	0.0137	0.6299	8.43%	1.86%	5.07%	1.51%	0%	91.57%	0.9457
Gpr101_p1	0.0027	0.0512	0.029	11.34%	2.46%	4.64%	4.24%	1.9%	86.76%	0.4396
Gpr101_p2	0.1456	0.1875	0.0944	10.09%	0.58%	6.15%	3.36%	10.77%	79.13%	0.6101
Gpr101_p3	0.0051	3.02E-5	1.16E-5	17.7%	1.84%	7.67%	8.19%	0%	82.3%	0.0049
Gpr101_p4	0.0963	0.649	0.5238	4.39%	0.64%	2.33%	1.42%	7.26%	88.35%	0.8955
Gpr101_p5	0.04	0.9848	0.8443	3.11%	1.37%	0.63%	1.11%	3.75%	93.14%	0.9907
Gpr101_p6	0.0941	0.4993	0.6208	4.44%	0.83%	2.02%	1.58%	0.66%	94.9%	0.9457
Gpr101_p7	0.0254	0.0021	0.2241	8.79%	1.19%	5.36%	2.24%	0%	91.21%	0.7744
Hnmt	0.8362	0.0026	0.7825	6.89%	0.01%	5.82%	1.05%	0%	93.11%	0.9788
Syt1	0.5351	0.8899	0.4543	3.03%	0.1%	1.13%	1.79%	3.82%	93.15%	0.8796
Syt1_p1	0.2291	0.1834	0.6724	5.28%	0.34%	3.79%	1.15%	4.46%	90.26%	0.9457
Syt1_p3	0.358	0.1425	0.3422	5.3%	0.23%	2.94%	2.13%	0%	94.7%	0.8712
Pspc1	0.0158	0.3046	0.2745	5.73%	1.43%	2.18%	2.13%	0.54%	93.73%	0.8313
Pspc1_p1	0.9132	0.0139	0.4295	7.9%	<0.01%	5.65%	2.24%	0%	92.1%	0.8712
Rnf11	0.3719	0.3903	0.685	3.28%	0.2%	1.86%	1.21%	<0.01%	96.72%	0.9457
Rnf11_p1	0.3738	0.1946	0.252	5.81%	0.23%	2.92%	2.66%	<0.01%	94.19%	0.8105
Rnf11_p2	0.1378	0.1066	0.9371	5.4%	0.73%	3.9%	0.78%	0%	94.6%	0.9962
Plxnc1	0.1244	0.2678	0.3548	6.06%	0.6%	3.48%	1.98%	4.48%	89.46%	0.8712

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Plxnc1_p1	0.0322	0.4876	0.4686	4.79%	1.24%	1.75%	1.8%	0%	95.21%	0.8886
Plxnc1_p2	0.1538	0.6781	0.8241	4.16%	0.75%	2.08%	1.33%	1.95%	93.89%	0.9877
Plxnc1_p3	0.2833	0.9444	0.9423	1.46%	0.3%	0.58%	0.59%	0%	98.54%	0.9962
Rgmb	0.0753	0.1973	0.0456	7.57%	0.82%	3.05%	3.71%	1.62%	90.81%	0.5072
Rgmb_p1	0.5595	0.645	0.5118	3.54%	0.1%	1.55%	1.89%	0%	96.46%	0.893
Cadps	0.1233	0.1075	0.2212	6.31%	0.63%	3.15%	2.53%	<0.01%	93.69%	0.7744
Cadps_p1	0.4468	0.2296	0.1291	6.86%	0.19%	3.03%	3.64%	0%	93.14%	0.6843
Cadps_p2	0.3335	0.6435	0.8127	3.04%	0.25%	1.79%	1%	2.51%	94.45%	0.9877
Ube2b	0.4301	0.0328	0.0923	6.85%	0.15%	3.71%	2.98%	0%	93.15%	0.6101
Ube2b_p1	0.6005	0.5539	0.8974	3.6%	0.08%	2.71%	0.81%	5.62%	90.78%	0.9961
Shroom2_p1	0.2131	0.4297	0.019	6.99%	0.35%	2.8%	3.83%	5.49%	87.53%	0.3834
Nptn.1	0.1259	0.06	0.7101	5.07%	0.58%	3.36%	1.14%	0%	94.93%	0.9457
B3galnt1	0.6395	0.7892	0.9256	1.86%	0.06%	1.1%	0.7%	<0.01%	98.14%	0.9962
B3galnt1_p1	0.5119	0.1698	0.3891	4.53%	0.11%	2.58%	1.84%	<0.01%	95.47%	0.8712
B3galnt1_p2	0.0278	0.1013	0.9235	5.55%	1.16%	3.78%	0.61%	2.39%	92.06%	0.9962
Calm2	0.7258	0.5677	0.1256	5.24%	0.03%	2.3%	2.91%	4.53%	90.23%	0.6742
Dnajc6	6.34E-5	0.0092	5e-04	19.17%	3.11%	10.98%	5.07%	12.53%	68.3%	0.0511
Plcb1.1	0.0061	0.543	0.3031	5.53%	1.83%	1.66%	2.04%	1.08%	93.4%	0.8427
Plcb1.1_p1	0.0012	0.0559	0.7102	8.28%	3.01%	3.95%	1.32%	0%	91.72%	0.9457
Plcb1.1_p2	0.4311	0.7482	0.6586	3.85%	0.17%	2.31%	1.38%	8.46%	87.69%	0.9457
Plcb1.1_p3	0.7066	0.7139	0.4786	3.31%	0.03%	1.7%	1.58%	4.21%	92.47%	0.8901
Plcb1.2	2e-04	0.0089	0.5357	13.21%	3.13%	8.72%	1.36%	7.63%	79.15%	0.905
Plcb1.2_p1	0.1193	0.7309	0.5876	3.24%	0.61%	1.23%	1.4%	0.92%	95.85%	0.9441
Plcb1.2_p2	0.9232	0.7616	0.3487	3.35%	<0.01%	1.21%	2.14%	0.53%	96.12%	0.8712

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Plcb1.2_p3	0.2463	0.0188	0.9763	9.48%	0.45%	8.48%	0.55%	5.47%	85.04%	0.9997
Plcb1.2_p4	0.442	0.0064	0.0747	8.63%	0.15%	5.12%	3.36%	<0.01%	91.37%	0.5766
Plcb1.2_p5	0.1246	0.6224	0.5179	4.16%	0.71%	1.59%	1.86%	0%	95.84%	0.8955
Srp9	0.0058	0.0288	0.202	7.94%	1.83%	3.76%	2.35%	<0.01%	92.06%	0.7639
Srp9_p1	0.1528	0.0814	0.7356	5.64%	0.6%	3.74%	1.29%	<0.01%	94.36%	0.9609
Pja2	0.7249	0.7964	0.9837	1.53%	0.03%	1.08%	0.41%	<0.01%	98.47%	0.9997
Pja2_p1	0.0136	0.0302	0.4848	6.83%	1.48%	3.77%	1.58%	<0.01%	93.17%	0.8917
Pja2_p2	0.2752	0.266	0.9973	2.99%	0.33%	2.44%	0.22%	<0.01%	97.01%	0.9997
Pja2_p3	0.3545	0.1078	0.4951	6.22%	0.28%	3.85%	2.09%	0%	93.78%	0.893
Isca1	0.1339	0.0371	0.0212	10.62%	0.61%	5.58%	4.44%	3.32%	86.06%	0.4019
Isca1_p1	0.9522	0.2373	0.417	4.51%	<0.01%	2.58%	1.93%	0.28%	95.21%	0.8712
Isca1_p2	0.2066	0.1432	0.2569	7%	0.4%	4.35%	2.25%	4.72%	88.28%	0.8105
Azin1	0.5168	0.7746	0.8	2.81%	0.14%	1.37%	1.29%	0%	97.19%	0.9877
Lrp12	4e-04	0.3855	0.3814	7.95%	3.65%	2.14%	2.16%	0%	92.05%	0.8712
Lrp12_p1	0.0101	0.025	0.6941	7.9%	1.91%	4.62%	1.36%	0%	92.1%	0.9457
Lrp12_p2	0.3463	0.3006	0.3457	4.65%	0.24%	2.28%	2.13%	0%	95.35%	0.8712
Nxph1	0.3667	0.0736	0.965	4.68%	0.24%	3.87%	0.57%	0%	95.32%	0.9962
Nxph1_p1	0.1833	0.8857	0.5225	3.86%	0.57%	1.31%	1.98%	3.7%	92.44%	0.8955
Slc25a1_p1	0.9723	0.2735	0.7753	6.04%	<0.01%	5.09%	0.95%	11.09%	82.86%	0.9788
Slc25a1_p2	0.0036	0.2292	0.4033	9.23%	1.69%	6.09%	1.44%	14.46%	76.32%	0.8712
Slc9a6	0.4667	0.8334	0.7282	2.36%	0.15%	0.98%	1.23%	<0.01%	97.64%	0.9609
Slc9a6_p1	0.0342	0.4426	0.7385	4.71%	1.19%	2.37%	1.15%	2.59%	92.7%	0.9609
Cltc	2e-04	0.0752	0.5056	9.48%	3.11%	4.96%	1.4%	5.19%	85.33%	0.893
Vav2	0.077	0.7348	0.5261	3.44%	0.79%	1.1%	1.54%	0%	96.56%	0.8959

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Vav2_p1	0.2318	0.1752	0.4028	5.6%	0.42%	3.03%	2.15%	<0.01%	94.4%	0.8712
Kif2a	0.0519	0.319	0.1779	5.46%	0.93%	2.01%	2.52%	0%	94.54%	0.7619
Susd2_p1	0.3467	0.4455	0.4928	3.56%	0.22%	1.72%	1.61%	0%	96.44%	0.893
Susd2_p2	0.4741	0.0277	0.3045	6.54%	0.14%	4.19%	2.22%	<0.01%	93.46%	0.8427
Grm7	0.0437	0.1216	0.066	7.57%	1.07%	3.01%	3.49%	0%	92.43%	0.5744
Grm7_p1	0.2865	0.0247	0.517	5.95%	0.28%	4.16%	1.51%	0.48%	93.57%	0.8955
Ubfd1	0.8776	0.3107	0.5081	3.66%	0.01%	2.08%	1.58%	0%	96.34%	0.893
Ubfd1_p1	0.006	0.7334	0.1473	5.61%	1.86%	1.08%	2.66%	0%	94.39%	0.7261
Ubfd1_p2	2.54E-7	0.1178	0.5806	10.25%	6.23%	2.7%	1.32%	<0.01%	89.75%	0.9432
Ubfd1_p4	0.0257	0.1241	0.5757	5.44%	1.23%	2.8%	1.41%	<0.01%	94.56%	0.9424
Ubfd1_p5	0.0912	0.6345	0.6866	3.24%	0.72%	1.32%	1.21%	0%	96.76%	0.9457
Mpped1	0.0204	0.4659	0.102	6.57%	1.24%	2.57%	2.76%	4.93%	88.49%	0.6101
Mpped1_p1	0.6774	0.2398	0.3526	5.63%	0.06%	3.02%	2.56%	0%	94.37%	0.8712
Mpped1_p2	0.0024	0.2101	0.1917	7.21%	2.21%	2.61%	2.38%	1.01%	91.78%	0.7639
App	0.0105	0.0511	0.371	9.94%	2.32%	4.95%	2.68%	0%	90.06%	0.8712
Syng3	0.9149	0.3487	0.6154	3.33%	<0.01%	1.97%	1.35%	<0.01%	96.67%	0.9457
Cdk17	0.7858	0.0464	0.0395	9.45%	0.02%	5.86%	3.58%	5.35%	85.2%	0.4927
Cdk17_p1	0.033	0.2496	0.1011	8.18%	1.45%	2.89%	3.83%	<0.01%	91.82%	0.6101
Cdk17_p2	0.0238	0.4576	0.8051	4.72%	1.21%	2.63%	0.89%	4.93%	90.34%	0.9877
Pcdh17	0.5732	0.0248	0.4497	6.19%	0.08%	4.29%	1.82%	0%	93.81%	0.8796
Pcdh17_p1	1.47E-8	0.047	0.3284	14.57%	8.59%	3.81%	2.15%	0%	85.44%	0.8712
Pcdh17_p3	0.9176	0.7498	0.4079	3.84%	<0.01%	1.43%	2.41%	<0.01%	96.16%	0.8712
Pcdh17_p4	0.995	0.1384	0.0762	6.36%	<0.01%	2.94%	3.42%	<0.01%	93.64%	0.5766
Pcdh17_p5	0.4082	0.0126	0.0245	8.29%	0.16%	4.28%	3.85%	0%	91.71%	0.4019

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Atp8a1	0.0965	0.8743	0.1109	5.05%	0.7%	1.39%	2.96%	6.17%	88.78%	0.6273
Atp8a1_p1	0.0036	0.9422	0.9958	3.06%	2.09%	0.74%	0.23%	2.44%	94.51%	0.9997
Atp8a1_p2	0.0576	0.8312	0.8666	3.14%	1.09%	1.07%	0.97%	0%	96.86%	0.9938
Atp8a1_p3	0.5322	0.158	0.4185	6.41%	0.09%	4.72%	1.6%	7.02%	86.56%	0.8712
Atp8a1_p4	0.516	0.8016	0.904	3.43%	0.09%	2.76%	0.58%	16.43%	80.14%	0.9962
Atp8a1_p5	0.6154	0.249	0.6161	4.39%	0.08%	2.71%	1.61%	0%	95.61%	0.9457
Atp8a1_p6	0.1102	0.7728	0.5788	4.89%	0.57%	3.06%	1.26%	16.98%	78.12%	0.9432
Socs5	0.5832	0.9467	0.3699	2.57%	0.08%	0.56%	1.93%	<0.01%	97.43%	0.8712
Socs5_p1	0.0761	0.0987	0.314	9.95%	0.62%	7.69%	1.63%	14.03%	76.02%	0.8427
Hmg1	0.0013	3.1E-7	0.0027	19.11%	2.62%	10.95%	5.54%	0%	80.89%	0.1426
Hmg1_p1	0.0246	0.1382	0.1014	10.27%	1.12%	6.48%	2.66%	11.69%	78.04%	0.6101
Pgap1	0.7216	0.2296	0.4934	6.42%	0.05%	4.09%	2.28%	2.59%	90.99%	0.893
Pgap1_p1	0.3568	0.9044	0.2016	4.64%	0.27%	1.28%	3.09%	4.58%	90.78%	0.7639
Pgap1_p3	0.1073	0.2416	0.2391	6.16%	0.76%	2.69%	2.7%	0%	93.84%	0.8022
Pgap1_p4	0.7189	0.5008	0.0742	4.82%	0.03%	1.58%	3.21%	<0.01%	95.18%	0.5766
Pgap1_p5	3e-04	0.9201	0.9615	4.47%	3.33%	0.65%	0.49%	0%	95.53%	0.9962
Ranbp6	0.5973	0.4734	0.1343	5.33%	0.08%	1.95%	3.29%	0%	94.67%	0.703
Ranbp6_p2	0.6069	0.3829	0.643	3.56%	0.07%	2.08%	1.41%	0.15%	96.28%	0.9457
Ahi1	0.6689	0.7743	0.2736	3.27%	0.05%	1.02%	2.2%	0%	96.73%	0.8313
Dnm3	0.0848	0.9497	0.256	3.94%	0.64%	1.35%	1.94%	12.91%	83.15%	0.8105
Dnm3_p1	0.7136	0.0012	0.3852	12.21%	0.03%	10.46%	1.72%	6.65%	81.14%	0.8712
Dnm3_p2	0.3259	0.2046	0.6053	6.04%	0.22%	4.6%	1.22%	7.92%	86.04%	0.9445
Dnm3.1_p3	0.109	0.0051	0.4098	7.87%	0.67%	5.31%	1.89%	<0.01%	92.13%	0.8712
Dnm3.1_p5	0.7425	0.7085	0.1207	4.4%	0.03%	1.25%	3.11%	<0.01%	95.6%	0.656

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Slc19a1	0.3533	0.563	0.7803	4.01%	0.22%	2.78%	1.01%	7.22%	88.77%	0.9788
Slc19a1_p1	0.1077	0.5897	0.8793	2.88%	0.65%	1.45%	0.77%	0.21%	96.91%	0.9959
Spsb4	0.3143	0.2449	0.6007	4.78%	0.3%	2.85%	1.63%	0.5%	94.72%	0.9445
Spsb4_p1	0.0264	0.0325	0.8987	7.22%	1.41%	4.98%	0.82%	1.25%	91.54%	0.9961
Spsb4_p2	0.5633	0.1465	0.4289	6.44%	0.08%	4.61%	1.75%	5.58%	87.98%	0.8712
Spsb4_p3	0.3235	0.2058	0.008	7.2%	0.24%	2.35%	4.61%	0%	92.8%	0.2437
Pum1	0.732	0.7011	0.0047	6.7%	0.03%	1.24%	5.43%	0%	93.3%	0.1852
Pum1_p1	0.3058	0.052	0.649	5.41%	0.28%	3.76%	1.37%	<0.01%	94.59%	0.9457
Por	0.8519	0.6179	0.928	2.32%	0.01%	1.69%	0.62%	2.13%	95.55%	0.9962
Nbea	0.2321	0.0511	0.1869	6.23%	0.35%	3.43%	2.46%	0%	93.77%	0.7619
Nbea_p1	0.0538	0.6131	0.8804	3.08%	0.94%	1.36%	0.77%	<0.01%	96.92%	0.9959
Nbea_p2	0.7394	0.1553	0.4487	7.37%	0.05%	4.5%	2.82%	0.3%	92.33%	0.8796
Wsb2	0.9121	0.785	0.9192	2.24%	<0.01%	1.35%	0.89%	0%	97.76%	0.9962
Rab3c	0.5903	0.3861	0.016	6.11%	0.07%	1.82%	4.22%	0%	93.89%	0.3399
Rab3c_p1	0.3588	0.1859	0.556	4.63%	0.23%	2.82%	1.59%	0.31%	95.06%	0.9226
Rab3c_p2	0.0055	0.4626	0.4742	6.16%	2.26%	1.97%	1.93%	0%	93.84%	0.8886
Rab3c_p3	0.2649	0.2551	0.4732	6.23%	0.28%	4.5%	1.46%	8.94%	84.83%	0.8886
Rab3c_p4	0.2417	0.1028	0.7631	4.36%	0.34%	2.98%	1.03%	<0.01%	95.64%	0.9788
Rab3c_p5	0.0121	0.8862	0.2036	5.14%	1.7%	0.81%	2.63%	0%	94.86%	0.7639
Rab3c_p6	0.033	0.8229	0.4027	4.21%	1.24%	0.99%	1.98%	<0.01%	95.79%	0.8712
Hipk1	0.506	0.8396	0.6768	2.85%	0.1%	1.61%	1.14%	7.39%	89.76%	0.9457
Hipk1_p2	0.6762	0.0071	0.0119	10.21%	0.04%	5.62%	4.55%	1.21%	88.57%	0.3086
Hipk1_p3	0.4493	0.884	0.9833	1.57%	0.14%	1.07%	0.36%	3.39%	95.04%	0.9997
Hipk1_p4	0.5746	0.2862	0.6357	3.54%	0.08%	2.15%	1.31%	<0.01%	96.46%	0.9457

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Prkacb_p1	0.8851	0.6	0.3956	5.19%	<0.01%	3.65%	1.54%	14.52%	80.29%	0.8712
Prkacb_p2	0.2123	0.4879	0.9997	2.34%	0.43%	1.8%	0.12%	<0.01%	97.66%	0.9997
Prkacb_p3	0.5627	0.8378	0.0965	4.08%	0.08%	1%	3%	1.28%	94.64%	0.6101
Dnaja1	0.006	0.0835	0.0117	11.74%	2.32%	3.86%	5.55%	0%	88.26%	0.3086
Dnaja1_p1	0.8911	0.5804	0.4512	4.14%	<0.01%	2.38%	1.75%	5.21%	90.65%	0.8796
Dnaja1_p2	0.4797	0.1678	0.1751	5.68%	0.13%	2.79%	2.75%	0%	94.32%	0.7619
Atmin	0.6364	0.9283	0.9991	0.97%	0.06%	0.75%	0.16%	0.69%	98.34%	0.9997
Atmin_p1	0.0692	0.5216	0.8998	4.14%	0.85%	2.55%	0.73%	4.95%	90.91%	0.9961
Atmin_p2	0.1166	0.5215	0.4757	4.25%	0.67%	1.8%	1.78%	0.69%	95.07%	0.8886
Htr7	0.5165	0.0636	0.1514	6.53%	0.11%	3.56%	2.85%	0%	93.47%	0.7361
Htr7_p1	0.1139	0.0145	0.693	7.19%	0.73%	5.09%	1.38%	0%	92.81%	0.9457
Uhrf2	0.962	0.5299	0.8632	2.72%	<0.01%	1.83%	0.89%	0.82%	96.47%	0.9938
Uhrf2_p1	0.0175	0.1714	0.8501	4.8%	1.4%	2.56%	0.83%	0%	95.2%	0.9907
Rnf14	0.9305	0.0738	0.2506	7.11%	<0.01%	4.19%	2.92%	0%	92.89%	0.8105
Zic4	0.3792	0.3562	0.9429	4.75%	0.17%	4.06%	0.51%	9.62%	85.63%	0.9962
Zic4_p1	0.6155	0.297	0.3009	5.07%	0.08%	2.51%	2.49%	0%	94.93%	0.8427
Zic4_p2	0.2842	0.6931	0.1741	4.11%	0.29%	1.27%	2.55%	0.67%	95.22%	0.7619
Crocc	0.3484	0.1654	0.4573	4.5%	0.22%	2.6%	1.68%	<0.01%	95.5%	0.8814
Nrep	0.4004	0.5309	0.2031	5.06%	0.16%	2.64%	2.25%	6.52%	88.42%	0.7639
Cul4b	0.079	0.9198	0.508	3.03%	0.78%	0.66%	1.59%	0%	96.97%	0.893
Cul4b_p1	0.5651	0.6337	0.3479	3.89%	0.09%	1.7%	2.1%	1.8%	94.31%	0.8712
Cul4b_p2	0.97	0.337	0.8321	3.17%	<0.01%	2.19%	0.97%	0%	96.83%	0.9877
Cul4b_p3	0.0804	0.5834	0.111	5.16%	0.75%	1.54%	2.87%	0.87%	93.97%	0.6273
Tmem30a_p1	0.7061	0.4268	0.8726	2.86%	0.04%	1.95%	0.87%	0.04%	97.1%	0.9959

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Tmem30a_p2	0.1292	0.7606	0.0062	7.06%	0.6%	1.36%	5.1%	2.2%	90.74%	0.22
Tmem30a_p3	0.6741	0.0377	0.7723	5.65%	0.05%	4.4%	1.2%	0%	94.35%	0.9788
Tmem30a_p4	0.0628	0.4597	0.3747	4.51%	0.86%	1.78%	1.87%	0.57%	94.92%	0.8712
Fem1b	2.68E-9	0.5939	0.6265	10.8%	8.18%	1.41%	1.22%	0.74%	88.46%	0.9457
Fem1b_p1	0.3682	0.9916	0.8922	1.51%	0.25%	0.36%	0.9%	0%	98.49%	0.9959
Fem1b_p2	0.0031	0.1327	0.658	6.63%	2.33%	2.96%	1.33%	<0.01%	93.38%	0.9457
Fem1b_p3	0.1279	0.993	0.2606	3.72%	0.7%	0.33%	2.68%	0%	96.28%	0.8105
Fem1b_p5	0.5986	0.4938	0.3323	4.01%	0.08%	1.75%	2.19%	<0.01%	95.99%	0.8712
Zdhhc13	0.0038	0.4974	0.0747	7.3%	2.22%	1.68%	3.41%	<0.01%	92.7%	0.5766
Peg10	0.1793	0.4435	0.6391	3.49%	0.45%	1.73%	1.3%	0%	96.51%	0.9457
Peg10_p2	0.0811	0.0011	0.7115	9.95%	0.95%	7.56%	1.44%	0%	90.06%	0.9457
Peg10_p3	0.0077	0.0847	0.3112	7.35%	1.87%	3.3%	2.18%	0%	92.65%	0.8427
Peg10_p4	0.1364	0.0965	0.3126	7.95%	0.49%	5.65%	1.81%	7.87%	84.18%	0.8427
Peg10_p5	1e-04	0.3267	0.2099	9.02%	3.79%	2.82%	2.41%	3.17%	87.81%	0.7651
Peg10_p6	0.4436	0.9187	0.2908	4.85%	0.19%	1.94%	2.72%	13.48%	81.67%	0.8427
Peg10.1	0.1064	0.8676	0.9853	2.45%	0.89%	1.08%	0.48%	0%	97.55%	0.9997
Peg10.1_p1	0.7797	0.106	0.7108	6.1%	0.02%	4.92%	1.16%	5.19%	88.72%	0.9457
Peg10.1_p2	0.4913	0.0718	0.3008	5.86%	0.13%	3.49%	2.25%	<0.01%	94.14%	0.8427
Nsf	0.4592	0.525	0.9724	4.95%	0.12%	4.43%	0.39%	15.89%	79.16%	0.9997
Hap1	2.74E-5	0.3825	0.0013	14.8%	3.49%	6.62%	4.69%	21.98%	63.22%	0.0915
Tmem66	0.1918	0.923	0.8867	2.02%	0.48%	0.71%	0.83%	0%	97.98%	0.9959
Serpini1	0.9715	0.0812	0.1038	8.37%	<0.01%	5.11%	3.26%	4.14%	87.49%	0.6112
Serpini1_p1	0.966	0.0976	0.2239	7.18%	<0.01%	4.81%	2.36%	4.69%	88.13%	0.7744
Serpini1_p2	0.6695	0.1626	0.2592	5.87%	0.04%	3.73%	2.1%	3.82%	90.31%	0.8105

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Myh10	0.189	0.6342	0.2994	4.53%	0.45%	1.86%	2.21%	2.99%	92.48%	0.8427
Myh10_p1	0.5524	0.3412	0.4455	3.87%	0.09%	2.07%	1.71%	0.4%	95.73%	0.8796
Slc39a11	0.0049	0.9812	0.9939	2.68%	1.99%	0.43%	0.27%	0.89%	96.43%	0.9997
Slc39a11_p1	0.0036	0.2468	0.4359	6.52%	2.25%	2.42%	1.85%	0%	93.48%	0.8719
Slc39a11_p3	0.0315	0.0977	0.0677	8.92%	1.03%	4.97%	2.93%	6.03%	85.04%	0.5744
Slc39a11_p4	2e-04	0.9401	0.7713	6.15%	3.86%	1.19%	1.1%	7.7%	86.15%	0.9788
Kcnd2	0.0028	0.4407	0.9215	6.77%	3.26%	2.57%	0.94%	0.28%	92.95%	0.9962
Kcnd2_p1	0.1966	0.9763	0.1811	3.85%	0.4%	1.01%	2.44%	11.73%	84.42%	0.7619
Kcnd2_p2	0.3368	0.1383	0.9572	4.2%	0.28%	3.31%	0.61%	0%	95.8%	0.9962
Kcnd2_p3	0.0022	0.0198	3e-04	13.27%	2.33%	4.11%	6.83%	0%	86.73%	0.0511
Lpin2	8e-04	0.6968	0.0351	10.36%	2.85%	3.69%	3.82%	16.96%	72.68%	0.4509
Lpin2_p1	0.3347	0.3943	0.6604	4.96%	0.23%	3.47%	1.25%	7.09%	87.96%	0.9457
Lpin2_p2	0.0308	0.3211	0.0482	7.16%	1.23%	2.18%	3.74%	0.14%	92.71%	0.5072
Mecp2	0.0948	0.8247	0.3113	3.68%	0.7%	0.9%	2.07%	<0.01%	96.32%	0.8427
Mecp2_p1	0.3585	0.1163	0.6381	4.39%	0.21%	2.88%	1.29%	0%	95.61%	0.9457
Mecp2_p3	0.9	0.611	0.6075	2.89%	<0.01%	1.52%	1.36%	0.99%	96.12%	0.9445
Mecp2_p4	0.132	0.2321	0.1201	6.15%	0.61%	2.49%	3.06%	<0.01%	93.85%	0.656
Mecp2_p6	0.0207	0.0128	0.9485	6.74%	1.42%	4.73%	0.58%	<0.01%	93.26%	0.9962
Mecp2_p7	0.0035	0.9234	0.7547	3.94%	2.07%	0.85%	1.03%	2.82%	93.23%	0.9755
Mecp2_p8	0.2527	0.266	0.3837	4.39%	0.33%	2.2%	1.86%	<0.01%	95.61%	0.8712
Mecp2_p9	0.9841	0.7242	0.4001	3.24%	<0.01%	1.23%	2.01%	<0.01%	96.76%	0.8712
Ran	0.293	0.0029	0.0668	9.99%	0.28%	6.37%	3.34%	1.3%	88.71%	0.5744
Ran_p1	0.2252	0.4682	0.9602	3.07%	0.45%	2.02%	0.61%	<0.01%	96.93%	0.9962
Cox14	0.3651	0.9106	0.9338	1.53%	0.21%	0.7%	0.62%	<0.01%	98.47%	0.9962

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Bicap	0.1165	0.0026	0.9473	7.89%	0.58%	6.79%	0.52%	2.39%	89.73%	0.9962
Bicap_p1	0.3152	0.0116	0.0235	20.56%	0.14%	18.25%	2.18%	27.9%	51.53%	0.4019
Wdr82	0.5984	0.8801	0.1807	3.74%	0.07%	1.29%	2.39%	5.91%	90.35%	0.7619
Wdr82_p1	0.4712	0.4746	0.7819	3.44%	0.14%	2.23%	1.07%	2.25%	94.32%	0.9788
Wdr82_p2	0.5805	0.2268	0.2939	4.52%	0.08%	2.34%	2.11%	<0.01%	95.48%	0.8427
Fam160b1	0.4801	0.2157	0.5357	7.62%	0.1%	6.29%	1.23%	14.55%	77.83%	0.905
Fam160b1_p1	0.9734	0.8169	0.8419	2.42%	<0.01%	1.25%	1.17%	<0.01%	97.58%	0.9907
Fam160b1_p2	0.0056	0.2788	0.1406	7.22%	2.03%	2.29%	2.9%	<0.01%	92.78%	0.7182
Fam160b1_p3	0.216	0.1888	0.2663	6.01%	0.36%	3.57%	2.07%	3.91%	90.09%	0.8181
Fam160b1_p5	0.3947	0.1314	0.3627	5.28%	0.2%	3.01%	2.07%	0%	94.72%	0.8712
Fam160b1_p6	0.7972	0.1586	0.1726	5.63%	0.02%	2.84%	2.77%	0%	94.37%	0.7619
Fam160b1_p7	0.947	0.3531	0.5733	5.1%	<0.01%	3.81%	1.29%	8.45%	86.44%	0.9424
Fam160b1_p8	0.1517	0.2354	0.6536	5.99%	0.51%	4.22%	1.26%	6.62%	87.39%	0.9457
Fam160b1_p9	0.0949	0.5299	0.6869	3.44%	0.7%	1.53%	1.2%	0%	96.56%	0.9457
Arf1	0.0015	0.3383	0.8556	7.02%	2.19%	4.11%	0.72%	9.6%	83.38%	0.9907
Arf1_p2	0.0222	0.5029	0.343	5.37%	1.4%	1.87%	2.1%	0.94%	93.69%	0.8712
Psd3_p1	0.5512	0.418	0.2767	5.44%	0.1%	2.87%	2.47%	3.82%	90.74%	0.8321
Psd3_p3	0.0629	0.8639	0.957	2.43%	0.96%	0.9%	0.57%	0%	97.57%	0.9962
Psd3_p4	0.8392	0.3364	0.0892	6.84%	0.01%	3.81%	3.02%	7.48%	85.68%	0.6101
Psd3_p5	0.9168	0.2492	0.3828	4.92%	<0.01%	2.7%	2.22%	<0.01%	95.08%	0.8712
Psd3_p7	0.3795	0.2382	0.0885	6.08%	0.21%	2.58%	3.3%	0.43%	93.49%	0.6101
Psd3_p8	0.8037	0.6519	0.8052	2.27%	0.02%	1.29%	0.96%	0%	97.73%	0.9877
Psd3_p9	0.4385	0.1516	0.9134	3.83%	0.16%	2.94%	0.73%	0%	96.17%	0.9962
Psd3_p10	0.1235	0.0426	0.6481	7.19%	0.77%	4.78%	1.64%	0.2%	92.61%	0.9457

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Pgs1.1	0.2334	0.0203	0.475	7.7%	0.4%	5.42%	1.87%	1.33%	90.98%	0.8886
2510039O18Rik	0.9649	0.2783	0.2842	4.68%	<0.01%	2.35%	2.33%	0%	95.32%	0.8333
2510039O18Rik_p1	0.0018	0.3082	0.3707	6.79%	2.58%	2.2%	2.01%	0%	93.21%	0.8712
2510039O18Rik_p2	0.2287	0.8179	0.9657	1.79%	0.37%	0.94%	0.48%	0%	98.21%	0.9962
Lingo3	0.7884	0.4077	0.912	2.76%	0.02%	2%	0.75%	0%	97.24%	0.9962
Zbtb5_p1	0.6298	0.7104	0.8576	2.18%	0.06%	1.29%	0.83%	0.88%	96.95%	0.9907
Zbtb5_p2	0.581	0.6364	0.8241	3.19%	0.08%	2.16%	0.94%	4.95%	91.87%	0.9877
Zbtb5_p3	0.7297	0.7934	0.6424	2.65%	0.03%	1.36%	1.26%	3.34%	94.01%	0.9457
Mapk9	0.5889	0.9307	0.2293	3.38%	0.08%	0.77%	2.53%	1.39%	95.23%	0.784
Mapk9_p1	0.0041	0.0616	0.7093	7.97%	2.08%	4.73%	1.16%	3.12%	88.91%	0.9457
Mapk9_p2	0.1368	0.3219	0.4294	4.95%	0.59%	2.49%	1.87%	1.26%	93.79%	0.8712
Mapk9.1	0.0015	0.5044	0.1921	8.2%	2.03%	4.17%	2%	14.67%	77.13%	0.7639
Mapk9.1_p1	0.3381	0.0376	0.1694	7.59%	0.24%	4.68%	2.67%	1.8%	90.61%	0.7619
Mapk9.1_p2	0.0595	0.4747	0.029	6.29%	0.87%	1.61%	3.82%	0%	93.71%	0.4396
Mapk9.1_p3	0.4623	0.0482	0.3804	5.48%	0.13%	3.5%	1.85%	<0.01%	94.52%	0.8712
Dcaf7	5e-04	0.0377	0.0596	10.59%	3.11%	4.06%	3.43%	0.63%	88.78%	0.5492
Dcaf7_p1	0.6107	0.1285	0.2841	5.41%	0.07%	3.03%	2.31%	0%	94.59%	0.8333
Dcaf7_p2	0.2326	0.7672	0.83	2.52%	0.4%	1.14%	0.99%	0%	97.48%	0.9877
Dcaf7_p3	0.7722	0.6439	0.5856	2.75%	0.02%	1.3%	1.43%	0%	97.25%	0.9441
Dcaf7_p4	0.1959	0.8031	0.3704	3.3%	0.42%	0.96%	1.92%	<0.01%	96.7%	0.8712
Vat1l	0.0657	0.1438	0.626	5.02%	0.83%	2.88%	1.3%	0.58%	94.41%	0.9457
Reep1	0.0872	0.069	0.2158	6.27%	0.72%	3.21%	2.34%	0%	93.73%	0.7744
Reep1_p1	0.5347	0.4383	0.4279	3.6%	0.1%	1.74%	1.76%	0%	96.4%	0.8712
Reep1_p2	0.0036	0.1655	0.0312	8.67%	1.97%	3.11%	3.59%	2.09%	89.23%	0.4415

PTRE-Seq - Effects on Translation Efficiency Cont'd

Gene.UTR_Peak	sequence	condition	interaction	R ²	sequence	condition	interaction	within samples	unexplained	FDR
Reep1_p3	0.7378	0.8131	0.9319	1.9%	0.03%	1.2%	0.67%	1.56%	96.54%	0.9962
Reep1_p4	0.0269	0.8156	0.1454	4.95%	1.19%	1.12%	2.64%	1.93%	93.13%	0.7252
Myo5a	0.4281	0.9587	0.8853	1.93%	0.21%	0.69%	1.02%	<0.01%	98.07%	0.9959
Myo5a_p1	0.0281	0.2048	0.2481	6.3%	1.29%	2.6%	2.42%	<0.01%	93.7%	0.8105
Myo5a_p2	0.0195	0.007	0.1967	8.31%	1.31%	4.64%	2.36%	0%	91.69%	0.7639
Myo5a_p3	0.6171	0.1128	0.008	9.17%	0.06%	4.93%	4.19%	6.46%	84.36%	0.2437
Myo5a_p4	0.1569	0.0635	0.4069	5.62%	0.49%	3.36%	1.77%	0.16%	94.22%	0.8712
Myo5a_p5	0.6729	0.1462	0.8315	7.34%	0.04%	6.56%	0.74%	12.67%	79.99%	0.9877
Myo5a_p6	0.9636	0.9273	0.8672	1.63%	<0.01%	0.74%	0.89%	0.55%	97.82%	0.9938
Myo5a_p7	0.4751	0.341	0.3838	4.15%	0.13%	2.17%	1.85%	0.86%	95%	0.8712
Myo5a_p8	0.2185	0.1607	0.046	6.98%	0.4%	2.79%	3.79%	<0.01%	93.02%	0.5072
Myo5a_p9	0.8096	0.741	0.3461	3.64%	0.02%	1.53%	2.09%	2.8%	93.56%	0.8712
8430427H17Rik	0.2992	0.1714	0.507	4.41%	0.27%	2.57%	1.57%	0%	95.59%	0.893
8430427H17Rik_p2	0.3032	0.9238	0.1717	3.81%	0.29%	0.7%	2.83%	0%	96.19%	0.7619
8430427H17Rik_p3	0.3465	0.2641	0.7167	3.59%	0.22%	2.23%	1.14%	<0.01%	96.41%	0.9497
8430427H17Rik_p5	0.1553	0.2367	0.9626	3.33%	0.51%	2.33%	0.49%	<0.01%	96.67%	0.9962
8430427H17Rik_p6	0.3231	0.4226	0.5014	3.93%	0.27%	1.93%	1.73%	0%	96.07%	0.893

Table S8: Oligos for PTRESeq Library Generation

name	sequence (5'-3')
pCMV_T7-F	tgacacgcgtGTTGACATTGATTATTGACTAGTTA
pCMV_T7-R	tgacggatccTCCCTATAGTGAGTCGTATTAATTT
pmrPTRE_AAV_Full_F	taagctagcctgtaccGGCATCCCTGTGACCCCTC
pmrPTRE_AAV_Full_R	ggtagcaggctagcttaCTTGACAGCTCGTCCATGCCGTAC
GFP-F	CCTACGGCGTGCAGTGCTTCAGC
GFP-R	CGGCGAGCTGCACGCTGCGTCCTC
pmrPTRE antisense	GGCACTGGAGTGGCAACT
pmrPTRE sense	gcatggacgagctgtacaag
KpnI_overhang_1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCATgta*c
KpnI_overhang_2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGGTgta*c
KpnI_overhang_3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTTCCCTgta*c
KpnI_overhang_4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCAGCTgta*c
KpnI_complement_1	/5phos/A*TGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3
KpnI_complement_2	/5phos/A*CCTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3
KpnI_complement_3	/5phos/A*GGAACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3
KpnI_complement_4	/5phos/A*GCTGCTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3
NheI_overhang	/5phos/c*tagAGATCGGAAGAGCACACGTCTG
NheI_complement	CAGACGTGTGCTCTTCCGATC*T

Chapter 5 References

1. Anderson, P. & Kedersha, N. Stress granules: the Tao of RNA triage. *Trends in biochemical sciences* **33**, 141–150 (2008).
2. Busch, A. & Hertel, K. J. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdisciplinary Reviews: RNA* **3**, 1–12 (2012).
3. Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nature reviews Molecular cell biology* **8**, 113 (2007).
4. Hoyle, N. P., Castelli, L. M., Campbell, S. G., Holmes, L. E. & Ashe, M. P. Stress-dependent relocalization of translationally primed mRNPs to cytoplasmic granules that are kinetically and spatially distinct from P-bodies. *J Cell Biol* **179**, 65–74 (2007).
5. Louis, I. V.-S., Dickson, A. M., Bohjanen, P. R. & Wilusz, C. J. CELFish ways to modulate mRNA decay. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **1829**, 695–707 (2013).
6. Xiong, H. Y. *et al.* The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).
7. Lee, J. & Cooper, T. Pathogenic mechanisms of myotonic dystrophy. *Biochemical Society Transactions* **37**, 1281–1286 (Dec. 2009).

8. Dasgupta, T. & Ladd, A. N. The importance of CELF control: molecular and biological roles of the CUG-BP, Elav-like family of RNA-binding proteins. *Wiley Interdisciplinary Reviews: RNA* **3**, 104–121 (2012).
9. Brimacombe, K. R. & Ladd, A. N. Cloning and embryonic expression patterns of the chicken CELF family. *Developmental Dynamics* **236**, 2216–2224 (2007).
10. Timchenko, L., Timchenko, N., Caskey, C. & Roberts, R. Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy. *Human molecular genetics* **5**, 115–121 (1996).
11. Wang, E. T. *et al.* Antagonistic regulation of mRNA expression and splicing by CELF and MBNL proteins. *Genome Research* **25**, 858–871 (Apr. 2015).
12. Moraes, K. C., Wilusz, C. J. & Wilusz, J. CUG-BP binds to RNA substrates and recruits PARN deadenylase. *RNA* **12**, 1084–1091 (June 2006).
13. Dougherty, J. D. *et al.* The disruption of Celf6, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors. *Journal of Neuroscience* **33**, 2732–2753 (2013).
14. Maloney, S. E., Khangura, E. & Dougherty, J. D. The RNA-binding protein Celf6 is highly expressed in diencephalic nuclei and neuromodulatory cell populations of the mouse brain. *Brain Structure and Function* **221**, 1809–1831 (2016).
15. Ladd, A. N., Nguyen, N. H., Malhotra, K. & Cooper, T. A. CELF6, a member of the CELF family of RNA-binding proteins, regulates muscle-specific splicing enhancer-dependent alternative splicing. *Journal of Biological Chemistry* **279**, 17756–17764 (2004).
16. Wagnon, J. L. *et al.* CELF4 regulates translation and local abundance of a vast set of mRNAs, including genes associated with regulation of synaptic function. *PLoS Genet* **8**, e1003067 (2012).

17. Singh, G., Charlet-B, N., Han, J. & Cooper, T. A. ETR-3 and CELF4 protein domains required for RNA binding and splicing activity in vivo. *Nucleic acids research* **32**, 1232–1241 (2004).
18. Doyle, J. P. *et al.* Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135**, 749–762 (2008).
19. Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172 (2013).
20. Vidaki, M. *et al.* A Requirement for Mena, an Actin Regulator, in Local mRNA Translation in Developing Neurons. *Neuron* **95**, 608–622.e5 (Aug. 2017).
21. Van Nostrand, E. L. *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature Methods* **13**, 508– (Mar. 2016).
22. Huppertz, I. *et al.* iCLIP: Protein–RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 (Feb. 2014).
23. Ascano, M., Hafner, M., Cekan, P., Gerstberger, S. & Tuschl, T. Identification of RNA–protein interaction networks using PAR-CLIP. *Wiley Interdisciplinary Reviews: RNA* **3**, 159–177 (Dec. 2011).
24. Corcoran, D. L. *et al.* PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biology* **12**, R79 (2011).
25. Kishore, S. *et al.* A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nature Methods* **8**, 559–564 (May 2011).
26. Lebedeva, S. *et al.* Transcriptome-wide Analysis of Regulatory Interactions of the RNA-Binding Protein HuR. *Molecular Cell* **43**, 340–352 (Aug. 2011).
27. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (Feb. 2013).

28. Moore, M. J. *et al.* Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nature Protocols* **9**, 263–293 (Jan. 2014).
29. Mukherjee, N. *et al.* Integrative Regulatory Mapping Indicates that the RNA-Binding Protein HuR Couples Pre-mRNA Processing and mRNA Stability. *Molecular Cell* **43**, 327–339 (Aug. 2011).
30. Sugimoto, Y. *et al.* Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biology* **13**, R67 (2012).
31. Ule, J. CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science* **302**, 1212–1215 (Nov. 2003).
32. Zhang, C. & Darnell, R. B. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nature Biotechnology* **29**, 607–614 (June 2011).
33. Uren, P. J. *et al.* Site identification in high-throughput RNA–protein interaction data. *Bioinformatics* **28**, 3013–3020 (Sept. 2012).
34. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multi-factor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**, 4288–4297 (Jan. 2012).
35. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (Nov. 2009).
36. Pablo, J. L., Wang, C., Presby, M. M. & Pitt, G. S. Polarized localization of voltage-gated Na⁺ channels is regulated by concerted FGF13 and FGF14 action. *Proceedings of the National Academy of Sciences* **113**, E2665–E2674 (2016).

37. Eiden, L. E., Schiøfer, M. K.-H., Weihe, E. & Schiøtz, B. The vesicular amine transporter family (SLC18): amine/proton antiporters required for vesicular accumulation and regulated exocytotic secretion of monoamines and acetylcholine. *Pflügers Archiv European Journal of Physiology* **447**, 636–640 (Feb. 2004).
38. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess over-representation of Gene Ontology categories in Biological Networks. *Bioinformatics* **21**, 3448–3449 (June 2005).
39. Ray, D. *et al.* Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nature biotechnology* **27**, 667 (2009).
40. Bailey, T. L. *et al.* MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research* **37**, W202–W208 (July 2009).
41. Bailey, T. L. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**, 1653–1659 (June 2011).
42. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *eng. Proceedings. International Conference on Intelligent Systems for Molecular Biology* **2**, 28–36 (1994).
43. McLeay, R. C. & Bailey, T. L. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, 165 (Apr. 2010).
44. Auweter, S. D., Oberstrass, F. C. & Allain, F. H.-T. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic acids research* **34**. reviewed, 4943–4959 (2006).
45. Harvey, R. F. *et al.* Trans-acting translational regulatory RNA binding proteins. *Wiley Interdisciplinary Reviews: RNA*. reviewed (2018).

46. Choi, H. S. *et al.* Poly (C)-binding proteins as transcriptional regulators of gene expression. *Biochemical and biophysical research communications* **380**, 431–436 (2009).
47. Scoumanne, A., Cho, S. J., Zhang, J. & Chen, X. The cyclin-dependent kinase inhibitor p21 is regulated by RNA-binding protein PCBP4 via mRNA stability. *Nucleic acids research* **39**, 213–224 (2010).
48. Gilks, N. *et al.* Stress Granule Assembly Is Mediated by Prion-like Aggregation of TIA-1. *Molecular Biology of the Cell* **15**, 5383–5398 (Dec. 2004).
49. Villalba, A., Coll, O. & Gebauer, F. Cytoplasmic polyadenylation and translational control. *Current Opinion in Genetics & Development. Differentiation and gene regulation* **21**, 452–457 (Aug. 2011).
50. Nakielnny, S. & Dreyfuss, G. The hnRNP C proteins contain a nuclear retention sequence that can override nuclear export signals. *en. The Journal of Cell Biology* **134**, 1365–1373 (Sept. 1996).
51. Shetty, S. Regulation of urokinase receptor mRNA stability by hnRNP C in lung epithelial cells. *en. Molecular and Cellular Biochemistry* **272**, 107–118 (Apr. 2005).
52. Hinnebusch, A. G., Ivanov, I. P. & Sonenberg, N. Translational control by 5â² – *untranslated regionsofeukaryoticmRNAs*. *Science* **352**, 1413–1416 (2016).
53. Ivshina, M., Lasko, P. & Richter, J. D. Cytoplasmic polyadenylation element binding proteins in development, health, and disease. *Annual review of cell and developmental biology* **30**, 393–415 (2014).
54. Piqué, M., López, J. M., Foissac, S., Guigó, R. & Méndez, R. A combinatorial code for CPE-mediated translational control. *Cell* **132**, 434–448 (2008).

55. Cottrell, K. A., Chaudhari, H. G., Cohen, B. A. & Djuranovic, S. PTRE-seq reveals mechanism and interactions of RNA binding proteins and miRNAs. en. *Nature Communications* **9**, 301 (Jan. 2018).
56. Heiman, M. *et al.* A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135**, 738–748 (2008).
57. Ladd, A. N. Multiple domains control the subcellular localization and activity of ETR-3, a regulator of nuclear and cytoplasmic RNA processing events. *Journal of Cell Science* **117**, 3519–3529 (July 2004).
58. Ladd, A. N., Charlet-B., N. & Cooper, T. A. The CELF Family of RNA Binding Proteins Is Implicated in Cell-Specific and Developmentally Regulated Alternative Splicing. *Molecular and Cellular Biology* **21**, 1285–1296 (Feb. 2001).
59. Ohsawa, N., Koebis, M., Mitsuhashi, H., Nishino, I. & Ishiura, S. ABLIM1 splicing is abnormal in skeletal muscle of patients with DM1 and regulated by MBNL, CELF and PTBP1. *Genes to Cells* **20**, 121–134 (2015).
60. Sureban, S. M. *et al.* Functional antagonism between RNA binding proteins HuR and CUGBP2 determines the fate of COX-2 mRNA translation. *Gastroenterology* **132**, 1055–1065 (2007).
61. Mukherjee, N. *et al.* Global target mRNA specification and regulation by the RNA-binding protein ZFP36. *Genome biology* **15**, R12 (2014).
62. Plass, M., Rasmussen, S. H. & Krogh, A. Highly accessible AU-rich regions in 3′ untranslated regions are hotspots for binding of regulatory factors. *PLoS computational biology* **13**, e1005460 (2017).
63. Jao, C. Y. & Salic, A. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proceedings of the National Academy of Sciences* **105**, 15779–15784 (2008).

64. Cosson, B. *et al.* Oligomerization of EDEN-BP is required for specific mRNA deadenylation and binding. *Biology of the Cell* **98**, 653–665 (2006).
65. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
66. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
67. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research* **41**, e108–e108 (2013).
68. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2013).
69. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models using lme4. *arXiv:1406.5823 [stat]*. arXiv: 1406.5823. (2016) (June 2014).
70. Nakagawa, S. & Schielzeth, H. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**, 133–142 (2013).
71. Hothorn T Bretz F, W. P. multcomp: Simultaneous tests and confidence intervals for general linear hypotheses in parametric models. *R package version 1.2-13*. (2012).

Conclusions and Future Directions

Summary

In this dissertation I have presented data along two parallel lines of inquiry regarding the function of the CELF6 RNA binding protein. First, I have presented additional data to our original 2013 study of this protein's association to behavior by characterizing the spectral and temporal features of the vocalization of *Celf6*^{-/-} mice. These data have not been published previously, but were presented at the International Meeting for Autism Research in 2013. At that time, I proposed that we use a conditional knockout strategy to narrow the phenotype of *Celf6*^{-/-} mice. Initially, before Susan Maloney in our group went on to fully describe where in the brain CELF6 is expressed, we hypothesized that the loss of CELF6 to the serotonergic cells would be sufficient to recapitulate the phenotype of the global null. This was not the case, and was also not the case when we expanded our search to include dopaminergic neurons.

In parallel, I was interested in understanding the biology of CELF6 which may contribute to the USV phenotype. When I uncovered that CELF6 primarily associated with 3'UTRs *in vivo* my immediate hypothesis was that it may, like CELF1, affect transcript stability or translation. PTRE-Seq enabled me to ask this question across hundreds of *in vivo*-identified regulatory elements. *In vitro*, CELF6 overexpression was not associated with changes to all of them. This is actually expected: some targets may be false posi-

tives or of low binding affinity, furthermore, *in vitro* the molecular context in which CELF6 is embedded is different from its *in vivo* context(s). Nevertheless, wherever CELF6 did correlate to changes in reporter abundance, these were overwhelmingly to reduce the levels of the reporter. These reductions were ablated by mutating *in silico* defined sequence motifs, many of which were UGU-containing elements. Translation efficiency of reporter library elements was altered in some cases, but some reporters went up and some went down and there were no generalizable trends. Thus on a case-by-case basis it seems that CELF6 can exert effects on translation, but that largely its effects are on mRNA abundance.

Future Directions

Does CELF6 enhance mRNA decay?

So far, I have been conservative in my language regarding the mechanism of CELF6's action on elements of the PTRE-Seq library, and thus on its 3'UTR targets. Nevertheless, the lack of generalizable effects on translation, and the apparent trends in reporter level reduction lead me to hypothesize that CELF6's actions are on destabilizing mRNA. This is supported by previous research on CELF1 which can associate and recruit the PARN de-adenylase which enhances mRNA degradation.

To replicate the findings of the reporter assay, I have subcloned example reporters shown in Chapter 5, Figures 3 and 4, as individual plasmids. This replication work is currently on-going. Additionally, using ethynyl uridine incorporation into nascent transcripts will allow us to calculate decay constants for each element in the library. Repeating the PTRE-Seq assay at multiple time points using this pulse labeling procedure can allow us to determine whether CELF6 condition interacts with the slope of mRNA decay, the hypothesis being that it quickens mRNA decay rates.

CELF1 association has been shown to enhance de-adenylase enzymatic activity as described in Chapter 4. Using co-immunoprecipitation, we can determine whether CELF6 can directly associate with PARN or another de-adenylase enzyme, and whether mutations of key domains in the CELF6 protein ablate this association and prevent target degradation.

Does CELF6 result in target repression *in vivo*?

It is clear from the literature that CELF proteins, like other RNA binding proteins, are versatile in their function. Thus the cell type and molecular context in which CELF6 is embedded likely determines the outcome on target molecules. It remains to be shown whether or not I can generalize my findings *in vitro* to the brain.

Our PTRE-Seq library was prepared in a construct which can be used to generate adeno-associated virus (AAV) for *in vivo* targeting. Currently, work is on-going in our group to validate and maximize recovery from massively parallel assays injected as virus *in vivo*. The CELF6 PTRE-Seq library can be introduced into regions exhibiting enriched CELF6 expression in wild-type and knockout animals. This can enable us to determine whether (a) CELF6 represses reporter elements *in vivo* as it does *in vitro*, and (b) whether these effects are cell type specific or whether they are true across all regions where CELF6 is expressed.

Does loss of CELF6 impact neuronal function *in vivo*?

A number of target mRNAs I identified in CLIP are regulators of synaptic function. One of most enriched species was the gene coding for the fibroblast growth factor 13 (FGF13). FGF13 is expressed in the noradrenergic neurons of the locus coeruleus (LC). However, unlike canonical FGFs, FGF13 is not a growth factor, but rather regulates localization of voltage-gated sodium channels. Mutations in the *FGF13* gene are associated with

cognitive impairment and an epilepsy-like seizure disorder, and seizures present in mice with disruption of *Fgf13* (Puranam et al., Journal of Neuroscience 2015).

The cited Maloney et al. 2016 study analyzing patterns of CELF6 expression in the brain found that the LC had the strongest levels of expression across all populations. Based on its phenotype *in vitro*, I would hypothesize that FGF13 protein levels would be reduced in the LC when CELF6 is present and that we would observe elevated levels in the knockout. These findings can be demonstrated by microscopy study with anti-FGF13 antibodies. Additionally, we can determine whether the localization of sodium channels has changed and whether firing patterns of LC neurons are altered in *Celf6*^{-/-} knockout mice. The monoaminergic neurotransmitter populations send projections widely throughout the brain affecting many systems. It is intriguing to hypothesize that changes to neuronal function in the LC underlie the USV or other behavioral phenotypes observed in these animals.

What are the molecular and circuit contributions to variability in behavior?

Tangential to the CELF6 story was the observation that USV production in neonates was highly variable from animal to animal, and that animals displayed inconsistent patterns of behavior from day to day. This is unlike other milestones from this time period such as righting reflex and weight gain. One cannot predict that measuring the loudest animal on day X, or the animal of highest pitch, or the animal of largest amount of response on that day will be likely to exhibit the same position in the population's distribution on a subsequent day. Although numerous environmental signals converge to affect USV, all of which cannot always be controlled, I still find it remarkable that across hundreds of animals that are otherwise genetically identical and reared in otherwise identical circumstances, that such variability can be observed.

As adults, although some measures of vocalization appeared to stabilize, a number of features primarily in the frequency domain remained unpredictable from test day to test day. These findings have two implications: (a) large sample sizes are often necessary to detect effects on neonatal USV, and (b) there are unknown or unmeasured factors which contribute to individual variability. Some of these factors may be genes and molecules, some environmental, and some may be stochasticity in circuit activity. The way this information is integrated to produce USV is not well understood. Because neonatal mouse USV is a robust milestone of development, is relatively easy to acquire, and can be analyzed in an automated fashion, a forward genetic screen in mice could be performed to identify genetic contributions to variability in USV. Work by groups such as Cori Bargmann and Benjamin de Bivort have begun to uncover genes and circuits that control variability in *C. elegans* and *D. melanogaster* respectively, but to my knowledge no such work has been performed in mice. Disorders of the nervous system, such as ASD, often present with highly idiosyncratic features from case to case, so a better understanding of the cellular and molecular factors contributing to idiosyncrasy may help us better understand the etiology of these disorders.

Broader Impact and Significance

In my graduate work, I have been able to gain insight into the functions of an important regulatory molecule affecting numerous targets in the mouse brain. The targets of CELF6 have not been previously defined, so it is my hope that this is of great benefit to researchers of CELF proteins. Additionally, I have uncovered a hierarchy of activity across CELF3-6. This adds to the growing literature of functional cooperativity, competition, and antagonism exhibited by other families of RNA binding proteins.

In addition to my studies on CELF6 protein, my pipeline of analysis of vocalization

has been used to study the effects of a number of proteins on behavior, including the forkhead box transcription factors FOXP1 & FOXP2. This dissertation also provides the accompanying software documentation for this MATLAB package.

A dense web of molecules in cells regulates cellular function. It is my hope that the data I have presented help to spur more research in how RNA binding proteins regulate neuronal function. Ultimately, these important molecules contribute to the overall architecture of the nervous system, allowing animals to integrate sensory information and respond to stimuli through a complex repertoire of behavior.

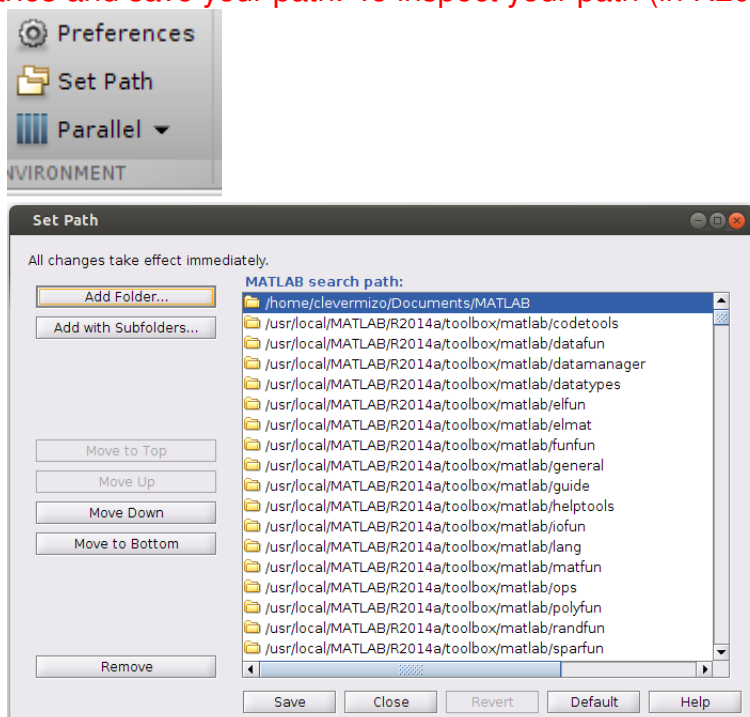
Appendices

Appendix 1

Using the MATLAB VocalizationFunctions Package

Installation & upgrades

Important note: It is very important that your default path in MATLAB does not contain any references to previous instances of the VocalizationFunctions package. Your package should be installed to your default MATLAB directory locally (usually “Documents/MATLAB”). We will not add these to the path (this will be done using the library() function). If there are any old entries from previous iterations make sure to remove these entries and save your path. To inspect your path (in R2013,R2014), click “Set Path”:



If MATLAB complains (usually this happens on Windows) that you cannot save your

path after making changes, restart MATLAB using Administrative privileges.

Installing for the first time

Downloading the Archives

The archives containing the package functions are stored on the genstorage server at the following address:

“//genstorage.wustl.edu/jdlab/Behavior/Behavior Assay Procedures and Protocols/Maternal Isolation Pup USV Recording”

There are two files you will need to copy:

- Rlike
- VocalizationFunctions-mm-dd-yyyy

Both are provided as either .zip or .tar.gz archives depending on your environment. Do not extract these on the server! Download them to your

Documents/MATLAB/

directory and then extract them there. Do this on your computer, not on an external disk, not anywhere else. To your Documents/MATLAB directory and nowhere else.

Installing the Rlike library

Unzip Rlike.tar.gz or Rlike.zip in the Documents/MATLAB directory. A number of files will appear:

- asvector.m
This function vectorizes a matrix. Useful for nested function calls.
- between.m
This function performs a logical test for values between a and b. Useful for readability.
- betwout.m
This function performs a logical test for values outside of a and b. Useful for readability.
- head.m
This function displays the first 10 rows of a dataset or other matrix-like object. head(X, n) displays the first n rows.

- `tail.m`
This function displays the last 10 rows of a dataset or other matrix-like object. `tail(X,n)` displays the last n rows.
- `nrow.m`
This function returns the number of rows. It is a call to `size(X,1)`.
- `ncol.m`
This function returns the number of columns. It is a call to `size(X,2)`.
- `library.m`
This function is executed as `library('LibraryName')`. *LibraryName* is the name of a folder under Documents/MATLAB. `library()` temporarily adds this folder and its subfolders to the MATLAB path during a session. This method avoids clashes between functions of the same name, and simplifies the path update procedure.

Because these are unzipped in the Documents/MATLAB directory they will always be available in your MATLAB path for use in any MATLAB session. If you unzip them elsewhere, this will not occur.

Installing the VocalizationFunctions library package

Unzip the `VocalizationFunctions-mm-dd-yyy.zip` or `VocalizationFunctions-mm-dd-yyyy.tar.gz` in the Documents/MATLAB directory. If you are using a graphical archive utility (e.g., WinZip, WinRAR, Mac Archive Manager), right-click and use an option for “Extract Here”. This will produce a top level `VocalizationFunctions` folder (which is also its library name), and three subfolders:

- `VocalizationFunctions`
 - Clean
This folder contains the functions for running a console-only version of the pipeline. The console-only version is recommended as it robustly checks files and data for inconsistencies and prevents crashes.
 - GUI
This folder contains functions for running the GUI frontend. The GUI version will no longer be updated after this version.
 - Misc
This folder contains many other functions that need to be catalogued, including Tim Holy’s legacy functions from Holy & Guo 2005.

This is it for installation. Before running any function it will be necessary to execute:

```
>> library('VocalizationFunctions')
```

from the MATLAB command prompt.

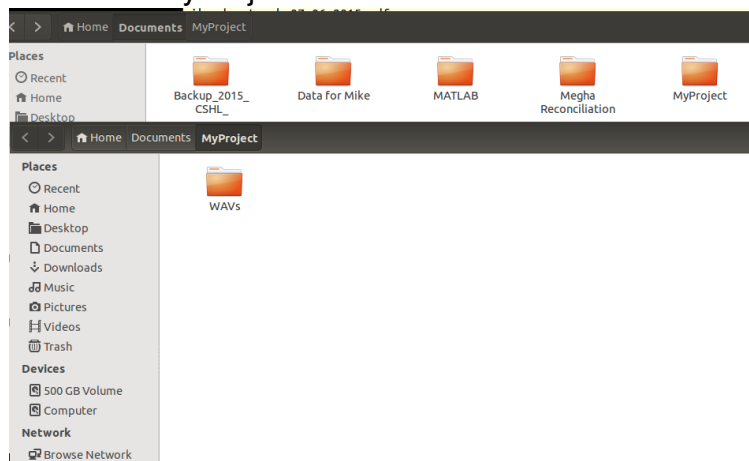
Upgrading to the newest version

Recent versions can contain fixes to bugs or cosmetic changes or both, and it is always a good idea to upgrade. To do so, download the most recent version of VocalizationFunctions-mm-dd-yyyy.tar.gz or .zip and unzip as in the previous section. It should overwrite existing functions in previous folders.

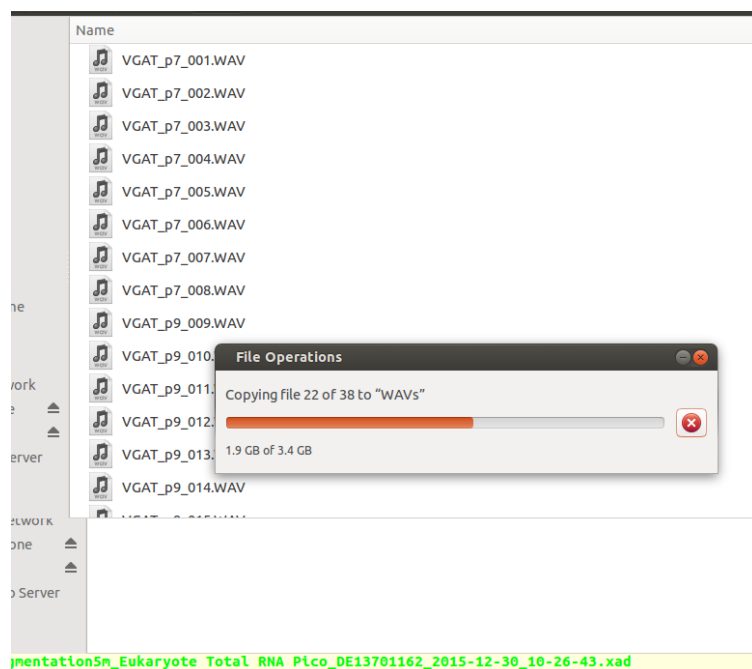
Running the GUI

Preparing your data

Create a destination folder for your project. In this example, I have created MyProject. Inside the MyProject folder I have created a folder calls WAVs.



Copy over the WAV files from their source to the WAVs folder:



Preparing a CSV for upload

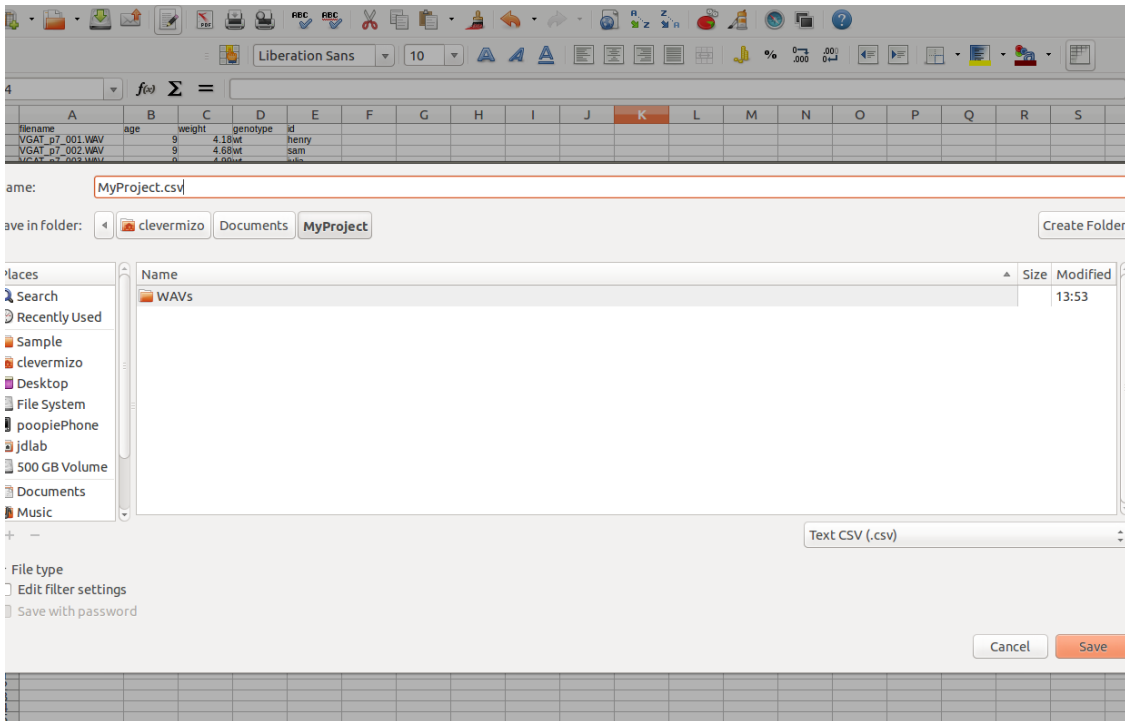
If you would like the output dataset to contain metadata (weight, sex, other variables), prepare a CSV file. The top line of your CSV should contain header names for each of the metadata variables you'd like to add. These names should not contain special characters: ~!@#\$\$%^&*()-+= and no spaces.

Under the header row you should make sure that the number of rows equals the number of files you want to process. It is **very important that the data in each row corresponds to the files in the order in which they appear in your WAVs folder**. Usually this is how the files are acquired in the first place. If you don't make sure this is the case, potentially the wrong files will be paired with the wrong data. (*This error is avoided in the console-only version.*) Make sure the CSV doesn't have empty cells. For Matlab use "NaN" as the "Not-A-Number" code for missing data where data are numeric. *NaN*, *Inf*, *-Inf* are the only non-numeric characters MATLAB will allow in a column otherwise filled with numbers.

Below shows an example CSV:

	A	B	C	D	E
1	filename	age	weight	genotype	id
2	VGAT_p7_001.WAV	9	4.18	wt	henry
3	VGAT_p7_002.WAV	9	4.68	wt	sam
4	VGAT_p7_003.WAV	9	4.99	wt	julia
5	VGAT_p7_004.WAV	9	3.47	ko	john
6	VGAT_p7_005.WAV	9	4.53	ko	jesus
7	VGAT_p7_006.WAV	9	4.85	ko	giacomo
8	VGAT_p7_007.WAV	9	4	ko	helen
9	VGAT_p7_008.WAV	9	4.94	wt	ida
10	VGAT_p9_009.WAV	9	3.99	wt	montana
11	VGAT_p9_010.WAV	9	3.07	wt	geoffrey
12	VGAT_p9_011.WAV	9	3.71	ko	cupcake
13	VGAT_p9_012.WAV	9	3.01	ko	beeswax
14	VGAT_p9_013.WAV	9	4.41	wt	misty
15	VGAT_p9_014.WAV	9	3.3	wt	sharon
16	VGAT_p9_015.WAV	9	4.8	wt	alyssa
17	VGAT_p9_016.WAV	9	3.54	wt	hamburger
18	VGAT_p9_017.WAV	9	3.17	ko	coffeecup
19	VGAT_p9_018.WAV	9	4.83	ko	horse
20	VGAT_p9_019.WAV	9	3.08	wt	cow
21	VGAT_p9_020.WAV	9	4.37	wt	chicken
22	VGAT_p9_021.WAV	9	3.34	ko	measles
23	VGAT_p9_022.WAV	9	4.76	ko	twinky
24	VGAT_p9_023.WAV	9	4.19	wt	calexica
25	VGAT_p9_024.WAV	9	4.61	wt	rhombus
26	VGAT_P9_025.WAV	9	3.51	ko	sasquatch
27	VGAT_P9_026.WAV	9	3.47	ko	ophelia
28	VGAT_P9_027.WAV	9	3.51	wt	hamlet
29	VGAT_P9_028.WAV	9	4.11	wt	juliet
30	VGAT_P9_029.WAV	9	4.62	ko	romeo
31	VGAT_P9_031.WAV	9	3.84	ko	william
32	VGAT_P9_032.WAV	9	3.81	wt	guillermo
33	VGAT_P9_033.WAV	9	4.96	wt	juan
34	VGAT_P9_035.WAV	9	3.13	wt	mercedes
35	VGAT_P9_036.WAV	9	3.09	wt	benzene
36	VGAT_P9_037.WAV	9	3.35	ko	circus
37	VGAT_P9_038.WAV	9	4.52	ko	trouble
38	VGAT_P9_039.WAV	9	4.4	wt	pleasant
39	VGAT_P9_040.WAV	9	3.18	wt	felicia
40					
41					
42					

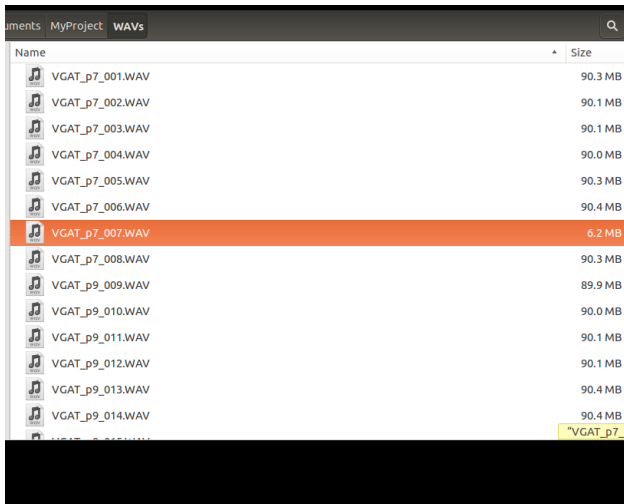
Save this CSV to your project folder:



NOTE: If you make a CSV in Windows, but do processing in Mac/Linux, or vice versa, errors can occur. This is because “under the hood” Windows and Unix use different “end-of-line” characters in text files. Make sure to save your CSV in the operating system environment you plan to do your processing in. This is as easy as opening up the Excel spreadsheet and saving to CSV.

Checking files for consistency

It is important that your files are of similar length. During processing a small chunk will be used for background calculations. If some of your files are abnormally short (such as an aborted recording) they will not have a chunk of adequate length and the processing pipeline will crash. You will note that VGAT_p7_007.WAV is 6.2 MB while most of the files are around ≈ 90 MB. This indicates you should remove this file (and its correspondingly line in the metadata CSV) before beginning:

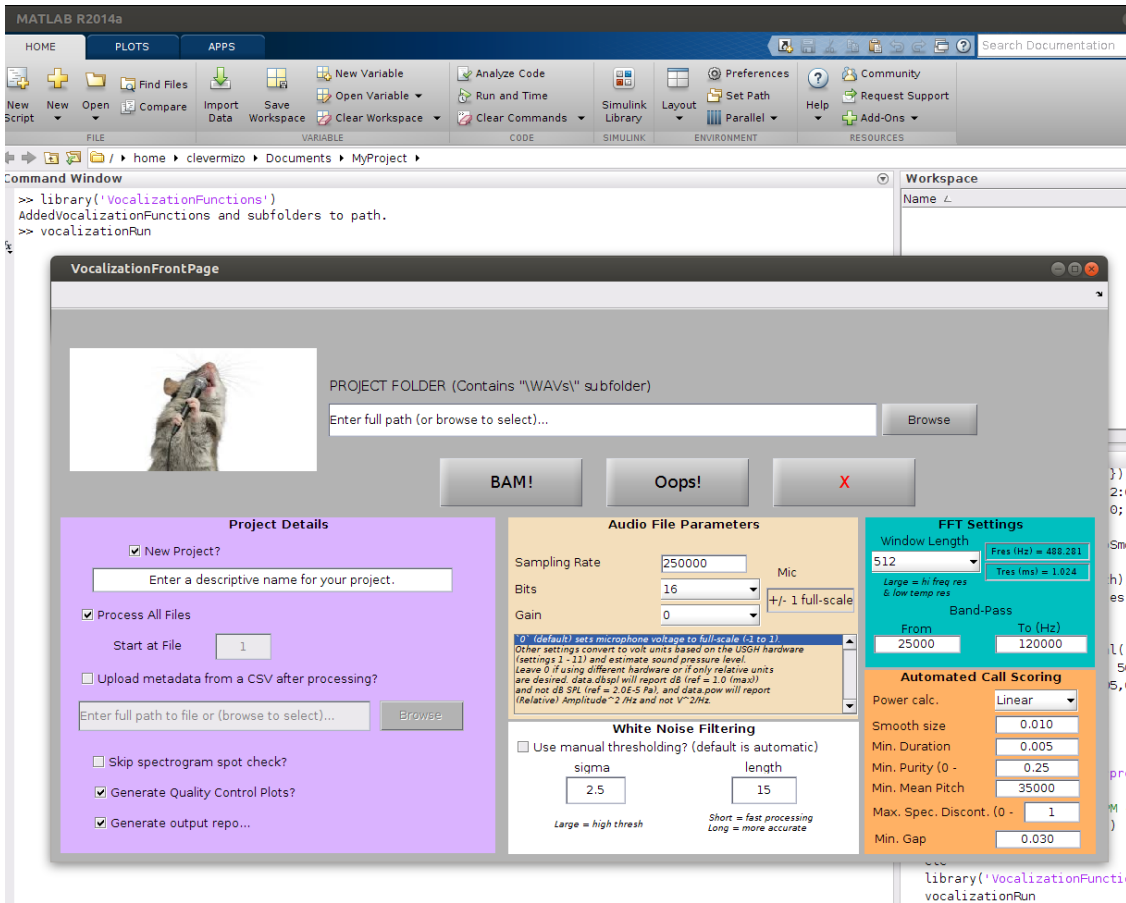


Starting the GUI

From the MATLAB command line, execute the following:

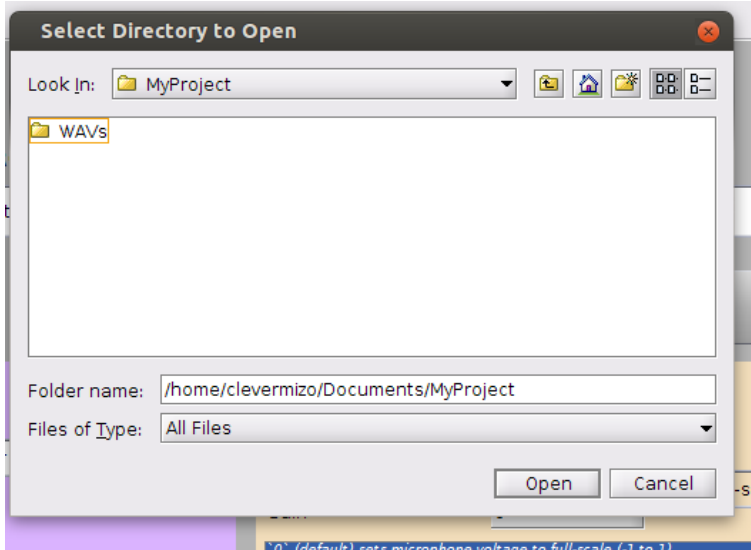
```
>> library('VocalizationFunctions')  
>> vocalizationRun
```

After typing each line, hit Enter. The first line loads the VocalizationFunctions package into the path.



Setting options for processing

From the GUI, under Project Folder, **Browse** to choose the project folder location.



Project details

Under Project details, type a name for your project. If this is a new project, leave Process All Files checked. If you have a CSV, click “Upload metadata from CSV...” and navigate and choose its location. The last three checkboxes are:

- Skip spectrogram spot check : If checked, you will not check the automatic scores of call start/end points in the spectrogram. If this is a new project, leave this unchecked. I like to verify about 10 % of files. Later it will ask how many files you want to check.
- Generate Quality Control Plots: Leave checked to generate QC variable histograms.
- Generate output report: Leave checked to generate output CSV with processed data.

Audio File Parameters

- Sampling Rate: Enter the sampling rate in kHz (default is 250000).
- Bits: Enter the bit depth. (default is 16 bits).
- Gain: Set the gain level. Default is 0. The default gain will return amplitude values in dB re 1.0. dB are a relative unit, where $1dB = 10 \cdot \log_{10} \left(\frac{amplitude^2}{referencevalue^2} \right)$. Each dB is $10 \log_{10}$ units relative to some reference power unit. “dB re 1.0” means the reference is the detector maximum. Since 1.0 is the detector max (read: “100%”), values are fraction, and dB are negative. “-20 dB re 1.0” , or -20 dB in common parlance means 20 dB units below the maximum, where the maximum is 0 dB.
If you choose the gain setting, voltage data will be reported in actual Volt units based upon estimated conversion factors from the manufacturer and dB data will be reported as dB SPL re $2 \mu Pa$, or dB SPL for short. Amplitude in Volts is converted to Pressure data in Pascals based upon manufacturer estimated microphone sensitivity for the Avisoft Bioacoustics CM16 microphone. These data are not true measurements of sound pressure because over time the microphone’s plate is differently sensitivity than initially. To know the true sound pressure, the microphone must be calibrated (using, surprise, surprise, a calibrated microphone! which in turn must be calibrated with another calibrated microphone...).
Unless you have strong feelings about what units to report, report the default (leave Gain set to 0) and no conversion will be performed.

White Noise Filtering

- Leave “Use manual thresholding?” unchecked. If checked, white noise levels must be approximated by eye. The pipeline will present a graph for each file of a (default) 15 second chunk of audio (frequencies on the x axis and \log_{10} unscaled FFT magnitude on the y-axis. A magic red line will be drawn based upon an automated

algorithm and the user can enter and select to change the position of this line. such that it approximates the noise floor level.

- sigma: The sensitivity of the automated noise threshold (default = 2.5). You can turn this knob up and down. The min value is 1. Larger numbers are more sensitive. If you feel that spectrograms are not being correctly scored, increase or decrease as desired.
- length: Default is 15 seconds. Processing the entire file for noise will be very laborious, so only a small chunk is chosen.

FFT Settings

- Window length: The fast Fourier transform is computed in 50% overlapping windows. The % overlap is not tweakable at this time. Changing the window length changes the temporal and spectral resolution which is shown in real time in the box at right. Big windows make for better spectral resolution, but poorer temporal resolution, and vice versa. 512 is the default. Each time point in the spectrogram will be 1.024 msec with a bin size of 512 samples.
- Band-pass filter: From (25000 Hz default) To (120000 Hz default). The high frequency spectrum is saved for mice, with non-white noise, but low frequency audio filtered out.

Automated Call Scoring

These are the parameters used to improve the automated scoring of USV call start and stop times.

- Power calc: Linear vs. Log. If your signal is very weak, log-scale scoring might improve signal detection accuracy. Linear is default.
- Smooth size: 0.01 seconds default. Frequency data is smoothed over 10 milliseconds. Helps smooth out “empty” spots where signal may drop below detection.
- Min. duration: 0.005 seconds default. The minimum duration for a USV call (for the so-called “whistle” calls).
- Min. purity: 0.25 default (**in practice I think 0.15 works better**). This says at any moment in time, within a call, some % (the “spectral purity”) must be concentrated in the frequency where the maximum occurs. This means you have a nice tall peak somewhere in the spectrum, and the rest of the audio can be alternatively distributed. Small numbers are “dirtier” more peaky spectra, and bigger number are “clean” spectra. A purity of 1.0 would be the Dirac delta function in the spectrum (100% of the power at the max, 0 elsewhere). I find pup calls are sometimes a lil dirty, so I use 0.15.

- Min mean pitch: Wherever the frequency of maximum power (“pitch”) is , it should be at least 35 kHz or greater (default).
- Max spec. discontinuity: Consecutive time bins within a call should have their peaks in mostly the same place. Values range from 0 (identical spectra from t to t+1) to 2 (spectra with peaks in opposite place from t to t+1). A value of 1 indicates consecutive spectra have a shape that is 50% the same. Gradual evolution over time suggest spectral discontinuity is on the range of 0:1. Discontinuities between 1:2 are very abrupt changes in the spectra. By the way, these occur whenever a pitch jump occurs, but this is taken care of by the next paramter.
- Min gap: 0.03 seconds default. Any areas between two calls where the signal is : (a) too dirty, (b) too discontinuous, (c) below detection, or would otherwise be scored as “not a call”, if these areas are 30 milliseconds are less, the two disjointed calls are merged and scored as a single call. This also allows the discontinuities evident in calls with pitch jumps to be scored as a single call despite violating the spectral discontinuity threshold.

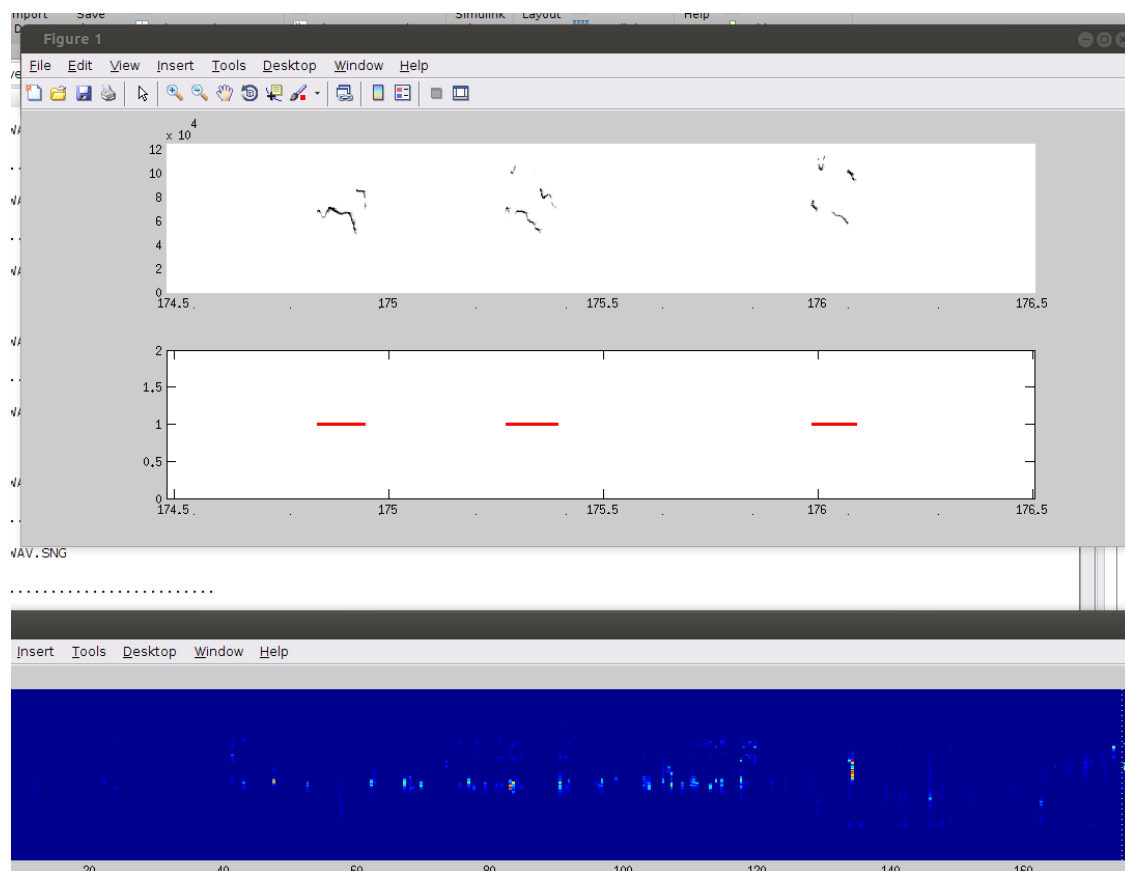
If you don't know what you're doing, I would change the spectral purity to 0.15 and leave the rest alone.

This is my filled out GUI:

The screenshot shows a software GUI with the following sections:

- Project Details (Purple Panel):**
 - New Project?
 - Project Name: myProject
 - Process All Files
 - Start at File: 1
 - Upload metadata from a CSV after processing?
 - Metadata File: /home/clevermizo/Documents/MyProject/MyProje
 - Skip spectrogram spot check?
 - Generate Quality Control Plots?
 - Generate output repo...
- Audio File Parameters (Yellow Panel):**
 - Sampling Rate: 250000
 - Bits: 16
 - Gain: 0
 - Mic: +/- 1 full-scale
- FFT Settings (Teal Panel):**
 - Window Length: 512
 - Freq (Hz) = 488.281
 - Tres (ms) = 1.024
 - Band-Pass: From 25000 To 120000
- Automated Call Scoring (Orange Panel):**
 - Power calc.: Linear
 - Smooth size: 0.010
 - Min. Duration: 0.005
 - Min. Purity (0 -): 0.15
 - Min. Mean Pitch: 35000
 - Max. Spec. Discont. (0 -): 1
 - Min. Gap: 0.030
- White Noise Filtering (White Panel):**
 - Use manual thresholding? (default is automatic)
 - sigma: 2.5
 - length: 15
 - Large = high thresh
 - Short = fast processing
 - Long = more accurate

Spectrogram Spot-Check



In the spectrogram view, the upper plot is the gray-scale spectrogram, and the lower plot shows the locations of calls as scored automatically from start time to stop time. The spectrogram can be navigated by selecting and using the arrow keys or arrow buttons in the lower “slider window”. Your goal is to make sure the automated scores overlap with calls you would score manually. If they don’t you may have restart and tweak some settings.

On subsequent runs if you think that the utilized settings do a good job, then you can check the box in the main GUI to skip this step.

When everything is complete, you should return the following at the MATLAB console:

```
VGAT_p9_024.WAV
Output file: VGAT_p9_024.WAV.SNG
.....
Importing metadata.....
Done importing metadata
Pipeline completed successfully.
x >>
```

Inspecting output reports

The first report that you will want to inspect is the <project name>_qcgraphs.pdf file. These are histograms which give you a sense of how different USV features are distributed.

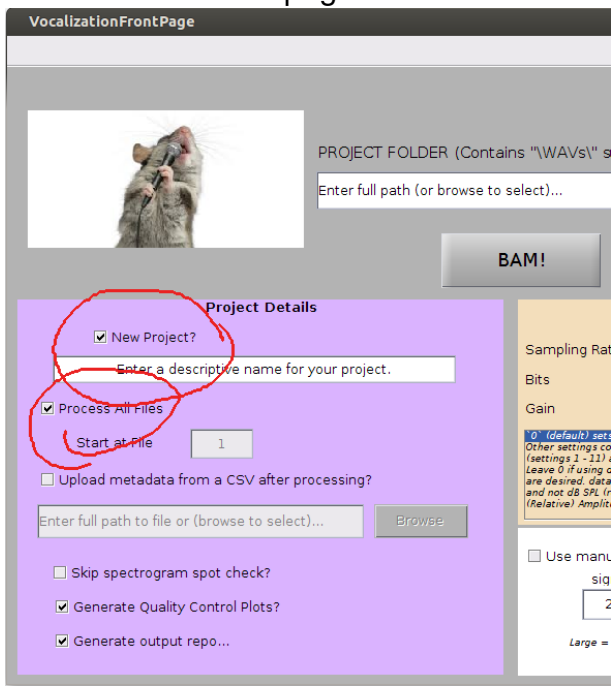
- **Noise Mean & Threshold:** These are actually a bit redundant, but they both are the estimates of the background white noise level in the spectrum, in the original FFT units and then rescaled to the units which the spectrograms are stored in (Voltage in the case that Gain>0 is used, or full scale where Gain=0). The threshold represents the magnitude level in the spectrum. FFT magnitudes below this value are set to 0 in the spectrum. The noise level is set on a per-file basis. You should be concerned however if the distribution appears bimodal - i.e., that some files appear to have a different noise level than other files. This is likely due to changes in the recording environment (such as switching boxes or rooms), or changes to the gain level on the the microphone's pre-amplifier. This may make results non-comparable as a change in the noise level or the sensitivity of the recording equipment will likely alter the number of calls detected.
- **Fraction of calls peaking:** The x-axis is file number, and the y-axis is the number of calls exceeding the maximum of the detector limit somewhere within the call. Peaking may not effect call counts, but will likely skew the computation of other features of USV such as features based on the pitch or amplitude envelope of the call.
- **Other features of USV:** Unlike the preceding two distributions, the other distributions are mostly informational. They inform you what each of these USV features looks like:
 - Call number: # of calls in each recording.
 - Call duration: Pooled duration times (milliseconds) over all calls.
 - Call duration averages: Average duration for each recording.
 - Pauses (<0.5 sec): Inter-call pause times. Pauses > 0.5 seconds are considered "silence" in the recording and not a pause before the next call.
 - Average Pause times per file.
 - Bout #: Number of bouts per recording, where a bout is a sequences of calls separated by pauses < 0.5 sec.
 - Fraction of bouts with only one call: # of bouts of length = 1.
 - Correlation between duration & subsequent pause length. This is usually around 0 but may be slightly negative.
 - Correlation between adjacent duration times: This is usually positive.
 - Correlation between adjacent pause times: This is also usually positive.

- Mean Pitch: The average pitch of each call, pooling all calls.
- Mean Pitch average: The average Mean Pitch for each file.
- Pitch Range: Pitch range observed pooled over all calls.
- Pitch Range file averages.
- Pitch slope: Hz/millisecond slopes in calls, usually center on 0 (flat calls).
- Pitch slope file averages.
- Spectral purity: “Noisiness” of the calls, with a minimum as defined by the processing parameters (e.g., 0.15).
- Spectral purity file averages.
- dB: Peak amplitude level pooled across all calls.
- dB file averages.
- Fraction with jumps (>10 kHz).

Adding new data to an existing project

This capability exists but is currently untested so use at your own risk. It should work. As of this issue, the GUI is deprecated, so I will not be supporting future updates to this feature.

From the GUI main page:



- Uncheck the box that says New Project. The GUI will check for the presence of a Vars subfolder containing initialization parameters and a “data.mat” file. If these don’t exist, it will tell you that it does not think there is actually an existing project.

- Uncheck Process All Files and change the Start at File to be the number of the next file to run, in numerical order from their listing in the WAVs folder, or in row order from your metadata CSV.

Appendix 2

Generalized RNA-Sequencing Library Preparation with Unique Molecular Identifiers

This protocol is written presuming a strip tube format. Make sure to have a Permagen Labware strip tube magnet, a 1.7 mL tube magnet as well to prepare beads, strip tube mini-centrifuge, and 10 μ L 8-channel multichannel pipettor, and 300 μ L 8-channel multichannel pipettor.

For samples deriving from CLIP or other pre-fragmented techniques, proceed to section "**Dephosphorylation Reaction**".

This protocol is modified from the protocol in:

Van Nostrand, Eric L., et al. "Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP)." *Nature methods* 13.6 (2016): 508.

Specific modifications include rRNA depletion & fragmentation for application to total RNA-Seq, as well as generalization of the RNA adapters to use a single adapter across samples for increased cost-effectiveness, using Illumina Index (Read #3) sequences for sample multiplexing.

rRNA Probe Annealing

Materials: NEBNext rRNA depletion kit (E6310)

RNA should be 10ng-1 μ g in a 12 μ L volume

The NEBNext rRNA depletion kit functions by hybridizing a proprietary mix of DNA probes complementary to rRNA. RNase H then degrades RNA:DNA hybrids, leaving unhybridized RNA intact. Afterwards, DNA probes are degraded by DNase I treatment. (Because the RNA is DNase treated at this stage, I

do not bother performing additional DNase I treatment & cleanup after RNA isolation.)

1. Per reaction assemble:
 - 1 μL NEBNext rRNA depletion solution
 - 2 μL Probe hybridization buffer
 - 12 μL RNA Sample
2. Vortex and spin down.
3. Thermal cycling program (with lid at 105 °C) is as follows:
 - 95 °C 2 min
 - 95-22 °C 0.1 °C/sec
 - 22 °C 5 min hold
 - * On our Biorad Thermal Cyclers, we have to set the following for second stage:
 - 731 total cycles, 1 sec
 - 95C start, -0.1deg/cycle
4. Spin and place on ice.

RNase H Digestion

1. Make a master mix. Per reaction add:
 - 2 μL NEBNext RNase H
 - 2 μL RNase H Reaction Buffer
 - 1 μL H₂O
2. Add 5 μL to 15 μL rRNA annealing reaction (20 μL final volume)
3. Vortex, spin, and heat to 37 °C for 30 minutes with lid at 40 °C

DNase I Digestion

1. Make a master mix. Per reaction add:
 - 5 μL DNase I Reaction Buffer
 - 2.5 μL DNase I (RNase-free)
 - 22.5 μL H₂O
2. Add 30 μL to 20 μL reaction (50 μL final volume)
3. Vortex, spin, and heat to 37 °C for 30 minutes with lid at 40 °C

Clean up with MyONE Silane Beads

Materials: MyONE Silane Beads (Thermo Scientific 37002D)

Buffer RLT (Qiagen, any RNA kit or Product # 79216)

5M NaCl, 100% EtOH, 75% EtOH, 80% EtOH (later steps in this protocol)

Strip tube magnet (Permagen Labware 0.2 mL PCR Strip Magnetic Separator)

1.7 mL tube magnet

MyONE Silane and **Agencourt RNAClean** are not the same technology. *MyONE Silane can purify small fragments given the right proportion of EtOH, Agencourt RNAClean bottoms out at 100 nt.*

1. Separate 20 μL of MyONE Silane beads per sample on magnet and remove storage buffer. (For 10 samples, separate 200 μL , etc. Use 1.7 mL tube for batch preparation.)
2. Wash beads in batch with 900 μL of Qiagen Buffer RLT.
3. Resuspend beads in 150 μL /sample of Qiagen Buffer RLT (3 starting sample volumes) and 5 μL per sample 5M NaCl. (For large numbers of samples, you may need to use a 15 mL conical. For 10 samples of 50 μL each, resuspend beads in 1500 μL Buffer RLT + 50 μL 5M NaCl.)
4. Split rRNA depletion reaction to two sets of strip tubes (50 μL \rightarrow 25 & 25). This is to ensure strips can accommodate total volume (you can also use 1.7 mL tubes, but the strip tube magnet is more convenient).
5. To each sample add. 77.5 μL Beads, RLT, NaCl with multichannel. Mix by pipetting up and down 10 times.
6. Add 154 μL 100% EtOH (1.5 mix volumes) to each strip tube with multichannel.
7. Mix by pipetting up and down and rotate samples at room temp for 15 minutes.
8. Separate on magnet for 30 seconds and remove supernatant.
9. Wash beads with 0.2 mL 75% EtOH. Pipette to fully resuspend and move to new strip. At this step, combine strips that were split in step (4).
10. Separate on magnet for 30 seconds. Remove wash with multichannel.
11. Wash 2 more times with 75% EtOH. Add wash buffer and let sit for 30 seconds on magnet and remove with multichannel. No need to resuspend.
12. Dry 5 minutes on magnet. Remove excess EtOH with vacuum or by pipette which may collect at bottom.
13. Resuspend in 10 μL of H₂O and let sit for 5 minutes off magnet. Then to clean up put back on magnet, separate, and move eluates to new strip tubes.

Optional: Resuspend 10 μ L and assess a small amount by Agilent TapeStation or Agilent Bioanalyzer to confirm loss of small (18S) and large (28S) rRNA peaks.

Note: Contamination with MyONE Silane beads does not appear to inhibit any downstream steps, so don't worry about a small amount of magnetic beads coming along. At the very end of library preparation, however, you do want to ensure libraries are bead-free before pooling for sequencer.

Fragmentation

If doing CLIP or another prep where RNA samples are already fragmented, use mix components in step (1) and skip to dephosphorylation reaction.

1. Per reaction assemble:
 - 2.5 μ L 10X PNK Buffer (Enzymatics Y9040)
 - 19.0 μ L rRNA-depleted sample RNA
2. Thermal cycler 94 °C (lid at 105 °C) for 5-15 minutes
3. Move to ice.

Dephosphorylation Reaction

RNase I digestion in CLIP, and heat based fragmentation for total RNA-Seq both leave 3' phosphates and 5' -hydroxyl. 3' phosphates must be removed before adapter ligation. Triton-X 100 is added to 1% based upon personal communication with NEB Tech Support that this improves dephosphorylase activity of T4 PNK to 90%.

1. Make a master mix. Per reaction add:
 - 2.5 μ L 10% Triton-X 100
 - 0.5 μ L T4 PNK
 - 0.5 μ L rRNasin
2. Add 3.5 μ L master mix to each sample (final volume 25 μ L).
3. Vortex briefly and spin down.
4. Heat to 37 °C for 30 minutes with lid at 40 °C

Clean up with MyONE Silane Beads (Abbreviated, see above for full protocol)

1. 20 μ L of MyONE Silane beads per sample.
2. Wash beads in batch with 900 μ L RLT.
3. Resuspend beads in 75 μ L/sample of Qiagen Buffer RLT (3 volumes, last step is 25 μ L) and 2.5 μ L/sample 5M NaCl.
4. Add 77.5 μ L NaCl/RLT/beads to samples.
5. Add 154 μ L EtOH (1.5 mix volumes) and mix.
6. Mix and rotate samples at room temp for 15 minutes.
7. Separate on magnet for 30 seconds and remove supernatant.
8. Wash with 0.2 mL 75% EtOH and move to new strip.
9. Wash 2 more times with 0.2 mL 75% EtOH.
10. Dry 5 minutes on magnet.
11. Remove residual EtOH.
12. Resuspend in 10.5 μ L of H₂O and let sit off magnet for 5 minutes.
13. Separate on magnet and move 10 μ L eluate to strip tubes containing A01m adapter. (See next section).

Optional: Resuspend in larger volume and assess a small amount by Agilent TapeStation to confirm size shift as a result of fragmentation. Elution volume is slightly larger than 10 μ L to ensure you can move volume safely to next set of tubes.

A01m Ligation

A01m Adapter:

/5Phos/rArGrArUrCrGrGrArArGrArGrCrGrUrCrGrUrGrUrArG/3SpC3/

(IDT: Purify at 250 nmol RNA Oligo scale using RNase-Free HPLC Purification, store at 200 μ M in H₂O, aliquoted, at -80° C. SpC3 is a molecule which is a carbon chain that can be added to oligos. It blocks any ligation at its own end, since there is no 3' OH group, thus adapter chains are not possible. Heat fragmentation (or digestion with RNase I) leaves 5' OH groups, thus chains of RNA fragments are also not possible.)

RNA Ligase High Concentration (NEB M0437)

1. Strip tube should contain 10.5 μL total: 10 μL dephosphorylated RNA fragments & 0.5 μL 40 μM A01m.
2. Heat to 65 $^{\circ}\text{C}$ for 2 minutes with lid at 105 $^{\circ}\text{C}$ to melt any secondary structure.
3. Place immediately on ice for 1 minute.
4. Make a master mix. Per reaction add:
 - 1.5 μL DMSO (100%)
 - 2.0 μL RNA Ligase Buffer (10x (Enzymatics buffer contains ATP))
 - 0.5 μL Promega rRNasin (40 U / mL)
5. Add 4.0 μL mix to each sample.
6. Vortex briefly and spin down.
7. Per reaction add 4 μL PEG8K (50%). (Cut pipette tip for easier pipetting.)
8. Vortex briefly and spin down.
9. Add 1.5 μL T4 RNA Ligase (Enzymatics L6050, 20 U/ μL).
10. Vortex briefly and spin down.
11. Tape down horizontally into a container and place on a shaker at 250 rpm for 2 hours at room temp.

Clean up with MyONE Silane Beads (Abbreviated, see above for full protocol)

NaCl is not added as there is NaCl in the RNA Ligase buffer. According to Eric Van Nostrand, the EtOH percentage is changed to favor larger fragments and not unligated adapter. My own experiments with MyONE Silane are somewhat inconclusive as to whether that matters.

1. Separate 10 μL of MyONE Silane beads per sample.
2. Wash beads in batch with 900 μL of Qiagen Buffer RLT.
3. Resuspend beads in 60 μL /sample of Qiagen Buffer RLT.
4. Add 60 μL beads/RLT to each sample and mix.
5. Add 52.5 μL EtOH (0.75 mix volumes) and mix.
6. Rotate samples at room temp for 15 minutes.

7. Separate on magnet for 30 seconds and remove supernatant.
8. Wash with 0.2 mL 75% EtOH. Resuspend and move to new tube.
9. Separate on magnet for 30 seconds. Wash 2 more times with 75% EtOH with re-suspending (30 seconds on magnet).
10. Dry 5 minutes on magnet.
11. Remove residual EtOH.
12. Resuspend in 7.5 μ L of H₂O and let sit for 5 minutes. Separate on magnet and move 7 μ L to a new set of strip tubes containing 1.5 μ L of 20 μ M AR17 primer.

Reverse Transcription

AR17 primer: ACACGACGCTCTTCCGA

Order as standard primer. Store in H₂O at 200 μ M at -20°C.

Working dilution is 20 μ M.

Thermo Superscript RT III First Strand Synthesis system (Thermo 18080051)

1. Strip tubes should contain 1.5 μ L AR17 (20 μ M) and 7 μ L A01m-ligated RNA
2. Heat to 65°C for 2 minutes with lid at 105°C
3. Place on ice for 1 minute.
4. Make a master mix. Per reaction add (total 11.5 μ L mix):
 - 2.0 μ L SSRTIII 10x Buffer
 - 2.0 μ L dNTPs (10 mM)
 - 4.0 μ L MgCl₂ (25 mM)
 - 2.0 μ L DTT (100 mM)
 - 0.6 μ L RnaseOUT
 - 0.9 μ L SSRTIII Enzyme

(Old kits fail!)
5. Add 11.5 μ L master mix to each 8.5 μ L sample (f.v. 20 μ L).
6. Vortex briefly and spin down.
7. Heat to 50°C for 45 minutes.

ExoSAP-It degrades primers and dNTPs, and thus only true RNA:cDNA hybrids remain intact.
8. Per reaction add 3.5 μ L ExoSAP-It.

9. Vortex briefly and spin down.
10. Heat to 37°C for 15 minutes.
11. Per reaction add 1 µL EDTA (0.5M). Vortex briefly and spin down.
12. Per reaction add 3 µL 1M NaOH. Vortex briefly and spin down.
13. Heat to 70°C for 12 minutes in thermal cycler to degrade RNA.
14. Per reaction add 3 µL 1M HCl to neutralize pH. (Final volume is 30.5 µL.)

Clean up with MyONE Silane Beads (Abbreviated, see above for full protocol)

1. Separate 10 µL of MyONE Silane beads per sample.
2. Wash beads in batch with 900 µL of Qiagen Buffer RLT.
3. Resuspend beads in 91.5 µL/sample of Qiagen Buffer RLT (3 starting sample volumes).
4. Add 91.5 µL Beads in Buffer RLT to each sample and mix.
5. 111 µL EtOH (0.91 mix volumes) and mix.
6. Rotate samples at room temp for 15 minutes.
7. Separate on magnet for 30 seconds and remove supernatant.
8. Wash with 0.2 mL 80% EtOH. Resuspend and move to new tube.
9. Separate on magnet for 30 seconds. Wash 2 more times with 80% EtOH with re-suspending (30 seconds on magnet).
10. Dry 5 minutes on magnet.
11. Remove residual EtOH.
12. Resuspend in 11 µL of H₂O and let sit for 5 minutes. Separate on magnet and move 10.5 µL to a new set of strip tubes containing 0.5 µL of 80 µM Rand103tr3 adapter.

Rand103tr3 Ligation

NEB has two protocols using T4 RNA Ligase High Concentration. The ligation protocol for a ssRNA oligo to an RNA molecule has a 2 hour incubation, but the protocol for ligating to DNA says to proceed overnight. T4 RNA Ligase may be less efficient with ssDNA than it is with RNA.

Rand103tr3 Adapter:

/5Phos/NNNNNNNNNAGATCGGAAGAGCACACGTCTG/3SpC3/

Purify at 100 nmol scale using PAGE Purification, and when asked for random Ns, I used "Machine Mixing" option (there are two options, hand mixing and machine mixing). Hand mixing might be better to ensure equimolar probabilities of random incorporation.

Store in H₂O at 200 μ M in -20°C. Working dilution is 80 μ M.

1. Strip tubes should contain 0.5 μ L Rand103tr3 (80 μ M) and 10.5 μ L cDNA
2. Heat to 65°C for 2 minutes with lid at 105°C.
3. Place on ice for 1 minute
4. Per reaction add (total 5.5 μ L mix):
 - 1.5 μ L DMSO (100%)
 - 2.0 μ L RNA Ligase Buffer (10x)
5. Add 3.5 μ L mix to each sample.
6. Vortex briefly and spin down.
7. Per reaction add 4 μ L PEG8K (50%). (Cut pipette tip for easier pipetting.)
8. Vortex briefly and spin down.
9. Add 1.5 μ L T4 RNA Ligase High Conc (20 U/ μ L).
10. Vortex briefly and spin down.
11. Tape down horizontally into a container and place on a shaker at 250 rpm overnight.

Note: Because MyONE Silane allows for purification of small things, some cDNA generated from free A01m adapter can make it through to this step, thus A01m:Rand103tr3 dimers are possible. The final PCR product of this adapter is 139 nt. This can be removed by size selection after PCR, and you can run a negative control (no starting RNA) through the protocol to verify the adapter. If you start with a high concentration of RNA in the protocol, it seems that very little of this gets made, but if the amount of RNA is limiting, it becomes more prevalent.

Clean up with MyONE Silane Beads

Unclear to me why the eCLIP protocol switches back to washes in 75% EtOH.

1. Separate 10 μ L of MyONE Silane beads per sample.
2. Wash beads in batch with 900 μ L of Qiagen Buffer RLT.
3. Resuspend beads in 60 μ L/sample of Qiagen Buffer RLT (3 starting sample volumes).
4. Add 60 μ L Beads in Buffer RLT to each sample and mix.
5. Add 60 μ L EtOH (0.75 mix volumes) and mix.
6. Rotate samples at room temp for 15 minutes.
7. Separate on magnet for 30 seconds and remove supernatant.
8. Wash with 0.2 mL 75% EtOH. Resuspend and move to new tube.
9. Separate on magnet for 30 seconds. Wash 2 more times with 75% EtOH with re-suspending (30 seconds on magnet).
10. Dry 5 minutes on magnet.
11. Remove residual EtOH.
12. Resuspend in 10 μ L of H₂O and let sit for 5 minutes.

Trial Library PCR

NEBNext Q5 Ultra II Q5 Master Mix (NEB M0544)

Universal Primer

AATGATACGGCGACCACCGAGATCTACTCTTTCCCTACACGACGCTCTTCCGATC*T

(The * is a phosphorothioate bond, which helps preserve the primer and prevent degradation. Make up at 100 μ M and use 10 μ M in reaction.)

Index Primers

Can be anything containing the Illumina Read2 priming site with an index that is also compatible with the adapter ligated template. Here is an SIC index primer (index in lower case)

CAAGCAGAAGACGGCATAACGAGATtccgtattaGTGACTGGAGTTCAGACGTGTGCTCTTCCGA

1. Make a master mix. Per sample:
 - 10 μ L Q5 Ultra Master Mix
 - 1 μ L 10 μ M NEBNext Universal primer
 - 7 μ L H₂O

2. In individual tubes, combine:
 - 1 μ L 10 μ M Index primer
 - 1 μ L adapter ligated cDNA
3. Add 18 μ L master mix to index primer/cDNA.
4. Run using Q5 Ultra PCR program
 - 98°C 30 sec
 - (98°C 10 sec
 - 55°C 15 sec
 - 65°C 60 sec) Repeat for N cycles and pull samples at desired cycle numbers and keep on ice. 65°C 5 min
 - 12°C Hold

Run 2.5% agarose gel/1X TBE. Determine cycle showing robust amplification of library and not high MW overamplification bands.

You don't want to run out of sample. Potentially use a test sample for this and you should be able to guess a couple different cycle numbers between 10 - 20. You can fine tune from there. Or another way to make a test sample is pull 1 μ L from all your libraries, and pool them together. Or, pull a small amount and dilute it in H₂O, then back-calculate using the base-2 logarithm of the dilution factor how many cycles to use for preparative PCR.

For 1 μ g of input RNA for total RNA-Seq, I found 10 cycles to be sufficient.

Preparative Library PCR

Cycle number to use should reflect proportional amount of cDNA input. Suppose I used 1 μ L in a test reaction in the previous step. If I use 8 μ L here, that is 8 fold more starting material, which is 3 base-2 logarithm units. If I determined 15 cycles in the previous step, then use $15-3 = 12$ cycles in the preparative PCR.

1. Make a master mix. Per sample:
 - 25 μ L Q5 Ultra Master Mix
 - 2.5 μ L 10 μ M NEBNext Universal primer
 - 12 μ L H₂O
2. In individual tubes, combine:
 - 2.5 μ L 10 μ M Index primer
 - 8 μ L adapter ligated cDNA
3. Add 39.5 μ L master mix to each sample.
4. Run using Q5 Ultra PCR program
 - 98°C 30 sec

(98°C 10 sec
55°C 15 sec
65°C 60 sec) x desired cycles
65°C 5 min
12°C Hold

SPRI Purification

Purification is a size selection step using altered polyethylene glycol concentration with Beckman Coulter AMPure XP beads (Beckman Coulter Product # A63881). This selects for things between 200 - 400 bp (I have empirically optimized this using a DNA ladder.) You add the beads directly from a well mixed container without washing them first. It works using the PEG in the bead storage buffer. This is preferable to gel purification because gel purification results in heavy loss of yield compared to magnetic bead based purification. It is also preferable to electroelution after gel purification because electroelution does not scale well to large numbers of samples.

Size selection is important because, especially for small amounts of starting material, the A01m:Rand103tr3 adapter dimer is a prevalent species, and will soak up a lot of reads.

1. Bring volume to 100 μ L with 50 μ L H₂O.
2. Add 80 μ L of AMPure XP. Mix 10 times by pipetting and incubate 5 minutes.
3. Separate on magnet.
4. Move *supernatant* to new tube.
5. Add 40 μ L of AMPure XP. Mix 10 times by pipetting and incubate 5 minutes.
6. Separate on magnet 2-3 minutes.
7. Wash 2x30s 80% EtOH. It is really important with SPRI beads not to disturb them on the magnet. Just add and let sit for thirty seconds. Remove with multichannel.
8. Dry 5 min.
9. Remove any residual ethanol which collects.
10. Elute in 10 μ L of 10 mM Tris pH 7.8.
11. Assay by TapeStation or Bioanalyzer.

Appendix 3

Translating Ribosome Affinity Purification - Optimized

This protocol is based on previous TRAP protocols as cited elsewhere in this manuscript. See:

1. Doyle, Joseph P., Joseph D. Dougherty, Myriam Heiman, Eric F. Schmidt, Tanya R. Stevens, Guojun Ma, Sujata Bupp et al. "Application of a translational profiling approach for the comparative analysis of CNS cell types." *Cell* 135, no. 4 (2008): 749-762.
2. Heiman, Myriam, Anne Schaefer, Shioaching Gong, Jayms D. Peterson, Michelle Day, Keri E. Ramsey, Mayte Suárez-Fariñas et al. "A translational profiling approach for the molecular characterization of CNS cell types." *Cell* 135, no. 4 (2008): 738-748.
3. Dougherty, Joseph D., Susan E. Maloney, David F. Wozniak, Michael A. Rieger, Lisa Sonnenblick, Giovanni Coppola, Nathaniel G. Mahieu et al. "The disruption of *Celf6*, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors." *Journal of Neuroscience* 33, no. 7 (2013): 2732-2753.

Protocol

Bead Prep

1. Add **60 μL x N IPs** of Streptavidin MyOne T1 **or 120 μL x N IPs** of Streptavidin M-280 beads to a tube. Put the tube on the magnet stand and give it a minute to separate. Remove storage buffer (0.1%BSA/1XPBS)
2. Prepare a binding mixture containing:
17 μL of 1 $\frac{\mu\text{g}}{\mu\text{L}}$ Protein L (17 μg) **x N IPs**
20 μL of 1.78 $\frac{\mu\text{g}}{\mu\text{L}}$ anti-EGFP clone 19F7 (36 μg) **x N IPs**

36 μL of 1 $\frac{\mu\text{g}}{\mu\text{L}}$ anti-EGFP clone 19F7 (36 μg) x **N IPs**

127 μL of 1X PBS (f.v. 200 μL) x **N IPs**

3. Resuspend beads in the antibody/Protein L mixture by pipetting.
4. Incubate beads and antibody/Protein L mixture at room temperature for at least 1 hour with end-over-end rotation (or up to overnight at 4C with end-over-end rotation).
5. Put beads on magnet stand and give it a minute to separate.
6. Discard supernatant.
7. Resuspend beads in 1 mL 1XPBS/0.1% BSA. Give it a minute in suspension and a minute on the stand to wash.
8. Repeat step 7 4 times (total of 5 washes).
9. Resuspend beads in 1 mL of Wash Buffer. Give it a minute in suspension and a minute on the stand to wash.
10. Repeat step 9 2 times (total of 3 washes).
11. After last wash, resuspend in 1.05x**N IPs**x100 μL Wash Buffer (5% more than the number of IPs.)
This step allows you to distribute 100 μL equally to all your IP tubes from the batch of beads. 5% extra volume ensures that you can do this equally. You will find if you resuspend your beads in 500 μL of lysis buffer and try to put 100 μL in each of 5 tubes, you will be unable to do so as the detergent in lysis buffer makes this difficult. You can also plan for N+1 IPs, but this works just as well.
12. Distribute 100 μL to **N** tubes equal to the number of actual IPs.
13. Keep on ice until you are ready to use them. Remove the Wash buffer you used to aliquot the beads before use.

Homogenization and Lysis

I use the (1mL) proportions for <half a brain, and the 2 mL proportions for a whole brain.

1. Dissect tissue or pellet cells.
2. Keep glass mortar(s) on ice and fill each with (1 mL | 2 mL) of lysis buffer (*1 use 1 mL for smaller dissections of brain or a pellet of cells from 6cm - 10cm dishes, and use 2 mL for a whole brain, or see the Big Protocol/Background document for recommendations about lysis buffer to mg tissue recommendations*).
3. Plop tissue into mortar.
4. Take drill fitted with teflon pestle and homogenize up and down 6 times at medium/high power (*see demo*).
5. After homogenization, pour off homogenate into a 1.7 mL tube (if homogenizing in 1 mL) or some larger size of tube that will accommodate your volume (2 mL tubes are nice).

6. Spin 2000 xg, 10 min, 4°C.
7. Remove (800 μL | 1.6 mL) of homogenate from Step 6 and move to a new tube and add:
(0.1 mL | 0.2 mL) 10% NP40 (final concentration 1%)
(0.1 mL | 0.2 mL) 300 mM DHPC (final concentration 30 mM)
8. Invert to mix and incubate on ice for 10 minutes.
9. Spin at 20,000xg, 15 minutes, 4C.
10. Measure total lysate. Take 0.1 volumes as Input sample and bring total volume to 250 μL with Wash Buffer. Add 750 μL Trizol LS and store at -80C until you are ready to extract RNA.
11. Take the remaining 0.9 volumes and resuspend your beads in it.
12. IP for 2 hours at 4C (or up to overnight).
13. After IP put the samples on the stand and let sit for a minute. Remove the supernatant and discard (you can also take this supernatant to compare to input if you like, in which case repeat step 10 above).
14. Resuspend in 1 mL of High Salt Wash Buffer. Let sit for a minute in suspension, then a minute on the stand.
15. Repeat step 14 3 more times (total of 4 washes).
16. Resuspend beads in 250 uL Wash Buffer. Add 750 μL of Trizol LS and store at -80C until you are ready to extract RNA.

RNA Extraction

1. Bring samples to room temperature if they have been stored at -80C and incubate at room temperature for 5 minutes. If you haven't done so already, take out the Glycoblue (stored at -20) and bring to room temperature.
2. Add 0.2 mL chloroform.
3. Shake vigorously by hand for 15 seconds.
4. Incubate for 7 minutes on bench.
At this point, you can use Phase Lock Gel Heavy tubes to help you separate phases later. Pellet Phase Lock Gel 12,000 xg, 30 seconds. Add your sample to the pelleted Phase Lock Gel. Mix well but do not vortex.
5. Centrifuge at 12,000xg, 15minutes, 4C.
6. Vortex the crap out of the Glycoblue and spin briefly. (The dye is not always perfectly distributed).

7. Add 1 μL of Glycoblue to new tubes equal to the number of samples you have.
8. Remove the aqueous phase to a new tube (the upper layer) from step (5) to the tubes with glycoblue and mix well.
9. Add 0.7 volumes of 100% isopropanol.
10. Incubate on the bench for 10 minutes.
11. Centrifuge at 4C, max speed ($\geq 12,000\text{g}$), for 15 minutes.
12. Pour off the supernatant. Do not pipette off the supernatant.
13. Add 1 mL of 80% EtOH. I like to dislodge the pellet and invert several times to wash.
14. Repeat the centrifugation in step 11.
15. Pour off the supernatant. Do not pipette off the supernatant.
16. Spin again to collect the EtOH you couldn't pour off.
17. Remove remaining supernatant using the pipette tip method (*see demo*).
18. Leave tubes open on your tube rack while you prepare the DNase treatment mix (3-5 minutes. Do not overdry!). DNase Treatment Mix:
 - 10 μL 10X Qiagen RDD Buffer x **n+1 samples**
 - 87 μL H₂O x **n+1 samples**
 - 3 μL Qiagen DNase I x **n+1 samples**
 - Final volume is 100 μL per sample (planning for 1 extra).**
19. Resuspend the pellet in 100 μL of DNase Treatment Mix.
20. Incubate at 37 C for 15 minutes.
21. Add 350 μL of Qiagen Buffer RLT to each sample (RNeasy Minelute kit). Mix by pipetting 5 times.
22. Add 250 μL of 100% EtOH to each sample (RNeasy Minelute kit). Mix by pipetting 5 times.
23. Pass volume (700 μL) over a Qiagen RNeasy Minelute kit by centrifugation at $\geq 10,000\text{g}$ for 15 seconds at room temp.
In subsequent steps after centrifugation you should dump out the flow-through, but it can help to have multiple collection tubes you can just move the column to in order to speed things up.
24. Add 500 μL of Buffer RPE to each column and centrifuge at $\geq 10,000\text{g}$ for 15 seconds at room temp.
25. Add 500 μL of 80% EtOH to each column and centrifuge at $\geq 10,000\text{g}$ for 2 minutes at room temp.

26. After you discard the last flow through dry the column by spinning for 1 minute at max speed at room temp.
27. Take a 1.7 mL tube and cut off the cap (or I just rip them off). This will collect your sample.
28. Put column into tubes from step 27. Add 15 μ L of RNase-free H₂O to the middle of each column (make sure you see the liquid getting onto the silica matrix).
29. Incubate on bench 1 minute.
30. Spin for 1 minute at max speed at room temp.
31. Congratulations you now have RNA. Assess quality and quantity of RNA by some method and wait till I distribute the next protocol.

Buffers and Reagents

Products Numbers

Description	Vendor	Product #	
strips of 8 x 200 ul tubes with individual caps	Bioexpress	T-3135-1	
2M KCl	ThermoFisher	AM9640G	
1M MgCl ₂	ThermoFisher	AM9530G	
SUPERase•In™ (20 U/ μ l)	ThermoFisher	AM2696	
07:0 PC (DHPC)	Avanti Polar Liquids	850306P	
rRnasin	Promega	N2515	
Complete mini (EDTA free)	Roche	4693159001	1 p
Cycloheximide	Sigma	C1988-1G	
1M DL-Dithiothreitol solution	Sigma	646563-10X.5ML	
HEPES, 1M Ph 7.4	Sigma	H0887-100ML	
RNase-Free DNase Set (50)	Qiagen	79254	
Nonstick, RNase-free Microfuge Tubes, 1.5 mL	ThermoFisher	AM12450	
Glycoblue	ThermoFisher	AM9515	
Ethanol 200 proof for molecular biology	Sigma	E7023-500ML	
Green Fastmix, Quanta Biosciences	VWR	101414-270	
qScript™ cDNA SuperMix, Quanta Biosciences	VWR	101414-100	
Mouse anti-GFP antibody clone 19F7	MSKCC	HtzGFP-19F7	
Mouse anti-GFP antibody clone 19C8	MSKCC	HtzGFP-19C8	
Dynabeads My one Streptavidin T1	ThermoFisher	65602	
Pierce™ Recombinant Protein L, Biotinylated	ThermoFisher	29997	
Bovine Serum Albumin (IgG-Free, Protease-Free)	Jackson ImmunoResearch	001-000-162	

1% BSA/1XPBS

Add IgG-Free Bovine Serum Albumen (100 mg per 10 mL) to 1X PBS and allow to rock gently for >10 minutes to go into solution.

0.1% BSA/1XPBS

Dilute 1% BSA/1XPBS 10-fold.

10% NP40

Carefully make up 10% v/v NP40 (IGEPAL CA-630, Sigma) by pipetting 100% NP40 into H₂O and let rock for >10 minutes to fully dilute.

300 mM DHPC

Resuspend 100 mg 07:0 PC (DHPC, Avanti polar lipids) in 692 μ L of H₂O. To avoid foaming, add H₂O and let it rock for a few minutes and then transfer to Eppendorf tubes.

100 mg/mL cycloheximide

Dissolve cycloheximide in 100% methanol at 100 mg/mL.

10X Roche cOmplete EDTA-free protease inhibitor

Crush tablet in a small volume (~300-500 μ L) of water. Pipette up and down to start to dissolve. Bring volume to 1 mL and mix until fully dissolved.

Wash Buffer

	units	start	final	df	1	10	25	mL
hepes ph 7.4	mM	1000	10	100	0.01	0.100	0.25	
kcl	mM	2000	150	13.3333333333	0.075	0.750	1.875	
mgcl2	mM	1000	5	200	0.005	0.050	0.125	
dtc	mM	1000	0.5	2000	0.0005	0.005	0.0125	
rnasin	%v/v	100	0.1	1000	0.001	0.010	0.025	
superase-in	%v/v	100	0.1	1000	0.001	0.010	0.025	
roche protease inh	X	25	1	25	0.040	0.400	0.625	
cyclohex	mg/mL	100	0.1	1000	0.001	0.010	0.025	
np40	%v/v	10	1	10	0.1	1.000	2.5	
				water	0.8815	8.815	22.0375	

Add DTT, RNasin, Superase-In, protease inhibitor, and cycloheximide just prior to use.

Homogenization Buffer

	units	start	final	df	1	5	10	mL
hepes ph 7.4	mM	1000	10	100	0.010	0.050	0.100	
kcl	mM	2000	150	13.3333333333	0.075	0.375	0.750	
mgcl2	mM	1000	5	200	0.005	0.025	0.050	
dtc	mM	1000	0.5	2000	0.0005	0.0025	0.005	
rnasin	%v/v	100	0.1	1000	0.001	0.005	0.010	
superase-in	%v/v	100	0.1	1000	0.001	0.005	0.010	
roche protease inh	X	25	1	25	0.040	0.200	0.400	
cyclohex	mg/mL	100	0.1	1000	0.001	0.005	0.010	
				water	0.8665	4.3325	8.665	

Add DTT, RNasin, Superase-In, protease inhibitor, and cycloheximide just prior to use.

High Salt Wash Buffer

	units	start	final	df	1	10	25	mL
hepes ph 7.4	mM	1000	10	100	0.010	0.100	0.250	
kcl	mM	2000	350	5.7142857143	0.175	1.75	4.375	
mgcl2	mM	1000	5	200	0.005	0.05	0.125	
dtc	mM	1000	0.5	2000	0.0005	0.005	0.0125	
rnasin	%v/v	100	0.1	1000	0.001	0.010	0.025	
superase-in	%v/v	100	0.1	1000	0.001	0.010	0.025	
roche protease inh	X	25	1	25	0.040	0.400	1.000	
cyclohex	mg/mL	100	0.1	1000	0.001	0.010	0.025	
np40	%v/v	10	1	10	0.100	1.000	2.500	
				water	0.6665	6.665	16.6625	

Add DTT, RNasin, Superase-In, protease inhibitor, and cycloheximide just prior to use.

Appendix 4

PTRE-Seq Sequencing Library Preparation Protocol

cDNA Synthesis

1. Add 2.5 μL of 12 μM pmrPTRE_Antisense1 to each of tube.
2. Add 8.0 μL of RNA/water (f.v. 10.5) to each tube containing 10 ng of total RNA or TRAP RNA.
3. Heat to 65°C for 2 minutes with lid at 105°C. Place immediately on ice for 1 minute.
4. Make mastermix, see spreadsheet for full calculations using strip tubes.:

	single(μL)
5x buffer	4
10 mM dntps	2
100 mM dtt	2
Promega rRNasin	0.6
ssrt iii	0.9

5. Split master mix into a strip tubes..
6. Add 9.5 μL of mastermix to each sample.
7. Incubate all at 50C for 45 minutes.
8. Add 3.5 μL of ExoSAP-It
9. Vortex & spin
10. Heat to 37C for 15 minutes. (f.v. 23.5)
11. Add 1 μL 0.5 M EDTA per reaction. Vortex & spin. (f.v. 24.5)
12. Add 3 μL 1 M NaOH. Vortex and Spin. (f.v. 27.5)

13. Heat to 70C for 12 minutes.
14. Add 3 μL 1M HCl. Vortex and Spin. (f.v. 30.5)
15. Clean up with MyONE Silane:
 - (a) Wash N*10 μL beads for N samples with 900 μL RLT
 - (b) Resuspend in 91.5 μL /sample buffer RLT.
 - (c) Add 91.5 μL to each sample and mix (f.v. 122)
 - (d) Add 111 μL 100% EtOH. Incubate 15 minutes. (f.v. 233)
 - (e) Wash MyONE silane style 3x80% Etoh (1 resus, 2 on mag)
 - (f) Dry 5. Remove EtOH. Elute in 12 μL 50 mM Tris pH 7.8 and incubate 5 minutes before clarifying.

ds cDNA amplification and plasmid DNA amplification

1. Dilute Plasmid DNA based on Nanodrop readings to $\sim 5 \text{ ng}/\mu\text{L}$
2. Seed 20 μL PCR reactions with 8 μL of cDNA samples or 8 μL of 5 $\text{ng}/\mu\text{L}$ DNA samples. Master mix:
 - 10.0 μL 2X Phusion HF Master Mix
 - 1.0 μL 10 μM PTRE UTR antisense 1
 - 1.0 μL 10 μM PTRE UTR sense 1
 - 8.0 μL cDNA
3. Add 12 μL of master mix to each.
4. Run following program:
 - 98°C 30s
 - 98°C 10s, 60°C 10s, 72°C 15s | 12 cycles (DNA) **or** 18 cycles (cDNA)
 - 72°C 10 min
 - 4°C ∞

AMPure 80/40 selection

1. Bring AMPure XP beads to room temperature ≥ 30 min.
2. Bring volume to 100 μL with H_2O (add 80 μL).
3. Vortex beads and pipette 80 μL and add to PCR products. Mix by pipetting up and down 10x.
4. Incubate on bench 5 minutes.

5. Put on strip tube magnet for 3 minutes.
6. Re-vortex beads and pipette 40 μL to a clean strip tube.
7. Harvest supernatant ($\sim 160 \mu\text{L}$ recoverable) and add to new tube with 40 μL beads in it. Mix by pipetting up and down 10x.
8. Incubate on bench 5 minutes.
9. Put on strip tube magnet 3 minutes.
10. Wash beads on magnet with 200 μL 80% EtOH for 30 seconds. Do not disturb pellet. Add wash, let sit, then remove.
11. Repeat for total of two washes.
12. Dry beads for 5 minutes and remove any excess EtOH with vacuum or by pipette.
13. Resuspend beads in 10.5 μL Buffer EB (so that it will be possible to recover 10 μL) and incubate on bench 5 minutes.
14. Add to magnet for 1 minute and harvest 10 μL .

NheI/KpnI Digest

1. Remove 10 μL of 80/40 purified PCR product and add to:
 - 2 μL 10X CutSmart Buffer
 - 1 μL NheI HF
 - 1 μL KpnI HF
 - 6 μL H₂O
2. Mix well, spin, and incubate in thermal cycler for 1 hour at 37°C with lid set to 40°C.

AMPure 100/50 selection

1. Bring volume to 100 μL with H₂O (80 μL).
2. Vortex beads and pipette **100** μL and add to digested products. Mix by pipetting up and down 10x.
3. Incubate on bench 5 minutes.
4. Put on strip tube magnet for 3 minutes.
5. Re-vortex beads and pipette **50** μL to a clean strip tube.

6. Harvest supernatant and add to new tube with **50 μL** beads in it. Mix by pipetting up and down 10x.
7. Incubate on bench 5 minutes.
8. Put on strip tube magnet 3 minutes.
9. For QC, save the first bead pellet. This is the first round of selection. Mostly selects out high MW stuff, but may select out a bit of desired product too.
10. Wash beads on magnet with 200 μL 80% EtOH for 30 seconds. Do not disturb pellet. Add wash, let sit, then remove.
11. Repeat for total of two washes.
12. Dry beads for 5 minutes and remove any excess EtOH with vacuum or by pipette.
13. Resuspend beads in 10.5 μL Buffer EB (so that it will be possible to recover 10 μL) and incubate on bench 5 minutes.
14. Add to magnet for 1 minute and harvest 10 μL .

Adapter Ligation

1. Combine 10 μL purified/digested PCR product in the follow reaction
 - 6 μL H₂O
 - 2 μL Enzymatics T4 DNA Ligase buffer 10X
 - 0.5 μL 1 μM NheI/P2 adapter
 - 0.5 μL 1 μM mix of KpnI/P1 adapters 1-4
 - 1 μL of Enzymatics T4 DNA Ligase
2. Incubate at 16 C for 1 hour
3. Purify products with Ampure 80/40 and elute in 10.5 μL EB, and harvest 10 μL .

Preparative PCR

1. Master mix:
 - 25 μL NEB Q5 Ultra II pcr master mix
 - 2.5 μL 10 μM universal primer
 - 12 μL H₂O
2. Aliquot 39.5 to each tube.
3. Add 2.5 μL of index primer.

4. Run with Q5 program for 9 cycles:
98°C 30 s
(98°C 10 s | 55°C 15 s | 65°C 1 min) X 9
65°C 5 min
12°C hold
5. Purify with 80/40 ampure XP (see above).