UNDERSTANDING THE IMPRINTING MECHANISM OF *UBE3A* FOR THERAPEUTIC INTERVENTION

A Dissertation

by

JADE MARIE BENJAMIN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Scott V Dindot
Committee Members,	David J Segal
	David W Threadgill
	C Jane Welsh
Intercollegiate Faculty Chair,	Dorothy Shippen

August 2017

Major Subject: Genetics

ABSTRACT

Human chromosome 15q11-q13 contains a cluster of imprinted genes that are associated with a number of neurological disorders that exhibit non-Mendelian patterns of inheritance, such as Angelman syndrome (AS) and Prader-Willi syndrome. Angelman syndrome is caused by the loss-of-expression of maternally inherited ubiquitin E3A protein ligase gene (*UBE3A*). Prader-Willi syndrome is caused by loss-of-function of paternally inherited *SNORD116* snoRNAs (small nucleolar RNAs), which are expressed as part of a long polycistronic transcriptional unit (PTU) comprised of *SNURF-SNRPN*, additional orphan C/D box snoRNA clusters, and the *UBE3A* antisense transcript (*UBE3A-AS*). The full-length transcript of PTU, including *UBE3A-AS*, is only expressed in neurons causing the imprinting of paternal *UBE3A*. Why this occurs in only neurons remains largely unknown. Furthermore, this neuron-specific imprinting adds additional difficulty for therapeutic intervention. In this dissertation, the imprinting mechanism of *UBE3A* is examined in detail, while an alternative high-throughput screening (HTS) method for drug discovery in neurons is developed.

A combination of bioinformatic and molecular analysis of the human and mouse PTU revealed that *UBE3A-AS/Ube3a-AS* is extensively processed via 5' capping, 3' polyadenylation and alternative splicing, suggesting that the antisense may have regulatory functions apart from imprinting *UBE3A* in neurons. Following this discovery, the transcripional profiles and processing of mouse paternal *Ube3a* was investigated as literature suggested that imprinted paternal *Ube3a*, unlike other imprinted genes, was transcribed up to intron 4. This analysis unveiled a fourth *Ube3a* isoform that terminates within intron 4. Moreover, expression of this isoform correlated with *Ube3a-AS* expression, suggesting alternative reasons for the imprinting of *Ube3a*. In addition to the analysis of the imprinting

of *Ube3a*, an alternative solution for drug discovery for central nervous system disorders was developed and validated. Here, an embryonic stem cell-derived neuronal culture system was developed for HTS and tested using the paternal *Ube3a*^{YFP} reporter cell-line. Using a known reactivator of paternal *Ube3a*, Topotecan - a topoisomerase inhibitor, as a positive control a proof-of-concept study demonstrated the utility of this method for HTS drug discovery. Collectively, these results advance the field and understanding of antisense lncRNAs and provide a versatile tool for drug discovery for neurological disorders.

DEDICATION

Faith, for tho art in heaven.

ACKNOWLEDGEMENTS

I would like to thank my committee members, Dr. Segal, Dr. Threadgill and Dr. Welsh, for their insight and guidance throughout the course of my doctoral research training. A special thank you goes to Dr. Jane Welsh for her support, constant encouragement and motivation, all of which are immeasurable. Additionally, I would like to thank Dr. Kevin Burgess for his helpful insight and guidance and Dr. Tracy Vemulapalli for stepping in at the last minute. Their help along with my chair Dr. Dindot was greatly appreciated in this long and length process.

I must also thank my program faculty and staff for their assistance. I will miss the familiarity of seeing Ms. Julia in her office. Additionally, I want to give thanks to the Department of Biochemistry and the Department of Veterinary Pathobiology staff for always greeting me with a welcoming smile. Finally, the Texas A&M University Interdisciplinary Genetics Program and the Texas A&M University College of Veterinary Medicine & Biomedical Sciences, Department of Veterinary Pathobiology were instrumental in supporting my research pursuits.

I would like to recognize: Mrs. Shrock my first grade teacher who managed to inscribe a love of reading, which managed to survive despite racist teachers actively working to tear me down a few years later. Ms. Jocelyn Sisson my tenth grade English teacher who gave me a backbone and didn't let me settle for anything less than perfection. Dr. Bilal M. M. Bomani for teaching me the importance of networking and interpersonal relationships. Dr. Danny Alge for teaching me the scientific method. And Dr. James Cooper for teaching me how to design and execute experiments.

I want to thank Dr. William Griffith and Ms. Annette Fincher, who took time out of their busy schedules to teach me electrophysology despite it not making it into this body

of work.

An immense thanks go to my fellow graduate students, colleagues and lab members: Dr. Sarah Black Christian, Dr. Cole McQueen, Dr. Paul R. Hillman, and Ms. Hillary Shaheen. Thank you all for your assistance. A special thank you goes to the Genetics Graduate Student Society for being an open place to talk with many late nights, free food, and much needed laughter. I wouldn't have been able to get through this without the camaraderie.

Finally, I cannot thank my family enough for their love and support. To my supportive family, my many aunts, uncles, and cousins who always asked about my research, I cannot thank you enough for the large amount of love and constant encouragement you all sent my way. I'd thank you each individually, but then I'd run over my page limit. To my little brother Victor and little sister Izzy, who understand the numerous holidays and birthdays I had to miss because of this work. To my older sister Tasha, who's support and understanding was unmeasurable. To my other brother Joshua, who was always in my corner and to this day is still my greatest hero. To my twin, Erica, for always lending an ear and wanting to fight for me at every turn during this difficult and trying process. To my father who will likely be the first to read this in it's entirety. And most importantly, to my beautiful, powerful, passionate mother whose life embodied Faith. She believed in all of her children with all of her heart to her very last breath. It is thanks to my mother and god that I am here today, and I could not thank her enough. Thank you mom, for everything. For the love, support, constant encouragement, and your persistent *fighting*. I will love you always.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Drs. Dindot [chair], Segal and Threadgill of the Department of Veterinary Pathobiology, College of Veterinary Medicine & Biomedical Sciences and Dr. Welsh of the Department of Veterinary Integrative Biosciences, College of Veterinary Medicine & Biomedical Sciences. Drs. Dindot and Threadgill have joint appointments in the Department of Molecular and Cellular Medicine, College of Medicine. Dr. Welsh holds a joint appointment in the Department of Neuroscience and Experimental Therapeutics, College of Medicine.

The RNA-seq hybrid mouse data analyzed for Chapters 2 and 3 was generated by Wellcome Trust Sanger Institute, Genome Research Limited (reg no. 2742969) `http://www.sanger.ac.uk/science/data/mouse-genomes-project`. Several primers used in Chapters 2 and 3 were designed by Drs. Sarah Black Christian and Paul Hillman. The B6D2F1 female mouse tissues used in Chapters 2 and 3 were dissected and flash frozen by Dr. Ryan Doan. Primary neuronal cultures used in Chapter 3 were established by Ms. Hillary Shaheen. Ms. Jean Kovar maintained the mouse colonies.

The image acquisition for high-throughput screening in Chapters 4 and Appendix F was conducted at Texas A&M Institute of Biosciences & Technology (IBT) in Houston, Texas. Dr. Clifford Stephan, Ms. Mary Sovieski, and Ms. Ivy Nuygen provided valuable technical support at the IBT. In addition, the Stephan group was in charge of changing media and adding drugs to the high-throughput data in Chapter 4.

The *Ube3a*$^{+/YFP}$ embryonic stem cells used in Chapter 4 were generated by TIGM (Texas A&M Institute for Genomic Medicine). Additionally, neuronal differentiation scale increase for high-throughput screening by Dr. Johnathan Ballard at TIGM.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a Texas A&M Institute for Genome Sciences and Society (formerly known as WSGI - Whole Systems Genomics initiative) fellowship and from Foundation for Angelman Syndrome Therapeutics (FAST).

NOMENCLATURE

ANOVA	Analysis of variance
AS	Angelman syndrome
BAM	Compressed binary version of SAM
BCF	Compressed binary version of VCF
BED	Browser extensible data
bp	Base pairs
C	Celsius
CAGE-seq	Cap analysis gene expression - sequencing
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
CpG	5’ C phosphate G 3’
CPM	Counts per million
CRISPR	Clustered regularly interspaced short palindromic repeats
CSV	Comma separated values
DEPC	Diethylpyrocarbonate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide
DTT	Dithiothreitol
Dup15q	Chromosome 15q duplication syndrome
E3	Ubiquitin-protein ligase
EDTA	Ethylenediaminetetraacetic acid
ES	Embryonic stem

EST(s)	Expressed sequence tag(s)
EtOH	Ethanol
FBS	Fetal Bovine Serum
FDR	False discovery rate
FPKM	Fragments per kilobase of transcripts per million mapped reads
GFP	Green fluorescent protein
GTF	Gene transfer format
gz	Gzip compressed file
h	Hour(s)
H_2O	Water
HCl	Hydrogen chloride
HTS	High-throughput screening
IGV	Integrative Genomics Viewer
IPW	Imprinted gene in the Prader-Willi syndrome region
kbp	Kilobase pairs
kDa	Kilodalton
l	Liter
lncRNA	Long non-coding RNA
M	Molarity
$MgCl_2$	Magnesium chloride
min	Minute(s)
ml	Milliliter(s)
mm	Millimeter(s)
mRNA	Messenger RNA
NaCl	Sodium chloride

NCBI	National Center for Biotechnology Information
ncRNA	Non-coding RNA
ng	Nanograms
nM	Nanomolar
PAF	Paraformaldehyde
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PE	Paired-end
polyA	Polyadenylation
PWS	Prader-Willi syndrome
qPCR	Quantitative PCR
RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
rpm	Revolutions per minute
RT-PCR	reverse-transcription PCR
s	Second(s)
SAM	Sequence alignment/map
SDS	Sodium dodecyl sulfate
SE	Single-end
SIM	Sandos Inbred Mice
snoRNA	Small nucleolar RNAs
SNP(s)	Single-nucleotide polymorphism
SNRPN	Small Nuclear Ribonucleoprotein Polypeptide N
SNURF	SNRPN Upstream Reading Frame

UCSC	University of California, Santa Cruz
μg	Microgram(s)
μl	Microliter(s)
VCF	Variant call format
wk	Week(s)
YFP	Yellow fluorescent protein

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. *UBE3A* IMPRINTING IN NEURONS: A UNIQUE MODEL FOR ANTISENSE LNCRNA REGULATION

The genomic instability of chromosome 15q11-q13 results in multiple disorders, two of which seems to be linked to ubiquitin ligase E3A protein (UBE3A) expression, also known as E6AP, [1–4]. The non-Mendelian inheritance of these disorders are a result of genomic imprinting, an epigenetic phenomenon that results in the differential expression of diploid alleles in a parent-of-origin specific manner [5]. With respect to *UBE3A*, it is clear that expression and dosage levels are important for human brain development and functionality. As such, the unique imprinting of *UBE3A* in only neurons is perplexing. Although it is possible that the imprint could be an innocent bystander, the fact that it has been evolutionarily constrained for over 100 million years suggests otherwise [6]. The focus of this review is to examine the functional significance of the imprinting mechanism of *UBE3A* and the roles antisense long non-coding RNA (lncRNA) play in regulation of the sense transcripts.

1.1 Diseases of Chromosome 15q11-q13

The long arm of human chromosome 15 is characterized by relatively frequent chromosome rearrangements, which are due to low-copy repeat elements at two proximal and one distal region of the 15q11-q13 region [7–10]. **Figure 1.1** depicts the breakpoint locations around the *HERC2* (HECT, homologous to the E6AP carboxyl terminus, and RDL, regulator of chromosome condensation 1 (RCC1) like domain, domain containing E3 ubiquitin protein ligase 2) duplicons and the percentage of usages within the large-scale deletions. This instability results in multiple structural abnormalities, including deletions, duplications, and translocations [1, 11–14] resulting in three distinct neurodevelopmental disorders: Prader-Willi syndrome (PWS), Chromosome 15q11-q13 Duplication syndrome

(Dup15q) and Angelman syndrome (AS).

1.1.1 Prader-Willi Syndrome

Prader-Willi syndrome (OMIM #176270) is characterized by neonatal hypotonia and failure to thrive, hyperphagia in early childhood leading to obesity, hypogonadism, short stature, behavior problems, and mild to moderate intellectual disability [15, 16]. The genetic or epigenetic mutations causing PWS are associated with the specific loss of paternal expression of the box C/D small nucleolar RNAs (snoRNAs) generated from *SNORD116* cluster (previously referred as *HBII-85*) in the brain [17–19]. The spectrum of mutations causing PWS include: (i) paternal interstitial deletions of 15q11-q13 region, (ii) maternal uniparental disomy of chromosome 15, (iii) genomic imprinting defects of the region, or (iv) loss-of-function mutations in *SNORD116* gene cluster [15].

1.1.2 Chromosome 15q11-q13 Duplication Syndrome

Chromosome 15q11-q13 duplication syndrome (OMIM #608636) is characterized by developmental delay, intellectual disability, early central hypotonia, seizures, and social impairment [20]. Additionally, duplication of 15q11-q13 is one of the most common genetic mutations observed in individuals diagnosed with autism spectrum disorder [4,20–23]. Chromosome 15q duplication syndrome is primarily caused by maternal duplications of chromosome 15q11-q13 [20, 24]. Currently, Dup15q is known to occur via one of two ways: interstitial duplication of 15q, or extra isodicentric chromosome of 15q [20, 25]. As the neurodevelopmental disorder is caused by maternal inheritance of duplications, it is linked to the overexpression of ubiquitin ligase E3A protein (UBE3A) [4].

1.1.3 Angelman Syndrome

Angelman syndrome (OMIM #105830) is a debilitating neurodevelopmental disorder characterized by severe intellectual disability, absent speech, ataxia, seizures, frequent

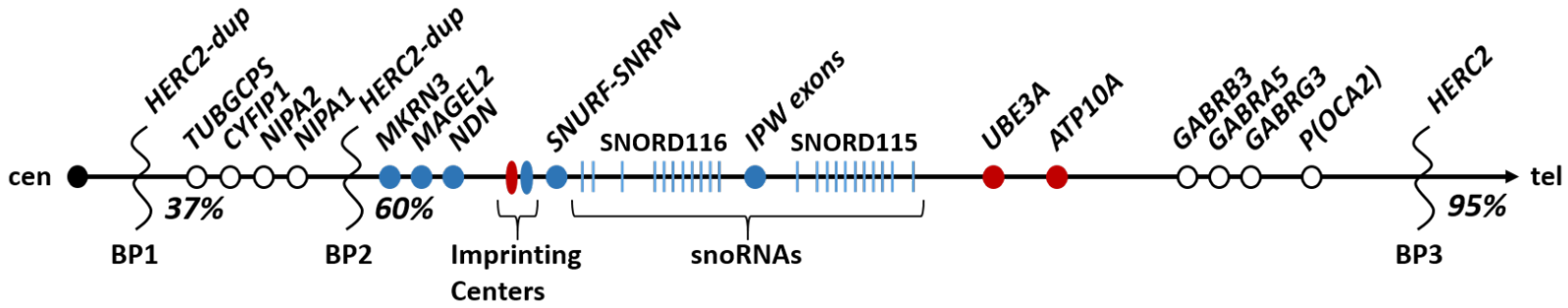


Figure 1.1: The frequent rearrangements of human chromosome 15q11-q13 region is due to recombination hotspots of the *HERC2* duplicons at the three breakpoints (BP). Paternally expressed genes in blue, maternally expressed genes in red, and biallelically expressed genes in white.

smiling and inappropriate laughter [26]. Although clinically distinct, AS shares a common pathogenesis with Dup15q, namely dysregulation of UBE3A protein. In contrast to Dup15q, the genetic or epigenetic mutations causing AS are associated with the specific loss of maternal expression of UBE3A protein in the brain [2, 3, 26]. The spectrum of mutations causing AS include: (i) maternal interstitial deletions of 15q11-q13 region, (ii) paternal uniparental disomy of chromosome 15, (iii) genomic imprinting defects of 15q11-q13, or (iv) loss-of-function mutations in the *UBE3A* gene [26].

1.2 Ubiquitin Ligase E3A Protein Gene

The *UBE3A* gene is located within the 15q11-q13 imprinted domain and encodes an E3 ubiquitin-protein ligase that is a central component of the ubiquitin proteasome system involving the successive action of E1 (ubiquitin-activating), E2 (ubiquitin-conjugating), and E3 ubiquitin-protein ligases activities [27, 28]. The UBE3A protein is a unique ligase as it can catalyze the formation of isopeptides via its HECT domain without the help of E2 proteins. In addition, UBE3A has been shown to function as a typical ubiquitin ligase and a transcriptional coactivator of steroid hormone receptors [29, 30]. Thus far, numerous cellular proteins have been shown to interact directly or indirectly with UBE3A suggesting that is has diverse cellular functions (**APPENDIX A, Tables A.1, A.2** and **A.3**).

1.2.1 UBE3A *in Neurons*

The non-Mendelian inheritance pattern of AS is due to genomic imprinting of *UBE3A* [2, 3]; however, *UBE3A*, unlike most imprinted genes, is expressed from both parental alleles in almost all cell types except for neurons [31–34]. Expression of the paternal *UBE3A* allele in neurons is inhibited by the antisense expression of a long polycistronic transcriptional unit that is comprised of *SNURF-SNRPN*, clusters of C/D box snoRNAs (*SNORD116* and *SNORD115*), and *UBE3A* antisense transcript (*UBE3A-AS*) [31, 33, 35–37]. **Figure 1.2** shows the polycistronic transcriptional unit transcribed from the paternal

allele (blue) with the overlapping antisense transcript silencing *UBE3A* (black).

1.3 Genomic Imprinting

Mammals are diploid organisms that inherit a chromosome set from each parent. As a result, the majority of genes are expressed from both parents; however, a subset of genes show parent-specific gene expression due to epigenetic modifications. **Figure 1.3** depicts normal biallelic expression compared to the monoallelic expression of imprinted genes. There are over a 100 of these imprinted genes in mouse and humans [38, 39]; the majority of which are found within gene clusters that house at least one non-coding RNA and several protein coding genes.

1.3.1 Imprinting Control Regions

Epigenetics is defined as heritable modifications of the genome that are not genetic changes (e.g. DNA and chromatin modifications) [40]. As imprinting is an epigenetic phenomenon, the imprinted genes must be able to acquire modifications, maintain imprinting status, and be re-established in the germline. For imprinted gene clusters, large-range *cis*-acting imprinting control regions (ICRs) are responsible for acquiring parental modification and maintaining imprinting status [41–43]. All ICRs have differential DNA methylation regions (DMR), which carry parental information [44].

The ICR for PWS and AS has a bipartite structure comprising the Prader-Willi and Angelman syndrome imprinting centers (PWS-IC and AS-IC); thus, giving this ICR bidirectional control of the cluster of imprinted genes in the 15q11-q13 region [45]. The PWS-IC is a positive regulatory element responsible for the establishment and maintenance of paternal gene expression [46], while AS-IC negatively regulate PWS-IC on the maternal chromosome [47].

Differential methylation is associated with a CpG island surrounding the promoter and exon 1 of *SNRPN* [48]. In somatic cells of human and mouse, PWS-IC is heavily CpG

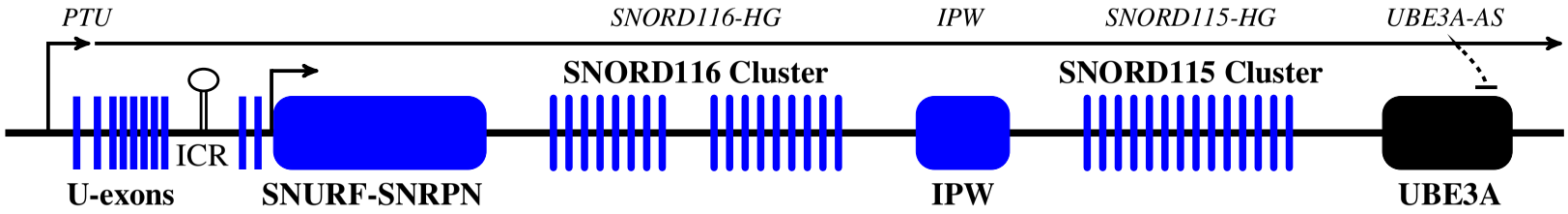


Figure 1.2: Paternal expression of long alternatively spliced transcript in neurons comprising *SNURF/SNRPN*, C/D box snoRNAs (*SNORD116* and *SNORD115*), *IPW*, and *UBE3A-AS* is controlled by unmethylated Prader-Willi syndrome imprinting center region (ICR). There some unknown mechanism *UBE3A-AS* inhibits expression of paternal *UBE3A*. Paternally expressed genes in blue and paternally silenced *UBE3A* in black.

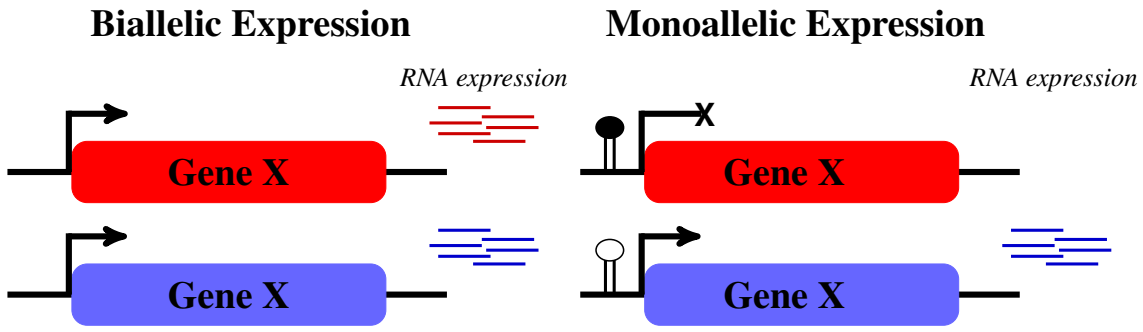


Figure 1.3: The majority of genes show biallelic expression with maternal (red) and paternal (blue) alleles expressed in all tissues. In a subset of genes via some epigenetic mechanism such as DNA methylation, gene expression is limited to monoallelic expression. 🍄 is methylated ICR, and 🍄 is unmethylated ICR.

methylated on the maternal chromosome, and is almost completely unmethylated on the paternal chromosome [49,50]. The AS-IC on the maternal allele confers DNA methylation and suppression of PWS-IC, but is methylated and inactive on the paternal allele. This results in PWS-IC regulation of the majority of the genes as paternal expression except for *UBE3A* and *ATP10A*, which are maternally expressed [47, 51].

1.4 Function of Imprinting

No matter the theory, the overall result of imprinting is the differential silencing of alleles. This imprinting does not happen by chance as it is often evolutionary conserved between mouse and humans. For example, the imprinted cluster within 15q11-q13 arose 105-180 million years ago [6]; and despite the disease phenotypes and genomic instability associated with the region, it remains conserved. For some reason, it is advantageous to imprint this gene cluster and many others. One method to explore what makes the imprinting of gene clusters advantageous, is to investigate the function of the imprinted genes. Are the genes dosage sensitive? Does the imprint increase regulatory function? For the 15q11-q13 imprinting cluster, these answers are not readily apparent.

1.4.1 Dosage Regulation

Why *Ube3a* became imprinted in neurons is unclear. The current theory is that *Ube3a-AS* evolved to reduce *Ube3a* expression in neurons to regulate dosage because termination of *Ube3a-AS* leads to increased levels of Ube3a expression in the brain [36, 52, 53]. When our laboratory rigorously tested this theory, we determined that the function of the imprinting mechanism was not to reduce the expression of Ube3a in the brain, as there was no correlation between the imprint and expression levels of *Ube3a/UBE3A* in mouse and human [54]. Additionally, while the paternal *Ube3a* allele was silenced, the maternal allele was upregulated during development so that overall Ube3a expression remained unchanged [54]. Altogether, the findings suggest that imprinting of Ube3a in neurons may have evolved for reasons other than to reduce Ube3a expression in neurons.

1.4.2 Regulatory Function

The evolution of *SNRPN* is a good example of a duplication event leading to tissue-specific imprinting with alternative function from its ancestor gene, *SNRPB/B*. Studies have shown that *Snrpn* originated from *Snrpb*, a small nuclear ribonucleoprotein (snRNP) gene encoding for SmB in mice, locus via a duplication event about 180-210 million years ago [6, 55]. While SmB is replaced by SmN in the brain, *SNRPB/B* is upregulate in the absence of *SNRPN* suggesting that the two genes are tightly regulated [55, 56]. Despite their similarities, *SNRPB/B* and *SNRPN* have distinct snRNP association along with tissue-specific expression, suggesting that *SNRPN* has evolved an alternative function compared to its ancestral gene [57].

Given the current findings, *Ube3a* does not appear to be imprinted as a mechanism to regulate dosage in neurons. With that in mind, it is possible that *Ube3a* is imprinted in neurons because of additional regulatory functions. While *Ube3a* is highly expressed in the brain, the imprint has no effect on overall expression [54]. Therefore, it is unlikely

that the imprint of *Ube3a* is directly linked to Ube3a protein expression; however, its long non-coding antisense RNA transcript, *Ube3a-AS*, arises only in neurons as part of the imprinting mechanism of *Ube3a*. Further study of this transcript may reveal the functional reason for imprinting Ube3a in neurons.

1.5 Long Non-Coding RNA

Over the last few decades, the scientific community has increasingly become fascinated with non-coding RNAs (ncRNAs), especially long non-coding RNAs (lncRNA), and with good reason. With the advancement of genomic sequencing technologies, numerous annotations and deep sequencing from multiple species have demonstrated that ncRNAs are more abundant than protein-coding genes [58, 59]. To put this in perspective, of the 75-90% of the human genome that is transcribed, only 3% is protein-coding [60–64]. More importantly, this is not transcriptional noise as these ncRNAs perform a myriad of functions by interacting with DNA, RNA, and proteins similar to protein-coding genes [65,66]. The importance of ncRNAs extends to genomic imprinting as well since all ICRs have at least one lncRNA expressed from the unmethylated parental chromosome [67–69]. Even with the clear importance of ncRNA, the function of the majority of ncRNAs, especially lncRNAs, is unknown.

Recently, the importance of ncRNAs in tissue- and developmental stage-specific gene expression has been extensively explored. Moreover, increasing evidence implements them in brain development, synaptic plasticity, and neurological disease, with the highest proportion of tissue-specific lncRNA expression in the brain [70, 71]. Long non-coding RNAs are less understood. With no common sequence or structure, classification is difficult. In general, lncRNA preform many similar roles as ncRNAs, like microRNAs and small interfering RNAs, and are separate from ncRNAs via size ($>$ 200 bp). A simplistic classification of lncRNAs uses loci-of-origin resulting in several different categories, three

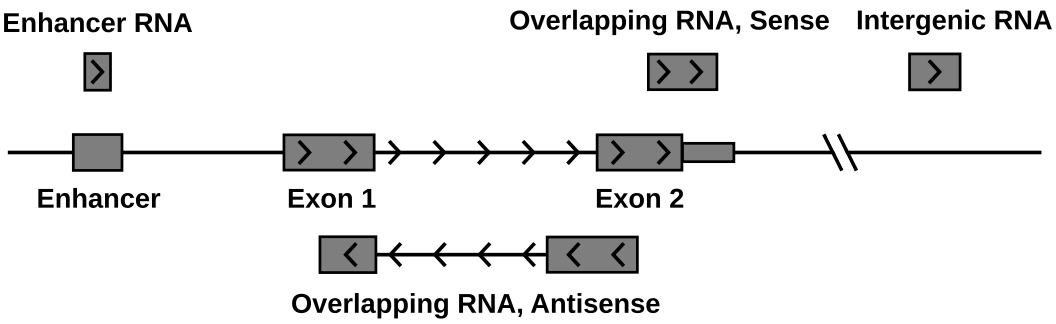


Figure 1.4: Schematic of lncRNAs based on loci-of-origin depicting an enhancer RNA, sense and antisense overlapping RNAs, and an intergenic RNA.

of which, enhancer RNAs, overlapping RNAs, and intergenic RNAs, are depicted in **Figure 1.4**. These three types of lncRNA are all present within ICRs [72–76], with some of the most well-studied lncRNAs in ICRs being antisense lncRNA.

1.6 Antisense lncRNA

Natural antisense transcripts are endogenous transcripts with complete or partial overlap of genes or ncRNA that can work in *cis* or *trans*. Sense/antisense pairings can be non-coding or protein-coding; however, the majority of pairs are non-coding antisense regulating protein-coding sense [74]. Of the protein-coding sense transcripts, a majority (70%) have antisense pairings, many of which are lncRNAs [77, 78]. Furthermore, sense/antisense origination is more likely to be conserved than gene pairs on the same strand, suggesting a conserved functional significance [79]. This functional conservation is observed in their diverse structure, expression pattern, and methods of regulation [80], such as direct regulation of transcription (i.e. transcriptional interference), epigenetic regulation (i.e. genomic imprinting), nucleus interactions (i.e. alternative splicing and termination), and cytoplasmic interaction (i.e. mRNA stability and masking microRNA binding sites).

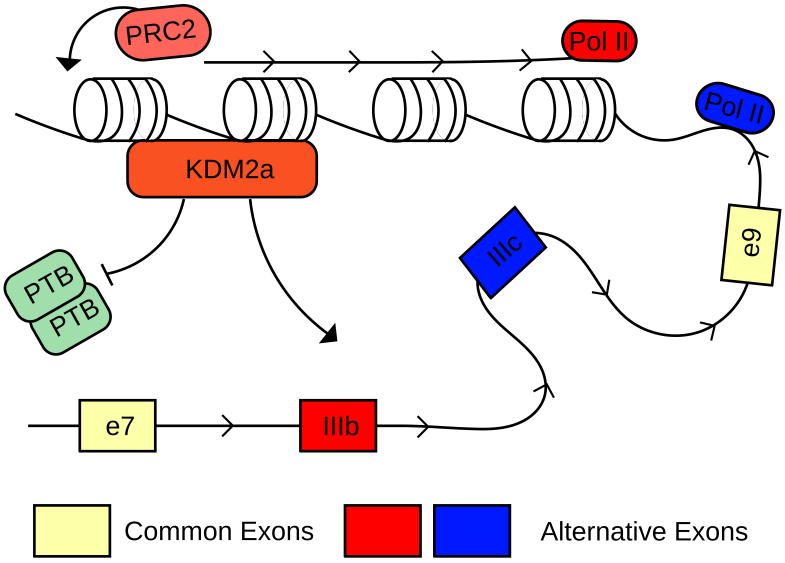


Figure 1.5: *FGFR2-AS* plays a central role in tissue-specific alternative splicing of *FGFR2* via chromatin remodeling. The *FGFR2* antisense transcript recruits PRC2 and KDM2a, which interfere with PTB repression of exon IIIb, resulting in exclusion of exon IIIc. PolII, polymerase II, antisense in red and sense strand in blue.

1.6.1 FGFR2-AS

Antisense transcript of *FGFR2* (*FGFR2-AS*), human fibroblast growth factor receptor 2, is a recently discovered antisense lncRNA of approximately 875 bp that starts 282 bp upstream exon IIIc of *FGFR2* gene [81]. *FGFR2* exhibits chromatin to tissue-specific alternative splicing via a chromatin-splicing adaptor system (MRG15), which recognizes H3K36me2,3 to inhibit inclusion of alternative splicing exon IIIb, and protein-protein interaction recruiting PTB (polypyrimidine tract binding protein) to exon IIIb (negative splicing regulatory element) [82, 83]. Only recently has *FGFR2* been shown to have an antisense transcript that does not appear to be spliced or polyadenylated [81]. Gonzalez *et al.* determined that this evolutionary conserved antisense transcript is located predominantly in the nucleus, where it plays a role in alternative splicing of *FGFR2* sense gene. Additionally, the group demonstrated that *FGFR2-AS* inhibits repression of exon IIIb by

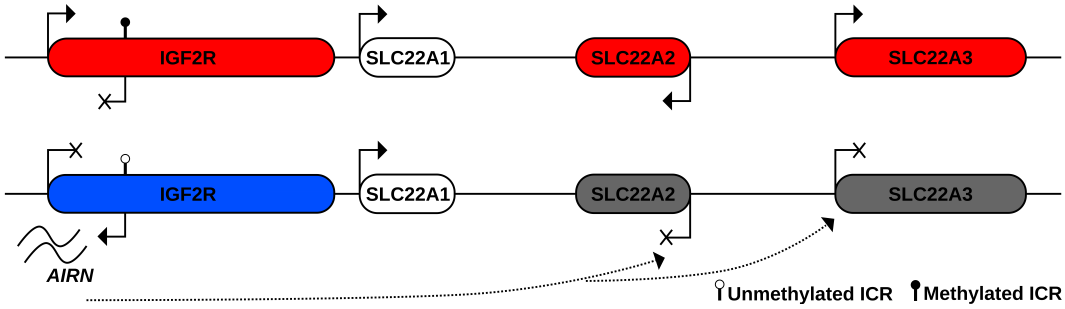


Figure 1.6: Transcription of *AIRN* regulates transcriptional gene silencing of the *IGF2R* gene cluster. Continuous *AIRN* transcription silences *IGF2R* by transcriptional overlap of *IGF2R* promoter. By some unknown mechanism, the *IGF2R* promoter is irreversibly methylated. Silencing of *SLC22A2* and *SLC22A3* occurs through *AIRN*-mediated recruitment of chromatin modifiers to their promoters. Maternal allele (red), paternal allele (blue), silenced genes (gray), and non-imprinting genes (white). Arrows denote direction of transcription.

interfering with PTB recruitment as shown in **Figure 1.5**. In doing so, *FGFR2-AS* recruits PRC2 (polycomb repressive complex 2) and KDM2a, a histone demethylase, to modulate splicing. Furthermore, demonstrating that this process is dependent on chromatin remodeling suggesting a central role in tissue-specific alternative splicing for *FGFR2-AS*.

1.6.2 Airn

Airn (antisense to *Igf2r* RNA non-coding), a 108 kbp paternally expressed lncRNA, is responsible for the silencing of the three maternally expressed protein-coding genes, *Igf2r*, *Slc22a2*, and *Slc22a3* within the *Igf2r* gene cluster. Expression of *Airn* is responsible for the silencing of overlapping *Igf2r* and non-overlapping *Slc22a2* and *Slc22a3* in *cis* [69] as shown in **Figure 1.6**. For *Slc22a2* and *Slc22a3*, *Airn* recruits chromatin modifiers in a sequence-specific manner to their promoters [84]. Interestingly, *Airn* must be continuously transcribed to silence overlapping *Igf2r* until the *Igf2r* promoter is irreversibly silenced by CpG methylation, which is sufficient to maintain the imprint [85, 86]. How chromatin and DNA modifiers are recruited to the *Igf2r* promoter, and what role the antisense transcript

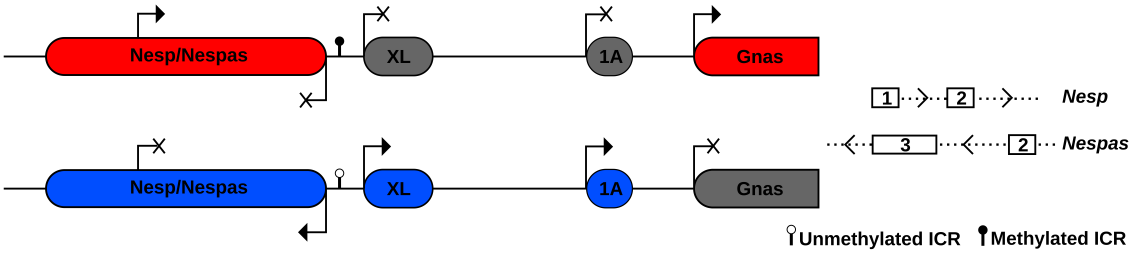


Figure 1.7: *Nespas* overlapping transcription occludes the *Nesp* promoter, promoting CpG methylation silencing paternal expression of *Nesp*. Maternal allele (red), paternal allele (blue), silenced genes (gray). Arrows denote direction of transcription. Zoom view of overlapping exons of *Nesp* and *Nespas*.

plays, if any, in its recruitment remains unclear.

1.6.3 Nespas

Nespas, a 27 kbp lncRNA, is a paternally expressed antisense lncRNA belonging to the *Gnas* imprinting cluster containing four sense transcripts: *Nesp*, *Gnasxl*, *Exon1A*, and *Gnas* [87–90]. *Nesp* is maternally expressed in all tissues [91, 92], and imprinted expression is controlled by paternally methylated *Nesp* DMR and maternally methylated *Nespas-Gnasxl* DMRs, which contains promoters for *Nespas* and *Gnasxl* [93, 94]. The paternal restricted expression of *Nespas* is due to methylation of the maternal *Nespas* promoter [95,96]. **Figure 1.7** depicts the silencing of paternal *Nesp* via *Nespas* transcriptional overlap of the *Nesp* promoter resulting in the recruitment of H3K4me3, which methylates the *Nesp* promoter [90, 97–100].

1.6.4 Nudt6

Nudt6 (nudix[nucleoside diphosphate linked moiety X]-type motif 6), also known as *Fgf-2* antisense transcript (*Fgf2-AS*), a protein-coding gene belonging to the cytosolic Nudix hydrolase gene family, is the transcribed antisense to *Fgf-2*, a heparin-binding growth factor involved in multiple physiological processes including cortical neurogene-

sis [101]. *Nudt6* has multiple isoforms localizing to mitochondria, nucleus and cytoplasm, with four of the isoforms producing proteins between 18 - 35 kDa [102]. The evolutionary conserved *Fgf-2/Nudt6* locus shows reciprocal expression that is tightly balanced via chromatin remodeling factors [103, 104]. Additionally, the partial overlap of *Fgf-2* 3' UTR (untranslated region) can inhibit *Fgf-2* mRNA expression in the absence of *Nudt6* translation by initiating Ago2-dependent pathways, reducing *Fgf-2* mRNA stability and translation efficiency [105] as shown in **Figure 1.8**. Altogether, *Nudt6* demonstrates both protein-coding and lncRNA function.

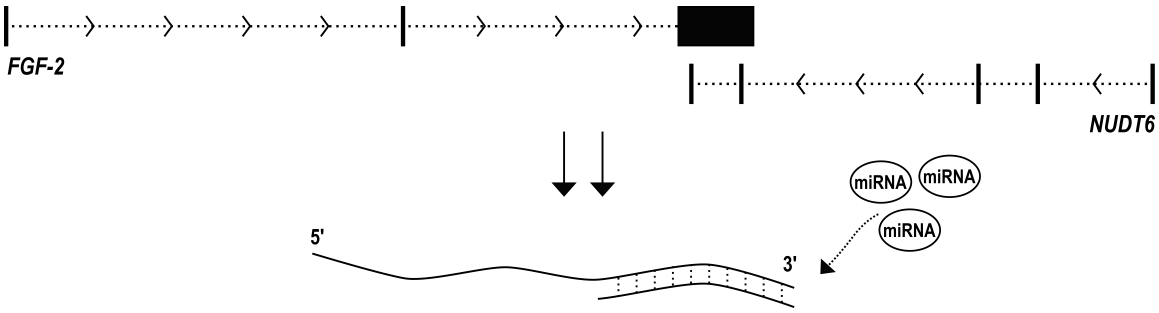


Figure 1.8: The partial overlap of *FGF-2* by the protein-coding gene *NUDT6* regulates *FGF-2* expression by forming double-stranded RNA duplexes reducing stability and translation efficiency. Arrows denote direction of transcription.

1.6.5 Kcnq1ot1

Kcnq1ot1, also known as *Lit1*, is a 92 kbp lncRNA that emerges from intron 11 of *Kcnq1* [106, 107]. The *Kcnq1* imprinting cluster encompasses 10-12 imprinted maternally expressed protein-coding genes. **Figure 1.9** shows that methylation of the maternal *Kcnq1ot1* promoter restricts expression of the antisense lncRNA to the paternal allele [108]. Currently, there are two hypotheses for the imprinting mechanism of the *Kcnq1* imprinting cluster: (1) direct silencing by *Kcnq1ot1* recruits and propagates repressive fac-

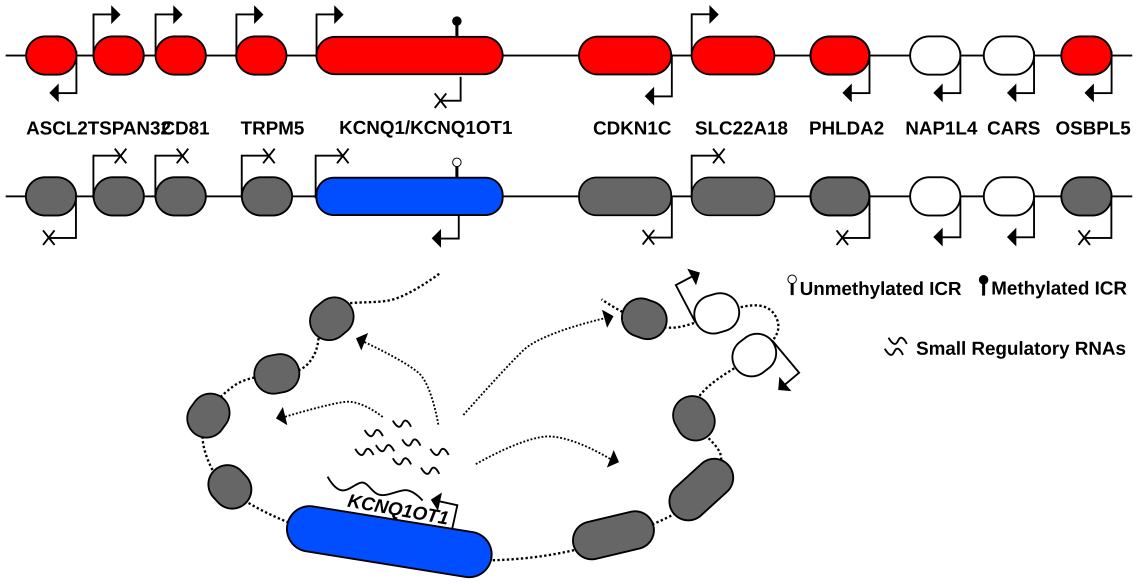


Figure 1.9: Long range silencing of the *KCNQ1* imprinting gene cluster is due to *KCNQ1OT1*. It is proposed that processing of *KCNQ1OT1* results in small regulatory RNAs, which interact with the imprinted genes causing silencing. Maternal allele (red), paternal allele (blue), silenced genes (gray), and non-imprinted genes (white). Arrows denote the direction of transcription.

tors [109], (2) the regulatory elements produced by the transcription of *Kcnq1ot1* recruits and propagates chromatin repressive factors [110, 111]. While there is evidence to support both hypotheses, recent work by Schultz *et al.* suggest that it is the regulatory elements generated from the transcription of *Kcnq1ot1* that are responsible for silencing the *Kcnq1* imprinting cluster [112].

By integrating publicly available sequencing data for the *Kcnq1ot1* region, Schultz *et al.* observed extensive processing of *Kcnq1ot1* resulting in enhancer- and promoter-associating RNAs. These sequences line up to the poly(A)-sequencing sites observed in the mm9 UCSC genome assembly [113] generated at Merck Research Laboratories. Additionally, Schultz *et al.* confirmed multiple independent ncRNAs transcribed from the *Kcnq1ot1* region using a combination of 5 rapid amplification of cDNA ends (RACE) and

chromosome conformation capture (3C) assay, and demonstrated that at least one of these RNAs directly interacted with the *Kcnq1* promoter in the heart. Furthermore, when this transcription-rich region was deleted, imprinting was lost in the *Kcnq1* imprinting cluster, altogether suggesting the region acts as a long-distance silencer.

1.6.6 BACE1-AS

The antisense transcript of *BACE1* (*BACE1-AS*), a ∼2 kbp lncRNA, is transcribed from the *BACE1* gene (beta-site amyloid precursor protein (APP)-cleaving enzyme 1), which is central to the pathogenesis of Alzheimer's disease. The lncRNA is a fully processed transcript that is highly expressed in Alzheimer's affected brains with two polyadenylated splice variants observed in human and mouse [114]. Additionally, *BACE1-AS* prevents microRNA-induced translational repression by competing with miR-485-5p binding of *BACE1* in a tightly regulated system for BACE1 protein [115]. When *BACE1-AS* binds to *BACE1*, it forms double-stranded RNA (dsRNA) duplexes that increase the stability of *BACE1* as shown in **Figure 1.10**. As a result, elevated levels of *BACE1-AS* increase *BACE1* expression creating a post-translational feed forward loop.

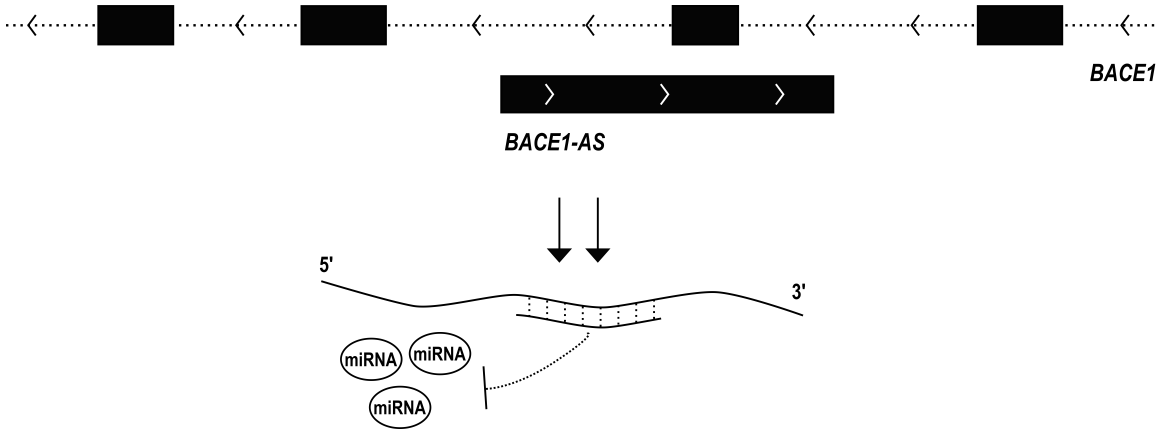


Figure 1.10: *BACE1-AS* increases stability of *BACE1* via dsRNA duplexes. This increase in *BACE1* stability results in increased protein levels of BACE1 creating a post-translational feed forward loop. Arrows denote the direction of transcription.

1.6.7 BDNF-AS

The antisense transcript to brain-derived neurotrophic factor (*BDNF-AS*) is a 191 kbp transcript with twelve splicing variants [116]. The first four exons of *BDNF-AS* are downstream *BDNF*, while the remaining exons overlap coding and introns of *BDNF* [116]. The *BDNF* gene, which plays an important role in peripheral neurons, neuron size, and arborisation, is a complex gene with 11 exons and 9 unique promoters resulting in 17 spliced transcripts with different 5' and 3' UTRs [116, 117]. The partially conserved antisense transcript forms dsRNA duplexes with *BDNF* mRNA in the brain resulting in the recruitment of EZH2 (enhancer of zeste homolog 2) and PRC2 to the promoter of *BDNF* [117] as depicted in **Figure 1.11**. Thus, knockdown of the antisense results in increased mRNA and protein levels of BDNF.

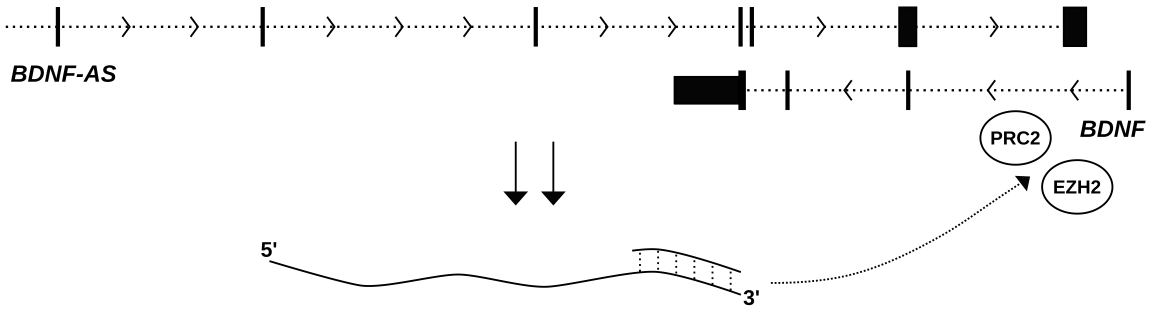


Figure 1.11: *BDNF-AS* regulates *BDNF* gene expression by forming double-stranded RNA resulting in the recruitment of chromatin modeling factors to the promoter of *BDNF*. Arrow denotes the direction of transcription.

1.6.8 Ube3a-AS

The *Ube3a* antisense transcript is a part of a paternally expressed large transcriptional unit ($>$ 1000 bp) that initiates upstream of the PWS-IC from the *Snurf/Snrpn* promoter [31, 33]. An unusual result of *Ube3a-AS* being a part of this large transcriptional unit is its

lack of a unique promoter. As such, its regulatory control for neuronal specific expression remains unclear. Moreover, how or why *Ube3a-AS* regulates *Ube3a* is still a mystery as very little is known about *Ube3a-AS*, besides the fact that it is sufficient to imprint *Ube3a* [35, 118, 119].

There is no methylation at the promoter of *Ube3a* or anywhere else within paternal *Ube3a* [118, 120, 121]. In fact, paternal *Ube3a* has been shown to be expressed; its transcription terminates between exon 4 and 5 [36, 122]. Although paternal *Ube3a* is not currently known to produce a transcript, this partial expression is in stark contrast to the other well-studied antisense lncRNAs. One explanation of *Ube3a* sense/antisense expression is that the transcriptional machinery for both transcripts collide causing transcription to terminate between exon 4 and 5 [36]. In this model depicted in **Figure 1.12A**, *Ube3a-AS* transcription would generate high levels of torsional stress leading to stalling of transcriptional elongation complexes and silencing of *Ube3a*.

With this model, it would be expected that transcription would stall at different places throughout *Ube3a*; however, this is not the case. In two independent studies, biallelic expression of *Ube3a* ends at one specific location [36, 122]. Additionally, Numata *et al.* demonstrated *Ube3a-AS* expression upstream *Ube3a* using their SNP analysis, which indicates that *Ube3a-AS* continues transcription beyond the suggested collision point. The termination of *Ube3a-AS* upstream *Ube3a* aligns with polyadenylation sites from Merck Research Laboratories poly(A)-sequencing data observed in mm9 UCSC Genome Browser [113]. Moreover, the transcriptional termination of biallelic expression of *Ube3a* also aligns with Merck poly(A)-sequencing data. Altogether, these observations suggest an alternative mechanism of silencing, where the transcription of *Ube3a-AS* leads to alternative polyadenylation within the intron between exon 4 and 5 of *Ube3a* that terminates transcription shown in **Figure 1.12B**.

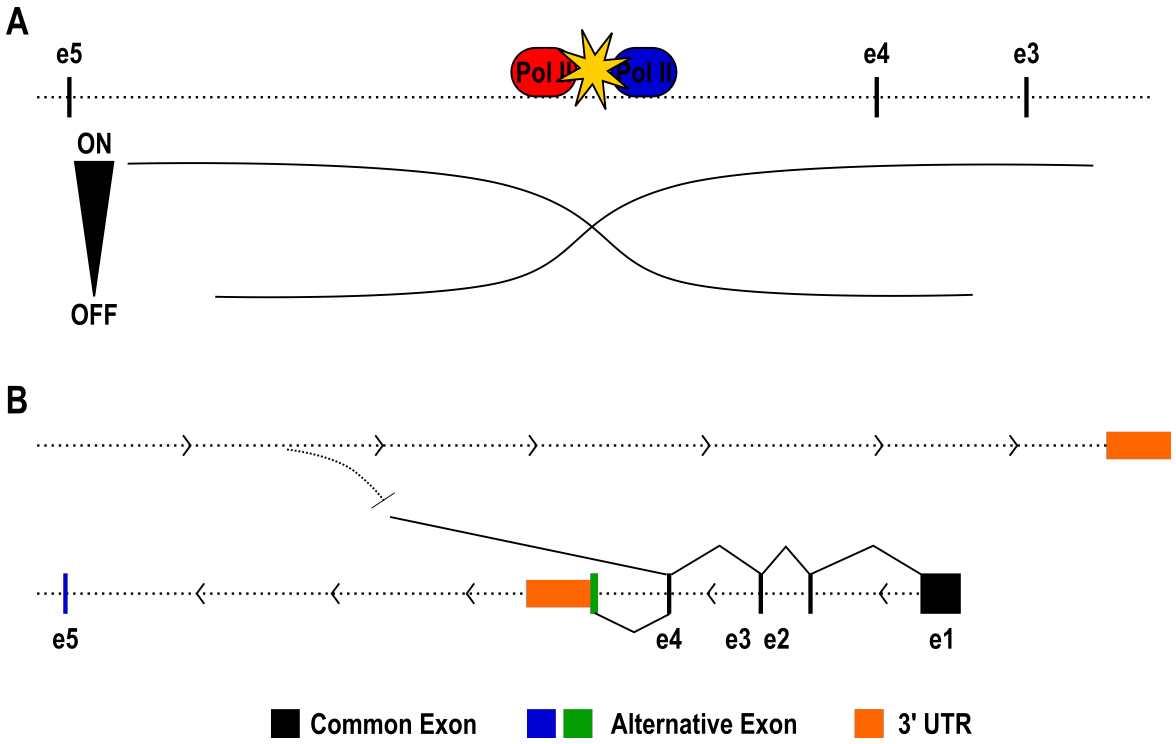


Figure 1.12: *UBE3A-AS* regulates paternal *UBE3A* expression in neurons. **A)** Transcriptional collision model for *UBE3A-AS* regulation of *UBE3A*. **B)** Purposed alternative splicing model for *UBE3A-AS* regulation of *UBE3A*. Polymerase II (Pol II), antisense Pol II in red and sense Pol II in blue. Arrows denote the direction of transcription.

1.7 Concluding Remarks

This paper has reviewed the importance of *UBE3A* and suggested a possible function for its antisense transcript outside of imprinting of *UBE3A* as a reason for its imprinting. Recent studies in our laboratory have demonstrated that *Ube3a* is not imprinted to regulate its gene expression in neurons [54]. Moreover, the imprinting of *Ube3a* has no overall effect on *Ube3a* expression suggesting that the importance of imprinting *Ube3a* may lie in its antisense transcript, *Ube3a-AS* [54]. Long non-coding RNAs, like *Ube3a-AS*, are diverse in structure and function. It is possible that *Ube3a-AS* also functions in a complex manner. It is clear from this review that more investigation is needed to elucidate *Ube3a-*

AS function and its connection to *Ube3a* imprinting.

The implications of *UBE3A-AS* having a function impacts therapeutic intervention for the diseases of the area, Angelman syndrome, Prader-Willi syndrome and Chromosome15q duplication syndrome. Specifically, for AS where the only current treatment options target the reactivation of paternal *UBE3A* via disruption of *UBE3A-AS* [36, 52, 123–127], any possible function of the antisense transcript will need to be extensively considered. As such, a better understanding of imprinting of *UBE3A* may help facilitate drug development for Angelman syndrome, while possibly mitigating the ramifications for transcriptional silencing of *UBE3A-AS*.

2. THE *UBE3A* ANTISENSE TRANSCRIPT UNDERGOES EXTENSIVE PROCESSING AND IS SPATIOTEMPORALLY REGULATED IN THE BRAIN

2.1 Overview

Human chromosome 15q11-q13 contains a cluster of imprinted genes that are associated with a number of neurodevelopmental disorders that exhibit non-Mendelian patterns of inheritance due to genomic imprinting, including Angelman syndrome (AS), Dup15q syndrome, and Prader-Willi syndrome (PWS). AS is caused by loss of the maternally inherited *UBE3A* allele, whereas PWS is caused by the loss of the paternally inherited *SNORD116* snoRNAs, which are expressed as part of a long polycistronic transcription unit (PTU) comprised of *SNRPN*, additional snoRNA clusters, and the *UBE3A* antisense transcript (*UBE3A-AS*). The PTU is imprinted with paternal-specific expression, and its antisense portion exclusively expressed in neurons. As a result, *UBE3A* is imprinted in neurons and biallelically expressed in all other cell-types. Why *UBE3A-AS* evolved to imprint *UBE3A* in neurons is largely unknown. In this study, we examined the transcriptional profiles and processing of the mouse and human antisense transcripts towards understanding the functional significance of *UBE3A* imprinting by *UBE3A-AS*. Our findings show that the *UBE3A-AS* is extensively processed via 5' capping, 3' polyadenylation, and alternative splicing, giving rise to a myriad of transcripts that are spatiotemporally regulated. Based on our findings, we propose that processing of the *UBE3A-AS* generates a diverse repertoire of regulatory RNAs in neurons.

2.2 Introduction

Human chromosome 15q11-q13 contains a cluster of genes that are associated with a number of neurodevelopmental disorders exhibiting non-Mendelian patterns of inheritance due to genomic imprinting. Angelman syndrome (AS) - characterized by intellectual

disability, ataxia, epilepsy, and an atypical happy disposition - is caused by mutations or epimutations leading to loss-of-function or loss-of-expression of the maternally inherited ubiquitin protein ligase E3A (*UBE3A*) gene [34, 128, 129]. Maternal-derived interstitial or isodicentric copy number gains of 15q11-q13 cause Dup15q syndrome, which is characterized by intellectual disability, ataxia, epilepsy, sleep disorder, and autism spectrum disorder [20, 130]. Although Dup15q is a contiguous gene disorder, overexpression of UBE3A in the brain is believed to be the principal mechanism underlying the symptoms associated with the condition [131]. Paternally inherited deletions of 15q11-q13, namely those involving the C/D box small nucleolar RNA (snoRNA) *SNORD116*, cause Prader-Willi syndrome (PWS), which is characterized by dysregulated hunger and satiety, thermoregulation, sleep disorder, and behavioral issues [132].

Genomic imprinting of the 15q11-q13 region is regulated by the AS and PWS imprinting centers (AS-IC and PWS-IC) [11, 51, 118, 133, 134]. Studies to date indicate that the AS-IC negatively regulates the PWS-IC [51], while the PWS-IC functions as an enhancer element that positively regulates the expression of genes in the region [46]. On the maternal chromosome, the AS-IC is active and thus represses the expression of the genes controlled by the PWS-IC. On the paternal chromosome, repressive histone modifications and DNA methylation inactivate the AS-IC allowing for the PWS-IC to regulate the expression of its target genes on the paternal chromosome, which including the polycistronic transcriptional unit (PTU) comprised of *SNURF/SNRPN*, and clusters of tandemly repeated C/D small nucleolar RNAs (*SNORD107, SNORD64, SNORD108, SNORD109A, SNORD116*, and *IPW*). In the brain, transcription extends downstream of *IPW* to include additional tandemly duplicated snoRNAs (*SNORD115* and *SNORD109B*) and the *UBE3A* antisense transcript (*UBE3A-AS*, also known as *UBE3A-ATS*). Likewise, the mouse functional equivalent of the PWS-IC regulates the expression of a PTU comprised of *Snurf/Snrpn*, clusters of C/D box snoRNAs (*Snord64, Snord116, Snord115*,

Ipw, and *Ube3a-AS*). But unlike in humans, expression of *Snord116, Ipw, Snord115,* and *Ube3a-AS* is brain-specific [31, 135]. As such, the imprinting of *UBE3A/Ube3a* is brain-specific, and aside from this role in imprinting there is no other function ascribed to it.

Studies in mouse and human have shown that expression of *Ube3a-AS/UBE3A-AS* transcript is both necessary and sufficient to silence expression of the paternal *Ube3a/UBE3A* allele *in cis* [36, 136]. But unlike most imprinted genes regulated by an antisense transcript, the *Ube3a-AS/UBE3A-AS* is believed to inhibit transcriptional elongation rather than transcriptional initiation, as the paternal *Ube3a* allele is modified with active epigenetic modifications, bound by RNA polymerase II, and transcribed to a region in intron 4 [36]. As such, Meng *et al.* (2013) proposed a collision model for the imprinting of *Ube3a* in neurons, wherein *Ube3a* and *Ube3a-AS* expression decreases within intron 4 due to collision of the RNA polymerases. This model, however, conflicts with reports detecting *Ube3a-AS* expression upstream *Ube3a* [122]. As such, this study sets out to investigate the expression profile of *Ube3a-AS/UBE3A-AS* as a means to understand the function of imprinting in neurons. Here, were report that *Ube3a-AS/UBE3A-AS* is a remarkably complex transcript that is extensively processed through 5' capping, alternative splicing, and 3' polyadenylation, which are differentially regulated among brain regions and during brain development.

2.3 Materials & Methods

2.3.1 Bioinformatics

2.3.1.1 Public data, genomes and annotations

Publicly available data

The analysis performed in this chapter was conducted with publicly available data downloaded from the European Nucleotide Archive, and can be viewed by the accession number (`http://www.ebi.ac.uk/ena/data/view/<accession>`). Mouse tis-

sue data was from 8 wk adults [137], while adult human data was of unknown age and origin for Human Protein Atlas (ERP003613) [138], and an average of 52.3 ± 7.9 year-old for the SRP072463 study [139]. The cellular populations in the mouse cerebral cortex dataset were purified with various purification methods [140]. Temporal hippocampal RNA-seq datasets were extracted from E18, P1, P10 and P30 mice [141]. A breakdown of tissue types, strain, and accession numbers is supplied in **APPENDIX B, Table B.1** for mouse data and **APPENDIX B, Table B.2** for human data. A complete list of publicly available RNA-seq datasets used in this chapter is provided in **Table 2.1**.

Table 2.1: Public Data: RNA-seq information

Study	**Instrument**	**Layout**	**Stranded**	**Species**
ERP000591	Illumina Genome Analyzer	PE	No	*Mus musculus*
SRP012040	Illumina HiSeq 2000	PE	Yes	*Mus musculus*
SRP033200	Illumina HiSeq 2000	PE	No	*Mus musculus*
SRP048593	Illumina HiSeq 2500	SE	Yes	*Mus musculus*
ERP003613	Illumina HiSeq 2000	PE	No	*Homo sapiens*
SRP072463	Illumina HiSeq 2000	PE	Yes	*Homo sapeins*

Genomes and annotation sets

Throughout this work, we used the February 2009, Genome Reference Consortium Human Reference 37 (GRCh37, hg19) human genome assembly [142] and the July 2007 finished mouse genome NCBI Build 37 assembly [143] (mm9). Annotations were collected from Illumina iGenomes collection for hg19 and mm9 last downloaded from UCSC on July 17, 2015 [144].

PolyA-seq data

PolyA-seq data files from Derti *et al.*, 2012 [145] were downloaded from UCSC. The sites clustered BED files from mouse and human (hg19) were separated by strand with `awk` (version 4.0). These files were than viewed with IGV (version 2.3.90 [146, 147]).

CAGE-seq data

CAGE (Capped Analysis of Gene Expression) sequencing bed files generated from the FANTOM5 consortium [148] were downloaded from `http://fantom.gsc.riken.jp/5/datafiles/latest/extra/CAGE_peaks/`. Similar to polyA-seq data, files were separated by strand with `awk` and viewed in IGV.

2.3.1.2 Data processing

Quality of downloaded raw fastq files were checked with FastQC [149] (version 0.11.5). As no read files failed initial quality control, adapter sequences and low quality reads (quality score ≤ 3) were trimmed with Trimmomatics [150] (version 0.36). Using the program's TruSeq3-PE-2.fa adapter file, and minimal length of 25. These trimmed paired- and single-end reads were used by Hisat2 [151,152] (version 2.0.4) to align to chromosome 15 (chr15) for human and chromosome 7 (chr7) for mouse data. The alignment was assisted with Hisat2 python provide extraction scripts for splice sites and exons within chr15 and chr7, human and mouse alignments respectively. The SAM file outputs were directly pipped into SAMtools [153] (version 1.3.1) to convert to BAM format (`samtools view`) and sorted (`samtools sort`) outputting only sorted BAM files to be indexed (`samtools index`) and merged (`samtools merged`). The sorted BAM files were assembled with StringTie [151, 154] (version 1.3.3). Annotation files using stranded, high-depth reads (SRP01204, SRP072463) were generated via `stringtie` *de novo* method for mouse and human datasets. These GTF annotation files were merged with `stringtie --merge`

using FPKM thresholds of 5 and 10 for mouse and human data respectively. Mouse data was additionally restricted with isoform fraction equal to 0.05. These annotations were used for downstream analysis like `featureCounts` [155] - an Rsubread [156] (version 1.24.1) utility.

2.3.1.3 Data analysis

Visual analysis

All visualization was conducted with IGV. Novel transcript annotation (GTF) in the antisense direction (minus - mouse, plus - human) were visualized along with polyA-seq and CAGE-seq brain-specific annotations (BED). For *UBE3A/Ube3a* antisense specific visualization, potential transcripts were extracted using a combination of `awk` and `grep`. For splice junction visualization, sorted BAM files merged by tissue and study plotted using the Sashimi plot function within IGV. Here, BAM files were limited to the antisense direction of *UBE3A/Ube3a*.

Differential transcript expression

For differential expression on the transcript and exon level, the edgeR [157, 158] (version 3.16.5) package download from Bioconductor [159, 160] (version 3.4, R [161] - version 3.3.2) was used in conjunction with `featureCounts`. The `featureCounts` produced read counts were used to generate an `DGEList` object for downstream analysis. The data was filtered based on counts per million (CPM) greater than 1 for 25% of the samples. Data was than normalized by library size with `calcNorm- Factors` an edgeR function and the negative binomial dispersion estimated by weighted likelihood empirical Bayes [157, 162]. A negative binomial generalized log-linear model was fitted to the data based on experimental design with `glmFit` [163]. Differential expression was statistically tested on the isoform/transcript and exon levels using `glmLRT` [163] for One-Way ANOVA-like statistics and `exactTest` [164] for pairwise comparisons of group

means. The `topTags` function [164, 165] was used to adjust p-values with the Benjamini & Hochberg method (FDR) [166]. Finally, `diffSpliceDGE` was used to test for differential splicing and exon usage.

SNP analysis

As F1 hybrid mice (maternal, C57BL/6J and paternal, DBA/J2) RNA-seq data was used, informative SNPs were extracted from the data using SAMtools `samtools mpileup` and BCFtools (version 1.3.1) snp calling and conversion functions (`bcftools call`, `bcftools view`) on sorted, indexed BAM files. A list of six SNPs from the region of interest was downloaded from the Mouse Genomes Project - Query SNPs [167, 168] after conversion of mm9 genomic coordinates to mm10 via LiftOver - an UCSC tool (`http://genome.ucsc.edu/cgi-bin/hgLiftOver`). The `-r` option for `samtools mpileup` was used to specify region of interest (chr7:66,439,800-66,808,000) for SNP identification.

2.3.2 Molecular

2.3.2.1 Animals

Mice were purchased from The Jackson Laboratory (Bar Harbor, ME) and housed in mouse facility under the specific-pathogen-free barrier conditions. All procedures were performed according to NIH guidelines and approved by the Texas A&M University Institutional Animal Care and Use Committee.

2.3.2.2 Rapid amplification of cDNA ends: 3' polyA

3' polyA rapid amplification of cDNA ends (3' RACE; 18373019, Life Technologies, Carlsbad, CA) was performed according to the manufacture's protocol using total RNA isolated from cortex and cerebellum of adult (10 wk) male C57BL/6J mice (000664, The Jackson Laboratory) and FirstChoice Human Brain Reference Total RNA (6050,

Life Technologies). Briefly, mouse RNA was isolated using TRIzol reagent (15596018, Thermo Fisher Scientific, Waltham, MA) according to the manufacturers protocol and then DNase treated using TURBO DNA-free kit (AM1907, Life Technologies). 3' RACE primers were designed using an annotation of *Ube3a-AS/UBE3A-AS* (personally communication). The primers were directed towards the 3' ends of *Ube3a-AS/UBE3A-AS* and upstream of the polyadenylated sites predicted by Derti *et al.* [145] (**APPENDIX B, Tables B.3 and B.4**). 3' RACE was performed using the AUAP universal primer and primers specific to the 3' ends of *Ube3a-AS/UBE3A-AS*. PCR amplicons were gel extracted (D4008, Zymo Research, Irvine, CA) and cloned into the PCR4-TOPO TA vector (K458001, Life Technologies). Vectors containing PCR inserts were identified by restriction digests of the plasmids (EcoRI-HF, R3101M, New England BioLabs Inc., Ipswich, MA). Sanger sequencing of the RT-PCR amplicons was performed at the Texas A&M University Gene Technologies Laboratory (`http://www.idmb.tamu.edu/gtl/`). Sequences were visualized in IGV (`BLAT`) and exported as BED files to be binded together with `paste` (Bash), sorted with `sortBed` and merged with `bedtools merge` - BEDtools [169] (version v2.25.0) utilities. The merged BED files were visualized with polyA- and CAGE-seq data in IGV.

2.3.2.3 Quantitative RT-PCR

Quantitative RT-PCR (qPCR) was used to examine differential expression of *Ube3a-AS* in cortex, cerebellum, and hippocampus of adult (10 wk) male mice (C57BL/6J, n = 4). Tissues were dissected and flash frozen with liquid nitrogen, and then RNA was isolated and processed as described above. Reverse transcription was performed using the SuperScript IV First Strand Synthesis System (Life Technologies) with oligo(dT) primer. The qPCR assays were performed in triplicate using SYBR-Green reagent (11760500, Life Technologies) on a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster

City, CA). A list of primers is provided in **APPENDIX B, Table B.4**. Statistical significance was determined for the qPCR using two-way ANOVA in R (`aov`) and Tukey's HSD multiple comparison post-hoc analysis was performed (`TukeyHSD`) with default parameters.

2.3.3 Charts

All charts were generated in R using the `ggplot2` library and the `pdf` function - `devtools` library.

2.4 Results

2.4.1 The antisense transcript of UBE3A/Ube3a is a highly processed transcript producing multiple isoforms

The genome annotation for mouse was generated using high read depth, stranded data, merged from cortex, cerebellum, and frontal lobe (C57BL/6J, n = 2). Similar, the human genome annotation was generated using high read depth, stranded data from merged Brodmann area 4 (motor cortex, n = 4). In both cases, several transcripts were assembled in the antisense direction of *UBE3A/Ube3a* (mouse = 18, human = 10) that overlapped the sense *UBE3A/Ube3a* (**Figure 2.1**). In the case of the mouse, several transcripts terminated upstream *Ube3a* (**Figure 2.1A**). Furthermore, both annotations revealed transcriptional interconnection between the upstream *SNORD115/Snord115* cluster[1].

To investigate processing of the antisense transcripts, we applied publicly available polyA-seq and CAGE-seq in conjunction with 3' RACE (**Figure 2.2**). In the antisense direction of *UBE3A/Ube3a*, polyadenylation sites were identified with several verified by 3' RACE sequence data (**APPENDIX B**). Analysis of antisense direction CAGE data revealed 5' capped sites within the antisense region. Combined with the polyadenylation data, this suggested that the antisense transcripts are being processed into smaller RNAs.

[1]The *Snord115* cluster is unannotated in mm9.

Figure 2.1: Antisense transcript of *Ube3a/UBE3A* is alternatively spliced in the brain. **A.** Schematic of mouse *Ube3a-AS*. Data generated from Pervouchine *et al.* (2015). **B.** Schematic of human *UBE3A-AS*. Data generated from Lin *et al.* (2016).

As the *UBE3A/Ube3a* region is imprinted, it was important that we identify allelic origin of expression. To this end, the hybrid mice - sequenced by Sanger Institute Collaboration - were used to identify informative SNPs in the region. Of the six SNPs in the region, five of them had approximately 97.8% expression from the paternal (DBA) allele at $\sim$52.1% coming from the reverse strand (**APPENDIX B, Table B.5**). Suggesting that these antisense transcripts are being expressed from the paternal allele.

In addition to the 3' and 5' processing of these antisense transcripts, we observed numerous splicing and alternative splicing in the antisense direction with Sashimi plots within the mouse cortex (**Figure 2.3A**), and within the human Brodmann area 4 (**Figure 2.3B**).

We used publicly available RNA-seq [137, 139], polyA-seq [145], CAGE-seq [148] data to characterize the expression patterns, splicing, 3' polyadenylation, and 5' capping of the antisense transcript of *UBE3A/Ube3a* in mouse and human. Altogether, these data indicate that the antisense transcripts for mouse and human are processed into multiple transcriptional units through alternative splicing, 5' capping, and 3' polyadenylation. The presence of 5' capped transcripts at exonic and intronic regions also suggests post-transcriptional processing.

2.4.2 The UBE3A-AS/Ube3a-AS *is brain-specific and highly expressed in neurons*

We next examined the expression profile and patterns of the antisense transcript among mouse and human tissues and among individual populations of mouse cerebral cortex cell-types using RNA-seq data [138, 140]. Alternative splicing in the antisense direction nearly disappeared completely in non-brain tissues in both mouse (**APPENDIX B, Figure B.1**) and human (**APPENDIX B, Figure B.2**). Analysis of isoform expression revealed that all of the antisense transcripts were downregulated in heart, liver and lung compared to hippocampus (mouse, **APPENDIX B, Figure B.3A-C**) or cortex (human, **APPENDIX B,**

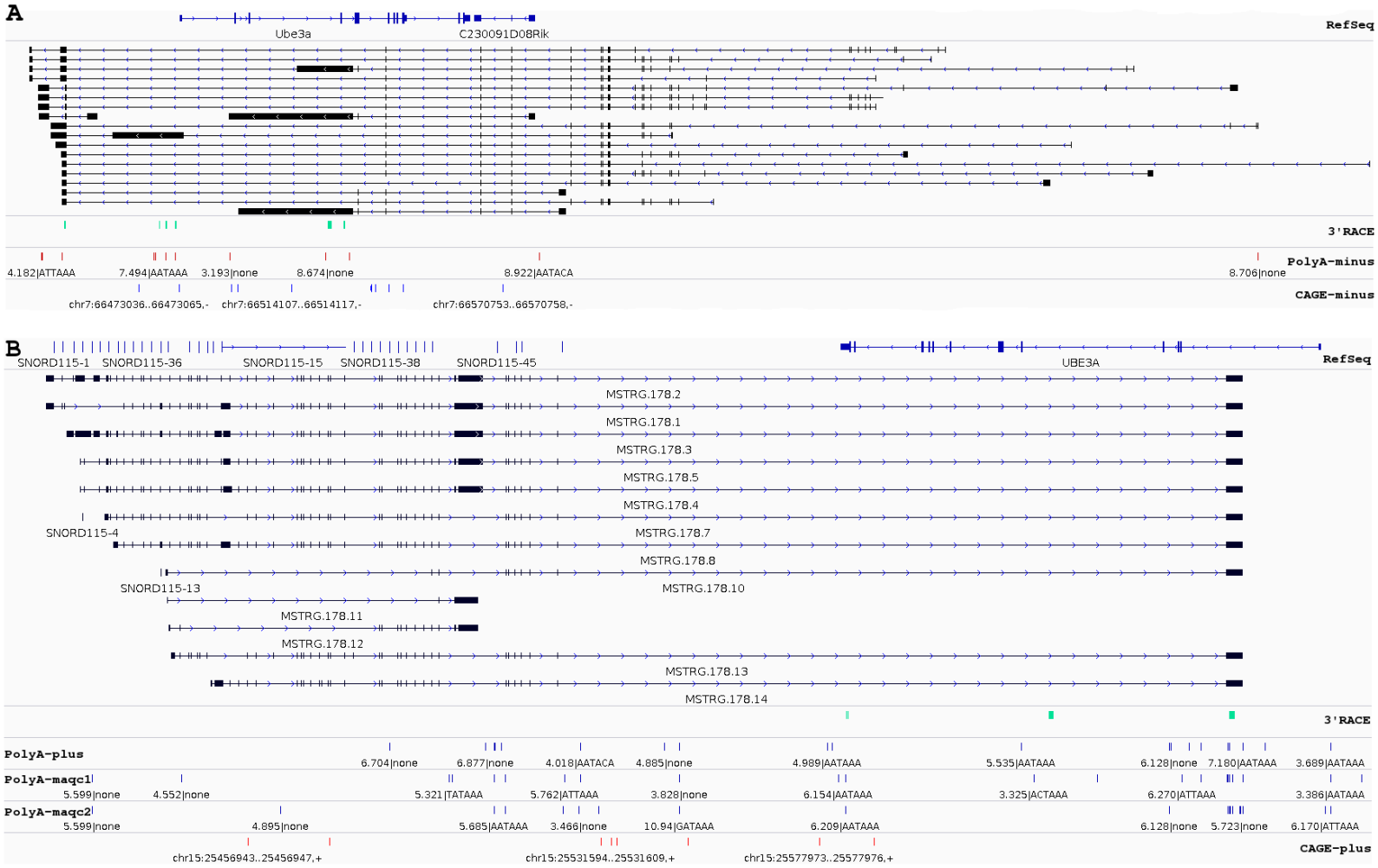


Figure 2.2: *Ube3a-AS/UBE3A-AS* is extensively processed in the brain. **A.** Schematic of mouse antisense transcript polyadenylation sites, 5' capping, and 3'RACE confirming several polyA sites. **B.** Schematic of human antisense transcript polyadenylation, 5' cappin, and 3'RACE. Polyadenylation data generated by Derti *et al.* (2012). 5' capping data generated by Lizio *et al.* (2015).

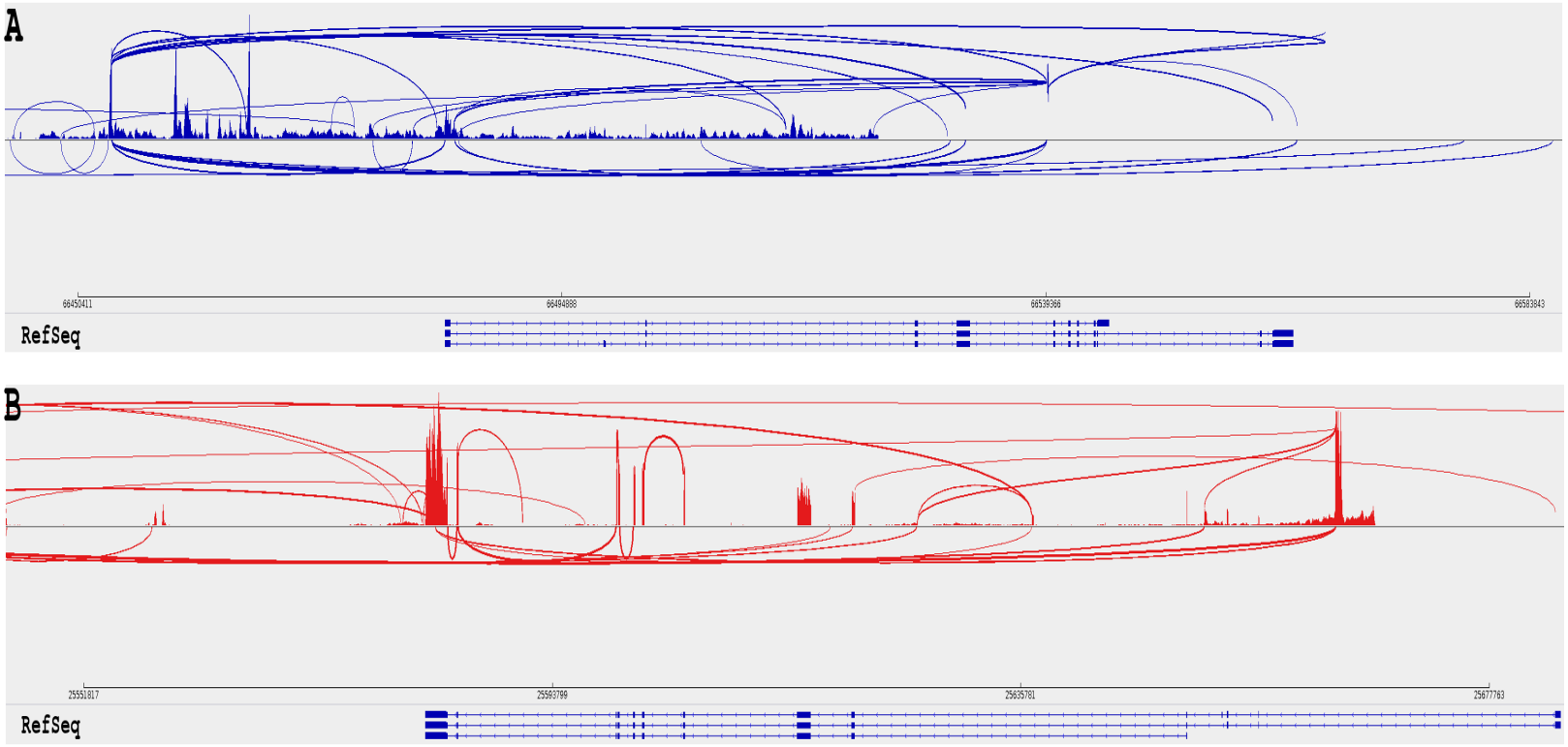


Figure 2.3: Sashimi plots demonstrating numerous splicing and alternative splicing events in the brain of **A.** mouse (cortex, n = 2) and **B.** human (BA4, n = 4). Minimum junction coverage = 5. Data generated from Pervouchine *et al.* (2015) and Lin *et al.* (2016).

Figure B.3D-F). Similar, *Ube3a-AS* transcripts were all downregulated in astrocytes, OPC (oligodendrocytes precursor cells), NFO (newly formed oligodendrocytes), MO (myelinating oligodendrocytes), microglia, and endothelial cells compared to neurons (**APPENDIX B, Figure B.4**, and **B.5**).

As coverage was low for the tissue RNA-seq data, we looked at exon usage between tissues and cell-types of cerebral cortex to determine significant changes in expression. Three general isoform categories were determined for mouse *Ube3a-AS* based on 3'RACE polyadenylation sites, and one for human *UBE3A-AS* based on 3'RACE polyadenylation (**Figure 2.4A**). Using this method, expression were significantly downregulated compared to the hippocampus in mouse (p-value < 0.001, FDR < 0.001; **Figure 2.4B**), and cortex in human (p-value < 0.001, FDR < 0.001; **Figure 2.4C**) with log2 fold-changes all below -2. Similarly, cell-type expression was also significantly downregulated compared to neurons in mouse (p-value < 0.001, FDR < 0.001; **Figure 2.4D**). Exon genomic position are listed in **APPENDIX B, Table B.6**.

Altogether, these data demonstrates that the antisense transcript of *Ube3a/UBE3A* is brain-specific, and that *Ube3a-AS* is also highly expressed in neurons compared to other cell-types in cerebral cortex.

2.4.3 The Ube3a-AS *is spatiotemporally regulated*

We next asked whether the antisense transcripts are differentially regulated among brain regions. Using the same general isoforms categories above (**Figure 2.4A**), we looked at the fold-change comparing cortex to cerebellum and frontal lobe, and frontal lobe compared to cerebellum (**Figure 2.5A**). *Ube3a-AS* isoform 1 (AS Iso1) and *Ube3a-AS* isoform 2 (AS Iso2) were significantly downregulated in cerebellum compared to cortex and frontal lobe (p-value < 0.001, FDR < 0.001), while *Ube3a-AS* isoform 3 (AS Iso3) was significantly upregulated in cerebellum compared to cortex and frontal lobe (p-value $<$

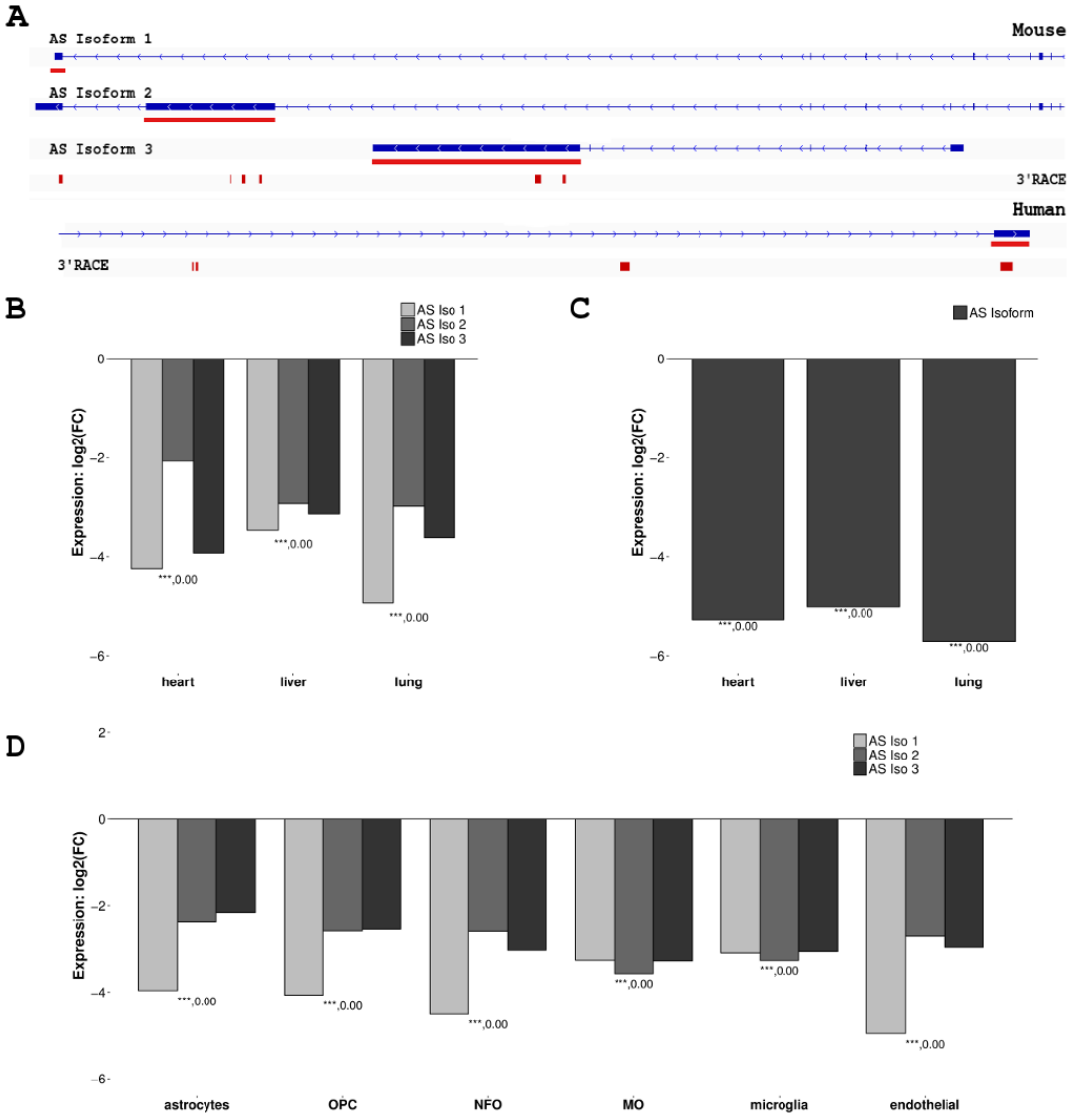


Figure 2.4: *Ube3a-AS/UBE3A-AS* demonstrates brain-specific differential expression and is upregulated in neurons. **A.** Schematic of the exons used for differential exon usage comparison. 3'RACE locations marked below annotations. **B.** Differential expression of *Ube3a-AS* isoforms comparing hippocampus to heart, liver, and lung. Data generated from Sanger Institute hydrid mice data. **C.** Differential expression of *UBE3A-AS* isoform comparing cortex to heart, liver, and lung. Data generated from Uhlen *et al.* (2015). **D.** The three mouse isoforms are downregulated in non-neuronal cell-types compared to neurons. Data generated from Zhang *et al.* (2014). P-value and FDR plotted in **B-C.** *** denotes p-value < 0.001. Abbreviations: OPC - oligodendrocytes precursor cells, NFO - newly formed oligodendrocytes, and MO - myelinating oligodendrocytes.

0.001, FDR < 0.001). There appeared to be no difference between cortex and frontal lobe expression, which is unsurprising given their locations in the brain.

To expand upon our findings that *Ube3a-AS* is differentially expressed among mouse brain regions, we used quantitative RT-PCR to examine the levels of the three general isoforms among adult mouse cerebellum, cortex, and hippocampus (C57BL/6, n=3; **Figure 2.5B**). The levels of each transcript were significantly different among the brain regions (ANOVA, F < 0.001), with significantly higher levels of relative expression for AS Iso1 in cortex and hippocampus compared to cerebellum (Tukey's HSD, p.adj < 0.01; p.adj < 0.01). AS Iso3 had significantly higher levels of relative expression in cerebellum compared to cortex or hippocampus (Tukey's HSD, p.adj < 0.001), which supported the RNA-seq analysis. AS Iso2 was virtually undetectable in all brain regions.

In addition to spatial analysis, we also wanted to see if *Ube3a-AS* was regulated during brain development in the hippocampus (**Figure 2.5C**). Using the general isoforms, we examined the fold-change comparing E18 to P1, P10, and P30 finding that AS Iso2 was significantly upregulated in P30 compared to E18 (p-value < 0.01, FDR < 0.05). When comparing P1 to P10 and P30, we found that expression of AS Iso1 was significantly upregulated compared to P10 and P30 (p-value < 0.01, FDR < 0.05), and AS Iso2 was significantly upregulated compared to P30 (p-value < 0.001, FDR < 0.001). AS Iso3 did not appear to be significantly upregulated.

Altogether, these findings indicate that *Ube3a-AS* is differentially regulated among brain regions and during brain development.

2.5 Discussion

In this study, we investigated expression profiles of the antisense transcript to *Ube3a/UBE3A*, and determined that the *Ube3a/UBE3A* antisense transcript is extensively processed in the brain with 5' capping, 3' polyadenylation, and alternative splicing. In addition to this,

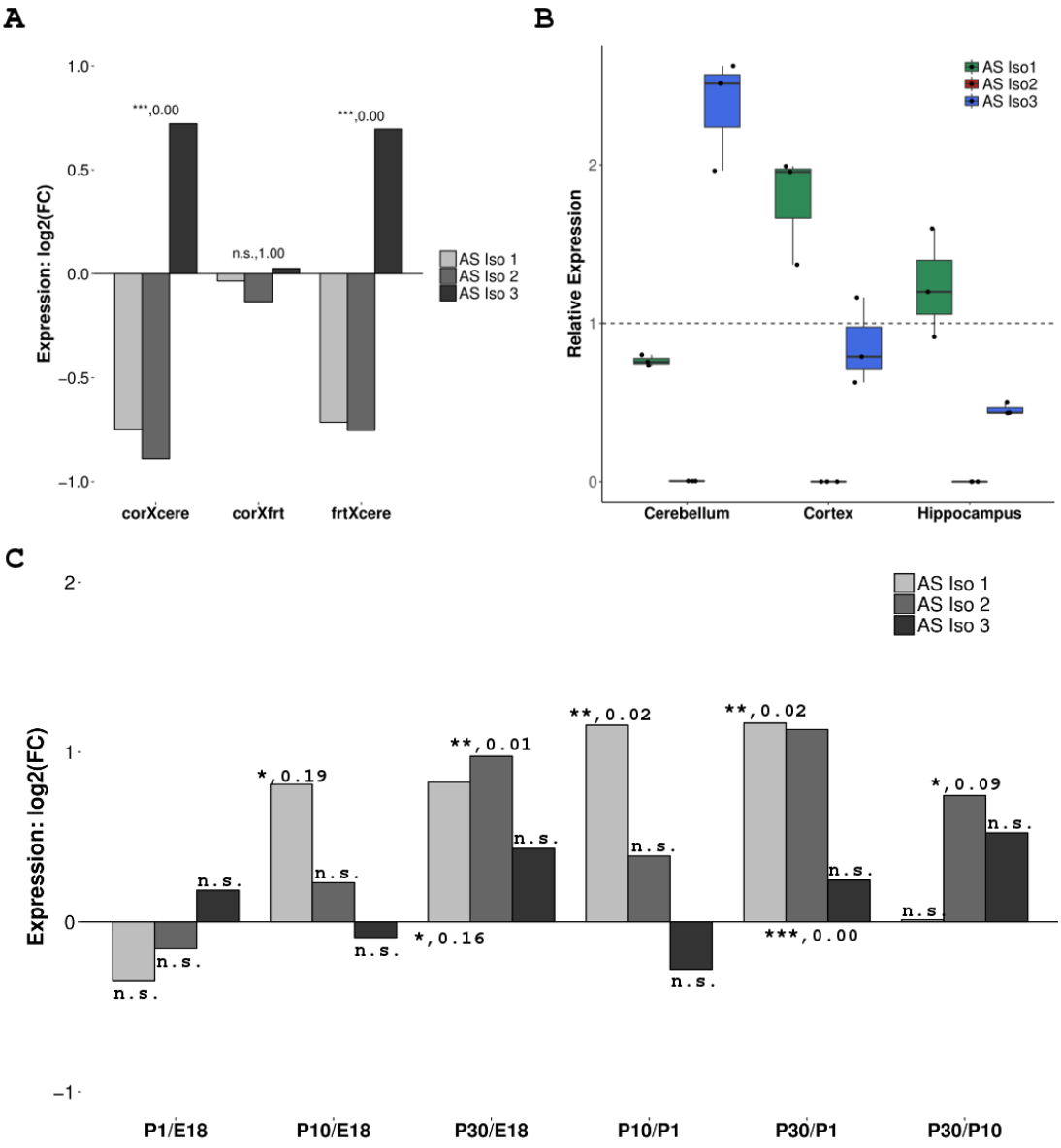


Figure 2.5: *Ube3a-AS* is spatiotemporally regulated in the brain. **A.** Log2 fold-change comparing cortex, cerebellum and frontal lobe (n = 2). Data generated from Pervouchine *et al.* (2015). **B.** qPCR relative expression comparing isoform expression between cortex, cerebellum and hippocampus (n = 3). **C.** Log2 fold-change comparing developmental timepoints (E18, P1, P10, and P30) within the hippocampus (n = 2). Data generated from You *et al.* (2015). P-value and FDR ploted in **A.** and **C.** *** denotes p-value < 0.001, ** denotes p-value < 0.01, and * denotes p-value < 0.05.

we demonstrated that *Ube3a-AS/UBE3A-AS* is brain-specific, and in mice, upregulated in neurons with paternal exclusive expression. Lastly, we found that *Ube3a-AS* is spatiotemporally expressed. Based on these findings, we propose that *Ube3a-AS/UBE3A-AS* is a highly processed transcript with potential functionality.

Studies to date indicate that the *UBE3A-AS* is transcribed as part of a long polycistronic transcriptional unit on the paternal chromosome [31, 37, 118]. Our results are consistent with this theory (**APPENDIX B, Figure B.6**); furthermore, we found that several of transcripts assembled were highly interconnected with the upstream *Snord115/SNORD115* gene. Additionally, we did not observe any 5' capping near the predicted 5' ends of the mouse antisense transcripts. The presence of 5' capped transcripts in *Ube3a-AS/UBE3A-AS* region lacked an active transcriptional start site and aligned either to exonic or intronic regions suggesting post-transcriptional modifications [170–172]. Furthermore, we identified and verified polyadenylation sites throughout both intronic and exonic regions of *Ube3a-AS/UBE3A-AS*. Altogether suggesting that the region is transcribed as 5' capping and 3' polyadenylation are often coupled with transcription to prevent degradation of the RNA transcript, facilitate nuclear export, and/or promote translation [173].

We also confirm that the *Ube3a-AS/UBE3A-AS* is primarily, if not exclusively, expressed in the brain. Splicing in the antisense direction was almost completely eliminated in non-brain tissues. Furthermore, upon examination of differential expression in cerebral cortex cell populations, we observed a drastic decrease in splicing in non-neuronal cell-types that are consistent with previous studies that imprinting of *Ube3a* is neuron-specific [34, 128]. By using exon level differential expression, we were able to see log2 fold-changes greater than 2, which was not apparent with transcript level differential expression. This could be due to the overlapping regions of *Ube3a/UBE3A* and the antisense transcripts. As such, general antisense isoforms were chosen near the 3' ends of the antisense transcript where 3'RACE had confirmed the polyadenylation site. Interestingly,

these isoforms appeared to also be differentially expressed within tissue and cell-type comparisons with isoform 1 having the, overall, highest differential expression.

In addition to being differentially expressed, we determined that the *Ube3a-AS* isoforms were also spatially regulated within the brain with isoform 3 upregulated in cerebellum, and isoform 1 and 2 upregulated in cortex, frontal lobe, and hippocampus. Furthermore, we determined that *Ube3a-AS* was also temporally regulated in the hippocampus with expression of isoform 1 during the P1 developmental time period. This was of interest as *Creb3l1* - a cAMP protein; *Kcnd2* - a potassium voltage-gated channel protein; *Stx3* - a syntaxin protein; and *Slc6a4* - a neurotransmitter protein are also temporally regulated in the brain at the P1 stage [174, 175]. Isoform 2 expressing during the P10 time period, where several enhancers and transcription factors like *Hes3* and *Atf6* are also temporally regulated in the brain [174, 175]. We did not observe temporal regulation with isoform 3; however, as the tissue examined was hippocampal and isoform 3 showed significant upregulation in cerebellum tissue. As such, it is possible that isoform 3 could be temporally regulated in cerebellum.

Altogether, these findings provide insight into the function of the *UBE3A-AS* and the function of neuron-specific imprinting of *UBE3A*. Processing of mRNA and ncRNA to generate shorter RNA transcripts often expand the functional capacity of the transcriptome, generating shorter RNAs and in some instances isoforms with coding potential. Here, we propose that *UBE3A-AS* is expressed in neurons for a regulatory function outside of the imprinting of *UBE3A*. Furthermore, these new insights offer clues as to how the antisense can be targeted for therapeutic intervention and raises potential ramifications of doing so.

3. ANTISENSE MEDIATED ALTERNATIVE SPLICING REGULATES IMPRINTING OF *Ube3a* IN NEURONS

3.1 Overview

Loss of the maternally inherited *UBE3A* allele causes Angelman syndrome, a debilitating neurological disorder associated with intellectual disability, absent speech, and ataxia. In both mouse and human, the *UBE3A* gene is imprinted with maternal-allelic expression in neurons of the CNS through the expression of the *UBE3A* antisense transcript (*UBE3A-AS*), which is both necessary and sufficient for establishing the imprint. Unlike most imprinted genes though, *UBE3A-AS* inhibits transcriptional elongation - rather than transcriptional initiation - of the paternal *UBE3A* allele. The mechanism by which this occurs is unknown. Here we show that mouse *Ube3a-AS* imprints *Ube3a* through alternative splicing and the use of an intronic alternative polyadenylation site. These findings provide insight into the functional significance of imprinting of *UBE3A* in neurons and also reveal novel strategies to reactivate expression of the paternal *UBE3A* as a therapy for individuals with Angelman syndrome.

3.2 Introduction

Human chromosome 15q11-q13 contains a cluster of imprinted genes that are associated with number of neurogenetic syndromes. Maternal derived mutations or epimutations leading to the loss of ubiquitin protein ligase E3A (UBE3A) gene cause Angelman syndrome, which is associated with intellectual disability, ataxia, epilepsy, and an atypical happy disposition [2, 3]. UBE3A is a member of the ubiquitin proteasome system, where it covalently attaches ubiquitin polypeptides to target proteins [176]; it also functions as a co-activator of nuclear steroid hormone receptors [177, 178]. The specific targets and pathways underlying the symptoms associated with AS, however, remain unclear. Paternal

derived mutations or epimutations leading to the loss of the C/D box *SNORD116* snoRNAs (small nucleolar RNAs) cause Prader-Willi syndrome (PWS), which is characterized by dysregulated hunger and satiety, thermoregulation, sleep-disorder, and behavioral issues [132]. Currently, the function of the *SNORD116* snoRNAs in the brain is poorly understood.

The *UBE3A* gene is located at the telomeric end of the 15q11-q13 imprinted region and is orientated in the opposite direction of the C/D box snoRNA clusters (*SNORD115* and *SNORD116*) and the *SNURF-SNRPN* gene. The *SNURF-SNRPN* gene and snoRNA clusters are expressed from the paternal allele as a long polycistronic transcriptional unit (PTU) that is also transcribed in the antisense direction across *UBE3A* - the 3' end of the PTU is hence referred to as the *UBE3A* antisense (*UBE3A-AS*, also known as *UBE3A-ATS*) transcript [37]. Recent studies have demonstrated that transcription of *UBE3A-AS* is both necessary and sufficient to silence expression of the *UBE3A* sense transcript in mice [36, 136]. Moreover, since the PTU is transcribed exclusively from the paternal allele and expressed only in neurons, *UBE3A* is imprinted with maternal-allelic expression in neurons and biallelically expressed in all other cell types [34, 37]. An orthologous region exists on mouse chromosome 7C where imprinting of *Ube3a* and the PTU is also conserved, which makes the mouse an excellent model for investigating the imprinting of *Ube3a*.

Currently, the mechanism by which *Ube3a-AS* inhibits expression of *Ube3a* is unclear. Whereas most antisense transcripts regulate expression of their sense counterparts by inhibiting transcriptional initiation [85, 99, 179], *Ube3a-AS* appears to inhibit transcriptional elongation of *Ube3a*. Meng *et al.* [36] reported that the paternal *Ube3a* allele is modified with active histone modifications, bound by RNA polymerase II, and transcribed up to a region in intron 4, where both *Ube3a* sense and antisense transcript levels diminish [35, 122]. Based on these observations, Meng *et al.* [36] proposed a transcriptional collision model of

genomic imprinting in which the *Ube3a* and *Ube3a-AS* transcriptional complexes collide, causing each to stall and dissociate from their respective template strands [36].

Recently, our laboratory detected high levels of *Ube3a-AS* transcripts as far as 40 kb upstream of *Ube3a* (**Chapter 2**), which is at odds with the transcriptional collision model. Based on this observation, we explored alternative mechanisms by which the antisense could inhibit transcriptional elongation of the paternal *Ube3a* allele. Here, we demonstrate the existence of a paternally expressed, short *Ube3a* isoform (isoform 4) that undergoes early termination through the use of an intronic alternative polyadenylation site in intron 4. Isoform 4 is polyadenylated and expressed exclusively in the brain from the paternal allele. Pharmacological inhibition of *Ube3a-AS* in mouse primary hippocampal neurons ablates the use of the intronic alternative polyadenylation site, resulting in reactivation of the paternal *Ube3a* allele. Based on these findings, we propose that *Ube3a-AS* regulates imprinting of *Ube3a* through alternative splicing and intronic alternative polyadenylation.

3.3 Materials & Methods

3.3.1 Bioinformatics

3.3.1.1 Public data, genomes and annotations

Publicly available data

The bioinformatic analysis performed in this chapter was conducted using publicly available data downloaded from the European Nucleotide Archive. Mouse tissue data was from 8 wk adults [137]. Topotecan treated neuron data were cultured cortical neurons (10 days *in vitro*) with a 72 h treatment on day 7 [53]. Finally, mouse cerebral cortex cellular populations data were purified using various methods specific to the cell-type [140]. A breakdown of tissue types, strain, and accession is supplied in **APPENDIX C, Table C.3**. A complete list of publicly available RNA-seq datasets used in this chapter is provided in **Table 3.1**.

Table 3.1: Public Data: RNA-seq information for *Ube3a* mechanism

Study	Instrument	Layout	Stranded	Strain
ERP000591	Illumina Genome Analyzer	PE	No	C57BL/6J x DBA/2J
SRP012040	Illumina HiSeq 2000	PE	Yes	C57BL/6J
SRP017966	Illumina HiSeq 2000	PE/SE	No	C57BL/6J x CASTEi/J
SRP033200	Illumina HiSeq 2000	PE	No	Multiple

Genomes and annotation sets

Throughout this work, the July 2007 finished NCBI Build 37 mouse genome assembly [143] (mm9) was used. Annotations were collected from Illumina iGenomes collection, using the July 17, 2015 UCSC annotations [144].

PolyA-seq and CAGE-seq data

PolyA-seq BED files [145] were downloaded from UCSC, while CAGE-seq BED files [148] were downloaded from `http://fantom.gsc.riken.jp/5/datafiles/latest/extra/CAGE_peaks/`. The polyA sites clustered datasets and CAGE-seq peak files was separated with `awk` (Bash version 4.0) and viewed with IGV (version 2.3.90 [146, 147]).

3.3.1.2 Data processing

The quality of the fastq files were checked with FastQC [149] (version 0.11.5), followed by adapter and low quality (quality score ≤ 3) sequence trimming with Trimmomatics [150] (version 0.36). The TruSeq3-PE-2.fa adapter file from Trimmomatics adapter file was used for adapter cutting and a minimal length of 25 was set. Paired- and single-end trimmed reads were than aligned with Hisat2 [151, 152] (version 2.0.4) to chromosome 7

(chr7) with the assistance of Hisat2 provided python scripts that extracted splice sites and exons from chr7. The SAM file were directly pipped into SAMtools [153] (version 1.3.1) to convert to BAM format (`samtools view`) and sorted (`samtools sort`). These files were merged (`samtools merge`) and indexed (`samtools index`) by biological replicas for viewing in IGV. StringTie [151, 154] (version 1.3.3) was used to assemble the sorted BAM files (unmerged). Individual annotation files for stranded, high-depth reads (SRP01204) was generated via `stringtie` using the *de novo* method, and merged (`stringtie --merge`) using FPKM threshold of 5, and isoform fraction of 0.05. This merged annotation file was used as a reference for the Rsubread [156] (version 1.24.1) function `featureCounts` [155] for downstream analysis.

3.3.1.3 Data analysis

Visual analysis

All visualization was conducted with IGV. Novel transcript annotation (GTF) on the forward strand were visualized along with polyA-seq and CAGE-seq brain-specific annotations (BED). For *Ube3a* specific visualization, potential transcripts were extracted using a combination of `awk` and `grep`. Sashimi plots - a utility within IGV - were used to visualize alternative splicing limited to the forward strand.

Differential expression analysis with edgeR

The Bioconductor [159, 160] (version 3.4, R [161] version 3.3.2) package, edgeR [157, 158] (version 3.16.5), was used to determine differential expression on the transcript and exon level. Read counts were generated with `featureCounts`, which was used to generate an edgeR `DGEList` object from count, group, and annotation information. Data was filtered (CPM ≥ 1, for 25% of samples) and normalized (`calcNormFactors`) before the dispersion was estimated [157, 162] and fitted to a negative binomial generalized log-linear model [163]. Isoform/transcript and exon levels differential expres-

sion was statistically tested with `exactTest` [164] for pairwise comparisons of group means and p-values were adjusted using Benjamini & Hochberg method (FDR) [166] with `topTags` [164, 165].

SNP analysis

Informative SNPs were extracted from hybrid mice (maternal, C57BL/6J and paternal, DBA/J2) via `samtools mpileup` and BCFtools (version 1.3.1) `bcftools call` and `bcftools view` on sorted, indexed, merged BAM files. A list of SNPs in the region was downloaded from the Mouse Genomes Project - Query SNPs [167, 168] and coordinates were converted from mm10 to mm9 using LiftOver - an UCSC tool (`http://genome.ucsc.edu/cgi-bin/hgLiftOver`). Region of interest, specified with the `-r` option for `samtools mpileup`, was chr7:66,439,800-66,808,000.

3.3.2 Molecular

3.3.2.1 Animals

Animals were housed under standard conditions in a pathogen-free mouse facility. All procedures performed according to NIH guidelines and approved by the Texas A&M University Institutional Animal Care and Use Committee (IACUC). The laboratory of Dr. Arthur Beaudet generated and provided *Ube3a*^{YFP} mouse model [128]. B6D2F1 (100006) hybrid mice were obtained from the Jackson Laboratory (Bar Harbor, ME). The *Ube3a*^{YFP} mice were maintained on C57BL/6J background (000664, The Jackson Laboratory).

3.3.2.2 Primary neuronal culture

This study involved the establishment of primary hippocampal neurons from the offspring of female wild-type, C57BL/6J, and male *Ube3a*$^{+/YFP}$ mice as previously described [180] with slight modifications. Briefly, hippocampi were dissected from P0-P2 mice and held on ice in hibernate medium (A1247501, Life Technologies, Carlsbad, CA)

supplemented with 2% B27 (17504044, Life Technologies) during surgery. Neurons were dissociated by trypsin treatment (10 min at 37°C and 600 rpm) using TrypLE (12604021, Life Technologies) and triturated with a glass Pasteur pipette in Neuron culture media consisting of Neuralbasal Media (21103049, Life Technologies) supplemented with 1% GlutaMAX (35050061, Life Technologies), 1% penicillin/streptomycin (15140122, Life Technologies), and 2% B27. Neurons were plated on 6-well cell culture plates coated with poly-l-ornithine (P0421, Sigma-Aldrich, St. Louis, MO) and laminin (23017015, Life Technologies). Typical plating density was one animal per well in a 6-well plate with cultures maintained at 37°C (95% O_2, 5% CO_2).

3.3.2.3 3' Rapid amplification of cDNA ends analysis

Total RNA was isolated from flash frozen cortex of adult 10-week old male C57BL/6J mice using TRIzol (15596018, Life Technologies) following the manufacturer's protocol. Gene specific primers (**APPENDIX C, Table C.2**) were designed to perform 3' RACE using AUAP universal primer (18373019, Life Technologies) following the manufacturer's protocol. PCR amplicons were cloned using the TOPO TA cloning kit (K458001, Life Technologies) and sequenced at the Texas A&M University Gene Technologies Laboratory. Sequences were visualized with IGV using the `BLAT` function and BED files were exported, binded together (`paste`, Bash), sorted (`sortBed`) and merged (`bedtools merge`) with BEDtools [169] (version v2.25.0). The merged BED files were visualized together with polyA- and CAGE-seq data in IGV.

Gene structure analysis

Gene structure prediction was performed using the GeneSeqer program [181]. Briefly, the genome annotation was converted to a FASTA file using `gffread -w` (`http://ccb.jhu.edu/software/stringtie/gff.shtml#gffread`). The Ube3a protein ESTs (expressed sequence tags) were downloaded from NCBI as FASTA files.

Using the `GeneSeqer` command for mouse, the forward strand was analyzed with the protein EST library for the predicted *Ube3a* isoform 4.

3.3.2.4 Cell culture treatment with Topotecan

Total RNA was isolated from *Ube3a*$^{+/YFP}$ cultured primary neurons using a PureLink kit (12183018A, Life Technologies). Topotecan hydrochloride (1672257, Sigma-Aldrich) was added to 3 ml total of Neuron culture media in 6-well plate (72 h treatment) at a final concentration of 300 nM in 1X TE buffer (12090015, Life Technologies).

3.3.2.5 Reverse-transcription and quantitative PCR analysis

Four month old B6D2F1 female mice were dissected and the lung, liver, kidneys, heart, ovaries, and cortex were flash frozen. RNA was isolated using TRIzol and treated with TURBO DNA-free kit (AM1907, Life Technologies). Complementary DNA was generated using SuperScript IV (18091050, Life Technologies) with oligo(dT) primers. RT-PCR was preformed on adult mouse tissues with forward primer for exon 4 and reverse primers for exon 4.1 and 5. SYBR-Green (11760500, Life Technologies) was used to assay mRNA expression level using the 7900HT Fast Real-Time PCR System (4351405, Applied Biosystems, Foster City, CA). Expression data were normalized using *ActB* and neuron expression was normalized using *Map2*. All primers listed in **APPENDIX C, Table C.2**. Statistical significance for qPCR expression data was determined using two-way ANOVA calculated in R (`aov`). Post-hoc analysis was performed simultaneously using Tukey's HSD multiple comparison (`TukeyHSD`) with default parameters.

3.3.3 Charts

Charts generated in R using `ggplot2` and `pdf`, a `devtools` function.

3.4 Results

3.4.1 Novel Ube3a isoform 4 expressed exclusively from the brain

While investigating the alternative splicing patterns of *Ube3a* in publicly available RNA-sequencing (RNA-seq) data sets, we identified a novel *Ube3a* isoform involving splicing between exon 4 and an unannotated exon (hereafter referred to as exon 4.1) in intron 4 (**Figure 3.1A**). The novel splice site was detected exclusively in transcriptome assemblies of mouse brain and not in any other organ or tissue (**APPENDIX C, Figure C.1**) and downregulated in other tissues (**APPENDIX C, Figure C.2**). Moreover, alternative splicing into exon 4.1 was only detected in neurons (**APPENDIX C, Figure C.3**) and the isoform downregulated in other cell-types (**APPENDIX C, Figure C.4**). 3'RACE and Sanger sequencing demonstrated that exon 4.1 was a terminal exon, aligning with the polyA-seq polyadenylation signal (**Figure 3.1B**), and revealing an in-frame coding sequence preceding a stop codon (**Figure 3.1C**). RT-PCR analysis (polyA enriched RNA) showed that *Ube3a* isoform 4 was expressed in adult mouse cortex but not heart, liver, kidney, and ovary (**Figure 3.1D**), confirming our RNA-seq analysis. Altogether, these data demonstrate the expression of a novel *Ube3a* isoform expressed in brain, hereafter referred to as *Ube3a* isoform 4.

3.4.2 Ube3a isoform 4 is paternally expressed

We next investigated the mRNA (polyA enriched) transcriptomes of adult hybrid (C57BL/6J x DBA/2J) mice hippocampus (n = 6) produced by the Sanger Institute to determine the allelic expression of *Ube3a* isoform 4. Analysis of informative single nucleotide variants (sense expressed transcripts) showed that *Ube3a* isoform 4 was primarily expressed from the paternal allele (**Table 3.2**, and **APPENDIX C, Figure C.5**).

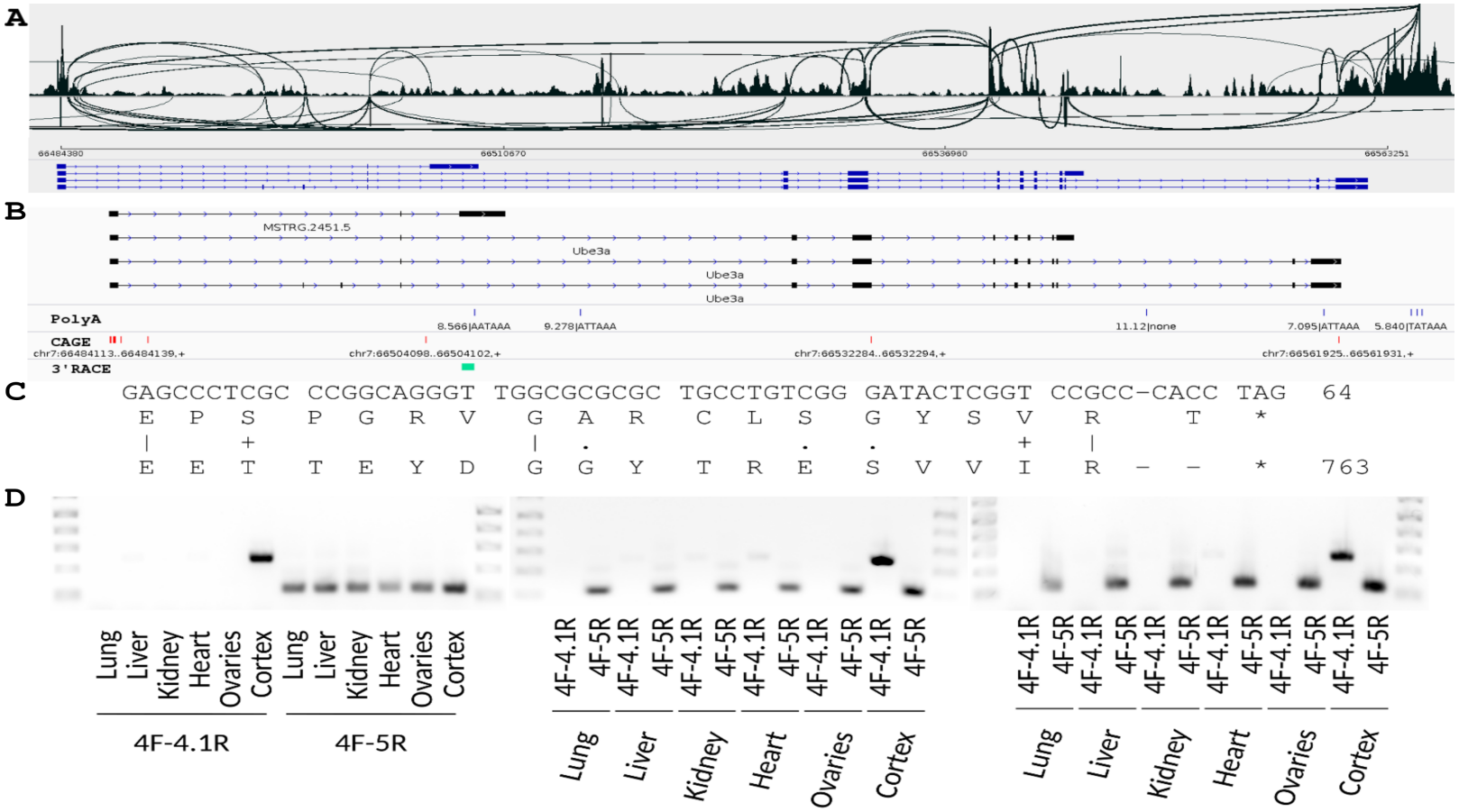


Figure 3.1: Identification of *Ube3a* isoform 4 in mouse brain. **A.** Schematic of alternative splicing into *Ube3a* exon 4.1 using merged cerebellum, cortex, and frontal lobe (n=2), stranded, paired-end RNA-seq data. Data generated from Pervouchine *et al.* (2015). **B.** Schematic of alternatively spliced transcripts of *Ube3a*. CAGE-seq 5' capping data produced by Lizio *et al.* (2015). PolyA-seq data produced by Derti *et al.* (2012). 3'RACE combined BLAT sequence . **C.** Sequence alignment of *Ube3a* exon 4.1 showing in-frame stop codon. **D.** RT-PCR showing that *Ube3a* isoform 4 is exclusively expressed in the mouse brain.

Table 3.2: SNP frequency for C57xDBA hybrid mice - exon 4.1 region

SNP	Location	Ref	Alt	Allelic Freq
T/G	Chr7:66507227	0	7	-1.00
C/T	Chr7:66508429	1	13	-0.93
C/A	Chr7:66508469	1	30	-0.96
G/A	Chr7:66509079	0	13	-1.00
A/T	Chr7:66509131	0	17	-1.00

3.4.3 Ube3a-AS regulates the expression of Ube3a isoform 4

Our findings prompted us to hypothesize that the antisense transcript of *Ube3a* regulated alternative splicing of *Ube3a* isoform 4. To test this, we first analyzed RNA-seq data generated by King *et al.*, (2013), which consists of transcriptomes of mouse primary neurons treated either with a vehicle (DMSO) or Topotecan, a topoisomerase inhibitor that reactivates expression of the paternal *Ube3a* allele by inhibiting expression of the *Ube3a-AS* transcript [52, 53]. Analysis of the transcriptomes revealed splicing into exon 4.1 in the control neurons but not the Topotecan treated neurons (**Figure 3.2A**). Analysis of isoform expression demonstrated significant downregulation of *Ube3a* isoform 4 in Topotecan treated neurons; however, there did not appear to be a significant fold-change in expression of the other three isoforms (**Figure 3.2B**).

An investigation into exon usage of *Ube3a* showed similar expression levels of *Ube3a* exon 4 in the control and Topotecan treated neurons; however, expression of exons 5 was significantly upregulated in the Topotecan treated neurons (**Figure 3.2C**), while it was significantly downregulated in exon 4.1. We then treated primary hippocampal neurons with Topotecan and vehicle (1X TE buffer) and used qPCR to quantify expression of *Ube3a*

isoform 4. Consistent with the RNA-seq analysis, the Topotecan treated neurons had significantly reduced relative expression levels of *Ube3a-AS* (TukeyHSD, p-value < 0.0001) and *Ube3a* isoform 4 (TukeyHSD, p-value < 0.001) compared to controls (**Figure 3.2D**). Taken together these findings demonstrate that *Ube3a-AS* regulates alternative splicing of *Ube3a* paternal sense expression.

3.5 Discussion

Here we demonstrate the existence of a brain-specific, paternally expressed *Ube3a* isoform that terminates in intron 4 and that is dependent on expression of the *Ube3a-AS* transcript. Based on these findings, we propose that *Ube3a-AS* inhibits transcriptional elongation of the paternal *Ube3a* allele through alternative splicing and the use of an intronic alternative polyadenylation signal. This notion is consistent with numerous reports, including imprinted genes, describing the antisense regulation of sense alternative splicing [182, 183]. Whether transcription of *Ube3a-AS* affects the elongation kinetics of the paternal *Ube3a* transcriptional complex, leading to inclusion of the exon 4.1, or whether the *Ube3a-AS* transcript induces inclusion of exon 4.1 by masking the downstream splice acceptor sites or regulatory elements is unknown and warrants further investigation.

We investigated the link between *Ube3a-AS* and *Ube3a* isoform 4 with Topotecan, a drug known to effect alternative splicing [184, 185]; and thus, this study does not clarify whether the reduction in *Ube3a* isoform 4 expression is due primarily to the decrease in *Ube3a-AS* expression or due to Topotecan. An alternative method to inhibition *Ube3a-AS* specifically with either antisense oligonucleotides, similar to Meng *et al.* (2013), or by using a PWS-IC paternal deletion transgenetic mouse model would answer this question. Nevertheless, we provide for the first time evidence linking *Ube3a-AS* expression to alternative splicing of the paternal *Ube3a* sense transcript.

Based on our findings, we envision at least three scenarios that could explain the func-

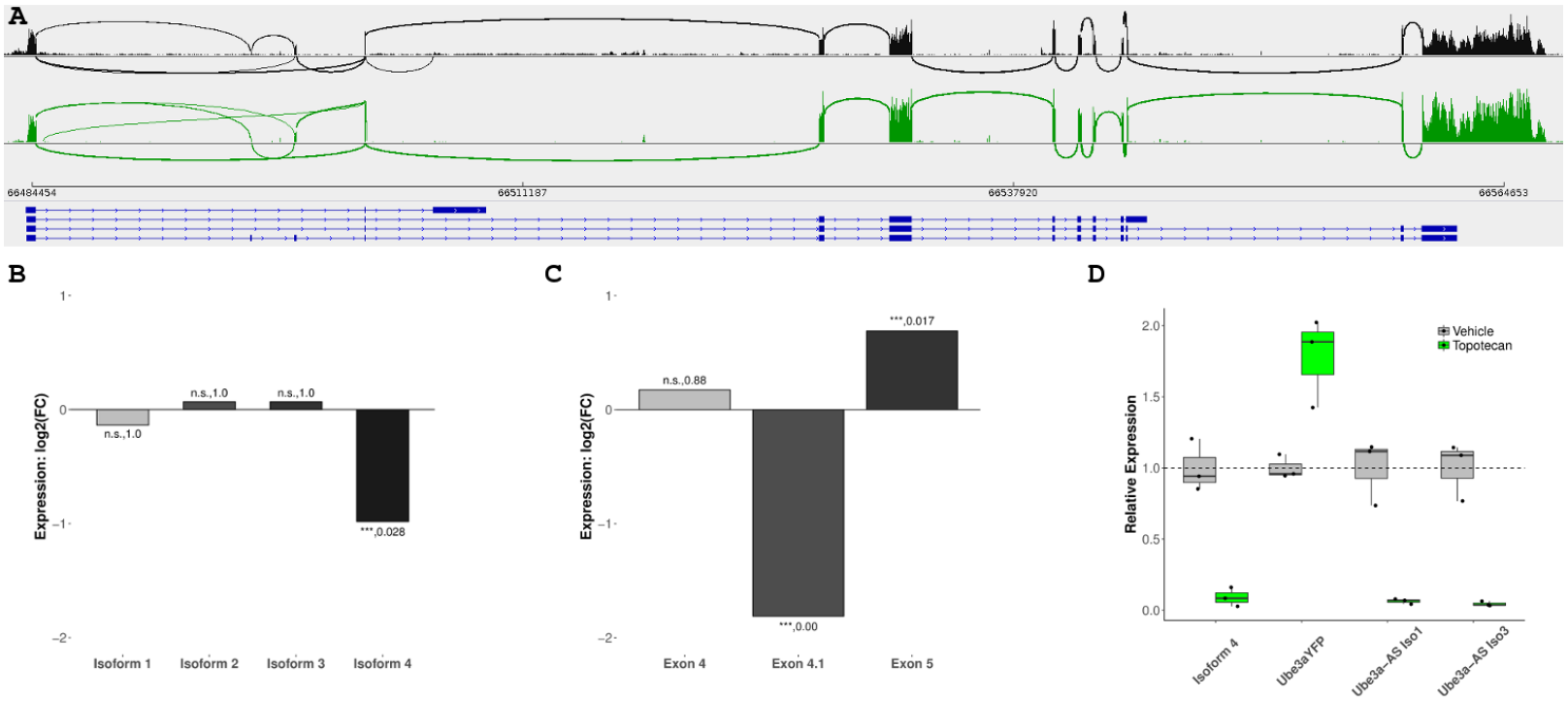


Figure 3.2: *Ube3a-AS* regulates expression of *Ube3a* isoform 4 in neurons. **A.** Schematic of *Ube3a* splicing events detected by RNA-seq in primary neurons treated with vehicle (DMSO) and Topotecan (300 nM) for 72 h. Data generated from King *et al.* (2013). **B.** Log2 fold-changed normalized expression of the four *Ube3a* isoforms comparing Topotecan to vehicle treated neurons. **C.** Log2 fold-change of exons 4, 4.1, and 5 comparing Topotecan to vehicle treated neurons. **D.** *Ube3a* isoform 4 and *Ube3a-AS* relative expression levels in primary neurons treated with vehicle (1X TE buffer) and Topotecan (300 nM) (n = 3). P-value and FDR plotted in **B** and **C**. *** denotes p-value < 0.001.

tional significance of *UBE3A* imprinting in neurons. First, isoform 4 may have a neuron-specific regulatory role. In this model, the expression of the antisense transcript expands the functionality of *Ube3a* specifically in neurons. This theory is further supported by our results showing temporal regulation of isoform 4 in hippocampal tissue (**APPENDIX C, Figure C.6**). Conversely, the antisense transcript may have neuron-specific regulatory functions. Given the recent studies in our laboratory demonstrating a remarkable complexity to *Ube3a-AS/UBE3A-AS* expression patterns, with more than a dozen alternatively spliced transcripts in mouse and at least ten alternatively spliced transcripts in human, suggests a regulatory role for *Ube3a-AS/UBE3A-AS* in the brain (**Chapter 2**). Although the functionality of these transcripts has yet to be investigation, their existence suggests a model in which reciprocal imprinting allows for the expression of both sense and antisense transcripts (i.e., the complementation model) [186, 187]. Finally, it is possible that both isoform 4 and the antisense transcripts have a regulatory role within the brain; and thus, increase the overall complexity of the treanscriptome in neurons.

In addition to proposing a new model for the imprinting of *Ube3a* in neurons, these results also apply to the current therapeutic strategies for Angelman syndrome - reactivation of the paternal *UBE3A* allele. Numerous laboratories, including ours, are actively pursuing strategies to reactive expression of the paternal *Ube3a/UBE3A* allele. The feasibility of this approach has been demonstrated using both pharmacological and epigenetic methodologies [52, 123, 125]. More importantly, reactivation of the paternal *Ube3a* allele has been shown to mitigate some of the phenotypes associated with the loss of *Ube3a* in an Angelman syndrome mouse model [125, 126]. The link between isoform 4 and reactivation of the paternal *Ube3a* allele offers a new target for AS therapeutics. Furthermore, it highlights potential ramifications for the knockdown of *Ube3a* isoform 4 and *Ube3a-AS*.

4. MOUSE EMBRYONIC STEM CELL-DERIVED NEURON HIGH-THROUGHPUT DRUG SCREEN ASSAY FOR ANGELMAN SYNDROME

4.1 Overview

High-throughput drug discovery efforts for neurological disorders often rely on the use of mouse primary neuronal cultures; however, establishing primary cultures from the rodent brain is labor intensive, expensive and provides only a limited supply of cells. Mouse embryonic stem cell-derived neurons are an ideal alternative to primary cultures, because large quantities of cells are easily generated and readily differentiated into neurons in vitro, providing an almost unlimited source of cells for high-throughput screening (HTS) assays. Here, we developed and validated an ES cell-based HTS method to identify new therapies for Angelman syndrome (AS), a severe neurodevelopmental disorder that is caused by loss of the maternally inherited *UBE3A* allele. In neurons, *UBE3A* is imprinted with maternal-specific expression, thus leaving the paternal allele transcriptionally inactive but genetically intact. As such, approaches to reactivate expression of the paternal allele are seen as a viable therapeutic option for AS. ES cells with a paternally inherited *Ube3a*^{YFP} reporter allele were generated to perform proof-of-concept HTS. Imprinted paternal *Ube3a*^{YFP} is reactivated after treatment with Topotecan, a topoisomerase inhibitor known to reactivate the silenced paternal allele. These initial results demonstrate the utility of ES-N to perform HTS to identify novel therapeutics for neurological disorders.

4.2 Introduction

High-throughput screening (HTS) of drug libraries, small molecule compounds, and biologicals is a powerful approach to identify new therapies for a wide rand of diseases [102, 188–194]. While most high-throughput drug discovery assays are relatively straightforward, those aimed at identifying new therapies for diseases of the central nervous sys-

tem (CNS) are difficult because of the challenges associated with using neuronal cell-lines. Indeed there are numerous sources of immortalized neuronal cell-lines (e.g., P19, SH-SY5Y neuroblastoma cells, NT2, PC12 cells, etc.), but genetically modified mouse primary neurons are the most appropriate cell-line for performing high-throughput drug discovery screens for monogenic disorders [195–203]. Methods currently used to establish mouse primary neurons involve immature neurons isolated from either prenatal (e.g., E15-E18) or early postnatal (e.g., P1-P2) brains. As such, scheduling experiments is entirely dependent on the breeding schedules and availability of mice.

The number of neurons obtained from the mouse brain is also rather limited; for example, current studies estimate that approximately 600,000 neurons per hippocampi and 800,000 neurons per cortices can be cultured per mouse brain [204]. Primary neuronal cultures are typically grown at 20,000 cells/well [52], approximately 7.68×10^6 cells would be needed for one 384-well plate. Mouse neural stem cells can be expanded *in vitro*, but they can only be maintained in an undifferentiated state for a finite period of time, and they yield low percentages of neurons after differentiation. Another challenge is that most HTS facilities do not accommodate experiments involving primary cell cultures because of the risks associated with contaminating other cell-lines (personal communication Clifford Stephen). Altogether, high-throughput drug discovery efforts for neurological disorders are met with numerous technical challenges that have likely impeded drug discovery efforts.

Mouse embryonic stem (ES) cell-derived neurons are an ideal alternative to primary neuronal cultures. Mouse ES cells rapidly divide *in vitro* and can be maintained in an undifferentiated state almost indefinitely. They are amendable to targeted genetic modifications or can easily be generated from existing transgenic models [205–208]. They are commonly used in HTS assays and accepted by most HTS facilities [209]. Importantly, ES cells can be reliably and efficiently differentiated into neuronal cell populations

that exhibit gene expression patterns, epigentic marks, and electrophysiological properties similar to mouse primary neurons [210–214]. Their use in high-throughput drug discovery assays, however, has been underutilized [215–219].

Angelman syndrome (AS) is a debilitating neurodevelopmental disorder characterized by severe intellectual disability, absent speech, ataxia, seizures, frequent smiling and inappropriate laughter [2, 3, 26]. The genetic or epigenetic mutations causing AS are associated with loss of maternal - but not paternal - expression of the ubiquitin ligase E3A protein gene (*UBE3A*). The non-Mendelian inheritance pattern of AS is due to genomic imprinting of *UBE3A* [2, 3]. In almost all cell-types, *UBE3A* is expressed from both parental alleles; however, in the brain, *UBE3A* is perferentially expressed from the maternal allele [34]. Expression of the paternal *UBE3A* allele is inhibited by the antisense expression of a long polycistronic transcription unit (PTU) that is comprised of *SNURF-SNRPN*, clusters of C/D box small nucleolar RNAs (*SNORD64, SNORD109, SNORD116,* and *SNORD115*), and the *UBE3A* antisense transcript (*UBE3A-AS*) [35, 36]). Studies in mice show that *Ube3a* is imprinted in neurons, including ES cell-derived neurons, and biallelically expressed in other cell-types of the brain [31–34], which is consistent with the neuron-specific expression of *UBE3A-AS*.

Currently, there are few treatment options for Angelman syndrome patients. Available treatments for those with Angelman syndrome focus on behavioral and physical therapies to minimize symptoms, along with drug therapies to control seizures and sleep disruption. Since the inactive paternal *UBE3A* allele is genetically intact but epigenetically silent, approaches to reactivate expression of the paternal *UBE3A* allele are therefore seen as viable therapeutic options to treat AS. In fact, recent studies in mice have shown that pharmacological, genetic, and epigenetic methods are all capable of reactivating expression of paternal *Ube3a* expression in primary neurons and the adult mouse [36, 52, 123, 125]. Importantly, there appears to be some degree of improvement of symptoms after reactiva-

tion [124–127].

In this study, we established and validated a fluorescence based HTS assay in ES cell-derived neurons (ES-N) with an *Ube3a*^{YFP} reporter allele. We demonstrate the utility of this approach for performing large-scale, high-throughput drug discovery assays to identify novel therapies to treat Angelman syndrome.

4.3 Materials and Methods

4.3.1 Animals

Animals housed under the standard conditions, pathogen-free mouse facility. All procedures performed according to NIH guidelines and approved by the Texas A&M University Institutional Animal Care and Use Committee (IACUC). The laboratory of Dr. Arthur Beaudet generated and provided *Ube3a*^{YFP} mouse model [128]. All mice maintained on C57BL/6J background (The Jackson Laboratories, Bar Harbor, ME).

4.3.2 Generation of Ube3a$^{+/YFP}$ *embryonic stem cells*

Mouse *Ube3a*$^{+/YFP}$ ES cells were established following standard methods [220–224]. Briefly, *Ube3a*$^{YFP/YFP}$ males were mated to three to four week old C57BL/6J superovulated female mice. Females were superovulated via 5 I.U. of pregnant mares serum gonadotropin (PMSG) in 0.1 ml of PBS followed by 5 I.U. of human chorionic gonadotropin (hCG) 46 h later. Embryos (E2.5) collected from oviducts were cultured overnight in one well of a 4-well plate (Thermo Scientific, Waltham, MA) in KSOM Evolve (IVFonline, Guelph, Canada) containing 1 mg/ml bovine serum albumin (Sigma-Aldrich, St. Louis, MO). The following day, individual embryos were transferred and cultured for four days in 30 μl microdrops of KORS+2i medium covered by mineral oil (37°C, 5% CO_2) [223,225]. After 96 h in KOSR-2i medium, outgrowths were trypsinized using 0.25% Trypsin-EDTA (Life Technologies, Carlsbad, CA) into single-cell suspensions and plated individually in 96-well, flat bottom, tissue culture treated plate (Corning Inc., Corning, NY) containing a

mitomycin-C inactivated SNL 76/7 feeder cells monolayer [226–229] (Applied StemCell, Inc., Milpitas, CA) at 50,000 cells/cm^2 and KOSR-2i preconditioned (>2 h) medium. Cells were incubated (37°C, 5% CO_2) for four days with daily changes of medium. To establish ES cell-lines, undifferentiated colonies were gradually expanded and genotyped using Jackson Laboratory genotyping primers for *Ube3a*^{YFP}: 5' TCAATGATAGGGA-GATAAAACA 3', 5' GAAAACACTAACATGGAGCTC 3', and 5' CTTGTGTAGCGC-CAAGTGC 3'. Of the six lines tested for ES cell growth, lines #2 and #10 were selected for further large-scale expansion resulting in 43 vials at 3.5×10^6 cells/vial each. The following studies were conducted using line #10 (**Appendix D, Figure D.1**).

4.3.3 Neuronal Differentiation

Retinoic acid (RA) based-induction methods [230–232] with slight modifications for high-throughput screening purposes directed differentiation of *Ube3a*$^{+/YFP}$ ES cells into neuronal cultures. Briefly, 1 vial of frozen *Ube3a*$^{+/YFP}$ ES cells (approx. 3.5×10^6 cells) were expanded in ES preconditioned (>2 h) medium (**Appendix D, Table D.1**) on 60-mm plates (Corning, Inc.) of mitomycin-C treated SNL feeder cells adhered to 0.1% gelatin (STEMCELL Technologies Inc., Cambridge, MA) for two passages. ES cells were separated from SNL feeder cells via a 30 min incubation at room temperature on T175 flasks (Thermo Scientific) in ES medium. Cells were counted and plated in suspension in CA medium (**Appendix D, Table D.1**) on six 100-mm bacteriological plates (Greiner Bio-One, Kremsmünster, Austria) at 4×10^6 cells/plate. The following day, the CA medium was changed and cellular aggregates (CAs) were split 1:2 by transferring the cell suspension into 50 ml tubes (1 plate/tube). The CAs settled via gravity (>= 3 min), and supernatant was removed and replaced with 20 ml of CA medium. The cells were gently pipetted to mix and plated at 10 ml per plate. After 48 h in suspension, CAs were split again (1:2) as described above and media changed to CA medium containing

retinoic acid (Sigma-Aldrich) at a final concentration of 0.5 mM. For the next four days, CA medium with RA was changed daily.

For the dissociation step, ES cell-derived neurons cellular aggregates were transported to HTS facility ($\sim$37°C, $\sim$2 h) or dissociated immediately in the lab. In both instances, cells were dissociated with 0.5% Trypsin-EDTA and plated in N2 medium, **Appendix D, Table D.1** on 384-well poly-d-lysine coated optical bottom plates (Thermo Scientific) at 24,000 cells/well using a microplate washer (Tecan Group Ltd, Männedorf, Switzerland) for HTS or 18 mm round coverslips (VWR, Radnor, PA) coated with poly-l-ornithine (Sigma-Aldrich) and laminin (Life Technologies) at 4×10^5 cells/well for imprint analysis. Stocks of 0.5 mg/ml poly-l-ornithine in 150 mM boric acid were diluted 1:5 in purified water to coat overnight at 4°C. After coverslips were rinsed three times with cell culture grade water, laminin (10 μg/μl) was added to the coverslips and incubated at 4°C overnight. After 48 h, media was changed to Complete medium, **Appendix D, Table D.1**. Every three days Complete medium was changed. Eight days post dissociation (DPD), compounds were added and maintained for 72 h. Topotecan hydrochloride (Sigma-Aldrich) was dissolved in water (10 mM stock) and then diluted to 3 μM into DMSO (Sigma-Aldrich) for HTS, while 10 μM stock in water was used for imprint analysis. Topotecan hydrochloride was added to the Complete medium at 300 nM final concentration along with vehicle (DMSO or water) using microplate washer for HTS or by-hand for imprint analysis as positive and negative controls, respectively.

4.3.4 Immunocytochemistry

Immunocytochemistry and staining were performed as described previously [233]. Briefly, cells were fixed with 4% paraformaldehyde (Sigma-Aldrich) for HTS or 4% paraformaldehyde / 4% sucrose (Sigma-Aldrich) for imprint analysis (10 min, RT). For high-throughput drug screening, cultures were blocked in 5% goat serum (Sigma-Aldrich) and

0.3% Triton X-100 (Sigma-Aldrich) for 15 min at 37°C, while imprint analysis blocked for 1 h at room temperature. Primary antibody was diluted in antibody buffer containing 5% goat serum and 0.3% Triton X-100 and incubated for 1 h at room temperature or for 30 min at 37°C for the high-throughput screen. Cells were washed three times before secondary antibody incubation with 0.1% Tween20 (Sigma-Aldrich) for 10 to 15 min at room temperature. Primary antibodies used here are as follows: anti-GFP (Novus Biologicals, Inc., Littleton, CO), anti-NeuN (EMD Millipore) and anti-βIII Tubulin (Sigma-Aldrich). Secondary antibodies used was Alexa Fluor 488, goat anti-rabbit, and Cy3 goat anti-mouse (Jackson Immuno Research Labs, West Grove, PA). Secondary antibodies were incubated in 5% goat serum and 0.3% Triton X-100 for 1 h at room temperature for imprint analysis or for 30 min at 37°C for high-throughput screening. TO-PRO-3 stain (Life Technologies) at 1:1000 dilution was used for nuclei staining. Images were captured using IN Cell 6000 (GE, Schenectady, NY), and confocal images were captured using Zeiss 510 META Confocal Microscope (Zeiss, Oberkochen, Germany).

4.3.5 Image analysis

4.3.5.1 Imprint analysis

For imprint analysis, FIJI (FIJI is just Image J, open source, `https://fiji.sc/`) was used in image preparation and to measure gray scale values of individual neurons [234, 235]. Positive βIII tubulin staining identified neurons. The Ube3a^{YFP} fluorescence, GFP staining, of each neuron was expressed as YFP intensities overlapping neuronal TO-PRO-3 staining. Two-tailed unpaired Student's t-test was used to determine statistical significance when comparing two groups. To determine statistical significance for the time course, Holm-Bonferroni method was used to control the family-wise error rate for multiple comparisons.

4.3.5.2 HTS analysis

Image analysis for high-throughput screening was performed using IN Cell Developer Toolbox 9.3.1. The TO-PRO-3, nuclei marker, and the NeuN, mature neuron marker, were used to generate a neuron specific overlapping mask, (**NeuN-Overlap**), which was used to measure median YFP intensity in each target. Several R scripts (**Appendix D**) were developed to analyze data rapidly for plate effects, within plate effects, and assay statistics. The Shapiro-Wilk test was used on randomly selected wells for each plate to determine normality of Vehicle and Topotecan treated cells. Normal Q-Q plots were used to visualize normality. To determine statistical significance between plates, one-sided unpaired Student's t-test was used (multiple comparisons, Holm-Bonferroni method), along with the strictly standardized mean difference (SSMD). To determine statistical significance within plate, two-tailed unpaired Student's t-test was used. To make batch analysis user friendly, a Shiny app was developed (`kj-benjamin90.shinyapps.io/angelman-hts-app/`).

4.3.6 Charts

Charts designed in R (version 3.2.2 [161]) programming using the `tikzDevice` and `pdf` functions from the `devtools` package. **Appendix D** shows sample scripts.

4.4 Results

4.4.1 Topotecan induces reactivation of silenced paternal Ube3a allele in ES cell-derived neurons

It has been established that ES-N can recapitulate the paternal imprint in ES cells [35, 236]; however, a time course for this paternal imprint has not been established to the best of our knowledge. Here, we verified paternal imprint of *Ube3a* via a 13 day time course to determine the days post dissociation (DPD) needed for the imprint to be established in this cell-line. **Figure 4.1A** shows that after 6 days in culture paternal Ube3a^{YFP} expression

significantly decreased (**Table 4.1**, Holm-Bonferroni) and continues to maintain a low level of expression. By 13 DPD the ES-N sustain the paternal *Ube3a* imprint as shown in **Figure 4.1B**. Thus, demonstrating that our cell-line also recapitulates the paternal imprint of *Ube3a* as early as 6 DPD.

In addition to developing a time course for *Ube3a* paternal imprinting, we demonstrate in **Figure 4.1C** that Topotecan, a drug proven to reactivate the paternal *Ube3a* in primary neurons via inhibition of topoisomerase (**Appendix D, Figure D.2**), could also reactivate the paternal allele in embryonic stem cell-derived neurons. This confirmed that Topotecan can act as a positive control for our embryonic stem cell-based HTS assay.

Table 4.1: P-values for time course analysis

	2 DPD	**4 DPD**	**6 DPD**	**8 DPD**	**12 DPD**
4 DPD	0.75	-	-	-	-
6 DPD	4.0×10^{-5}	7.3×10^{-7}	-	-	-
8 DPD	3.5×10^{-5}	6.2×10^{-7}	1.00	-	-
12 DPD	2.2×10^{-8}	3.4×10^{-10}	0.24	0.24	-
13 DPD	4.4×10^{-8}	6.9×10^{-10}	0.28	0.28	1.00

4.4.2 Plate effect is not observed for the NeuN-Overlap method

To determine if image acquisition time effects the YFP intensity, plate statistics were calculated and plotted in **Figure 4.2**. Over the course of roughly 12 h, assay stability remains the same with large separation between Vehicle and Topotecan. P-values from the Student's T test were significant for all plates with no correlation over time, **Table 4.2**. For quality control of assay, the z factor (*Z Factor*) was calculated, **Table 4.2**; however, since the z factor is based on a normal distribution and the majority of randomly selected wells

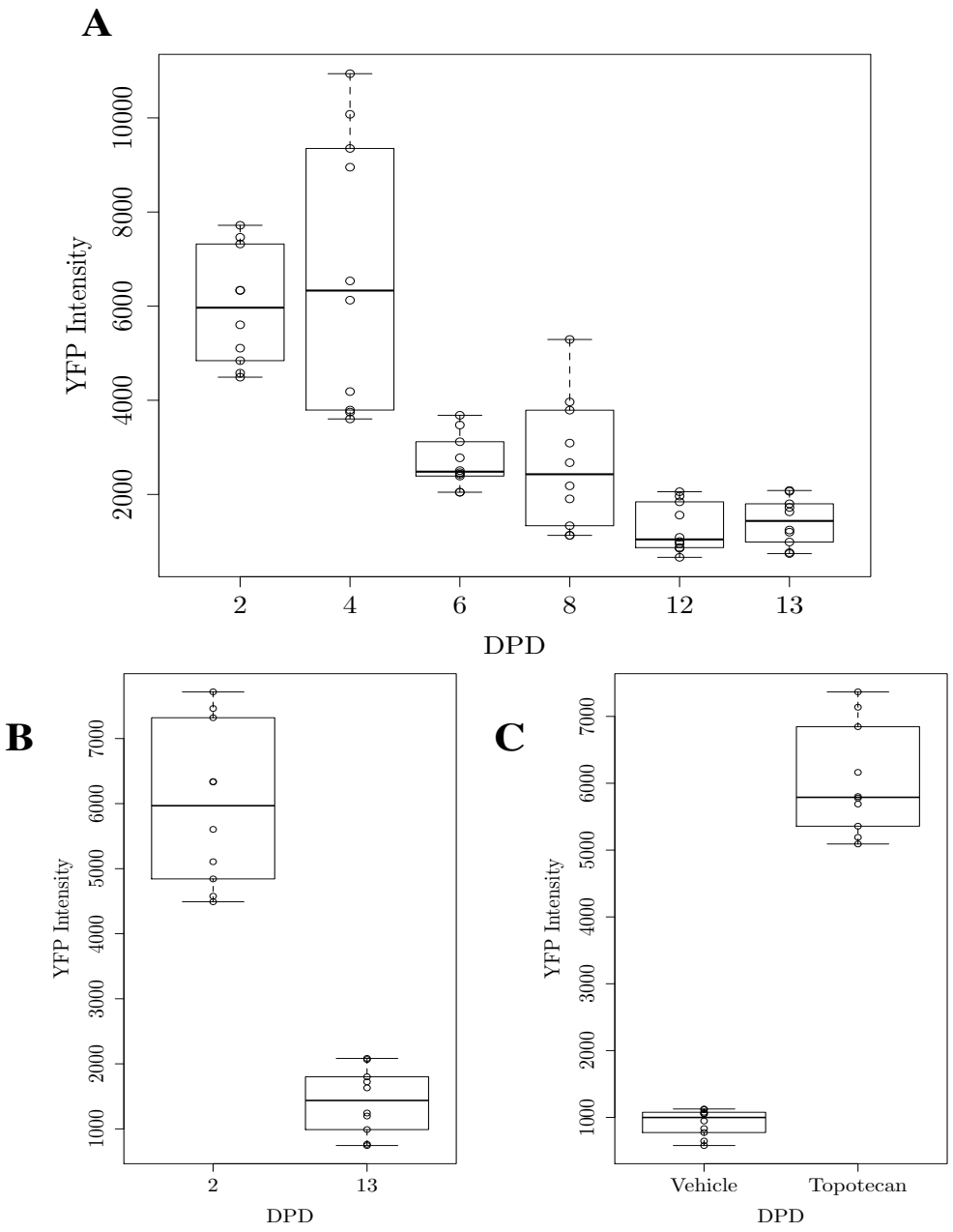


Figure 4.1: Topotecan induces reactivation of paternal *Ube3a* allele in ES cell-derived neurons. **A)** Boxplots of time course analysis of imprinted neurons (N = 10). **B)** Boxplots of *Ube3a*^{YFP} ES cell-derived neurons at 2 and 13 days post dissociation (DPD) demonstrating the imprinting of paternal *Ube3a*. **C)** Boxplots of ES cell-derived neurons at 13 DPD with vehicle (water) or Topotecan (300 nM) treatment demonstrating the reactivation of paternal *Ube3a*. N = 15 neurons.

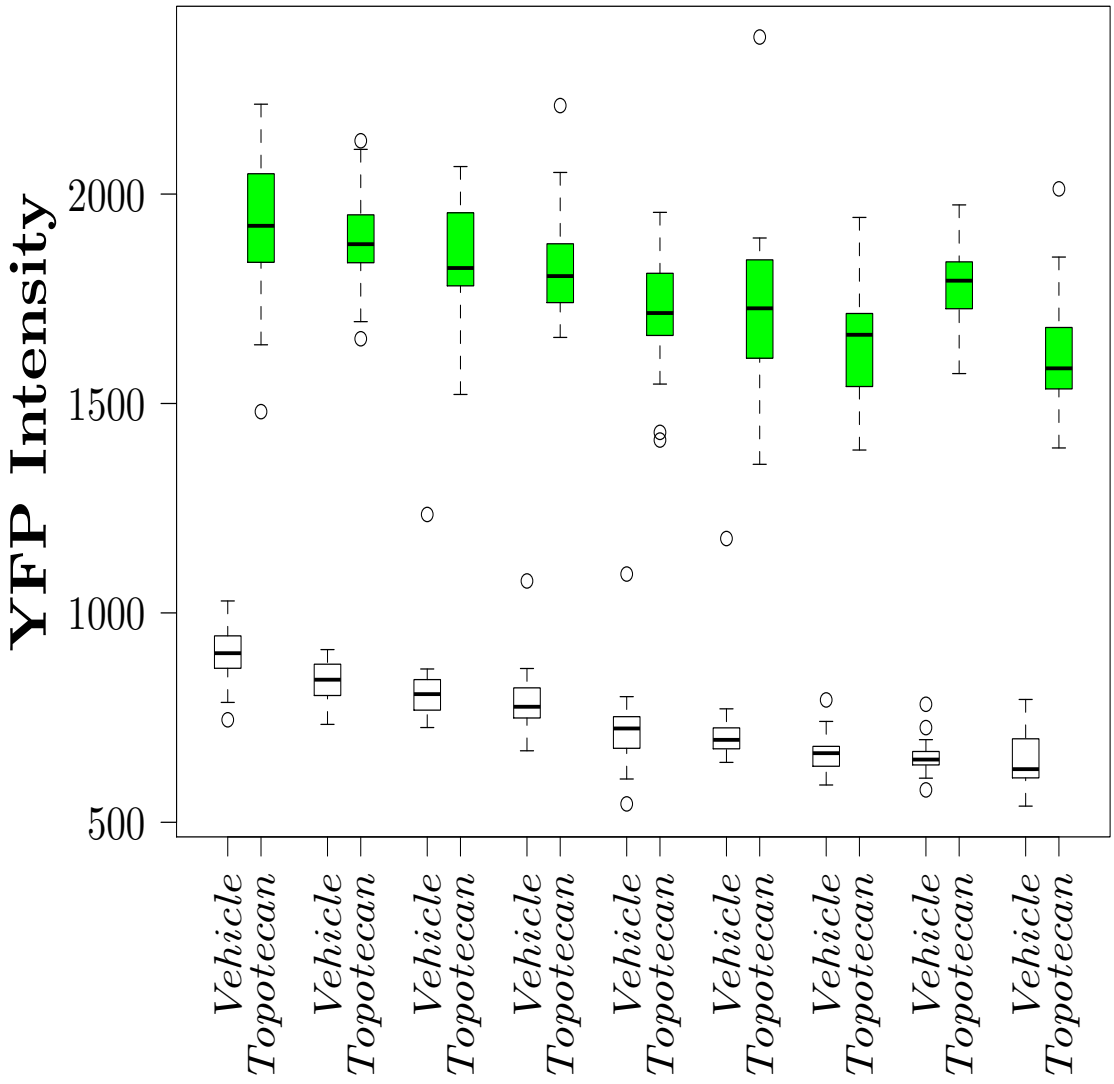


Figure 4.2: The decrease in Ube3a^{YFP} intensity as a function of time does not effect separation of Topotecan intensity from Vehicle.

for each plate had non-normal distributions (Shapiro-Wilk, p-value < 0.05, **Appendix D, Table D.2**), the strictly standardized mean difference (SSMD, $\hat{\beta}$) was also calculated,

Table 4.2. Q-Q plots shown in **Appendix D**.

$$Zfactor = 1 - \frac{3(\tilde{\sigma}_p + \tilde{\sigma}_n)}{|\tilde{\mu}_p - \tilde{\mu}_n|},$$

where $\tilde{\sigma}$ and $\tilde{\mu}$ are sample standard deviations and sample means, respectively, for positive (p) and negative (n) controls [237, 238].

$$\hat{\beta} = \frac{\tilde{X}_p - \tilde{X}_n}{1.4826\sqrt{\tilde{s}_p^2 + \tilde{s}_n^2}},$$

where $\tilde{X}$ and $\tilde{s}$ are medians and median absolute deviations in the positive and negative controls, respectively [238, 239].

Although assay quality remains the same over time, there is a slight decrease in YFP intensity over time. As a high-throughput assay, plates are often run in large quantities. To determine a theoretical number of plates that can be run at one time, linear regression models were estimated as a function of time for Vehicle YFP intensity ($R^2 = 0.9645$) and Topotecan intensity ($R^2 = 0.7768$). Each plate has roughly 90 min separating them as determined from acquisition log information.

$$y(Topotecan) = -0.38t + 1911 \text{ and } y(Vehicle) = -0.36t + 879$$

Set max Vehicle to equal Topotecan to determine when time adversely effects the assay.

$$879 = -0.38t + 1911$$

$$879 - 1911 = -0.38t$$

$$-1032 = -0.38t$$

$$t = \frac{1032}{0.38} = 2716 \text{ min}$$

At roughly 90 min image acquisition time, **NeuN-Overlap** method could ideally run 30 plates assuming antibody decay rate is not exponential. Using more conservative parameters, three times the mean standard deviation of both Topotecan (T) and Vehicle (V), the stability of data acquisition as a function of time is calculated below:

$$y(T) = -0.38t + 1911 - (3 \times 139.3) \text{ and } y(V) = -0.36t + 879 + (3 \times 67.0)$$

$$1080 = -0.38t + 1493$$

$$1080 - 1493 = -0.38t$$

$$t = \frac{413}{0.38} = 1087 \text{ min}$$

At roughly 90 min image acquisition time, **NeuN-Overlap** method can conceivably run 12 plates.

Table 4.2: P-values for NeuN-Overlap plate effect analysis

	T-test	**Z factor**	**SSMD**
Plate-1	1.43×10^{-27}	0.335	5.89
Plate-2	3.08×10^{-36}	0.526	9.28
Plate-3	2.68×10^{-31}	0.324	7.54
Plate-4	8.24×10^{-38}	0.436	8.37
Plate-5	1.13×10^{-30}	0.311	7.49
Plate-6	6.37×10^{-21}	0.136	5.87
Plate-7	3.24×10^{-26}	0.433	6.98

Table 4.2: *Continued...*

	T-test	Z factor	SSMD
Plate-8	1.14×10^{-34}	0.629	13.65
Plate-9	1.63×10^{-28}	0.389	7.34

4.4.3 Well position effects Ube3a^{YFP} intensity

To determine if well position effected YFP intensity measured, wells with vehicle from column 1 were compared with vehicle from column 24, and Topotecan wells from column 2 were compared with Topotecan wells from column 23. The p-values from the two-tailed Student's T-test are shown in **Table 4.3**. Only two plates for Vehicle show well position effects, while Topotecan wells have nearly half showing significant differences between positions within the plate.

Table 4.3: P-values (T-test) for NeuN-Overlap well effect analysis

	Vehicle	Topotecan
Plate-1	0.005	0.170
Plate-2	0.001	0.017
Plate-3	0.249	0.021
Plate-4	0.117	0.507
Plate-5	0.234	< 0.001
Plate-6	0.218	0.228
Plate-7	0.114	0.003
Plate-8	0.285	0.382
Plate-9	0.500	0.194

4.5 Discussion

In this chapter, the high-throughput screening method using embryonic stem cell-derived neuronal cultures is validated via proof-of-concept screening for Angelman syndrome. Embryonic stem cells were generated from reporter mice *Ube3a*$^{+/YFP}$ [128] and expanded. The expansion of these ES cells provides a large pool of available ES cells at the same time point for further differentiation. The neurons derived from these ES cells were significantly more than those that can be collected from primary cultures with on average 600×10^6 from 6 initial neuronal induction plates. For neurons obtained from embryonic mice, the generation of animals presents a major bottleneck for high-throughput screening. Even with recent advances in neuron mini-cultures, 1 million cells can only plate little more than four 384-well plates [240]. To compare, mini-cultures would still require more than 200 embryos with excellent culturing techniques to generate similar numbers.

As research has shown that paternal *Ube3a* is imprinting during development, the time point of imprinting of paternal *Ube3a* in culture was determined for the HTS assay. We determined that within six days Ube3a^{YFP} significantly decreased. For the HTS assay, drugs were added 7 DPD for convenience of HTS facility. Nine 384-well optical-bottom plates were run with Vehicle, columns 1 & 24, and Topotecan, columns 2 & 23, added to the plate using a 300 drugs/small molecule compounds drug loading model. At roughly 30 min/antibody, the total plate image acquisition time was approximately 90 min/plate, which is longer than most HTS that commonly only image with a single antibody. As such, there was some concern that antibody intensity would significantly decrease over-time; however, using the NeuN-Overlap, which is specific to mature neurons, there was no significant decrease observed. Even so, it is this author's recommendation that only 12 plates be run in one sitting assuming that acquisition time is approximately 90 min.

All plates showed significant difference between Vehicle and Topotecan controls, **Ta-**

ble 4.2; however, only two plates had a z factor of ≥ 0.5, which indicates an *Excellent* assay [237]. This may be due to the outliers observed in **Figure 4.2** as the z factor calculation is not robust to outliers [241]. As an alternative, the strictly standardized mean difference method specific to outliers robustness was also calculated per plate with an average $\hat{\beta}$ of approximately 8.04, where $\hat{\beta} > 7.0$ is an *Excellent* assay for HTS small molecule assays [238]. Altogether using the outliers robust SSMD method, this HTS assay scores an *Excellent* for quality control.

Finally, this work is directly applicable for performing a high-throughput screen for Angelman syndrome. Currently, the only drug known to reactivate the paternal *Ube3a* allele is Topotecan. This drug, however, is extremely toxic and currently approved as a chemotherapeutic. Therefore, an alternative therapeutic drug is still needed. Furthermore with this ES cell-derived neuronal culture HTS assay, it will be possible to conduct an RNA interference screen for pathway analysis of the imprint of *Ube3a*. With such a screen, additional targets for therapeutic intervention for Angelman syndrome can possibly be determined. In summary, this data shows that ES-N model can dramatically increase the scale of screening studies for neurological disorders.

5. CONCLUSION

Recent studies show that the human brain has one of the most complex expression patterns of the body. The expression patterns of the imprinted region 15q11-13, containing *UBE3A* and *UBE3A-AS*, are no exception. The studies herein explore the unique imprinting of paternal *UBE3A/Ube3a* in neurons. Specifically, we were able to demonstrate that both mouse and human *Ube3a-AS/UBE3A-AS* are extensively processed via bioinformatics and molecular analysis; demonstrating spatiotemporal regulation of *Ube3a-AS*. For both mouse and humans, we sequenced 3' RACE clones verifying polyadenylation sites within the antisense region of *Ube3a/UBE3A*. This finding suggesting that dosage sensitive *UBE3A* may be imprinted in neurons so that sense and antisense transcripts are co-expressed in the brain. Additionally, the temporal regulation of *Ube3a-AS* suggest it may play a role in the developing brain. Following these experiments, we demonstrated the existence of a fourth paternal-specific *Ube3a* isoform in mice generated via alternative polyadenylation at novel exon 4.1. In the investigating of this novel isoform, we demonstrated that its expression corresponded with *Ube3a-AS* expression suggesting an alternative mechanism of paternal *Ube3a* imprinting. Simultaneously, we developed a high-throughput screening method in embryonic stem cell-derived neurons; demonstrating its potential for drug discovery for neurodevelopmental disorders like Angelman syndrome. Taken together, these results suggest the lncRNA, *UBE3A-AS*, may have additional regulatory functions outside of the imprinting of *UBE3A* and that the novel isoform 4 may also serve as a regulatory RNA. Moreover, they also propose an alternative mechanism of imprinting for *Ube3a*, wherein by some unknown mechanism *Ube3a-AS* causes termination and polyadenylation of paternal *Ube3a* transcription. Finally, these results setup future drug discovery experiments for Angelman syndrome.

5.1 *Ube3a-AS* demonstrates complex expression within the brain

The antisense transcript of *UBE3A*, identified in 1998, originates from the 3' end of a large polycistronic transcriptional unit consisting of *SNURF/SNRPN*, *IPW*, and a cluster of snoRNAs [31–33, 37]; however, *UBE3A-AS* is only expressed in neurons. While *Ube3a-AS* is sufficient to repress the expression of *Ube3a* in neurons [35, 36], our findings showed temporal regulation only for *Ube3a-AS* and *Ube3a* isoform 4, suggesting *Ube3a-AS* expression is unrelated to the imprinting of *Ube3a* and therefore may be a by-product of another regulatory process. In fact, additional work in our laboratory showed that maternal Ube3a expression increased as paternal expression decreased keeping overall Ube3a expression constant [54]. Further investigation into *UBE3A-AS/Ube3a-AS* expression revealed that the antisense transcript was alternatively spliced generating dozens of transcripts in the mouse and at least ten transcripts in the human with several polyadenylation sites and 5' capping events. In the mouse, we confirmed RNA-seq analysis that these isoforms were differentially expressed between brain regions. Furthermore, we found that *Ube3a-AS* generalized isoforms 1 and 2 along with *Ube3a* isoform 4 were temporally regulated in the hippocampus, showing upregulation during post-natal development.

These findings contradict current *Ube3a-AS* theories mainly because of differing approaches in analyzing the 15q11-13 imprinted region. Here, we used powerful high-throughput sequencing technologies, specifically stranded paired-end RNA-seq, to generate transcriptional profiles where sense and antisense transcripts could be distinguished. Additionally, we used the hybrid C57xDBA mice from public data to further distinguish allelic contributions. As such, we were able to analyze the imprinted region in a more detailed and rigorous manner than other studies, and in doing so our findings showed spatiotemporally regulation of *Ube3a-AS*.

Spatiotemporal regulation used to generated specialized tissues during development in

a highly dynamic environment like the brain. For instance, genes associated with temporal regulation during early embryonic development are grouped into categories involving neuron differentiation, axonogenesis, and forebrain development, which are often connected to proper morphological growth [242]. Furthermore, temporally regulated genes associated with late post-natal development involve the regulation of synaptic transmission, behavior, and learning and memory [242]. As such, the upregulation of *Ube3a-AS* during post-natal development, is in agreement with the imprinting of *Ube3a* and the phenotypes associated with neurodevelopmental disorders of the imprinted 15q11-q13 region. Moreover, the uniqueness of cerebellum temporal gene expression is also reflected in our results with differential expression of the isoforms within the brain regions [242–244]. As such, it is also possible that the generalized *Ube3a-AS* isoform 3 could be temporally regulated in the cerebellum, where its expression is upregulated. Altogether, these results suggest a regulatory function for *UBE3A-AS*.

The implications of *UBE3A-AS* having additional functions potentially impacts the current therapeutic strategy for Angelman syndrome. Since the inactive paternal *UBE3A* allele is genetically intact but epigenetically silent, recent studies have targeted the paternal *Ube3a* for reactivation via disruption of *Ube3a-AS* [36, 52, 123–126]; however, if *UBE3A-AS* has additional regulatory functions, disruption of the lncRNA could adversely effect other pathways that *UBE3A-AS* may play a role in. It is for this reason, that *UBE3A-AS* should be intensively studied to better understand any potential ramification of disrupting its expression as a treatment for Angelman syndrome.

5.2 *Ube3a-AS* generates a paternal, neuron-specific, *Ube3a* isoform - isoform 4

Previous investigations into the expression of *Ube3a-AS* revealed partial expression of paternal *Ube3a* [35, 122]. This phenomenon along with ChIP-on-chip experiments confirming that both *Ube3a* promoters are enriched with chromatin modification (histone H3

lysine 4 trimethylation), bound by RNA polymerase II, and actively transcribed gave rise to the collision model as a mechanism of paternal *Ube3a* [36] imprinting in neurons. In this model, RNA polymerase II from *Ube3a* sense-strand and *Ube3a-AS* antisense-strand collide causing a gradual decrease in expression resulting in partial paternal *Ube3a* expression, **Figure 1.12A**. As noted in **Chapter 1**, this model implies that *Ube3a-AS* is not expressed past the collision within intron 4 of *Ube3a*, which is not in agreement with Numata *et al.* SNP data showing paternal RNA expression upstream the *Ube3a* primer [122]. From our *Ube3a-AS* isoform annotations (**Figure 2.1**) several assembled transcripts aligned with Numata *et al.* SNP data upstream the *Ube3a* promoter, suggesting that *Ube3a-AS* RNA polymerase does not stall within *Ube3a* gene. Furthermore, we identified and verified a fourth *Ube3a* isoform that terminates within intron 4 in agreement with previous studies [35, 122]. Altogether suggesting an alternative model of imprinting *alternative polyadenylation*, wherein *Ube3a-AS* though some unknown mechanism leads to alternative exon usage, exon 4.1, and termination of *Ube3a* transcription.

Alternative polyadenylation (APA) is a widespread phenomenon within humans that generates isoforms with alternative 3' ends. Moreover, APA events occur commonly during development and cellular differentiation in a tissue-specific manner [245]. For example, BDNF, or brain-derived neurotrophic factor, is one such brain-specific sense/antisense gene that involves APA to produce short and long 3' UTR isoforms [246, 247]. Although unlike the proposed *Ube3a/Ube3a-AS* model, BDNF APA sites are within the 3' UTR. Even so, this type of alternative polyadenylation, terminal exon APA usage hypothesized in **Figure 1.12B** is known to be associated with splicing and APA machinery [248]. In the *Ube3a/Ube3a-AS* example, a weak 5' splicing site within the long intron 4 leads to dynamic competition between splicing and polyadenylation [249], which causes the alternative polyadenylation within intron 4 - paternal *Ube3a* isoform 4.

With the understanding that tissue-specific APA is not a unique phenomenon, we pro-

posed three possible reasons for the imprinting of *Ube3a* in neurons. The first involves the importance of paternal *Ube3a* isoform 4 caused by APA. Here, isoform 4, like many APA isoforms, could change microRNA binding sites [250], or potentially code for a new protein [251,252]. Additionally, changes in 3' UTR have effects to stability, cellular localization, and translation efficiency [253, 254], all of which effect RNA functionality. The second reason could be that *Ube3a-AS* has tissue-specific regulatory functions and the imprint evolved to express both *Ube3a* and *Ube3a-AS*. Finally, it is possible that *Ube3a-AS* and *Ube3a* isoform 4 expression have regulatory roles in brain development. Nevertheless, there are clear ramifications for current Angelman syndrome drug therapy strategies as discussed above.

Even so, the revelation of alternative polyadenylation generating paternal *Ube3a* isoform 4 opens the door for an alternative approach to AS drug therapies like exon skipping. Exon skipping is a form of RNA splicing used to restore reading frame within a gene and can be experimentally done via antisense-mediation. Antisense-mediation uses antisense oligonucleotides to hybridize to a sense target sequence leading to RNase H cleaving and gene-specific knockdown [255]. For exon skipping purposes, the antisense oligonucleotides are modified to act on pre-mRNA splicing by blocking splicing signals and induces exon skipping [256]. Currently, this technique is being used in phase III clinical trials held by Sarepta Therapeutics for Duchenne muscular dystrophy (DMD) [257]. As DMD is often caused by mutation that disrupt the open reading frame, exon skipping is used to modulate splicing of *DMD* gene to restore the reading frame [258–260]. Since *Ube3a* isoform 4 uses an alternative exon, exon 4.1, that terminates transcription, antisense oligonucleotides can be designed to skip exon 4.1 and possible produce a full-length paternal *Ube3a* transcript.

5.3 Embryonic stem cells are a versatile source for high-throughput screening

In **Chapter 4**, embryonic stem cells were used to generate neurons ultimately for HTS for Angelman syndrome. These ES cells were relatively easy to generate and expand for large-scale studies unlike primary neurons, which have a finite number of cells per animal. Additionally, ES cells from *Ube3a*$^{+/YFP}$ mice housed in the laboratory were produced within months demonstrating the modularity of the HTS method. While ES cell-derived neuronal cultures generate significantly more cells than primary cultures, there are other sources of immortalized neuronal cell-lines (e.g., P19, SH-SY5Y neuroblastoma cells, NT2, PC12 cells, etc.) that could potentially do the same; however, iPS (induced pluripotent stem) cells are the only immortalized cell-line surging in popularity for HTS purposes [261–263].

One of the major benefits of using human iPS cells over ES cells (human or mouse) is the ability to use patient-specific cells. As these cells are derived from postnatal somatic cells [264, 265], they can be collected directly from diseased individuals for more biological complex drug screenings. While the advantages of using iPS cells over ES cell appear abundantly clear, especially for translational medicine applications, there are still several concerns with the use of iPS cells including functionality and chromosomal aberrations and genetic modifications. For example, recent studies have shown variability in iPS neurons differentiation efficiency compared to more consistent differentiation using ES cells [266]. Moreover, high-resolution genetic and epigenetic analysis revealed differences between iPS and ES cells including DNA methylation and expression profiles [267, 268], which could effect the ability of iPS cells to recapitulate certain diseases accurately. Although it is not clear what these results mean, more research is required to understand which stem cell-line is better suited for HTS.

Moving forward with HTS in ES cell-derived neurons, three key recommendations are

provided. **1)** Test multiple lots of NeuN antibody before purchasing in bulk (> 20 vials).[1] **2)** Run no more than 12-13 plates in one setting.[2] **3)** Always add positive (Topotecan) and negative (Vehicle) controls to each plate for quality control.

5.4 Future studies

5.4.1 Investigation of Ube3a-AS/UBE3A-AS predicted transcripts

5.4.1.1 PacBio

While we were able to several polyadenylation sites of *UBE3A-AS/Ube3a-AS*, the majority of the predicted transcripts have yet to be identified. Moreover, we were unable to verify full-length transcripts with the short-read RNA-seq. To circumvent short-read problems for isoform sequencing, full-length cDNA sequences can be generated using PacBio long-read technologies to determine the exact number of isoforms, the degree of interconnection between upstream snoRNAs, and exon usage between isoforms.

5.4.1.2 circRNAs

In both human and mouse assemblies of the antisense region, transcripts that appeared to be circular RNAs (circRNAs) visually were detected. These transcripts showed significant brain-specific expression compared to the generalized isoform categories for human and mouse. First, bioinformatic algorithms like STAR circRNA function [269] or by following Memczak *et al.* protocol [270] to predict circRNA and cross-referenced with the databases of annotated circRNAs [271]. Their expression can then be verified with qPCR techniques as described in Li *et al.* (2017) [272]. Briefly, total RNA would be digested with RNase R and purified with phenol-chloroform extraction to be used in cDNA reaction and subsequent circRNA specific qPCR assays [270]. Results from these experiments

[1] NeuN antibody varied greatly depending on lot number

[2] Assuming image acquisition is approximate 90 min/plate, otherwise, plate number should be adjusted accordingly.

would provide additional support for regulatory functions of *UBE3A-AS/Ube3a-AS*.

5.4.2 Additional verification of Ube3a-AS control of Ube3a isoform 4 expression

As Topotecan is known to effect alternative splicing [184,185], additional experiments that are *Ube3a-AS* specific could also be conducted. One means of doing this is by using antisense oligonucleotides (ASOs) like Meng *et al.* (2013) [36]. These ASOs can be used to treated primary neuronal cultures and qPCR can be conducted similarly to Topotecan treated neurons in **Chapter 3**. In addition to ASOs, a PWS-IC deletion transgenetic mouse model could be used. For this approach, RNA would extracted from transgenetic mouse brains and expression could be measured with RT-PCR. Results from these experiments would provide additional support for the control of alternative polyadenylation of *Ube3a* isoform 4 by *Ube3a-AS*.

5.4.3 Molecular analysis of Ube3a-AS and Ube3a isoform 4

5.4.3.1 RNA stability analysis

Recent studies investigated the processing of *Ube3a-AS* determined that *Ube3a-AS* was an atypical RNAPII transcript [35]; however, we found that the antisense primers listed in the publication targeted introns of the spliced *Ube3a-AS* predicted by our RNA-seq analysis. As such, the processing of *Ube3a-AS* should be re-evaluated for polyadenylation, cellular localization, and stability. To test polyadenylation, processed mRNA can be compared to total RNA with quantitative PCR similar to Meng *et al.* (2012) [35]. To test localization of sense and antisense transcript, RNA FISH (fluorescence *in situ* hybridization) can be used on generalized *Ube3a-AS* isoforms and *Ube3a* isoform 4. This would allow for visualization of these transcripts at the sub-cellular level in a quantifiable manner. Stellaris offers several protocols for performing RNA FISH on their website (`https://www.biosearchtech.com/support/resources/stellaris-protocols`). Finally, RNA stability can be determined via half-life experiments on actinomycin D treated pri-

mary neuronal cultures as previously described [35].

5.4.3.2 Temporal regulation of the antisense transcripts

The temporal regulation analysis was limited to mouse hippocampal neurons, as such, addition molecular experiments could be conducted to look at temporal regulation in mouse cerebellum with qPCR techniques. Furthermore, temporal regulation in the human brain can also be investigated with publicly available RNA-seq data [273]. Both analysis could be used in conjunction with the Allen Brain Atlas to determine possible pathways for the *Ube3a-AS/UBE3A-AS*.

5.4.3.3 RNA Co-Immunoprecipitation

In addition to the temporal expression analysis, RNA co-immunoprecipitation to elucidate possible RNA-protein interactions for *Ube3a-AS* and *Ube3a* isoform 4. Jedamzik and Eckmann (2009) provide a full protocol for analyzing RNA-protein complexes by RNA co-immunoprecipitation [274]. Results from this study could be also be used in pathway analysis.

5.4.4 Exon skipping of Ube3a isoform 4 as a therapeutic strategy for Angelman syndrome

The identification of a paternal-specific *Ube3a* isoform generated from alternative polyadenylation offers the potential of an alternative therapeutic strategy for Angelman syndrome, exon skipping. While not a trivial strategy, there is a comprehensive protocol book for exon skipping [275]. In brief, potential splice sites and cis-regulatory elements (i.e. exonic and intronic splicing enhancer sequences) would first be identified via open-source software like ESEfinder 3.0 [276, 277]. Following identification, a splicing functional assay based on minigenes would be used, where the exonic fragment surround *Ube3a* exon 4.1 is amplified and cloned into a splicing competent minigene vector. If suc-

cessful, the antisense oligonucleotides used in the *in vitro* splicing assay can be optimized for animal models. Furthermore, the protocol book edited by Aartsma-Rus has several protocols to help with troubleshooting and optimization of exon skipping.

5.4.5 High-throughput screening assays

We developed and verified the ES cell-derived neuronal culture HTS method (**Chapter 4**). With this frame work in place, the next step is to perform a small molecule drug screen. Following a small molecule screen, hits would be verified via a primary neuron assay and drug dosage profile created for each positive hits. After which, animals can be used to further verify results. In addition to this small molecule screening, an RNA-interference HTS can also be run. Despite numerous studies into *UBE3A*, little is known about its pathway in neurons. An RNA-interference screen could give some insight into the pathways involved with the imprint of *Ube3a* in neurons.

REFERENCES

[1] Christian SL, Fantes JA, Mewborn SK, Huang B, Ledbetter DH. Large genomic duplicons map to sites of instability in the Prader-Willi/Angelman syndrome chromosome region (15q11-q13). Hum Mol Genet. 1999;8(6):1025–1037.

[2] Kishino T, Lalande M, Wagstaff J. UBE3A/E6-AP mutations cause Angelman syndrome. Nat Genet. 1997;15(1):70–73. doi:10.1038/ng0197-70.

[3] Matsuura T, Sutcliffe JS, Fang P, Galjaard RJ, Jiang YH, Benton CS, et al. De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. Nat Genet. 1997;15(1):74–77. doi:10.1038/ng0197-74.

[4] Nurmi EL, Bradford Y, Chen YH, Hall J, Arnone B, Gardiner MB, et al. Linkage disequilibrium at the Angelman syndrome gene UBE3A in autism families. Genomics. 2001;77(1-2):105–113. doi:10.1006/geno.2001.6617.

[5] Russo VEA, Martienssen RA, Riggs AD. Epigenetic mechanisms of gene regulation. Cold Spring Harbor monograph series,. Plainview, N.Y.: Cold Spring Harbor Laboratory Press; 1996.

[6] Rapkins RW, Hore T, Smithwick M, Ager E, Pask AJ, Renfree MB, et al. Recent assembly of an imprinted domain from non-imprinted components. PLos Genet. 2006;2(10):e182. doi:10.1371/journal.pgen.0020182.

[7] Amos-Landgraf JM, Ji Y, Gottlieb W, Depinet T, Wandstrat AE, Cassidy SB, et al. Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. The American Journal of Human Genetics. 1999;65(2):370–386. doi:10.1086/302510.

[8] Ji Y, Rebert NA, Joslin JM, Higgins MJ, Schultz RA, Nicholls RD. Structure of the highly conserved HERC2 gene and of multiple partially duplicated paralogs in human. Genome Res. 2000;10(3):319–329.

[9] Ji Y, Walkowicz MJ, Buiting K, Johnson DK, Tarvin RE, Rinchik EM, et al. The ancestral gene for transcribed, low-copy repeats in the Prader-Willi/Angelman region encodes a large protein implicated in protein trafficking, which is deficient in mice with neuromuscular and spermiogenic abnormalities. Hum Mol Genet. 1999;8(3):533–542.

[10] Pujana MA, Nadal M, Gratacos M, Peral B, Csiszar K, Gonzalez-Sarmiento R, et al. Additional complexity on human chromosome 15q: identification of a set of newly recognized duplicons (LCR15) on 15q11-q13, 15q24, and 15q26. Genome Res. 2001;11(1):98–111.

[11] Buiting K, Dittrich B, Gross S, Lich C, Farber C, Buchholz T, et al. Sporadic imprinting defects in Prader-Willi syndrome and Angelman syndrome: implications for imprint-switch models, genetic counseling, and prenatal diagnosis. The American Journal of Human Genetics. 1998;63(1):170–180. doi:10.1086/301935.

[12] Jiang YH, Wauki K, Liu Q, Bressler J, Pan Y, Kashork CD, et al. Genomic analysis of the chromosome 15q11-q13 Prader-Willi syndrome region and characterization of transcripts for GOLGA8E and WHCD1L1 from the proximal breakpoint region. BMC Genomics. 2008;9:50. doi:10.1186/1471-2164-9-50.

[13] Pujana MA, Nadal M, Guitart M, Armengol L, Gratacos M, Estivill X. Human chromosome 15q11-q14 regions of rearrangements contain clusters of LCR15 duplicons. Eur J Hum Genet. 2002;10(1):26–35. doi:10.1038/sj.ejhg.5200760.

[14] Webb T. Inv dup(15) supernumerary marker chromosomes. J Med Genet. 1994;31(8):585–594.

[15] Cassidy SB, Schwartz S, Miller JL, Driscoll DJ. Prader-Willi syndrome. Genetics in Medicine. 2012;14(1):10–26. doi:10.1038/gim.0b013e31822bead0.

[16] Goldstone AP. Prader-Willi syndrome: advances in genetics, pathophysiology and treatment. Trends Endocrinol Metab. 2004;15(1):12–20.

[17] Bortolin-Cavaille ML, Cavaille J. The SNORD115 (H/MBII-52) and SNORD116 (H/MBII-85) gene clusters at the imprinted Prader-Willi locus generate canonical box C/D snoRNAs. Nucleic Acids Research. 2012;40(14):6800–6807. doi:10.1093/nar/gks321.

[18] Ding F, Li HH, Zhang SW, Solomon NM, Camper SA, Cohen P, et al. SnoRNA Snord116 (Pwcr1/MBII-85) Deletion Causes Growth Deficiency and Hyperphagia in Mice. PLoS ONE. 2008;3(3):e1709. doi:10.1371/journal.pone.0001709.

[19] Duker AL, Ballif BC, Bawle EV, Person RE, Mahadevan S, Alliman S, et al. Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. Eur J Hum Genet. 2010;18(11):1196–1201. doi:10.1038/ejhg.2010.102.

[20] Battaglia A. The inv dup (15) or idic (15) syndrome (Tetrasomy 15q). Orphanet Journal of Rare Diseases. 2008;3(1):30. doi:10.1186/1750-1172-3-30.

[21] Bolton PF, Dennis NR, Browne CE, Thomas NS, Veltman MWM, Thompson RJ, et al. The phenotypic manifestations of interstitial duplications of proximal 15q with special reference to the autistic spectrum disorders. American Journal of Medical Genetics. 2001;105(8):675–685. doi:10.1002/ajmg.1551.

[22] Cook EH, Lindgren V, Leventhal BL, Courchesne R, Lincoln A, Shulman C, et al. Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. Am J Hum Genet. 1997;60(4):928–934.

[23] Schroer RJ, Phelan MC, Michaelis RC, Crawford EC, Skinner SA, Cuccaro M, et al. Autism and maternally derived aberrations of chromosome 15q. Am J Med Genet. 1998;76(4):327–336. doi:10.1002/(sici)1096-8628(19980401)76:4<327::aid-ajmg8>3.0.co;2-m.

[24] Browne CE, Dennis NR, Maher E, Long FL, Nicholson JC, Sillibourne J, et al. Inherited Interstitial Duplications of Proximal 15q: Genotype-Phenotype Correlations. The American Journal of Human Genetics. 1997;61(6):1342–1352. doi:10.1086/301624.

[25] Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, Willsey AJ, et al. Common genetic variants, acting additively, are a major source of risk for autism. Mol Autism. 2012;3(1):9. doi:10.1186/2040-2392-3-9.

[26] Williams CA, Beaudet AL, Clayton-Smith J, Knoll JH, Kyllerman M, Laan LA, et al. Angelman syndrome 2005: updated consensus for diagnostic criteria. Am J Med Genet A. 2006;140(5):413–418. doi:10.1002/ajmg.a.31074.

[27] Huibregtse JM, Scheffner M, Beaudenon S, Howley PM. A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. Proc Natl Acad Sci U S A. 1995;92(7):2563–2567.

[28] Scheffner M, Nuber U, Huibregtse JM. Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. Nature. 1995;373(6509):81–83. doi:10.1038/373081a0.

[29] Kumar S, Talis AL, Howley PM. Identification of HHR23A as a substrate for E6-associated protein-mediated ubiquitination. J Biol Chem. 1999;274(26):18785–18792.

[30] Sun J, Zhou W, Kaliappan K, Nawaz Z, Slingerland JM. ERalpha phosphorylation at Y537 by Src triggers E6-AP-ERalpha binding, ERalpha ubiquitylation, promoter occupancy, and target gene expression. Mol Endocrinol. 2012;26(9):1567–1577. doi:10.1210/me.2012-1140.

[31] Landers M, Bancescu DL, Le Meur E, Rougeulle C, Glatt-Deeley H, Brannan C, et al. Regulation of the large (similar to 1000 kb) imprinted murine Ube3a antisense transcript by alternative exons upstream of Snurf/Snrpn. Nucleic Acids Res. 2004;32(11):3480–3492. doi:10.1093/nar/gkh670.

[32] Rougeulle C, Cardoso C, Fontes M, Colleaux L, Lalande M. An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript. Nat Genet. 1998;19(1):15–16. doi:10.1038/ng0598-15.

[33] Runte M, Kroisel PM, Gillessen-Kaesbach G, Varon R, Horn D, Cohen MY, et al. SNURF-SNRPN and UBE3A transcript levels in patients with Angelman syndrome. Hum Genet. 2004;114(6):553–561. doi:10.1007/s00439-004-1104-z.

[34] Yamasaki K, Joh K, Ohta T, Masuzaki H, Ishimaru T, Mukai T, et al. Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of Ube3a. Hum Mol Genet. 2003;12(8):837–847.

[35] Meng L, Person RE, Beaudet AL. Ube3a-ATS is an atypical RNA polymerase II transcript that represses the paternal expression of Ube3a. Hum Mol Genet. 2012;21(13):3001–3012. doi:10.1093/hmg/dds130.

[36] Meng L, Person RE, Huang W, Zhu PJ, Costa-Mattioli M, Beaudet AL. Truncation of Ube3a-ATS unsilences paternal Ube3a and ameliorates behavioral defects in the Angelman syndrome mouse model. PLos Genet. 2013;9(12):e1004039. doi:10.1371/journal.pgen.1004039.

[37] Runte M, Huttenhofer A, Gross S, Kiefmann M, Horsthemke B, Buiting K. The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. Human Molecular Genetics. 2001;10(23):2687–2700. doi:10.1093/hmg/10.23.2687.

[38] Radford EJ, Ferrón SR, Ferguson-Smith AC. Genomic imprinting as an adaptative model of developmental plasticity. FEBS Lett. 2011;585(13):2059–2066. doi:10.1016/j.febslet.2011.05.063.

[39] Santoro F, Barlow DP. Developmental control of imprinted expression by macro non-coding RNAs. Semin Cell Dev Biol. 2011;22(4):328–335. doi:10.1016/j.semcdb.2011.02.018.

[40] Bird A. Perceptions of epigenetics. Nature. 2007;447(7143):396–398. doi:10.1038/nature05913.

[41] Edwards CA, Ferguson-Smith AC. Mechanisms regulating imprinted genes in clusters. Curr Opin Cell Biol. 2007;19(3):281–289. doi:10.1016/j.ceb.2007.04.013.

[42] Edwards CA, Rens W, Clarke O, Mungall AJ, Hore T, Graves JA, et al. The evolution of imprinting: chromosomal mapping of orthologues of mammalian imprinted domains in monotreme and marsupial mammals. BMC Evol Biol. 2007;7:157. doi:10.1186/1471-2148-7-157.

[43] Spahn L, Barlow DP. An ICE pattern crystallizes. Nat Genet. 2003;35(1):11–12. doi:10.1038/ng0903-11.

[44] Ciccone DN, Chen T. Histone lysine methylation in genomic imprinting. Epigenetics. 2009;4(4):216–220.

[45] Nicholls RD, Jong MT, Glenn CC, Gabriel J, Rogan PK, Driscoll DJ, et al. Multiple imprinted genes associated with Prader-Willi syndrome and location of an imprint-

ing control element. Acta Genet Med Gemellol (Roma). 1996;45(1-2):87–89.

[46] Nicholls RD, Knepper JL. Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. Annu Rev Genomics Hum Genet. 2001;2:153–175. doi:10.1146/annurev.genom.2.1.153.

[47] Brannan CI, Bartolomei MS. Mechanisms of genomic imprinting. Current Opinion in Genetics & Development. 1999;9(2):164–170. doi:10.1016/s0959-437x(99)80025-2.

[48] Blaydes SM, Elmore M, Yang T, Brannan CI. Analysis of murine Snrpn and human SNRPN gene imprinting in transgenic mice. Mammalian Genome. 1999;10(6):549–555. doi:10.1007/s003359901042.

[49] Glenn CC, Saitoh S, Jong MT, Filbrandt MM, Surti U, Driscoll DJ, et al. Gene structure, DNA methylation, and imprinted expression of the human SNRPN gene. Am J Hum Genet. 1996;58(2):335–346.

[50] Shemer R, Birger Y, Riggs AD, Razin A. Structure of the imprinted mouse Snrpn gene and establishment of its parental-specific methylation pattern. Proc Natl Acad Sci U S A. 1997;94(19):10267–10272.

[51] Kantor B, Shemer R, Razin A. The Prader-Willi/Angelman imprinted domain and its control center. Cytogenet Genome Res. 2006;113(1-4):300–305. doi:10.1159/000090845.

[52] Huang HS, Allen JA, Mabb AM, King IF, Miriyala J, Taylor-Blake B, et al. Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. Nature. 2012;481(7380):185–189. doi:10.1038/nature10726.

[53] King IF, Yandava CN, Mabb AM, Hsiao JS, Huang HS, Pearson BL, et al. Topoisomerases facilitate transcription of long genes linked to autism. Nature.

2013;501(7465):58–62. doi:10.1038/nature12504.

[54] Hillman PR, Christian SGB, Doan R, Cohen ND, Konganti K, Wang X, et al. Genomic imprinting does not reduce Ube3alevels expression levels; 2017.

[55] Gray TA, Smithwick MJ, Schaldach MA, Martone DL, Graves JA, McCarrey JR, et al. Concerted regulation and molecular evolution of the duplicated SNRPB'/B and SNRPN loci. Nucleic Acids Res. 1999;27(23):4577–4584.

[56] Nahkuri S, Taft RJ, Korbie DJ, Mattick JS. Molecular Evolution of the HBII-52 snoRNA Cluster. J Mol Biol. 2008;381(4):810–815. doi:10.1016/j.jmb.2008.06.057.

[57] Huntriss JD, Latchman DS, Williams DG. The snRNP core protein SmB and tissue-specific SmN protein are differentially distributed between snRNP particles. Nucleic Acids Res. 1993;21(17):4047–4053.

[58] Clark MB, Amaral PP, Schlesinger FJ, Dinger ME, Taft RJ, Rinn JL, et al. The reality of pervasive transcription. PLoS Biol. 2011;9(7):e1000625. doi:10.1371/journal.pbio.1000625.

[59] Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. Nat Rev Genet. 2007;8(6):413–423. doi:10.1038/nrg2083.

[60] Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. Science. 2005;309(5740):1559–1563. doi:10.1126/science.1112014.

[61] Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74. doi:10.1038/nature11247.

[62] Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. Nature. 2012;489(7414):101–108. doi:10.1038/nature11233.

[63] Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009;458(7235):223–227. doi:10.1038/nature07672.

[64] Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science. 2007;316(5830):1484–1488. doi:10.1126/science.1138341.

[65] Mattick JS, Makunin IV. Small regulatory RNAs in mammals. Hum Mol Genet. 2005;14 Spec No 1:R121–32. doi:10.1093/hmg/ddi101.

[66] Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet. 2009;10(3):155–159. doi:10.1038/nrg2521.

[67] Lee JT, Bartolomei MS. X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell. 2013;152(6):1308–1323. doi:10.1016/j.cell.2013.02.016.

[68] Mohammad F, Mondal T, Kanduri C. Epigenetics of imprinted long noncoding RNAs. Epigenetics. 2009;4(5):277–286.

[69] Sleutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature. 2002;415(6873):810–813. doi:10.1038/415810a.

[70] Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22(9):1775–1789. doi:10.1101/gr.132159.111.

[71] Washietl S, Kellis M, Garber M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. Genome Res. 2014;24(4):616–628. doi:10.1101/gr.165035.113.

[72] Bartolomei MS, Zemel S, Tilghman SM. Parental imprinting of the mouse H19 gene. Nature. 1991;351(6322):153–155. doi:10.1038/351153a0.

[73] Darrow EM, Chadwick BP. Boosting transcription by transcription: enhancer-associated transcripts. Chromosome Res. 2013;21(6-7):713–724. doi:10.1007/s10577-013-9384-6.

[74] Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. Antisense transcription in the mammalian transcriptome. Science. 2005;309(5740):1564–1566. doi:10.1126/science.1112009.

[75] Wevrick R, Francke U. An imprinted mouse transcript homologous to the human imprinted in Prader-Willi syndrome (IPW) gene. Hum Mol Genet. 1997;6(2):325–332.

[76] Wevrick R, Kerns JA, Francke U. Identification of a novel paternally expressed gene in the Prader-Willi syndrome region. Hum Mol Genet. 1994;3(10):1877–1882.

[77] Beiter T, Reich E, Williams RW, Simon P. Antisense transcription: a critical look in both directions. Cellular and Molecular Life Sciences. 2009;66(1):94–112. doi:10.1007/s00018-008-8381-y.

[78] Carlile M, Swan D, Jackson K, Preston-Fayers K, Ballester B, Flicek P, et al. Strand selective generation of endo-siRNAs from the Na/phosphate transporter gene Slc34a1 in murine tissues. Nucleic Acids Research. 2009;37(7):2274–2282. doi:10.1093/nar/gkp088.

[79] Dahary D, Elroy-Stein O, Sorek R. Naturally occurring antisense: transcriptional leakage or real overlap? Genome Res. 2005;15(3):364–368. doi:10.1101/gr.3308405.

[80] Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. Nat Rev Genet. 2013;14(12):880–893. doi:10.1038/nrg3594.

[81] Gonzalez I, Munita R, Agirre E, Dittmer TA, Gysling K, Misteli T, et al. A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. Nat Struct Mol Biol. 2015;22(5):370–376. doi:10.1038/nsmb.3005.

[82] Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of alternative splicing by histone modifications. Science. 2010;327(5968):996–1000. doi:10.1126/science.1184208.

[83] Zhang P, Du J, Sun B, Dong X, Xu G, Zhou J, et al. Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. Nucleic Acids Res. 2006;34(22):6621–6628. doi:10.1093/nar/gkl989.

[84] Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science. 2008;322(5908):1717–1720. doi:10.1126/science.1163802.

[85] Latos PA, Pauler FM, Koerner MV, Senergin HB, Hudson QJ, Stocsits RR, et al. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science. 2012;338(6113):1469–1472. doi:10.1126/science.1228110.

[86] Santoro F, Mayer D, Klement RM, Warczok KE, Stukalov A, Barlow DP, et al. Imprinted Igf2r silencing depends on continuous Airn lncRNA expression and is not restricted to a developmental window. Development. 2013;140(6):1184–1195. doi:10.1242/dev.088849.

[87] Kelsey G, Bodle D, Miller HJ, Beechey CV, Coombes C, Peters J, et al. Identification of imprinted loci by methylation-sensitive representational difference analysis: application to mouse distal chromosome 2. Genomics. 1999;62(2):129–138. doi:10.1006/geno.1999.6022.

[88] Peters J, Wroe SF, Wells CA, Miller HJ, Bodle D, Beechey CV, et al. A cluster of oppositely imprinted transcripts at the Gnas locus in the distal imprinting region of mouse chromosome 2. Proc Natl Acad Sci U S A. 1999;96(7):3830–3835.

[89] Williamson CM, Ball ST, Nottingham WT, Skinner JA, Plagge A, Turner MD, et al. A cis-acting control region is required exclusively for the tissue-specific imprinting of Gnas. Nat Genet. 2004;36(8):894–899. doi:10.1038/ng1398.

[90] Williamson CM, Skinner JA, Kelsey G, Peters J. Alternative non-coding splice variants of Nespas, an imprinted gene antisense to Nesp in the Gnas imprinting cluster. Mamm Genome. 2002;13(2):74–79. doi:10.1007/s00335-001-2102-2.

[91] Egrie JC, Browne JK. Development and characterization of novel erythropoiesis stimulating protein (NESP). Nephrology Dialysis Transplantation. 2001;16(SUPPL. 3):3–13.

[92] Peters J, Holmes R, Monk D, Beechey CV, Moore GE, Williamson CM. Imprinting control within the compact Gnas locus. Cytogenet Genome Res. 2006;113(1-4):194–201. doi:10.1159/000090832.

[93] Coombes C, Arnaud P, Gordon E, Dean W, Coar EA, Williamson CM, et al. Epigenetic properties and identification of an imprint mark in the Nesp-Gnasxl domain of the mouse Gnas imprinted locus. Mol Cell Biol. 2003;23(16):5475–5488.

[94] Williamson CM, Turner MD, Ball ST, Nottingham WT, Glenister P, Fray M, et al. Identification of an imprinting control region affecting the expression of all tran-

scripts in the Gnas cluster. Nat Genet. 2006;38(3):350–355. doi:10.1038/ng1731.

[95] Bliek J, Verde G, Callaway J, Maas SM, De Crescenzo A, Sparago A, et al. Hypomethylation at multiple maternally methylated imprinted regions including PLAGL1 and GNAS loci in Beckwith-Wiedemann syndrome. European Journal of Human Genetics. 2009;17(5):611–619. doi:10.1038/ejhg.2008.233.

[96] Holmes R, Williamson C, Peters J, Denny P, Arakawa T, Carninci P, et al. A comprehensive transcript map of the mouse Gnas imprinted complex. Genome Res. 2003;13(6 B):1410–1415. doi:10.1101/gr.955503.

[97] Ball ST, Kelly ML, Robson JE, Turner MD, Harrison J, Jones L, et al. Gene dosage effects at the imprinted Gnas cluster. PLoS ONE. 2013;8(6):e65639. doi:10.1371/journal.pone.0065639.

[98] Tibbit JC, Williamson MC, Mehta S, Ball TS, Chotalia M, Nottingham TW, et al. Antisense activity across the Nesp promoter is required for Nespas-mediated silencing in the imprinted Gnas cluster. Non-Coding RNA. 2015;1(3). doi:10.3390/ncrna1030246.

[99] Williamson CM, Ball ST, Dawson C, Mehta S, Beechey CV, Fray M, et al. Uncoupling antisense-mediated silencing and DNA methylation in the imprinted Gnas cluster. PLos Genet. 2011;7(3):e1001347. doi:10.1371/journal.pgen.1001347.

[100] Wroe SF, Kelsey G, Skinner JA, Bodle D, Ball ST, Beechey CV, et al. An imprinted transcript, antisense to Nesp, adds complexity to the cluster of imprinted genes at the mouse Gnas locus. Proc Natl Acad Sci U S A. 2000;97(7):3342–3346. doi:10.1073/pnas.050015397.

[101] Tiberi L, Vanderhaeghen P, van den Ameele J. Cortical neurogenesis and morphogens: diversity of cues, sources and functions. Curr Opin Cell Biol.

2012;24(2):269–276. doi:10.1016/j.ceb.2012.01.010.

[102] Zhang SC, Barclay C, Alexander LA, Geldenhuys L, Porter GA, Casson AG, et al. Alternative splicing of the FGF antisense gene: differential subcellular localization in human tissues and esophageal adenocarcinoma. J Mol Med (Berl). 2007;85(11):1215–1228. doi:10.1007/s00109-007-0219-9.

[103] Knee R, Li AW, Murphy PR. Characterization and tissue-specific expression of the rat basic fibroblast growth factor antisense mRNA and protein. Proc Natl Acad Sci U S A. 1997;94(10):4943–4947.

[104] McEachern LA, Murphy PR. Chromatin-remodeling factors mediate the balance of sense-antisense transcription at the FGF2 locus. Mol Endocrinol. 2014;28(4):477–489. doi:10.1210/me.2013-1220.

[105] MacFarlane LA, Gu Y, Casson AG, Murphy PR. Regulation of fibroblast growth factor-2 by an endogenous antisense RNA and by argonaute-2. Mol Endocrinol. 2010;24(4):800–812. doi:10.1210/me.2009-0367.

[106] Lee MP, Hu RJ, Johnson LA, Feinberg AP. Human KVLQT1 gene shows tissue-specific imprinting and encompasses Beckwith-Wiedemann syndrome chromosomal rearrangements. Nat Genet. 1997;15(2):181–185. doi:10.1038/ng0297-181.

[107] Smilinich NJ, Day CD, Fitzpatrick GV, Caldwell GM, Lossie AC, Cooper PR, et al. A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. Proc Natl Acad Sci U S A. 1999;96(14):8064–8069.

[108] Mancini-Dinardo D, Steele SJ, Levorse JM, Ingram RS, Tilghman SM. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev. 2006;20(10):1268–1282. doi:10.1101/gad.1416906.

[109] Thakur N, Tiwari VK, Thomassin H, Pandey RR, Kanduri M, Gondor A, et al. An antisense RNA regulates the bidirectional silencing property of the Kcnq1 imprinting control region. Mol Cell Biol. 2004;24(18):7855–7862. doi:10.1128/MCB.24.18.7855-7862.2004.

[110] Golding MC, Magri LS, Zhang L, Lalone SA, Higgins MJ, Mann MR. Depletion of Kcnq1ot1 non-coding RNA does not affect imprinting maintenance in stem cells. Development. 2011;138(17):3667–3678. doi:10.1242/dev.057778.

[111] Shin JY, Fitzpatrick GV, Higgins MJ. Two distinct mechanisms of silencing by the KvDMR1 imprinting control region. EMBO J. 2008;27(1):168–178. doi:10.1038/sj.emboj.7601960.

[112] Schultz BM, Gallicio GA, Cesaroni M, Lupey LN, Engel N. Enhancers compete with a long non-coding RNA for regulation of the Kcnq1 domain. Nucleic Acids Res. 2015;43(2):745–759. doi:10.1093/nar/gku1324.

[113] Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. Genome Res. 2002;12(6):996–1006. doi:10.1101/gr.229102.

[114] Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, et al. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of -secretase. Nat Med. 2008;14(7):723–730. doi:10.1038/nm1784.

[115] Faghihi MA, Zhang M, Huang J, Modarresi F, Van der Brug MP, Nalls MA, et al. Evidence for natural antisense transcript-mediated inhibition of microRNA function. Genome Biol. 2010;11(5). doi:10.1186/gb-2010-11-5-r56.

[116] Pruunsild P, Kazantseva A, Aid T, Palm K, Timmusk T. Dissecting the human BDNF locus: bidirectional transcription, complex splicing, and multiple promoters. Genomics. 2007;90(3):397–406. doi:10.1016/j.ygeno.2007.05.004.

[117] Modarresi F, Faghihi MA, Lopez-Toledano MA, Fatemi RP, Magistri M, Brothers SP, et al. Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. Nat Biotechnol. 2012;30(5):453–459. doi:10.1038/nbt.2158.

[118] Chamberlain SJ, Brannan CI. The Prader-Willi syndrome imprinting center activates the paternally expressed murine Ube3a antisense transcript but represses paternal Ube3a. Genomics. 2001;73(3):316–322. doi:10.1006/geno.2001.6543.

[119] Powell WT, Coulson RL, Crary FK, Wong SS, Ach RA, Tsang P, et al. A Prader-Willi locus lncRNA cloud modulates diurnal genes and energy expenditure. Hum Mol Genet. 2013;22(21):4318–4328. doi:10.1093/hmg/ddt281.

[120] Dindot SV, Person R, Strivens M, Garcia R, Beaudet AL. Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. Genome Res. 2009;19(8):1374–1383. doi:10.1101/gr.089185.108.

[121] Kishino T. Imprinting in neurons. Cytogenet Genome Res. 2006;113(1-4):209–214. doi:10.1159/000090834.

[122] Numata K, Kohama C, Abe K, Kiyosawa H. Highly parallel SNP genotyping reveals high-resolution landscape of mono-allelic Ube3a expression associated with locus-wide antisense transcription. Nucleic Acids Res. 2011;39(7):2649–2657. doi:10.1093/nar/gkq1201.

[123] Bailus BJ, Pyles B, McAlister MM, O'Geen H, Lockwood SH, Adams AN, et al. Protein delivery of an artificial transcription factor restores widespread Ube3a expression in an Angelman syndrome mouse brain. Molecular Therapy. 2016;24(3):548–555. doi:10.1038/mt.2015.236.

[124] Elgersma Y. Genetic engineering cures mice of neurological deficits: prospects for treating Angelman syndrome. Pharmacogenomics. 2007;8(6):539–541. doi:10.2217/14622416.8.6.539.

[125] Meng L, Ward AJ, Chun S, Bennett CF, Beaudet AL, Rigo F. Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. Nature. 2015;518(7539):409–412. doi:10.1038/nature13975.

[126] Shi SQ, Bichell TJ, Ihrie RA, Johnson CH. Ube3a imprinting impairs circadian robustness in Angelman syndrome models. Curr Biol. 2015;25(5):537–545. doi:10.1016/j.cub.2014.12.047.

[127] Silva-Santos S, van Woerden GM, Bruinsma CF, Mientjes E, Jolfaei MA, Distel B, et al. Ube3a reinstatement identifies distinct developmental windows in a murine Angelman syndrome model. J Clin Invest. 2015;125(5):2069–2076. doi:10.1172/JCI80554.

[128] Dindot SV, Antalffy BA, Bhattacharjee MB, Beaudet AL. The Angelman syndrome ubiquitin ligase localizes to the synapse and nucleus, and maternal deficiency results in abnormal dendritic spine morphology. Human Molecular Genetics. 2008;17(1):111–8. doi:10.1093/hmg/ddm288.

[129] Sutcliffe JS, Jiang YH, Galijaard RJ, Matsuura T, Fang P, Kubota T, et al. The E6-Ap ubiquitin-protein ligase (UBE3A) gene is localized within a narrowed Angelman syndrome critical region. Genome Res. 1997;7(4):368–77.

[130] Battaglia A, Gurrieri F, Bertini E, Bellacosa A, Pomponi MG, Paravatou-Petsotas M, et al. The inv dup(15) syndrome: a clinically recognizable syndrome with altered behavior, mental retardation, and epilepsy. Neurology. 1997;48(4):1081–6.

[131] Scoles HA, Urraca N, Chadwick SW, Reiter LT, Lasalle JM. Increased copy number for methylated maternal 15q duplications leads to changes in gene and protein expression in human cortical samples. Mol Autism. 2011;2(1):19. doi:10.1186/2040-2392-2-19.

[132] Sahoo T, del Gaudio D, German JR, Shinawi M, Peters SU, Person RE, et al. Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. Nat Genet. 2008;40(6):719–21. doi:10.1038/ng.158.

[133] Buiting K, Lich C, Cottrell S, Barnicoat A, Horsthemke B. A 5-kb imprinting center deletion in a family with Angelman syndrome reduces the shortest region of deletion overlap to 880 bp. Hum Genet. 1999;105(6):665–6.

[134] Buiting K, Barnicoat A, Lich C, Pembrey M, Malcolm S, Horsthemke B. Disruption of the bipartite imprinting center in a family with Angelman syndrome. Am J Hum Genet. 2001;68(5):1290–4. doi:10.1086/320120.

[135] Le Meur E, Watrin F, Landers M, Sturny R, Lalande M, Muscatelli F. Dynamic developmental regulation of the large non-coding RNA associated with the mouse 7C imprinted chromosomal region. Dev Biol. 2005;286(2):587–600. doi:10.1016/j.ydbio.2005.07.030.

[136] Martins-Taylor K, Hsiao JS, Chen PF, Glatt-Deeley H, De Smith AJ, Blakemore AI, et al. Imprinted expression of UBE3A in non-neuronal cells from a Prader-Willi syndrome patient with an atypical deletion. Hum Mol Genet. 2014;23(9):2364–73. doi:10.1093/hmg/ddt628.

[137] Pervouchine DD, Djebali S, Breschi A, Davis CA, Barja PP, Dobin A, et al. Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression. Nature Communications. 2015;6:5903. doi:10.1038/ncomms6903.

[138] Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419–1260419. doi:10.1126/science.1260419.

[139] Lin L, Park JW, Ramachandran S, Zhang Y, Tseng YT, Shen S, et al. Transcriptome sequencing reveals aberrant alternative splicing in Huntington's disease. Human Molecular Genetics. 2016;25(16):3454–3466. doi:10.1093/hmg/ddw187.

[140] Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, et al. An RNA-Sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. The Journal of Neuroscience. 2014;34(36):11929–11947. doi:10.1523/JNEUROSCI.1860-14.2014.

[141] You X, Vlatkovic I, Babic A, Will T, Epstein I, Tushev G, et al. Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. Nature Neuroscience. 2015;18(4):603–610. doi:10.1038/nn.3975.

[142] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921. doi:10.1038/35057062.

[143] Chinwalla AT, Cook LL, Delehaunty KD, Fewell GA, Fulton LA, Fulton RS, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002;420(6915):520–562. doi:10.1038/nature01262.

[144] Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2015 update. Nucleic Acids Research. 2014;43(D1):D670–D681. doi:10.1093/nar/gku1177.

[145] Derti A, Garrett-Engele P, MacIsaac KD, Stevens RC, Sriram S, Chen R, et al. A quantitative atlas of polyadenylation in five mammals. Genome Research. 2012;22(6):1173–1183. doi:10.1101/gr.132563.111.

[146] Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics. 2013;14(2):178–192. doi:10.1093/bib/bbs017.

[147] Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative Genomics Viewer. Nature Biotechnology. 2011;29(1):24–26. doi:10.1038/nbt.1754.

[148] Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biology. 2015;16(1):1–14. doi:10.1186/s13059-014-0560-6.

[149] Andrews S. FastQC: A Quality Control tool for High Throughput Sequence Data; 2010. Website. Available from: `http://www.bioinformatics.babraham.ac.uk/projects/fastqc/`.

[150] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–2120.

[151] Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature Protocols. 2016;11(9):1650–1667. doi:10.1038/nprot.2016.095.

[152] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Meth. 2015;12(4):357–360. doi:10.1038/nmeth.3317.

[153] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–2079.

[154] Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotech. 2015;33(3):290–295. doi:10.1038/nbt.3122.

[155] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2013;30(7):923–930. doi:10.1093/bioinformatics/btt656.

[156] Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Research. 2013;41(10):e108–e108. doi:10.1093/nar/gkt214.

[157] Chen Y, Lun ATL, Smyth GK. Differential Expression Analysis of Complex RNA-seq Experiments Using edgeR. In: Statistical Analysis of Next Generation Sequencing Data. Springer Nature; 2014. p. 51–74. Available from: https://doi.org/10.1007%2F978-3-319-07212-8_3.

[158] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26(1):139–140. doi:10.1093/bioinformatics/btp616.

[159] Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nature Protocols. 2013;8(9):1765–1786. doi:10.1038/nprot.2013.099.

[160] Huber, W , Carey, J V, Gentleman, R , et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nature Methods. 2015;12(2):115–121.

[161] R Core Team. R: A Language and Environment for Statistical Computing; 2016. Available from: `https://www.R-project.org/`.

[162] Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. The Annals of Applied Statistics. 2016;10(2):946–963. doi:10.1214/16-AOAS920.

[163] McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Research. 2012;40(10):4288–4297. doi:10.1093/nar/gks042.

[164] Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics. 2007;9(2):321–332. doi:10.1093/biostatistics/kxm030.

[165] Robinson MD, Smyth GK. Moderated statistical tests for assessing differences in tag abundance. Bioinformatics. 2007;23(21):2881–2887. doi:10.1093/bioinformatics/btm453.

[166] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289–300.

[167] Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 2011;477(7364):289–294. doi:10.1038/nature10413.

[168] Yalcin B, Wong K, Agam A, Goodson M, Keane TM, Gan X, et al. Sequence-based characterization of structural variation in the mouse genome. Nature. 2011;477(7364):326–329. doi:10.1038/nature10432.

[169] Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–842. doi:10.1093/bioinformatics/btq033.

[170] Fejes-Toth K, Sotirova V, Sachidanandam R, Assaf G, Hannon GJ, Kapranov P, et al. Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. Nature. 2009;457(7232):1028–1032. doi:10.1038/nature07759.

[171] Kiss DL, Oman K, Bundschuh R, Schoenberg DR. Uncapped 5 ends of mRNAs targeted by cytoplasmic capping map to the vicinity of downstream CAGE tags. FEBS Letters. 2015;589(3):279–284. doi:10.1016/j.febslet.2014.12.009.

[172] Mercer TR, Dinger ME, Bracken CP, Kolle G, Szubert JM, Korbie DJ, et al. Regulated post-transcriptional RNA cleavage diversifies the eukaryotic transcriptome. Genome Research. 2010;20(12):1639–1650. doi:10.1101/gr.112128.110.

[173] Hocine S, Singer RH, Grunwald D. RNA Processing and Export. Cold Spring Harbor Perspectives in Biology. 2010;2(12):a000752–a000752. doi:10.1101/cshperspect.a000752.

[174] for Brain Science AI. Allen Developing Mouse Brain Atlas; 2008. Available from: http://developingmouse.brain-map.org/.

[175] Sunkin SM, Ng L, Lau C, Dolbeare T, Gilbert TL, Thompson CL, et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. Nucleic Acids Research. 2012;41(D1):D996–D1008. doi:10.1093/nar/gks1042.

[176] Scheffner M, Huibregtse JM, Vierstra RD, Howley PM. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell.

1993;75(3):495–505. doi:10.1016/0092-8674(93)90384-3.

[177] Dhananjayan SC, Ramamoorthy S, Khan OY, Ismail A, Sun J, Slingerland J, et al. WW domain binding protein-2, an E6-associated protein interacting protein, acts as a coactivator of estrogen and progesterone receptors. Molecular Endocrinology. 2006;20(10):2343–54. doi:10.1210/me.2005-0533.

[178] Khan OY, Fu G, Ismail A, Srinivasan S, Cao X, Tu Y, et al. Multifunction steroid receptor coactivator, E6-associated protein, is involved in development of the prostate gland. Molecular Endocrinology. 2006;20(3):544–59. doi:10.1210/me.2005-0110.

[179] Umlauf D, Goto Y, Cao R, Cerqueira F, Wagschal A, Zhang Y, et al. Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. Nature Genetics. 2004;36(12):1296–1300. doi:10.1038/ng1467.

[180] Kavalali ET, Klingauf J, Tsien RW. Activity-dependent regulation of synaptic clustering in a hippocampal culture system. Proceedings of the National Academy of Sciences. 1999;96(22):12893–12900. doi:10.1073/pnas.96.22.12893.

[181] Brendel V, Xing L, Zhu W. Gene structure prediction from consensus spliced alignment of multiple ESTs matching the same genomic locus. Bioinformatics. 2004;20(7):1157–1169. doi:10.1093/bioinformatics/bth058.

[182] MacIsaac JL, Bogutz AB, Morrissy AS, Lefebvre L. Tissue-specific alternative polyadenylation at the imprinted gene Mest regulates allelic usage at Copg2. Nucleic Acids Research. 2011;40(4):1523–1535. doi:10.1093/nar/gkr871.

[183] Morrissy AS, Griffith M, Marra MA. Extensive relationship between antisense transcription and alternative splicing in the human genome. Genome Research. 2011;21(8):1203–1212. doi:10.1101/gr.113431.110.

[184] Shkreta L, Froehlich U, Paquet ER, Toutant J, Elela SA, Chabot B. Anticancer drugs affect the alternative splicing of Bcl-x and other human apoptotic genes. Molecular Cancer Therapeutics. 2008;7(6):1398–1409. doi:10.1158/1535-7163.mct-08-0192.

[185] Powell WT, Coulson RL, Gonzales ML, Crary FK, Wong SS, Adams S, et al. R-loop formation atSnord116mediates topotecan inhibition ofUbe3a-antisenseand allele-specific chromatin decondensation. Proceedings of the National Academy of Sciences. 2013;110(34):13938–13943. doi:10.1073/pnas.1305426110.

[186] Kaneko-Ishino T, Kohda T, Ono R, Ishino F. Complementation hypothesis: the necessity of a monoallelic gene expression mechanism in mammalian development. Cytogenetic and Genome Research. 2006;113(1-4):24–30. doi:10.1159/000090811.

[187] Kaneko-Ishino T. The Regulation and Biological Significance of Genomic Imprinting in Mammals. Journal of Biochemistry. 2003;133(6):699–711. doi:10.1093/jb/mvg090.

[188] Aubi O, Flydal MI, Zheng HX, Skjaerven L, Rekand I, Leiros HKS, et al. Discovery of a specific inhibitor of pyomelanin synthesis in Legionella pneumophila. Journal of Medicinal Chemistry. 2015;58(21):8402–8412. doi:10.1021/acs.jmedchem.5b01589.

[189] Holbeck SL. Update on NCI in vitro drug screen utilities. European Journal of Cancer. 2004;40(6):785–793. doi:http://dx.doi.org/10.1016/j.ejca.2003.11.022.

[190] Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, Garyantes T, et al. Impact of high-throughput screening in biomedical research. Nature Reviews: Drug Discovery. 2011;10(3):188–195. doi:10.1038/nrd3368.

[191] Seto SW, Kiat H, Lee SMY, Bensoussan A, Sun YT, Hoi MPM, et al. Zebrafish models of cardiovascular diseases and their applications in herbal medicine research. European Journal of Pharmacology. 2015;768:77–86. doi:10.1016/j.ejphar.2015.10.031.

[192] Stewart AM, Grieco F, Tegelenbosch RAJ, Kyzar EJ, Nguyen M, Kaluyeva A, et al. A novel 3D method of locomotor analysis in adult zebrafish: Implications for automated detection of CNS drug-evoked phenotypes. Journal of Neuroscience Methods. 2015;255:66–74. doi:10.1016/j.jneumeth.2015.07.023.

[193] Stone VN, Parikh HI, El-rami F, Ge XC, Chen WH, Zhang Y, et al. Identification of Small-Molecule Inhibitors against Meso-2, 6-Diaminopimelate Dehydrogenase from Porphyromonas gingivalis. PLoS One. 2015;10(11):24. doi:10.1371/journal.pone.0141126.

[194] White JR, Lee JM, Young PR, Hertzberg RP, Jurewicz AJ, Chaikin MA, et al. Identification of a potent, selective non-peptide CXCR2 antagonist that inhibits interleukin-8-induced neutrophil migration. Journal of Biological Chemistry. 1998;273(17):10095–10098. doi:10.1074/jbc.273.17.10095.

[195] Bernstock JD, Lee YJ, Peruzzotti-Jametti L, Southall N, Johnson KR, Maric D, et al. A novel quantitative high-throughput screen identifies drugs that both activate SUMO conjugation via the inhibition of microRNAs 182 and 183 and facilitate neuroprotection in a model of oxygen and glucose deprivation. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism. 2016;36(2):426–41. doi:10.1177/0271678x15609939.

[196] Berry BJ, Akanda N, Smith AST, Long CJ, Schnepper MT, Guo X, et al. Morphological and Functional Characterization of Human Induced Pluripotent Stem

Cell-Derived Neurons (iCell Neurons) in Defined Culture Systems. Biotechnology Progress. 2015;31(6):1613–1622. doi:10.1002/btpr.2160.

[197] Eglen RM, Gilchrist A, Reisine T. An overview of drug screening using primary and embryonic stem cells. Combinatorial Chemistry & High Throughput Screening. 2008;11(7):566–572. doi:10.2174/138620708785204108.

[198] Gordon KR, Wang Y, Allbritton NL, Taylor AM. Magnetic Alignment of Microelements Containing Cultured Neuronal Networks for High-Throughput Screening. Journal of Biomolecular Screening. 2015;20(9):1091–1100. doi:10.1177/1087057115598609.

[199] Kaltenbach LS, Bolton MM, Shah B, Kanju PM, Lewis GM, Turmel GJ, et al. Composite Primary Neuronal High-Content Screening Assay for Huntington's Disease Incorporating Non-Cell-Autonomous Interactions. Journal of Biomolecular Screening. 2010;15(7):806–819. doi:10.1177/1087057110373392.

[200] Mazzio E, Georges B, McTier O, Soliman KFA. Neurotrophic Effects of Mu Bie Zi (Momordica cochinchinensis) Seed Elucidated by High-Throughput Screening of Natural Products for NGF Mimetic Effects in PC-12 Cells. Neurochemical Research. 2015;40(10):2102–2112. doi:10.1007/s11064-015-1560-y.

[201] Pouton CW, Haynes JM. Embryonic stem cells as a source of models for drug discovery. Nature Reviews Drug Discovery. 2007;6(8):605–616. doi:10.1038/nrd2194.

[202] Schulte J, Sepp KJ, Wu C, Hong P, Littleton JT. High-Content Chemical and RNAi Screens for Suppressors of Neurotoxicity in a Huntington's Disease Model. Plos One. 2011;6(8). doi:10.1371/journal.pone.0023841.

[203] Stewart E, Shelat A, Bradley C, Chen X, Federico S, Thiagarajan S, et al. Development and characterization of a human orthotopic neuroblastoma xenograft.

Developmental Biology. 2015;407(2):344–355. doi:10.1016/j.ydbio.2015.02.002.

[204] Beaudoin GMJ, Lee SH, Singh D, Yuan Y, Ng YG, Reichardt LF, et al. Culturing pyramidal neurons from the early postnatal mouse hippocampus and cortex. Nature Protocols. 2012;7(9):1741–1754. doi:10.1038/nprot.2012.099.

[205] Gonzalez F, Zhu Z, Shi ZD, Lelli K, Verma N, Li QV, et al. An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell Stem Cell. 2014;15(2):215–26. doi:10.1016/j.stem.2014.05.018.

[206] Vasileva EA, Shuvalov OU, Garabadgiu AV, Melino G, Barlev NA. Genome-editing tools for stem cell biology. Cell Death Dis. 2015;6:e1831. doi:10.1038/cddis.2015.167.

[207] Wang Y, Zhang WY, Hu S, Lan F, Lee AS, Huber B, et al. Genome editing of human embryonic stem cells and induced pluripotent stem cells with zinc finger nucleases for cellular imaging. Circ Res. 2012;111(12):1494–503. doi:10.1161/CIRCRESAHA.112.274969.

[208] Zhu Z, Gonzalez F, Huangfu D. The iCRISPR platform for rapid genome editing in human pluripotent stem cells. Methods Enzymol. 2014;546:215–50. doi:10.1016/B978-0-12-801185-0.00011-8.

[209] Desbordes SC, Placantonakis DG, Ciro A, Socci ND, Lee G, Djaballah H, et al. High-throughput screening assay for the identification of compounds regulating self-renewal and differentiation in human embryonic stem cells. Cell Stem Cell. 2008;2(6):602–612. doi:10.1016/j.stem.2008.05.010.

[210] Benninger F, Beck H, Wernig M, Tucker KL, Brüstle O, Scheffler B. Functional integration of embryonic stem cell-derived neurons in hippocampal slice cultures. Journal of Neuroscience. 2003;23(18):7075–7083.

[211] Garcia I, Huang L, Ung K, Arenkiel BR. Tracing Synaptic Connectivity onto Embryonic Stem Cell-derived Neurons. Stem Cells. 2012;30(10):2140–2151. doi:10.1002/stem.1185.

[212] Kim JH, Auerbach JM, Rodriguez-Gomez JA, Velasco I, Gavin D, Lumelsky N, et al. Dopamine neurons derived from embryonic stem cells function in an animal model of Parkinson's disease. Nature. 2002;418(6893):50–56.

[213] Weick JP, Liu Y, Zhang SC. Human embryonic stem cell-derived neurons adopt and regulate the activity of an established neural network. Proceedings of the National Academy of Sciences, USA. 2011;108(50):20189–20194. doi:10.1073/pnas.1108487108.

[214] Young GT, Gutteridge A, Fox HDE, Wilbrey AL, Cao L, Cho LT, et al. Characterizing Human Stem Cell-derived Sensory Neurons at the Single-cell Level Reveals Their Ion Channel Expression and Utility in Pain Research. Molecular Therapy. 2014;22(8):1530–1543. doi:10.1038/mt.2014.86.

[215] Breier JM, Radio NM, Mundy WR, Shafer TJ. Development of a high-throughput screening assay for chemical effects on proliferation and viability of immortalized human neural progenitor cells. Toxicological Sciences. 2008;105(1):119–133. doi:10.1093/toxsci/kfn115.

[216] Koike-Yusa H, Li YL, Tan EP, Velasco-Herrera MD, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nature Biotechnology. 2014;32(3):267–273. doi:10.1038/nbt.2800.

[217] Sadek H, Hannack B, Choe E, Wang J, Latif S, Garry MG, et al. Cardiogenic small molecules that enhance myocardial repair by stem cells. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(16):6063–6068. doi:10.1073/pnas.0711507105.

[218] Tumarkin E, Tzadu L, Csaszar E, Seo M, Zhang H, Lee A, et al. High-throughput combinatorial cell co-culture using microfluidics. Integrative Biology. 2011;3(6):653–662. doi:10.1039/c1ib00002k.

[219] Yoshikawa T, Piao YL, Zhong JH, Matoba R, Carter MG, Wang YX, et al. High-throughput screen for genes predominantly expressed in the ICM of mouse blastocysts by whole mount in situ hybridization. Gene Expression Patterns. 2006;6(2):213–224. doi:10.1016/j.modgep.2005.06.003.

[220] Behringer R, Gertsenstein M, Nagy KV, Nagy A. Manipulating the mouse embryo : a laboratory manual. Cold Spring Harbor, New York : Cold Spring Harbor Laboratory Press, [2014] Fourth edition.; 2014. Available from: http://lib-ezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat03226a&AN=tamum.92441&site=eds-live.

[221] Conner DA. Mouse Embryonic Stem (ES) Cell Isolation. Current Protocols in Molecular Biology. 2001;52(23.4):23.4.1–23.4.9. doi:10.1002/0471142727.mb2304s52.

[222] Czechanski A, Byers C, Greenstein I, Schrode N, Donahue LR, Hadjantonakis AK, et al. Derivation and characterization of mouse embryonic stem cells from permissive and nonpermissive strains. Nature Protocols. 2014;9(3):559–574. doi:10.1038/nprot.2014.030.

[223] Luo C, Zuñiga J, Edison E, Palla S, Dong W, Parker-Thornburg J. Superovulation Strategies for 6 Commonly Used Mouse Strains. Journal of the American Association for Laboratory Animal Science. 2011;50(4):471–478.

[224] Nagy A. Manipulating the mouse embryo : a laboratory manual. Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, [2003] Third

edition.; 2003. Available from: http://lib-ezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat03226a&AN=tamum.49300&site=eds-live.

[225] Gertsenstein M, Nutter LMJ, Reid T, Pereira M, Stanford WL, Rossant J, et al. Efficient Generation of Germ Line Transmitting Chimeras from C57BL/6N ES Cells by Aggregation with Outbred Host Embryos. PLoS One. 2010;5(6):e11260. doi:10.1371/journal.pone.0011260.

[226] McMahon AP, Bradley A. The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. Cell. 1990;62(6):1073–1085.

[227] Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. Nature. 2007;448(7151):313–317. doi:10.1038/nature05934.

[228] Takahashi K, Okita K, Nakagawa M, Yamanaka S. Induction of pluripotent stem cells from fibroblast cultures. Nature Protocols. 2007;2(12):3081–3089.

[229] Takahashi K, Narita M, Yokura M, Ichisaka T, Yamanaka S. Human Induced Pluripotent Stem Cells on Autologous Feeders. PLoS One. 2009;4(12):e8067. doi:10.1371/journal.pone.0008067.

[230] Bibel M, Richter J, Lacroix E, Barde YA. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. Nature Protocols. 2007;2(5):1034–1043. doi:10.1038/nprot.2007.147.

[231] Kiris E, Nuss JE, Burnett JC, Kota KP, Koh DC, Wanner LM, et al. Embryonic stem cell derived motoneurons provide a highly sensitive cell culture model for botulinum neurotoxin studies, with implications for high-throughput drug discovery. Stem Cell Research. 2011;6(3):195–205. doi:10.1016/j.scr.2011.01.002.

[232] Wu CY, Whye D, Mason RW, Wang W. Efficient differentiation of mouse embryonic stem cells into motor neurons. Journal of Visualized Experiments: JoVE. 2012;(64):e3813. doi:10.3791/3813.

[233] Watase K, Weeber EJ, Xu B, Antalffy B, Yuva-Paylor L, Hashimoto K, et al. A Long CAG Repeat in the Mouse Sca1 Locus Replicates SCA1 Features and Reveals the Impact of Protein Solubility on Selective Neurodegeneration. Neuron. 2002;34(6):905–919. doi:10.1016/S0896-6273(02)00733-X.

[234] Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9(7):676–82. doi:10.1038/nmeth.2019.

[235] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671–5.

[236] Kohama C, Kato H, Numata K, Hirose M, Takemasa T, Ogura A, et al. ES cell differentiation system recapitulates the establishment of imprinted gene expression in a cell-type-specific manner. Human Molecular Genetics. 2012;21(6):1391–1401. doi:10.1093/hmg/ddr577.

[237] Zhang JH, Chung TDY, Oldenburg KR. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. Journal of Biomolecular Screening. 1999;4(2):67–73. doi:10.1177/108705719900400206.

[238] Zhang XD. Optimal high-throughput screening : practical experimental design and data analysis for genome-scale RNAi research. Cambridge: Cambridge University Press; 2011.

[239] Zhang XD. Illustration of SSMD, z Score, SSMD*, z* Score, and t Statistic for Hit Selection in RNAi High-Throughput Screens. Journal of Biomolecular Screening.

2011;16(7):775–785. doi:10.1177/1087057111405851.

[240] Niedringhaus M, Dumitru R, Mabb AM, Wang Y, Philpot BD, Allbritton NL, et al. Transferable neuronal mini-cultures to accelerate screening in primary and induced pluripotent stem cell-derived neurons. Scientific Reports. 2015;5:8353. doi:10.1038/srep08353.

[241] Sui Y, Wu Z. Alternative Statistical Parameter for High-Throughput Screening Assay Quality Assessment. Journal of Biomolecular Screening. 2007;12(2):229–234. doi:10.1177/1087057106296498.

[242] Liscovitch N, Chechik G. Specialization of Gene Expression during Mouse Brain Development. PLoS Computational Biology. 2013;9(9):e1003185. doi:10.1371/journal.pcbi.1003185.

[243] Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, et al. Genome-wide atlas of gene expression in the adult mouse brain. Nature. 2006;445(7124):168–176. doi:10.1038/nature05453.

[244] Zapala MA, Hovatta I, Ellison JA, Wodicka L, Rio JAD, Tennant R, et al. Adult mouse brain gene expression patterns bear an embryologic imprint. Proceedings of the National Academy of Sciences. 2005;102(29):10357–10362. doi:10.1073/pnas.0503357102.

[245] Zhang H, Lee JY, Tian B. Biased alternative polyadenylation in human tissues. Genome Biology. 2005;6(12):1–13. doi:10.1186/gb-2005-6-12-r100.

[246] Timmusk T, Palm K, Metsis M, Reintam T, Paalme V, Saarma M, et al. Multiple promoters direct tissue-specific expression of the rat BDNF gene. Neuron. 1993;10(3):475–489. doi:10.1016/0896-6273(93)90335-O.

[247] Lau AG, Irier HA, Gu J, Tian D, Ku L, Liu G, et al. Distinct 3'UTRs differentially regulate activity-dependent translation of brain-derived neurotrophic factor (BDNF). Proceedings of the National Academy of Sciences. 2010;107(36):15945–15950. doi:10.1073/pnas.1002929107.

[248] Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. Nat Rev Genet. 2013;14(7):496–506.

[249] Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008;456(7221):470–476. doi:10.1038/nature07509.

[250] Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating Cells Express mRNAs with Shortened 3' Untranslated Regions and Fewer MicroRNA Target Sites. Science. 2008;320(5883):1643–1647. doi:10.1126/science.1155390.

[251] Alt FW, Bothwell ALM, Knapp M, Siden E, Mather E, Koshland M, et al. Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3′ ends. Cell. 1980;20(2):293–301. doi:10.1016/0092-8674(80)90615-7.

[252] Tian B, Pan Z, Lee JY. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome Research. 2007;17(2):156–165. doi:10.1101/gr.5532707.

[253] Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. Annual Review of Biochemistry. 2010;79(1):351–379. doi:10.1146/annurev-biochem-060308-103103.

[254] Andreassi C, Riccio A. To localize or not to localize: mRNA fate is in 3′UTR ends. Trends in Cell Biology. 2009;19(9):465–474. doi:10.1016/j.tcb.2009.06.001.

[255] Hausen P, Stein H. Ribonuclease H. An Enzyme Degrading the RNA Moiety of DNA-RNA Hybrids. European Journal of Biochemistry. 1970;14(2):278–283. doi:10.1111/j.1432-1033.1970.tb00287.x.

[256] Kole R, Sazani P. Antisense effects in the cell nucleus: modification of splicing. Current opinion in molecular therapeutics. 2001;3(3):229–234.

[257] Kole R, Krieg AM. Exon skipping therapy for Duchenne muscular dystrophy. Advanced Drug Delivery Reviews. 2015;87:104–107. doi:http://dx.doi.org/10.1016/j.addr.2015.05.008.

[258] Hoffman EP, Brown J Robert H, Kunkel LM. Dystrophin: The protein product of the duchenne muscular dystrophy locus. Cell. 1987;51(6):919–928. doi:10.1016/0092-8674(87)90579-4.

[259] Monaco AP, Bertelson CJ, Liechti-Gallati S, Moser H, Kunkel LM. An explanation for the phenotypic differences between patients bearing partial deletions of the {DMD} locus. Genomics. 1988;2(1):90–95. doi:http://dx.doi.org/10.1016/0888-7543(88)90113-9.

[260] Koenig M, Beggs AH, Moyer M, Scherpf S, Heindrich K, Bettecken T, et al. The molecular basis for Duchenne versus Becker muscular dystrophy: Correlation of severity with type of deletion. American Journal of Human Genetics. 1989;45:498–506.

[261] Chambers SM, Qi Y, Mica Y, Lee G, Zhang XJ, Niu L, et al. Combined small-molecule inhibition accelerates developmental timing and converts human pluripotent stem cells into nociceptors. Nature Biotechnology. 2012;30(7):715–720. doi:10.1038/nbt.2249.

[262] Crompton LA, Byrne ML, Taylor H, Kerrigan TL, Bru-Mercier G, Badger JL, et al. Stepwise, non-adherent differentiation of human pluripotent stem cells to generate basal forebrain cholinergic neurons via hedgehog signaling. Stem Cell Research. 2013;11(3):1206–1221. doi:http://dx.doi.org/10.1016/j.scr.2013.08.002.

[263] Engle SJ, Vincent F. Small Molecule Screening in Human Induced Pluripotent Stem Cell-derived Terminal Cell Types. The Journal of Biological Chemistry. 2014;289:4562–4570.

[264] Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007;131(5):861–872. doi:10.1016/j.cell.2007.11.019.

[265] Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, et al. Induced pluripotent stem cell lines derived from human somatic cells. Science. 2007;318(5858):1917–1920. doi:10.1126/science.1151526.

[266] Hu BY, Weick JP, Yu J, Ma LX, Zhang XQ, Thomson JA, et al. Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. Proceedings of the National Academy of Sciences. 2010;107(9):4335–4340. doi:10.1073/pnas.0910012107.

[267] Chin MH, Mason MJ, Xie W, Volinia S, Singer M, Peterson C, et al. Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures. Cell Stem Cell. 2009;5(1):111–123. doi:10.1016/j.stem.2009.06.008.

[268] Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat Biotech. 2009;27(4):353–360. doi:10.1038/nbt.1530.

[269] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2012;29(1):15–21. doi:10.1093/bioinformatics/bts635.

[270] Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature. 2013;495(7441):333–338. doi:10.1038/nature11928.

[271] Guo JU, Agarwal V, Guo H, Bartel DP. Expanded identification and characterization of mammalian circular RNAs. Genome Biology. 2014;15(7). doi:10.1186/s13059-014-0409-z.

[272] Li L, Guo J, Chen Y, Chang C, Xu C. Comprehensive CircRNA expression profile and selection of key CircRNAs during priming phase of rat liver regeneration. BMC Genomics. 2017;18(1). doi:10.1186/s12864-016-3476-6.

[273] Liu SJ, Nowakowski TJ, Pollen AA, Lui JH, Horlbeck MA, Attenello FJ, et al. Single-cell analysis of long non-coding RNAs in the developing human neocortex. Genome Biology. 2016;17(1). doi:10.1186/s13059-016-0932-1.

[274] Jedamzik B, Eckmann CR. Analysis of RNA-Protein Complexes by RNA Coimmunoprecipitation and RT-PCR Analysis from Caenorhabditis elegans. Cold Spring Harbor Protocols. 2009;2009(10):pdb.prot5300–pdb.prot5300. doi:10.1101/pdb.prot5300.

[275] Aartsma-Rus A, editor. Exon Skipping. Humana Press; 2012. Available from: http://dx.doi.org/10.1007/978-1-61779-767-5.

[276] Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Human Molecular Genetics. 2006;15(16):2490–2508. doi:10.1093/hmg/ddl171.

[277] Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: a web resource to identify exonic splicing enhancers. Nucleic Acids Research. 2003;31(13):3568–3571. doi:10.1093/nar/gkg616.

[278] Shimoji T, Murakami K, Sugiyama Y, Matsuda M, Inubushi S, Nasu J, et al. Identification of annexin A1 as a novel substrate for E6AP-mediated ubiquitylation. Journal of Cellular Biochemistry. 2009;106(6):1123–35. doi:10.1002/jcb.22096.

[279] Jensen L, Farook MF, Reiter LT. Proteomic Profiling in Drosophila Reveals Potential Dube3a Regulation of the Actin Cytoskeleton and Neuronal Homeostasis. PLoS ONE. 2013;8(4):e61952. doi:10.1371/journal.pone.0061952.

[280] Thomas M, Banks L. Inhibition of Bak-induced apoptosis by HPV-18 E6. Oncogene. 1998;17(23):2943–54. doi:10.1038/sj.onc.1202223.

[281] Oda H, Kumar S, Howley PM. Regulation of the Src family tyrosine kinase Blk through E6AP-mediated ubiquitination. Proc Natl Acad Sci U S A. 1999;96(17):9557–62.

[282] Gossan NC, Zhang F, Guo B, Jin D, Yoshitane H, Yao A, et al. The E3 ubiquitin ligase UBE3A is an integral component of the molecular circadian clock through regulating the BMAL1 transcription factor. Nucleic Acids Research. 2014;42(9):5765–5775. doi:10.1093/nar/gku225.

[283] Chan AL, Grossman T, Zuckerman V, Campigli Di Giammartino D, Moshel O, Scheffner M, et al. c-Abl phosphorylates E6AP and regulates its E3 ubiquitin ligase activity. Biochemistry. 2013;52(18):3119–29. doi:10.1021/bi301710c.

[284] Gao S, Fang L, Phan LM, Qdaisat A, Yeung SCJ, Lee MH. COP9 signalosome subunit 6 (CSN6) regulates E6AP/UBE3A in cervical cancer. Oncotarget. 2015;6(29):28026–28041. doi:10.18632/oncotarget.4731.

[285] Mortensen F, Schneider D, Barbic T, Sladewska-Marquardt A, Kühnle S, Marx A, et al. Role of ubiquitin and the HPV E6 oncoprotein in E6APmediated ubiquitination. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(32):9872–9877. doi:10.1073/pnas.1505923112.

[286] Margolis SS, Salogiannis J, Lipton DM, Mandel-Brehm C, Wills ZP, Mardinly AR, et al. EphB-mediated degradation of the RhoA GEF Ephexin5 relieves a developmental brake on excitatory synapse formation. Cell. 2010;143(3):442–55. doi:10.1016/j.cell.2010.09.038.

[287] Galligan JT, Martinez-Noël G, Arndt V, Hayes S, Chittenden TW, Harper JW, et al. Proteomic analysis and identification of cellular interactors of the giant ubiquitin ligase HERC2. Journal of Proteome Research. 2015;14(2):953–966. doi:10.1021/pr501005v.

[288] Kuhnle S, Kogel U, Glockzin S, Marquardt A, Ciechanover A, Matentzoglu K, et al. Physical and functional interaction of the HECT ubiquitin-protein ligases E6AP and HERC2. Journal of Biological Chemistry. 2011;286(22):19410–6. doi:10.1074/jbc.M110.205211.

[289] Martinez-Noel G, Galligan JT, Sowa ME, Arndt V, Overton TM, Harper JW, et al. Identification and proteomic analysis of distinct UBE3A/E6AP protein complexes. Molecular and Cellular Biology. 2012;32(15):3095–106. doi:10.1128/MCB.00201-12.

[290] Lee D, Kwon JH, Kim EH, Kim ES, Choi KY. HMGB2 stabilizes p53 by interfering with E6/E6AP-mediated p53 degradation in human papillomavirus-positive HeLa cells. Cancer Letters. 2010;292(1):125–32. doi:10.1016/j.canlet.2009.11.015.

[291] Liu X, Yuan H, Fu B, Disbrow GL, Apolinario T, Tomaic V, et al. The E6AP ubiquitin ligase is required for transactivation of the hTERT promoter by the human papil-

lomavirus E6 oncoprotein. Journal of Biological Chemistry. 2005;280(11):10807–16. doi:10.1074/jbc.M410343200.

[292] Niebler M, Qian X, Höfler D, Kogosov V, Kaewprag J, Kaufmann AM, et al. Post-Translational Control of IL-1 via the Human Papillomavirus Type 16 E6 Oncoprotein: A Novel Mechanism of Innate Immune Escape Mediated by the E3-Ubiquitin Ligase E6-AP and p53. PLoS Pathogens. 2013;9(8):e1003536. doi:10.1371/journal.ppat.1003536.

[293] Kuhne C, Banks L. E3-ubiquitin ligase/E6-AP links multicopy maintenance protein 7 to the ubiquitination pathway by a novel motif, the L2G box. J Biol Chem. 1998;273(51):34302–9.

[294] Jung HM, Phillips BL, Chan EKL. MiR-375 activates p21 and suppresses telomerase activity by coordinately regulating HPV E6/E7, E6AP, CIP2A, and 14-3-3. Molecular Cancer. 2014;13(1):80. doi:10.1186/1476-4598-13-80.

[295] Mishra A, Godavarthi SK, Jana NR. UBE3A/E6-AP regulates cell proliferation by promoting proteasomal degradation of p27. Neurobiology of Disease. 2009;36(1):26–34. doi:10.1016/j.nbd.2009.06.010.

[296] Mishra A, Jana NR. Regulation of turnover of tumor suppressor p53 and cell growth by E6-AP, a ubiquitin protein ligase mutated in Angelman mental retardation syndrome. Cell Mol Life Sci. 2008;65(4):656–66. doi:10.1007/s00018-007-7476-1.

[297] Jeong KW, Kim HZ, Kim S, Kim YS, Choe J. Human papillomavirus type 16 E6 protein interacts with cystic fibrosis transmembrane regulator-associated ligand and promotes E6-associated protein-mediated ubiquitination and proteasomal degradation. Oncogene. 2007;26(4):487–99. doi:10.1038/sj.onc.1209837.

[298] Birch SE, Kench JG, Takano E, Chan P, Chan AL, Chiam K, et al. Expression of E6AP and PML predicts for prostate cancer progression and cancer-specific death. Annals of Oncology. 2014;25(12):2392–7. doi:10.1093/annonc/mdu454.

[299] Louria-Hayon I, Alsheich-Bartok O, Levav-Cohen Y, Silberman I, Berger M, Grossman T, et al. E6AP promotes the degradation of the PML tumor suppressor. Cell Death Differ. 2009;16(8):1156–66. doi:10.1038/cdd.2009.31.

[300] Mishra A, Dikshit P, Purkayastha S, Sharma J, Nukina N, Jana NR. E6-AP promotes misfolded polyglutamine proteins for proteasomal degradation and suppresses polyglutamine protein aggregation and toxicity. Journal of Biological Chemistry. 2008;283(12):7648–56. doi:10.1074/jbc.M706620200.

[301] Nasu J, Murakami K, Miyagawa S, Yamashita R, Ichimura T, Wakita T, et al. E6AP ubiquitin ligase mediates ubiquitin-dependent degradation of peroxiredoxin 1. Journal of Cellular Biochemistry. 2010;111(3):676–85. doi:10.1002/jcb.22752.

[302] Zaaroor-Regev D, de Bie P, Scheffner M, Noy T, Shemer R, Heled M, et al. Regulation of the polycomb protein Ring1B by self-ubiquitination or by E6-AP may have implications to the pathogenesis of Angelman syndrome. Proceedings of the National Academy of Sciences. 2010;107(15):6788–93. doi:10.1073/pnas.1003108107.

[303] Lee SY, Ramirez J, Franco M, Lectez B, Gonzalez M, Barrio R, et al. Ube3a, the E3 ubiquitin ligase causing Angelman syndrome and linked to autism, regulates protein homeostasis through the proteasomal shuttle Rpn10. Cellular and Molecular Life Sciences. 2014;71(14):2747–2758. doi:10.1007/s00018-013-1526-7.

[304] Greer PL, Hanayama R, Bloodgood BL, Mardinly AR, Lipton DM, Flavell SW, et al. The Angelman Syndrome protein Ube3A regulates synapse development by ubiquitinating arc. Cell. 2010;140(5):704–16. doi:10.1016/j.cell.2010.01.026.

[305] Nakagawa S, Huibregtse JM. Human scribble (Vartul) is targeted for ubiquitin-mediated degradation by the high-risk papillomavirus E6 proteins and the E6AP ubiquitin-protein ligase. Mol Cell Biol. 2000;20(21):8244–53.

[306] Mishra A, Maheshwari M, Chhangani D, Fujimori-Tonou N, Endo F, Joshi AP, et al. E6-AP association promotes SOD1 aggresomes degradation and suppresses toxicity. Neurobiology of Aging. 2013;34(4):1310 e11–23. doi:10.1016/j.neurobiolaging.2012.08.016.

[307] Wu W, Sun Z, Wu J, Peng X, Gan H, Zhang C, et al. Trihydrophobin 1 phosphorylation by c-Src regulates MAPK/ERK signaling and cell migration. PLoS One. 2012;7(1):e29920. doi:10.1371/journal.pone.0029920.

[308] Yang Y, Liu W, Zou W, Wang H, Zong H, Jiang J, et al. Ubiquitin-dependent proteolysis of trihydrophobin 1 (TH1) by the human papilloma virus E6-associated protein (E6-AP). Journal of Cellular Biochemistry. 2007;101(1):167–80. doi:10.1002/jcb.21164.

[309] Zheng L, Ding H, Lu Z, Li Y, Pan Y, Ning T, et al. E3 ubiquitin ligase E6AP-mediated TSC2 turnover in the presence and absence of HPV16 E6. Genes Cells. 2008;13(3):285–94. doi:10.1111/j.1365-2443.2008.01162.x.

[310] Nuber U, Schwarz SE, Scheffner M. The ubiquitin-protein ligase E6-associated protein (E6-AP) serves as its own substrate. Eur J Biochem. 1998;254(3):643–9.

[311] Cummings CJ, Reinstein E, Sun Y, Antalffy B, Jiang Y, Ciechanover A, et al. Mutation of the E6-AP ubiquitin ligase reduces nuclear inclusion frequency while accelerating polyglutamine-induced pathology in SCA1 mice. Neuron. 1999;24(4):879–92.

[312] Wong EY, Tse JY, Yao KM, Tam PC, Yeung WS. VCY2 protein interacts with the HECT domain of ubiquitin-protein ligase E3A. Biochem Biophys Res Commun. 2002;296(5):1104–11.

[313] Mani A, Oh AS, Bowden ET, Lahusen T, Lorick KL, Weissman AM, et al. E6AP mediates regulated proteasomal degradation of the nuclear receptor coactivator amplified in breast cancer 1 in immortalized cells. Cancer Research. 2006;66(17):8680–6. doi:10.1158/0008-5472.CAN-06-0557.

[314] Chakraborty M, Paul BK, Nayak T, Das A, Jana NR, Bhutani S. The E3 ligase ube3a is required for learning in Drosophila melanogaster. Biochemical and Biophysical Research Communications. 2015;462(1):71–77. doi:10.1016/j.bbrc.2015.04.110.

[315] Picard N, Charbonneau C, Sanchez M, Licznar A, Busson M, Lazennec G, et al. Phosphorylation of activation function-1 regulates proteasome-dependent nuclear mobility and E6-associated protein ubiquitin ligase recruitment to the estrogen receptor beta. Molecular Endocrinology. 2008;22(2):317–30. doi:10.1210/me.2007-0281.

[316] Godavarthi SK, Dey P, Maheshwari M, Jana NR. Defective glucocorticoid hormone receptor signaling leads to increased stress and anxiety in a mouse model of Angelman syndrome. Human Molecular Genetics. 2012;21(8):1824–34. doi:10.1093/hmg/ddr614.

[317] Jung SY, Malovannaya A, Wei J, O'Malley BW, Qin J. Proteomic analysis of steady-state nuclear hormone receptor coactivator complexes. Molecular Endocrinology. 2005;19(10):2451–65. doi:10.1210/me.2004-0476.

[318] Low D, Chen KS. UBE3A regulates MC1R expression: a link to hypopigmentation in Angelman syndrome. Pigment Cell & Melanoma Research. 2011;24(5):944–52. doi:10.1111/j.1755-148X.2011.00884.x.

[319] Gopinathan L, Hannon DB, Smith Iii RW, Peters JM, Vanden Heuvel JP. Regulation of peroxisome proliferator-activated receptors by e6-associated protein. PPAR Research. 2008;2008:1–8. doi:10.1155/2008/746935.

[320] Ramamoorthy S, Dhananjayan SC, Demayo FJ, Nawaz Z. Isoform-specific degradation of PR-B by E6-AP is critical for normal mammary gland development. Molecular Endocrinology. 2010;24(11):2099–113. doi:10.1210/me.2010-0116.

[321] Srinivasan S, Nawaz Z. E3 ubiquitin protein ligase, E6-associated protein (E6-AP) regulates PI3K-Akt signaling and prostate cell growth. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2011;1809(2):119–27. doi:10.1016/j.bbagrm.2010.08.011.

[322] Mulherkar SA, Sharma J, Jana NR. The ubiquitin ligase E6-AP promotes degradation of alpha-synuclein. Journal of Neurochemistry. 2009;110(6):1955–64. doi:10.1111/j.1471-4159.2009.06293.x.

[323] Kuhnle S, Mothes B, Matentzoglu K, Scheffner M. Role of the ubiquitin ligase E6AP/UBE3A in controlling levels of the synaptic protein Arc. Proceedings of the National Academy of Sciences. 2013;110(22):8888–93. doi:10.1073/pnas.1302792110.

[324] Singhmar P, Kumar A. Angelman syndrome protein UBE3A interacts with primary microcephaly protein ASPM, localizes to centrosomes and regulates chromosome segregation. PLoS One. 2011;6(5):e20397. doi:10.1371/journal.pone.0020397.

[325] Mullegama SV, Alaimo JT, Chen L, Elsea SH. Phenotypic and Molecular Convergence of 2q23.1 Deletion Syndrome with Other Neurodevelopmental Syndromes Associated with Autism Spectrum Disorder. International Journal of Molecular Sciences. 2015;16(4):7627. doi:10.3390/ijms16047627.

[326] Reiter LT, Seagroves TN, Bowers M, Bier E. Expression of the Rho-GEF Pbl/ECT2 is regulated by the UBE3A E3 ubiquitin ligase. Human Molecular Genetics. 2006;15(18):2825–35. doi:10.1093/hmg/ddl225.

[327] Sominsky S, Kuslansky Y, Shapiro B, Jackman A, Haupt Y, Rosin-Arbesfeld R, et al. HPV16 E6 and E6AP differentially cooperate to stimulate or augment Wnt signaling. Virology. 2014;468:510–523. doi:10.1016/j.virol.2014.09.007.

[328] Fu J, Frazee AC, Collado-Torres L, Jaffe AE, Leek JT. ballgown: Flexible, isoform-level differential expression analysis; 2016.

[329] Lunn JS, Sakowski SA, Hur J, Feldman EL. Stem cell technology for neurodegenerative diseases. Annals of Neurology. 2011;70(3):353–361. doi:10.1002/ana.22487.

[330] Hao J, Li TG, Qi X, Zhao DF, Zhao GQ. WNT/beta-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. Developmental Biology. 2006;290(1):81–91. doi:10.1016/j.ydbio.2005.11.011.

[331] Ogawa K, Nishinakamura R, Iwamatsu Y, Shimosato D, Niwa H. Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells. Biochemical and Biophysical Research Communications. 2006;343(1):159–166. doi:10.1016/j.bbrc.2006.02.127.

[332] Clevers H. Wnt/beta-Catenin Signaling in Development and Disease. Cell. 2006;127(3):469–480. doi:10.1016/j.cell.2006.10.018.

[333] MacDonald BT, Tamai K, He X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. Developmental Cell. 2009;17(1):9–26. doi:10.1016/j.devcel.2009.06.016.

[334] Wray J, Kalkan T, Gomez-Lopez S, Eckardt D, Cook A, Kemler R, et al. Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. Nature Cell Biology. 2011;13(7):838–845. doi:10.1038/ncb2267.

[335] Smith AG, Heath JK, Donaldson DD, Wong GG, Moreau J, Stahl M, et al. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature. 1988;336(6200):688–690. doi:10.1038/336688a0.

[336] Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP, et al. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature. 1988;336(6200):684–687. doi:10.1038/336684a0.

[337] Doetschman TC, Eistetter H, Katz M, Schmidt W, Kemler R. The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. Journal of embryology and experimental morphology. 1985;87:27–45.

[338] Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings of the National Academy of Sciences of the United States of America. 1981;78:7634–7638.

[339] Manthorpe M, Engvall E, Ruoslahti E, Longo FM, Davis GE, Varon S. Laminin promotes neuritic regeneration from cultured peripheral and central neurons. The Journal of Cell Biology. 1983;97(6):1882–1890. doi:10.1083/jcb.97.6.1882.

[340] Clark P, Britland S, Connolly P. Growth cone guidance and neuron morphology on micropatterned laminin surfaces. Journal of Cell Science. 1993;105(1):203–212.

[341] Namba T, Funahashi Y, Nakamuta S, Xu C, Takano T, Kaibuchi K. Extracellular and intracellular signaling for neuronal polarity. Physiological Reviews.

2015;95(3):995–1024. doi:10.1152/physrev.00025.2014.

APPENDIX A

KNOWN UBE3A PROTEIN INTERACTIONS

Table A.1: Ubiquitin Functions

Protein	**Description**	**Localization**	**Cell Type**	**Ref**
Annexin A1	Inhibit proliferation; anti-inflammatory	Cytosol; nuclear	HEK293T; C33A	[278]
ATPα	ATP hydrolysis (Na/K ions)	Cytosol	Flies	[279]
Bak	Pro-apoptotic	Nuclear, ER, mitochondria	Human cell lines	[280]
Blk	Tyrosine kinase	Golgi; cytosol	Human T cells; yeast	[281]
BMAL1	Circadian clock TF	Cytosol	Mice; HEK293T; NIH3T3; flies	[282]
c-Abl	Non-receptor tyrosine kinase	Cytosol	HeLa; HEK293T; E6AP null MEF	[283]
CSN6	Tumorigenesis	Cytosol; nuclear	Human cell lines	[284]

Table A.1: *Continued...*

Protein	**Description**	**Localization**	**Cell Type**	**Ref**
E6 (HPV)	Oncoprotein	Cytosol	H1299-shE6AP	[285]
Ephexin5	Excitatory synapse development	Cytosol	Mice	[286]
HERC2	E3 ubiquitin ligase	Cytosol	HEK293T	[287, 288]
HHR23A/B	Ubiquitin-binding DNA repair	Cytosol; nuclear	Human cells; HEK293T	[29, 288]
HIR1AN	Asparagine hydroxylase	Cytosol	HEK293T	[289]
HMGB2	Non-histone nuclear protein	Nuclear	HeLa; MCF7; H1299; HCT116	[290]
hTERT	Telomerase enzyme	Nuclear	Primary HFK; HeLa; NIH3T3; E6AP null MEF	[291]
IL-1β	Immune	Cytosol	Human cell lines	[285, 292]

Table A.1: *Continued...*

Protein	**Description**	**Localization**	**Cell Type**	**Ref**
MCM7	DNA replication	Chromosome	HeLa; yeast	[293]
miR-375	Micro RNA	Nuclear	Human cell lines	[294]
NEURL4	Regulation of centrosome	Cytosol	HEK293T	[289]
p27	Cyclin-dependent kinase inhibitor	Chromosome	Mice	[295]
p53	Cell-cycle checkpoint	Chromosome	Mice; human cell lines	[296]
PIST	Golgi/post-Golgi trafficking	Golgi	HEK293T	[297]
PML	Tumor suppressor	Cytosol; nuclear	HEK293T; Human study	[298, 299]
Polyglutamine proteins	Pathological poly-Q expansions	Cytosol; nuclear	Mouse cell lines	[300]
Prx1	Antioxidant peroxidase	Cytosol	HEK293T	[301]
RING1b/ PRC1	Ubiquitin ligase; gene expression	Chromosome	HeLa	[302]

Table A.1: *Continued...*

Protein	**Description**	**Localization**	**Cell Type**	**Ref**
Rpn10	Proteasome-shuttling factor	Cytosol	BG2 neuronal cells; flies	[303]
Sacsin	Synaptic development	Cytosol	HEK293T	[304]
Scribble	Tumor suppression; cell-cycle checkpoint	Chromosome	Flies; HEK293T	[305]
SOD1	Antioxidant enxyme	Cytosol	Cos-1; Neuro2a	[306]
TH1	NELF complex; inhibits MEK/ERK signaling	Cytosol	HeLa; HepG2; yeast	[307, 308]
Tuberin	mTOR pathway	Cytosol; nuclear	HEK293T	[309]
Ube3a	Self-ubiquitination	Cytosol; nuclear	Plasmids	[310]
Ubiquilin	Ubiquitin-binding trafficking	Cytosol	Rats; mice; HeLa	[311]
VCY2	Testis specific	Nuclear	Human testicular tissue; yeart	[312]

Table A.2: Co-activator Functions

Protein	Description	Localization	Cell Type	Ref
AIB1	Steroid receptor co-activator	Cytosol; nuclear	Cancer cell lines; flies	[313]
Androgen	Hormone response; gene transcription	Cytosol; nuclear	Mice; HeLa; PC3	[178]
Derailed	Receptor tyrosine kinase (WNT signaling)	Cytosol	Flies	[314]
Estrogen receptor	Hormone response; gene transcription	Cytosol; nuclear	HEK293T	[177, 315]
Glucocorticoid receptor	Hormone response	Cytosol; nuclear	Mice	[316]
Golgin-160	Golgi membrane associated	Cytosol; golgi	HeLa	[317]
Highwire	Putative E3 Ub-ligase	Cytosol	HeLa	[317]
MC1R	Skin pigmentation	Cytosol; chromosome	Mice	[318]
PPAR	Lipid and glucose metabolism	Cytosol; nuclear	Mice; FaO	[319]
PR-B	Hormone response	Cytosol; nuclear	T47D; mice	[320]

Table A.2: *Continued...*

Protein	**Description**	**Localization**	**Cell Type**	**Ref**
Progestrone	Hormone response	Cytosol; nuclear	HeLa; MCF7; yeast	[177]
RhoA-PI3K-AKT	Growth survival	Cytosol; nuclear	Mice; HeLa; PC3	[178, 321]

Table A.3: Indirect Regulation

Protein	**Description**	**Localization**	**Cell Type**	**Ref**
Actin	Cytoskeleton	Cytosol	Flies	[279]
α-synculein	Unfolded protein; Lewy cells	Prenuclear	Neuro2a; Cos-7	[322]
Arc	Synaptic protein	Cytosol; nuclear	HEK293T	[323]
ASPM	Microcephaly-associated protein	Centrosome	HEK293T; human tissue; yeast	[324]
MAPK6 (ERK3)	Extracellular signal-regulated kinase 3	Cytosol	HEK293T	[289]
MBD5	Methy-CpG-binding 5	Nuclear	Patient cell lines	[325]
Pbl/ECT2	Neural development	Cytosol	Flies; mice	[326]
Wnt/β-catenin	Stem cell pluripotency; cell-fate decisions	Cytosol; nuclear	HEK293T	[327]

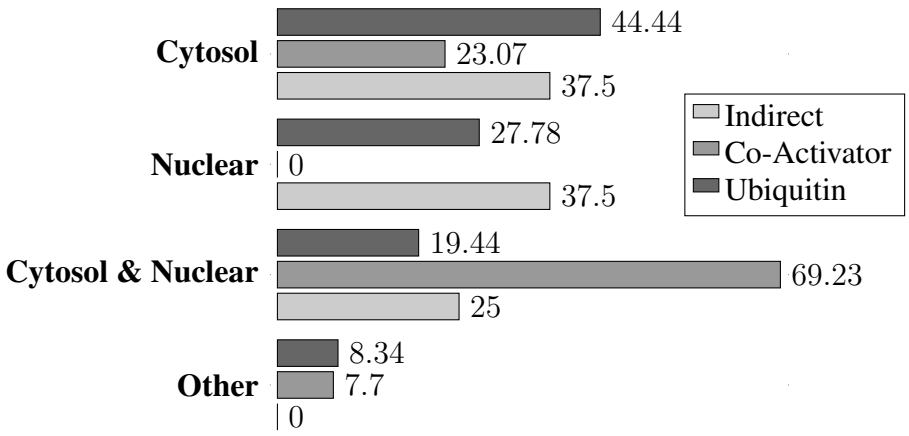


Figure A.1: Localization of UBE3A protein interactions (%)

APPENDIX B

SUPPLEMENTAL DATA - CHAPTER TWO

B.1 Methods extended

Table B.1: Mouse tissue information by study

Study	Strain	Tissue	Accession	Read count (million)
ERP000591	C57BLxDBA	heart	ERR032227	31.0
			ERR032228	8.6
			ERR032229	30.1
			ERR032238	29.5
			ERR032230	27.8
			ERR032231	31.2
		hippocampus	ERR032232	18.8
			ERR032233	33.6
			ERR032234	23.8
			ERR032235	23.2
			ERR032236	22.6
			ERR032237	32.7
		liver	ERR032203	28.6
			ERR032204	30.0
			ERR032205	29.3
			ERR032206	30.8

Table B.1: *Continued...*

Study	Strain	Tissue	Accession	Read count (million)
			ERR032207	31.3
			ERR032208	31.4
		lung	ERR032221	11.7
			ERR032222	25.8
			ERR032223	23.3
			ERR032224	12.1
			ERR032225	28.4
			ERR032226	18.3
SRP012040	C57BL/6J	cerebellum	SRR567488	151
			SRR567489	145
		cortex	SRR567480	156
			SRR032481	166
		frontal lobe	SRR567478	186
			SRR567479	159
SRP033200	Aldh1l1-EGFP	astrocytes	SRR1033783	29.6
			SRR1033784	32.0
	NA	neurons	SRR1033785	37.9
			SRR1033786	33.9
		oligodendrocyte precursor cells	SRR1033787	32.2
			SRR1033788	32.5

Table B.1: *Continued...*

Study	Strain	Tissue	Accession	Read count (million)
		newly formed oligodendrocytes	SRR1033789	32.1
			SRR1033790	30.5
		myelinating oligodendrocytes	SRR1033791	33.4
			SRR1033792	29.7
		microglia	SRR1033793	29.2
			SRR1033794	30.0
	Tie2-EGFP	endothelial cells	SRR1033795	36.5
			SRR1033796	33.8
SRP048593	C57BL/6J	E18-hippo	SRR1772425	41.3
			SRR1772429	34.9
		P1-hippo	SRR1772426	38.5
			SRR1772430	34.6
		P10-hippo	SRR1772427	34.1
			SRR1772431	41.1
		P30-hippo	SRR1772428	43.4
			SRR1772432	41.8

Table B.2: Human tissue information by study

Study	Tissue	Accession	Read count (million)
ERP003613	cortex	ERR315432	17.6
		ERR315455	24.3
		ERR315432	28.5
	heart	ERR315384	16.7
		ERR315413	16.6
		ERR315356	17.7
	liver	ERR315327	6.1
		ERR315451	15.1
		ERR315463	15.4
	lung	ERR315341	29.3
		ERR315346	25.4
		ERR315424	8.8
SRP072463	Brodmann area 4	SRR3306830	171
		SRR3306831	152
		SRR3306832	162
		SRR3306833	163

Table B.3: *UBE3A-AS* Primer List: ’3 RACE

Primer Name	**Sequence**
Isoform 1, pa1 Fwd	GACAGGATGGATAGATGGACAAG
Isoform 1, pa1 nested Fwd	TGTAGCCAGTAGACCTATACTTTAGA
Isoform 1, pa2 Fwd	CGTAAGACAACTGGCCTTTGA
Isoform 1, pa2 nested Fwd	AACGCTGCTGTGGAATCTATAA
Isoform 2 Fwd	CAGCTGCCAGAAAGTGAAGA
Isoform 2 nested Fwd	CAGTGAACGCCAAACAAGTAAG
Isoform 3 Fwd	CCTTGGGAGAGTAGTTCTGTTG
Isoform 3 nested Fwd	GGCTCAACCTCAAGCAGTAATA

Table B.4: *Ube3a-AS* Primer List

Primer Name	**Sequence**	**Reference**
ActB Fwd	GGCTGTATTCCCCTCCATCG	[35]
ActB Rev	CCAGTTGGTAACAATGCCATGT	[35]
Isoform 1 Fwd	GGCTCTACGAGAAGCTGACTG	
Isoform 1 nxt[1]Fwd	GATGCAGAGAATTACAGCCAAC	
Isoform 1 Rev	TCTGTGTCAGTCAGCTTCTCG	
Isoform 1 qPCR Rev	GTTGCCATCACCTTCAGTTC	
Isoform 2, pa1 Fwd	CTGGAACCCACTCTGTAGAC	
Isoform 2, pa1 nxt Fwd	ACCAGGCTGGCTTTGAAT	
Isoform 2, pa1 Rev	TTACACACACACACGCCTATTA	

[0]nxt: nested

Table B.4: *Continued...*

Primer Name	Sequence	Reference
Isoform 2, pa2 Fwd	GGCAAGTGGGCTCATATTCT	
Isoform 2, pa2 nxt Fwd	CAGTCTCCATACATCCTCCTCTA	
Isoform 2, pa2 Rev	TCTTCCTGAGGTCCTGAGTT	
Isoform 2, pa3 Fwd	TGCAACACTTCCCTATTCATTTC	
Isoform 2, pa3 nxt Fwd	CAGGCCAAGGGTCTCTAAATC	
Isoform 2, pa3 Rev	GATGGTTCTTTGAGAAGATAAACCAG	
Isoform 3 Fwd	TGAGGCCAGATTGTTTGGGT	
Isoform 3 nested Fwd	CTCCCAGAAATAGAATTGGGATCA	
Isoform 3 Rev	GCCCTGCATTTGGCATTCAA	
Isoform 3 qPCR Fwd	GCTACATGCTAGGCCCTAATG	
Isoform 3 qPCR Rev	ATGGAGTTCTCTTGACCAAGTC	
Exon 5.1 Fwd	TGTTGAGGTTGGAGGATTGTC	
Exon 5.2 Fwd	GACCTACACTCACTGGGCAC	
Exon 6 Fwd	TGCTTCTTTTTGTTGCTTTCAA	
Exon 7 Fwd	ACAAGCAACGTTGGGAGAAC	
Exon 10A Fwd	CCTGCCTACCTAGCATCAGG	
Exon 10B Fwd	CAGCTCGCATACCCAATTTT	
Exon 12A Fwd	TTAACGCCAAGTTCGGTTTC	
Exon 12B Fwd	CTGAAGTGTTAATTCGCTGGA	

B.2 3' RACE - sequences

B.2.1 Mouse

```
>F02_ATS1-1
TGAGCCGCTCGTATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTCATTTAAAATTTATTTT
```

TTCCTTTTAAAAAATTGATTATTTACATACCAAATGTTGACCTCCCTCCATTCGCAGTGTTCTTCATGCCATCCCCTTCCCTTCACCTCACCTCTGAAAGAGTGCCTTCCATCCCAACCC
CTGGGCATCGCCCTTCTCTGGGGCATCAAGTCTATACAGGATTAGGCACATCCTCTCCCACTAGGGCCAGATTAGGCTGTCAGTCCTTAGCTATGTATGTGCTGGGAGCCTTAGACCAGC
CAGTGCATGCTCTTTGTTGGTGGCTTAGTCCCTGGGGTCCAGGTTAGTTGGCCCTGTTGGTCTTCCTGTGGGATTGCAATCCCCTTCAGTTCCTTCAGCCTTTCCCCTAACTCTTCCATA
GGGATCCCTAACCCTAACCCAATAGTTGGCTGTAATTCTCTGCATCAAGGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTT
ACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTT
GCGCAGCCTATACGTACGGCAGTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCGGGGCGACGGATGGTGATCCC
CCTGGCCAGTGCACGTCTGCTGTCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATATGGCCAGTGTGCCGGTCTCCGT
TATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGACATCAAAAAACGCCATTAACCTGATGTTTCTGGGGAATAATAAATGTCAGCATGAGATTATCAAAGGATTCTCACCTT
AGATCTTTTCACGTAGAAGGCCAGTCCGGCAGAACGGTGGCCTGACCGGATGATGTCAGCTACTGGACTATCCTGGCAACAAAGA

>G02_ATS1-2
TGCGCCGCTCGCATTAACCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTGGTTTTTTAATAATTTAT
TTTAAATTTTATTCTTCCCTTTTGTTATACACTGATAAGCTGTTAAACTCATTTAAATATTTTTTAAAAAATTATTTTATCTATTTGCATTCCAAATCTTGTTCCTCCTCCTGGTCTTCC
CCCCAGAGTTCTTCACCCTATTTCCCTTCCCCTTTGCCTCTGAGAGGGAGTTCTCCCACCCACCCCAAATGACCCACCCACCCATCCCCACCTCACCTCACTTTACTGCCCCATCCCTGG
GTTATCCCCCTTCCCTTGGGCATCAAGTCTCCATAGGATTAGGTTCATCCTCTCCCATTGAGGCTAGACAAAGCAGTTCTCTGCTACATATATGCCAGGGGCCATGGGCCAGTCCATGTA
TGTTCTTTGGTTGCTCCCATGGGTCCAGGTTAGTTGGCACTGTTGGTTTTCCTATGGTAGTTGCCATCACCTTCAGTTCTTTCAATCCTTCATCCAACTCTTCCATAGGGGTGTCCAACC
TCAGTCCAATGGTTGGCTGTAATTCTCTGCATCAAGGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGAC
TGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTATAC
GTACGGCAGTTTAAAGTTTACACCCTATAAAAGAGAGAGCCGTTATCGTCTGTTTTGTGGATGTACAGAGTGATATTTATTGACACGCCGGGGCGACGGATGGTGATCCCCCTGGCCAGT
GCACGTCTGCTGTCAGATAAAGTCTCCCGTGAACTTTACCGGTGGTTGCATATCGGGGAATGAAGCTTGCGCATGATGACCAACGATATGCCAGTGTGCCAGTCTCGTTATCGGGAGGAG
TGCCTGATCTCAGCACGCGAATGGACTCAATGCATTTCACCTGATGTTCTGAGAATATTTAG

>H02_ATS1-3
TGCTTTCACTCACCGCAGCCACTGTATTAACCCTCACTAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTTGGT
TTTTTAATAATTTATTTTAAATTTTATTCTTCCCTTTTGTTATACACTGATAAGCTGTTAAACTCATTTAAATATTTTTTAAAAAATTATCTTATCTATTTGCATTCCAAATCTTGTTCC
TCCTCCTGGTCTTCCCCCCAGAGTTCTTCACCCTATTTCCCTTCCCCTTTGCCTCTGAGAGGGAGTTCTCCCACCCACCCCAAATGACCCACCCACCCATCCCCACCTCACCTCACTTTA
CTGCCCCATCCCTGGGTTATCCCCCTTCCCTTGGGCATCAAGTCTCCATAGGATTAGGTTCATCCTCTCCCATTGAGGCTAGACAAAGCAGTTCTCTGCTACATATATGCCAGGGGCCAT
GGGCCAGTCCATGTATGTTCTTTGGTTGCTCCCATGGGTCCAGGTTAGTTGGCACTGTTGGTTTTCCTATGGTAGTTGCCATCACCTTCAGTTCTTTCAATCCTTCATCCAACTCTTCCA
TAGGGGTGTCCAACCTCAGTCCAATGGTTGGCTGTAATTCTCTGCATCAAGGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTT
TTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAG
TTGCGCAGCCTATACGTACGGCAGTTTAAGGTTTACACCCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGGATGTACAGAGTGATATTATTGACACGCCGGGGCGACGGATGGTGA
TCCCCTGGCCAGTGCACGTCTGCTGTCAGAATAAAGTCTCCCGTGGACCTTACCCGGTGGGTGGCAATATCGGGGGATGGAAAGGCTGGGCGGCATGGATGGACCACCCGATATGGGCAA
GGTGGTGCGGTCTACGTATCGGGGAAGAAGTTGGCCTGAATCTCAGGCCCCCGGAATGACTCAAAACGGCATTACCTGAATGGTTTCTTGA

>A03_ATS2-1
TGAGCCGCTCGATTAACCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTGTGTTTGTCAAATCCCATT
TACATAAACTCACTTTACTGCCCCATCCCTGGGTTATCCCCCTTCCCTTGGGCATCAAGTCTCCATAGGATTAGGTTCATCCTCTCCCACTGAGGCTAGACAAAGCAGTTCTCTGCTACA
TATATGCCAGGGGCCATGGGCCAGTCCATGTATGCTCTTTGGTTGCTCCCATGGGTCCAGGTTAGTTGGCACTGTTGGTTTTCCTATGGTAGTTGCCATCACCTTCAGTTCTTTCAATCC
TTCATCCAACTCTTCCATAGGGGTGTCCAACCTCAGTCCAATGGTTGGCTGTAATTCTCTGCATCAAGGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGTATTACA
ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCG
ATCGCCCTTCCCAACAGTTGCGCAGCCTATACGTACGGCAGTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCGG
GGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGCTGTCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATATGG
CCAGTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCCGCGAAAATGACATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAAATGTCAGGCATGAGATTA
TCAAAAAGGATTCTTTCACCCTAGATCCCTTTCACGTAGAAAGCCCAGTCCGCAAGAAACGGTGCTGACCCCGAATGAATGTCAGCTACTGGGCTATTCTGACAAGGAAACCGCAGCGCA
AGAGAAGGCAGTAGGCTTGGCATTGGCTACATTGCGATAGGCTAGACCTTGCCGCTTTATGACGCAGGCACCAGAGTGCCAGCCTAGTGGGGCGC

>B03_ATS2-2
TGCCGCCGCTAGCATTAACCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCCGCGTCGACTAGTACTTTTTTTTTTTTTTTTTGTGTTTGTCAAATCCCATT
TACATAAACTCACTTTACTGCCCCATCCCTGGGTTATCCCCCTTCCCTTGGGCATCAAGTCTCCATAGGATTAGGTTCATCCTCTCCCATTGAGGCTAGACAAAGCAGTTCTCTGCTACA
TATATGCCAGGGGCCATGGGCCATTCCATGTATGTTCTTTGGTTGCTCCCATGGGTCCAGGTTAGTTGGCACTGTTGGTTTTCCTATGGTAGTTGCCATCACCTTCAGTTCTTTCAATCC
TTCATCCAACTCTTCCATAGGGGTGTCCAACCTCAGTCCAATGGTTGGCTGTAATTCTCTGCATCAAGGGCAAATTCCCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGTATTACA
ATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAAAGGCCCGCACCG

ATCGCCCTTCCCAACAGTTGCGCAGCCTATACGTACGGCAGTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCGG
GGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGCTGTCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATATGG
CCAGTTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGGCTGATTCTCAGCCCACCCGCGAAAATTGACATTCAAAACGCCATTAACCCTGATTGTTTCTGGGAAATATAAATGGTCAGGC
ATGAGATTATTCAAAAAGGATTCTTTCACCTAGATCCTTTTCACGTAAGAAAGCCCAGTTCCGCAAGAAACGGGTGCTTGAACCCGGATGAATTGTTCAGCTTACTGGCTATTCTTGACA
GGCAAATCGCAAGCGCAAAGAGAAGGCAAGGTTAGCTTGCAATGGCTTACATTGCCGATTAGCCTAGACTGCGTTTTATGCACAGGCAAGCGAAACGGAAAATTGGCTCAG

>C03_ATS2-3
TTCAGCCGCTCTATTAACCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTGTGTTTGTCAAATCCCAT
TTACATAAACTCACTTTACTGCCCCATCCCTGGGTTATCCCCCTTCCCTTGGGCATCAAGTCTCCATAGGATTAGGTTCATCCTCTCCCATTGAGGCTAGACAAAGCAGTTCTCTGCTAC
ATATATGCCAGGGGCCATGGGCCAGTCCATGTATGCTCTTTGGTTGCTCCCATGGGTCCAGGTTGGTTGGCACTGTTGGTTTTCCTATGGTAGTTGCCATCACCTTCAGTTCTTTCAATC
CTTCATCCAACTCTTCCATAGGGGTGTCCAACCTCAGTCCAATGGTTGGCTGTAATTCTCTGCATCAAGGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGTATTAC
AATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACC
GATCGCCCTTCCCAACAGTTGCGCAGCCTATACGTACGGCAGTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCG
GGGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGCTGTCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATATG
GCCAGTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAATGACATCAAAAACGCCATTTAACCTGATGTTCTGGGGAATATAAATGTCAGGCATGAGATTA
TCAAAAAAGGATTCTTCACCTAGATCCTTTTCACGTAGAAGCCCAGTCCGCAGAAACGGTGCTGACCCGATGAATGTCAGCTACTTGGCCTATCTGACAGGAAACGCAGCGCAAGAGAAA
GCAAGGTAGCCTGCATGGGACTTACATGCCGATAGCCTAGACTGCGTATGGAGCAGCGAACCGGAAATTG

>C05_ATS8-1
TGCTCGCCGCTAGTATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGATGCAGAGAATTACAGCCAACAATAACTCAAAAATAATAACTCACCAAACAA
AAACTAAACCCAACCTAATATTTTCCACCCTAGCTATCATAAGCACAATAACCCTACCCCTAGCCCCCCAACTAATTACCTAAAAAAAAAAAAAAAAAAGTACTAGTCGACGCGTGGCCAA
GGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCT
TGCAGTACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTATACGTACGGCAGTTTAAGGTTTACACCTATAAAAGAGA
GAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCGGGGCGACGGATGGTGATCCCCCT

>D05_ATS8-2
TTCTACGCCGCTCGCATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTGGGGTTTTAAAAAAATTTA
TTAAAAATTTTATTCTTCCCTTTTGTTATACACTGATAAGGTGTAAAACTCATTTAAATATTTTTAAAAAAATTATTTTATCTATTTGCTTTCCAAATTTTGTTCCTTCTCCTGGTCTTC
CCCCCAGATTTTTTCACCCTTTTTCCCTTCCCCTTTGCCTGTGAGGGGGAGTTCTCCCACCCACCCCAAATGACCCACCCACCCATTCCCACCTCACCTCACTTTACTGCCCCATCCCTG
GGTTATCCCCCTTCCCTTGGGCACCAAGTCTCCAAGGATTAGGTTCATCCTTTCCCATTGAGGCTAGACAAAGCACTTCTTATCTACATATATTCCAGGGGCCATGGGCCAGTCCATGTA
TGTTCTTTGGTTGCTCCCATGGGTCCCGGTTACTTGGCACTGTTGGTTTTCCTATGGTAGTTGCCATCGGGGTCAGTTCTTTCAATCCTTCATCCAGGTTTTCCATAGGGGTGTCCAACC
TGAATCCAATGGTTGGCTGTTATTTTCTGCCTAAAGGGCGAATTCTTGGCCGCTAAATTCAATTCGCCCTATAGTGAGTGGTATGACAATTCACTGGCCGTCCTTTTACAACGTCCTGAC
TGGAAAAAACCCTGGCGTCACCCAACTTAATTGCCTTACAGCACATCCCCCCTTTCCCCCTTCTGGACTAATAGCCAAAGAGGCCCGCACCCGATCGCTCCTTGTCTACAGTTGTACAAG
CCTATACCTACGGCAGTTTTAAGGGTTTACACCTCATAAGAGATAGAGCCGATACTCCCGTGGTTTACGGAATGGTACAG

>C05_Ex14-111F
TTCCTTGTACGACTCTATAGGGCGAATTGGGCCCTCTAGATGCAGGAACGAGCGGCCGCCAGTGTGATGGATATCGGGGAGAAAAAAGCCCTTGGCCAGAGCGACTAGTACTTTTTTTTT
TTTTTTTTTTTTTTTTTTAATTTCAGGATGGAAAAATTTGACATAGATGGAAAAAATTTTTCTTTTGGAGGAAAAAATATCTTAACCAAATGGTGTGGGGACAAGTGGTTTGTTGTTATG
GAGAAAAATGAAACTGACCCATATTATTCAGCTTTCAAAACAACAACAACAACAACAACAATGCCATCTCCAAAACCCAAAAAACAAAACCAAAAAAAAATATTTTAAAAAAAAAAAATC
TTTTTTTAACTCTGAAACACTGAGGGGGCTGGAGAGATGACTCAAAGGTTAAGAACACTTTTTGTTTTTCCTGAGGTCCTGAGTTCAATTCCCAGCAACTGCATGGTGGCTTACAACCAT
TTTTAATGGGATCTGATGCTCTCTTCTGGTGTGTCTGACAAGAGCGACAGTGTACTCATACAAATAAAATAAATAAATCTTAAAACAAAACAAGAAAACCACTGAAATACTGAAACTGTT
AGAAGAAAACATAGTGCCCTCTAACATGTAGATGTTGGAATGCTCTTAGAGGAGGATGTTTGGAGACTGAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGG
TACCAAGCTTGATGCATAGCTTTGAGTATTCTATAGTGTCACCCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGGGTGAAATTGTTATCCCCCTCCCAATTCCCCCACAAC
AATACGAGCCGGAAACCATAAAAGTGTAAAGCCCTGGGGGTGCCCTAATGAGTGAAGCCAAACTCACATTTATTTGGCGTTGCGCTTCCCTGCCCCGCTTTTCCAGTCGGGAAAACCCTG
GTGGTGCCCACTTGCATTTAATGAATCCGCCCCAACCCGCGGGGGGGAGAGGGCGGGTTTGGCTTATTGGGGCGCTCTTTTCCCTTTTCCTTGCGCTTACTCTGACCTCGGTTCGGGGCC
TC

>D05_Ex14-112F
TTTTCCGTATACGCTCTATAGGGCGAATTGTGCCCTCTAGAGGGGAACGAGCGGGGGAGCGTGATGGATATCTGGATGGAAGCCTGTCGGTGGGGAACATCCTCCTCTAAGAGCATTCCA
ACATCTACATGTTAGAGGGCACAGGAACTTCTAACAGTTAAGATTTCAATTGCTCTCCTTGTTTTGTTTTAAGATTTATTTATTTTATTTGTATGAGTACACTGTTTTCTCTTGTCAGAC
ACACCACCACGAGAGTATCTGATCCCATTACATATGGTTGTAAGCCACCATGCAGTTGCTGGGAATTGAACTCAGGACCTCAGGAAGTTTCATTCTTTGTTCTTAACCTTTTGAGTTATC
TCCCCCCCCCCTTTTTTTTTCAGGGTTTTTTGGGGTTTTTTTTTTTGTTTTTTTTTTTTTGGTCTCGCTTTTTTTTTTTTACCCTGTTTTTTTTTTTTTTTTTT

>G05_Ex14-211F

CGAAAACGCTACTATAGGGCGAATTGGGCCCTCTAGATGCAGGACGAGCGGCCGCCAGTGTGATGGATATCTGCAGAAAAGCCCTTACCAGGGAGGCTTTGAATTCAGAGATTCGCTTTC
TGAATGCTGGGATTAAAGGCTGCACACTTAAAATTAAAAAAAATAGGCGTTTGTTGTGTAAACTATTACAGGATAAATAATAAGTGCAACTGTTATATTATTATTTAAGATACTTTAAAT
TTTTTCCCTCCTTTTTTTTTCTCCATCCATTAAAGATCCTCATTTGGAATACCATTGTATTATATGAAAGGGGCCAGTAATAAATGTTTTTCTTTTTAAAGAGTTTTTTTTGCATATGAG
GGTAAACACAAAATTTTTTTCTGAAAATCATTAGACTGCTAAGAAAATTTATTTTTCTTTTTTCTGTTGCAGCATGTGAATTCGCTTTCAGCTATTTTTTTTCCTACATAAAAGTTTGAA
AATGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGTACTAGTCGACGCGTGGCCAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCCAGCTCGGTACCAAG
CTTGATGCATAGCTTGAGTTTTCTATAGTGTCCCCT

>H05_Ex14-212F

CTGAAACCGCTCTATAGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTT
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGCCCTTCATTTTCAGACTTTTATTTAGGGGGGAAAAGCTGAAAGCGGCTTCACATGCTGGAACAGTTTATGTTCTTT
ACTTTTTTTAGCAGGCTAATGATTTTCAATTTTGGGGGTGGGTTTAAAAAGTTTGCAAAAATTACTTTTTAAAGGGGTACATTTTTTACTGGGCCCTTTCATATAAAACAATGGGGGTGG
GAAAGGGGGTTTTTAAAGGGGGGGGAAAAAAAAAAAATAGAAAAAATTTAAAGGTTTTTAAATTTGATTTTACAGTTGCAATTATTATTTATCCCGTGATAGTTTACCAAACCCCACCCC
CAATATTTTAAATTTAAAAATTTCAAGGCTTAATCCCAGCATCAAAAACGGAATCTTGAATCAAAGGAGCTGGTAGGCCATTCCAGCCCTCTGCGGCCTGACAGTTGATCCGCGTCGCT
CCAAGCTGAGCATAGCTGGGTATCGAACTGGCACCTATATAGCTGCGCGACTATGGCAAGCTGTCTCCGTGCGGCATGTGTCATCGCC

>A06_Ex14-221F

AGGGAAGAAATAAAATTTCGTTAAGGATCTATAGGGCGAATGGGCCCTCTGAGCAGGCACGAGCGGCCGCCAGTGTGATGGATATGGGGAAAAAGAGAACTTACCAGGGAGGCTCGAATT
CAGAGATTCGCTTTCTGAATGCTGGGATGAAAGGCAAGAGGGGGGAAAAAATTAAAAACCAATAGGCTTTTTTTGTGCGTAAACTAGGTACAGGATAAATAATAAGTGCAACTGTTATAT
TTTTTGTTTAAGATACTTTAAATTTTTCCCCCCTCTTTTTTTTCTCCATCCATTAAAGATCCTTATTTGGAATACCATTGTATTATATGAAAGGGGCCAGTAATAAATGTTTTTTTTTTT
AAAGAGTTTTTCTTTGTTTTTTTATCCCCCCCCTTTTTTTTTTTCTGAAAATCATTAGACTGCTAAGAAAATTTAAGAATATTTTTTTGGGGGGGCATGTTTTAGTTTTTTTTCAGATAT
TACTCCCCCCCCCCCAAAATTTTTTTTATGAAAAAAAAAAAAAAAAAGTACTAGCCGACGCGTGGCCAAGGGCGAATTCCAGCCCACTGCCGGCCGCTACTAGTGGATCCGAGCTCGGTACC
AAGCTTGATGCATAGCTTGAGTTTCCTATAGCGCCACCTAAATAGC

>D06_Ex14-302F

ATGGACGCTCTATAGGGCGAATTGGGCCCTCTAGAGCATGCACGAGCGGCCCCCAGTGTGATGGAATCTGCAGAAAACGCCCTTCAGGCAAGGGTCTCTAAATCTGGTTTATCTTCTCAA
AGAACCATCTAATAGATTTGCCGACTCTTTTTATTATTTTCACTGTTTATATTTCATTGTTTTCTGCTCTGACTTTTATTATTTCTCATGATTGACTACATTATTTATTTAGTTTGTTTT
GATTGGTTTTTCCAAAAAAAAAAAAAAAAGTACTACTCGACCCGTGGCCAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGTTTTTTTTCTCGGTTCCTTTTTTTTTTTCATAG
CTTGAGTATTTTTTAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCTTTTTTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCTCACTATCTACAAGCCGGAAGCATAAACC

>D03_ATS3-1

TGAGCCGCTCGATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGTCTCTTTCTGTGTCTCAGCTTCCTCATCCTTAGGATGCCAGTATATTTCAGAATT
CAAATGATTTAAGAATTACAAGTGTTTAGACTATTTCAGTAAATTTTGGAAGAGGTAGGAAAAAAATAGGAGGTGAAGATGAGCAAAATGCATTGTATCATATATGAAATCCTCAATTAA
TGGAATAGTATGTATTACCTGCTACATGCTAGGCCCTAATGGTAAGTATCAGATGCTATTGCAATTGCTATAATTTCTATCATTTGCATGAGTATATGACTTGGTCAAGAGAACTCCATT
GCACTAAATCTGAAAAATTAAAATCTAGGCTTAATTCCAAAATTTGAAATTTTAGCATTACTAAAGGGGAGACAGCAGAGAAAATGTTTGAAACTGACCAGCTTTTTTGCACCAAGTTAC
AAGTAGTAGCCCACTAATACTAATAATTATATTAAATTAAATAAACAGAGGTCATAATCAACCATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAGGTACTAGTCGACGCGTGGCCAAGGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC
CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTATACGTACGGCAG
TTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTTATCGTCTGTTTTGTGGATGTTACAGAGTGATTATTATTGACACGCCGGGGGCGACGATGGTGATCCCCCTTGACCAGTGCACGTT
CTTGCTTGTCAGATAAAGTTCTCCCGTGAAACTTTACCCCGGTTGGTGCATTATCGGGATGAAAGCTTGCGCATGATTGACCACCCGATTATGACCAGTTGTGCCGTTCTTCGGTTATTC
GGGAGAGTGGCTTGTATCTCTCAGCCCCCGGCAAATGGACCTTCAAAACGCCATTAACTTGATGTTTCTGGGGAAATTATAATTGTGTCTC

>E03_ATS3-2

TGAGCCGCTCGATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGTCTCTTTCTGTGTCTCAGCTTCCTCATCCTTAGGATGCCAGTATATTTCAGAATT
CAAATGATTTAAGAATTACAAGTGTTTAGACTATTTCAGTAAATTTTGGAAGAGGTAGGAAAAAAATAGGAGGTGAAGATGAGCAAAATGCATTGTATCATATATGAAATCCTCAATTAA
TGGAATAGTATGTATTACCTGCTACATGCTAGGCCCTAATGGTAAGTATCAGATGCTATTGCAATTGCTATAATTTCTATCATTTGCATGAGTATATGACTTGGTCAAGAGAACTCCATT
GCACTAAATCTGAAAAATTAAAATCTAGGCTTAATTCCAAAATTTGAAATTTTAGCATTACTAAAGGGGAGACAGCAGAGAAAATGTTTGAAACTGACCAGCTTTTTTGCACCAAGTTAC
AAGTAGTAGCCCACTAATACTAATAATTATATTAAATTAAATAAACAGAGGTCATAATCAACCATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGTACTAGTCGACGCGT
GGCCAAGGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA
TCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTATACGTACGGCAGTTTAAGGGTTTACACCTATA
AAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCGGGGCCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGCTGTCAGATAAAGTCTCCCGT

GAACCTTTACCCGGTGGTGCATATCCGGGGATGAAAGCTGGCGCATGATGACCACCGATATGCCAGTGTGCCAGTCTCGTTATCGGGCAGAGTGGCCTGATCTCAGCACCGCGTAATTGA
CATTCAAAACGCCATGACTTGAATGTTCTGGGCATTCATGTCAGGCATGAATATCAGAAGATTTCTGTCAAACT

>F03_ATS3-3
TGAGCCGCTCGATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTTATTTTAAGGTGACAAAGC
ATTTTACTTTTGGCTAGTTACATTATGATAGAGAACATACAGGAAATATACAAGTACTTGAAATTTACCATGTTACATTATAAATTATGGACATAGACATTTAAGTTTTATGCAATTTTC
ACACATGAAACAAAATATAATTTTTTAATGATTTAAGACTATAAAATATTCTAAACTTCAACTGTAAAAAAAGTAGATTTGACACATGAGGTATGATTTAGTGACCAGTCCTATGTATGG
ATTAGCATTGGAAGGTTATCAAAAGCTATTATTAGTATGTATGTTTCCAACTTTAGAGGCCTTTTACATGTCCTTAATGAATTTATATGCCCTGCATTTGGCATTCAATGGTATCATTTA
TAATGAGAAAAGAATAGCTTATTATTGTTAATTAATCATATTCAATCATATTAACCAATTTAATTAACTTAATATTAATAAACGCACATTAATCAAATTAGTCAACTTAATATGGTTGAT
TATGACCTCTGTTTATTTAATTTAATATAATTATTAGTATTAGTGGGCTACTACTTGTAACTTGGTGCAAAAAAGCTGGTCAGTTTCAAACATTTTCTCTGCTGTCTCCCCTTTAGTAAT
GCTAAAATTTCAAATTTTGGAATTAAGCCTATATTTTAATTTTTCAGATTTAGTGCAATGGAGTTCTCTTGACCAGGTCATATACTCATGCAAATGATAGAAATTATAGCAAGTTGCAAT
AGCATCTGATACTTACCATTAGGGGCCTAGCATGTAGCAGGTCATACATACTATTCGATTAATCTGACGATTTCATATATGGATACGAATTGCCATTATTGGCTCATCATTGCACCCCTC
CTGACTTTATCTTCCCTACCCTCTTTCCATATATCTCACCTGCAAATCAGTACATAAATACACCGTGAGATAAGTCTAAACCTCCATTGTGGAAATTCCTGAATCATACTGGACCATTCC
TAGGATGCAGGCAAGTCTGGAGAGCCACTGAGAGCTCACGGACCGAAATCGGCGAGGCCCCG

>G03_ATS4-1
TGCAGCCGCTCGCATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTTTAAGGTGACAAAGCA
TTTTACTTTTGGCTAGATACATTATGATAGAGAACATACAGGAAATATACAAGTACTTGAAATTTACCATGTTACATTAGAAATTATGGACATAGACATTTAAGTTTTATGCAATTTTCA
CACATGAAACAAAATATAATTTTTTAATGATTTAAGACTATAAAATATTCTAAACTTCAACTGTAAAAAAAGTAGATTTGACACATGAGGTATGATTTAGTGACCAGTCCTATGTATGGA
TTAGCATTGGAAGGTTATCAAAAGCTATTATTAGTATGCATGTTTCCAACTTTAGAGGCCTTTTACATGTCCTTAATGAATTTATATGCCCTGCATTTGGCATTCAATGGTATCATTTAT
AATGAGAAAAGAATAGCTTATTATTGTTAATTAATCATATTCAATCATATTAACCAATTTAATTAACTTAATATTAATAAACGCACATTAATCAAATTAGTCAACTTAAAATGGTTGATT
ATGACCCTATGAATTATTTAATTTAATATAATTTTTAGTA

>H03_ATS4-2
TGCAGCCGCTCGCATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTTTCATTTTAAGGT
GACAAAGCATTTTACTTTTGGCTAGATACATTATGATAGAGAACATACAGGAAATATACAAGTACTTGAAATTTACCATGTTACATTAGAAATTATGGACATAGACATTTAAGTTTTATG
CAATTTTCACACATGAAACAAAATATAATTTTTTAATGATTTAAGACTATAAAATATTCTAAACTTCAACTGTAAAAAAAGTAGATTTGACACATGAGGTATGATTTAGTGACCAGTCCT
ATATATGGATTAGCATTGGAAGGTTATCAAAAGCTATTATTAGTATGTATGTTTCCAACTTTAGAGGCCTTTTACATGTCCTTAATGAATTTATATGCCCTGCATTTGGCATTCAATGGT
ATCATTTATAATGAGAAAAGAATAGCTTATTATTGTTAATTAATCATATTCAATCATATTAACCAATTTAATTAACTTAATATTAATAAACGCACATTAATCAAATTAGTCAACTTAATA
TGGTTGATTATGACCTCTGTTTATTTAATTTAATATAATTATTAGTATTAGTGGGCTACTACTTGTAACTTGGTGCAAAAAAGCTGGTCAGTTTCAAACATTTTCTCTGCTGTCTCCCCT
TTAGTAATGCTAAAATTTCAAATTTTGGAATTAAGCCTAGATTTTAATTTTTCAGATTTAGTGCAATGGAGTTCTCTTGACCAAGTCATATACTCATGCAAATGATAGAAATTATAGCAA
TTGCAATAGCATCTGATACTTACCATTAGGGCCTAGCATGTAGCAGGTAAATACATACTATTTCCATTAATTGAGGGATTTCATATATGATACAATGCATTTTTGCTCATCTTCACCCTC
CTATTTTTTTTCCTTACCCTCTTCCAAAATTTACTGGAAATAGTCCTAAAACACCTTGTATTTCTAAAATCATTTGGATTTCTGGAAATATACCTGGCATCCTAAGGATGAAGGAAAGCT
GGACACACAGAGGAAGAACAGGGCGAAATTCGCGGGCGCTTAATCATTCGCCTATGAGTCGATTACAATTCACTGGCCGGTCCGATTTCA

>A04_ATS4-3
TGAGCCGCTCGTATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGTCTCTTTCTGTGTCTCAGCTTCCTCATCCTTAGGATGCCAGTATATTTCAGAAT
TCAAATGATTTAAGAATTACAAGTGTTTAGACTATTTCAGTAAATTTTGGAAGAGGTAGGAAAAAAATAGGAGGTGAAGATGAGCAAAATGCATTGTATCATATATGAAATCCTCAATTA
ATGGAATAGTATGTATTACCTGCTACATGCTAGGCCCTAATGGTAAGTATCAGATGCTATTGCAATTGCTATAATTTCTATCATTTGCATGAGTATATGACTTGGTCAAGAGAACTCCAT
TGCACTAAATCTGAAAAATTAAAATCTAGGCTTAATTCCAAAATTTGAAATTTTAGCATTACTAAAGGGGAGACAGCAGAGAAAATGTTTGAAACTGACCAGCTTTTTTGCACCAAGTTA
CAAGTAGTAGCCCACTAATACTAATAATTATATTAAATTAAATAAACAGAGGTCATAATCAACCATATTAAGTTGACTAATTTGATTAATGTGCGTTTATTAATATTAAGTTAATTAAAT
TGGTTAATATGATTGAATATGATTAATTAACAATAATAAGCTATTCTTTTCTCATTATAAATGATACCATTGAATGCCAAATGCAGGGCATATAAATTCATTAAGGACATGTAAAAGGCC
TCTAAAGTTGGAAACATACATACTAATAATAGCTTTTGATAACCTTCCAATGCTAATCCATACATAGGACTGGTCACTAAATCATACCTCATGTGTCAAATCTACTTTTTTTTTACAGTTG
AAGTTTAGAATATTTTATAGTCTTAAATCATTAAAAAATTATATTTTGTTTCATGTGTGAAAATTGCATAAAACTTAAATGTCTATGTCCATAATTTCTAATGTAACATGGTAAATTTCA
AGTACTTGTATATTTCCTGTATGTTCTCTATCATAATGTATCTAGCCAAAAGTAAAATGCTTTGTCACCTTGAATGAACAATAACTATAATAGAGATAATATAAAAGACAATACACTATA
GTGGCTCATTGTCGCCGTTGTGCCAAGGCCAATTTTGCGCGCGCGCGAAATAACACTAATGTCGCGCCCA

>B04_ATS5-1
TGAGCCGCTCGATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTTAAGGTGACAAAGCATTT
TACTTTTGGCTAGATACATTATGATAGAGAACATACAGGAAATATACAAGTACTTGAAATTTACCATGTTACATTAAAAATTATGGACATAGACATTTAAGTTTTATGCAATTTTCACAC
ATGAAACAAAATATAATTTTTTAATGATTTAAGACTATAAAATATTCTAAACTTCAACTGTAAAAAAAGTGGATTTGACACATGAGGTATGATTTAGTGACCAGTCCTATGTATGGATTA
GCATTGGAAGGTTATCAAGAGCTATTATTAGTATGTATGTTTCCAACTTTAGAGGCCTTTTACATGTCCTTAATGAATTTATATGCCCTGCATTTGGCATTCAATGGTATCATTTATAAT

GAGAAAAGAATAGCTTATTATTGTTAATTAATCATATTCAATCATATTAACCAATTTAATTAACTTAATATTAATAAACGCACATTAATCAAATTAGTCAACTTAATATGGTTGATTATG
ACCTCTGTTTATTTAATTTAATATAATTATTAGTATTAGTGGGCTACTACTTGTAACTTGGTGCAAAAAAGCTGGTCAGTTTCAAACATTTTCTCTGCTGTCCCCCCTTTAGTAATGCTA
GAATTTCAAATTTTGGAATTAAGCCTAGATTTTAATTTTTCAGATTTAGTGCAATGGAGTTCTCTTGACCGAGTCATATACTCATGCAAATGATAGAAATTATAGCAATTGCAATAGCAT
CTGATACTTACCATTAGGGCCTAGCATGTAGCAGGTAATACATACTATTCCATTAATTGAGGATTTCATATATGATACAATGCATTTTGCTCATCTTCACCCTCCCTATTTTTTTTCCTA
CCCTCTTTCAAAATTTACTGAAATAGTCTAAAACACTTGGTAAGTCATTAAATCATTCGAATTTCTGACATAATACCTGCATCCCTACGGATGAGGAAGCTGACTACACCAGCAAGAGAC
AAGCGACGAATGCACTGCCGCTTAAATTACAAATTCGCCTCTATCAGTGGAGGTGCGGTATTTAACCAATATTCCAACTTGGAGACACT

>C04_ATS5-2
TGAGCCGCTCGCTTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTTTTTAAGGTGACAAAGCAT
TTTACTTTTGGCTAGATACATTATGATAGAGAACATACAGGAAATATACAAGTACTTGAAATTTACCATGTTACATTAGAAATTATGGACATAGACATTTAAGTTTTATGCAATTTTCAC
ACATGAAACAAAATATAATTTTTTAATGATTTAAGACTATAAAATATTCTAAACTTCAACTGTAAAAAAAGTAGATTTGACACATGAGGTATGATTTAGTGACCAGTCCTATGTATGGAT
TAGCATTGGAAGGTTATCAAAAGCTATTATTAGTATGTATGTTTCCAACTTTAGAGGCCTTTTACATGTCCTTAATGAATTTATATGCCCTGCATTTGGCATTCAATGGTATCATTTATA
ATGAGAAAAGAATAGCTTATTATTGTTAATTAATCATATTCAATCATATTAACCAATTTAATTAACTTAATATTAATAAACGCACATTAATCAAATTAGTCAACTTAATATGGTTGATTA
TGACCTCTGTTTATTTAATTTAATATAATTATTAGTATTAGTGGGCTACTACTTGTAACTTGGTGCAAAAAAGCTGGTCAGTTTCAAACATTTTCTCTGCTGTCTCCCCTTTAGTAATGC
TAAAATTTCAAATTTTGGAATTAAGCCTAGATTTTAATTTTTCAGATTTAGTGCAATGGAGTTCTCTTGACCAAGTCATATACTCATGCAAATGATAGAAATTATAGCAATTGCAATAGC
ATCTGATACTTACCATTAGGGCCTAGCATGTGGCAGGTAATACATACTATTCCATTAATTGAGGATTTCATATATGATACAATGCATTTTGCTCATCTTCACCTCCCTATTTTTTTCCTA
CCTCTTCCAAATTTACTGAATAGTCTAAACACTTGTAATTCTTAATCATTTGGATTCTGAAATATACTGGCATCCTTAGAATGAGAGCTGAGACACAGAAGAGACAGGCCGATCGCGCCG
CTATTCAATTCGCTATAGTGAGTCGTATACATTCACTGGCTGCGTCAACGTCTGAAGTGGAAAAC

>D04_ATS5-3
TGCAGCCGCTCGCATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTAAGG
TGACAAAGCATTTTACTTTTGGCTAGATACATTATGATAGAGAACATACAGGAAATATACAAGTACTTGAAATTTACCATGTTACATTAGAAATTATGGACATAGACATTTAAGTTTTAT
GCAATTTTCACACATGAAACAAAATATAATTTTTTAATGATTTAAGACTATAAAATATTCTAAACTTCAACTGTAAAAAAAGTAGATTTGACACATGAGGTATGATTTAGTGACCAGTCC
TATGTATGGATTAGCATTGGAAGGTTATCAAAAGCTATTATTAGTATGTATGTTTCCAACTTTAGAGGCCTTTTACATGTCCTTAATGAATTTATATGCCCTGCATTTGGCATTCAATGG
TATCATTTATAATGAGAAAAGAATGGCTTATTATTGTTAATTAATCATATTCAATCATATTAACCAATTTAATTAACTTAATATTAATAAACGCACATTAATCAAATTAGTCAACTTAAT
ATGGTTGATTATGACCTCTGTTTATTTAATTTAATATAATTATTAGTATTAGTGGGCTACTACTTGTAACTTGGTGCAAAAAAGCTGGTCAGTTTCAAACATTTTCTCTGCTGTCTCCCC
TTTAGTAATGCTAAAATTTCAAATTTTGGAATTAAGCCTAGATTTTAATTTTACAGATTTAGTGCAATGGAGTTCTCTTGACCAAGTCATATACTCATGCAAATGATAGAAATTATAGCA
ATTGCAATAGCATCTGATACTTACCATTAGGGCCTAGCATGTAGCAGGTAATACATACTATTCCATTAATTGGAGGATTTCATATATGATACAATGCATTTTGCTCATCTTCACCCTCCT
ATTTTTTTCCCTACCCTCTTCCAAATTTACTGAAATAGTCTAAAACACTTGTAATTCTTAAATCATTTGATTCTGAAATATACTGGGCATCCTTAGGGATGGAGGGAAGCTGGAACCACA
GAAAGGAGAACAGGGCGGAACTTCCGGCCCGCTAATCATCGCCCTAAGTGGAGTCGTATAACTCTGCCGGTCGATTACACACGCGTCCTGC

>E04_ATS6-1
TGCAGCCGCTCGATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTCTTTTTTTAAGAAAGCAG
GGTAAAATTTTGCAATAGTTATCTGAAAATTACATTTTTTCCTTCTTACATTCTGCCCTCTCTACGTCAGGAATTAGATATTTTCATACTGCTTAGGGACGTAACTTACCAGTAATAGTC
TTTACTCAGCTCTTGTGGGGTTTCGTTTCATTTTGTTTTGTGCAAGTTAAGTGGCACTATTCTTGTACTTTTAATACTCACACTTCTATTTCCCCAAATAGTCTAAACACTTGTAATTCT
TAAATCATTTGAATTCTAAAATATACTGGCATCCTAAGGATGAGGAAGCTGAGACACAGAAAGAGACAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCT
AAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAA
CTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAA
GTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGG
TGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGT
TGTCAGAAGTAAGTTTGGCCGCAGTGTTATCACTCATGGGTTATGGGCAGCACTGCAATAAATTCTTCTTAACTGTCATGCCCATCCCGTAAGAATGCTTTTTCTTGTGAACTGGTGAAG
TTACTTCAACCCAAAGTCATTTCTGAAGAAATAAGTGTAATGCGCCGAACCGGATTGCTCTGCCGACGTCAAATACGGAAATATAAATACACCGGCGGG

>F04_ATS6-2
ATTACGCCCGCTCGTATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTCTTTTTTTAAGAAAG
CAGGGTAAAATTTTGCAATAGTTATCTGAAAATTACATTTTTTCCTTCTTACATTCTGCCCTCTCTACGTCAGGAATTAGATATTTTCATACTGCTTAGGGACGTAACTTACCAGTAATA
GTCTTTACTCAGCTCTTGTGGGGTTTCGTTTCATTTTGTTTTGTGCAAGTTAAGTGGCACTATTCTTGTACTTTTAATACTCACACTTCTATTTCCCCAAATAGTCTAAACACTTGTAAT
TCTTAAATCATTTGAATTCTAAAATATACTGGCATCCTAAGGATGAGGAAGCTGAGACACAGAAAGAGACAAGGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGTA
TTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCG
CACCGATCGCCCTTCCCAACAGTTGCGCAGCCTATACGTACGGCAGTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTTGTGGATGTACAGAGTGATATTATTGACA
CGCCGGGGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGCTGTCAGATAAAGTCTCCCCGTGAACTTTACCCGGTGGTTGCATATCGGGGATGAAAAGCCTGGCGCATGATGACC
ACCGGATATGGGCCAGTTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCCTGATCTTAACCCACCGCGAAAATGACATCAAAAAACGCCTTTAACCTTGATGTTTTTGGGGGAATTATA

AATGTTCAGGCATGGAGATTATTCAAAAAAGGGAT

>G04_ATS6-3

TTCCCGCCGCCTCGTATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTCTTTTTTAAGAAA
GCAGGGTAAAATTTTGCAATAGTTATCTGAAAATTACATTTTTTCCTTCTTACATTCTGCCCTCTCTACGTCAGGAATTAGATATTTTCATACTGCTTAGGGACGTAACTTACCAGTAAT
AGTCTTTACTCAGCTCTTGTGGGGTTTCGTTTCATTTTGTTTTGTGCAAGTTAAGTGGCACTATTCTTGTACTTTTAATACTCACACTTCTGTTTCCCCAAATAGTCTAAACACTTGTAA
TTCTTAAATCATTTGAATTCTAAAATATACTGGCATCCTAAGGATGAGGAAGCTGAGACACAGAAAGAGACAAGGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGT
ATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCC
GCACCGATCGCCCTTCCCAACAGTTTGCGCAGCCTATACGTACGGCAGTTTAAGGTTTACACCCTATAAAAGAGAGAGCCCGTTATCGTCTGTTTTGTGGATGTACAGAGTGATATTTAT
TGACACGCCGGGGCGACGGATGGTGATCCCCCTGGCCCAGTGCACGTCTGCTTGTCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATG
ACCACCCGATTTGGGCCAGTGTGCCGGTCTTCCGTTATCGG

>H04_ATS7-1

CCCCCCCCCGCCTAGTATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTTTTTTATTTT
AAGGTGACAAAGCATTTTACTTTTGGCTAGATACATTATGATAGAGAACATACAGGAAATATACAAGTACTTGAAATTTACCATGTTACATTAGAAATTATGGACATAGACATTTAAGCT
TTATGCAATTTTCACACATGAAACAAAATATAATTTTTTAATGATTTAAGACTATAAAATATTCTAAACTTCAACTGTAAAAAAAGTAGATTTGACACATGAGGTATGATTTAGTGACCA
GTCCTATGTATGGATTAGCATTGGAAGGTTATCAAAAGCTATTATTAGTATGTATGTTTCCAACTTTAGAGGCCTTTTACATGTCCTTAATGAATTTATATGCCCTGCATTTGGCATTCA
ATGGTATCATTTATAATGAGAAAAGAATAGCTTATTATTGTTAATTAATCATATTCAATCATATTAACCAATTTAATTAACTTAATATTAATAAACGCACATTAATCAAATTAGTCAACT
TAATATGGTTGATTATGACCTCTGTTTATTTAATTTAATATAATTATTAGTACTAGTGGGCTACTACTTGTAACTTTGGTGCAAAAAAGCTGGTCAGTTTTCAAACATTTTCTCTGCTTG
TCTCCCCTTTTAGTAATGCTAAAATTTTCAAATTTTTGGAATTAAAGCCTAGATTTTAATTTTTTTCAAGATTTAGTGCAAATGGGAGTTTCTCTTTGCCCAAGTCATATACCTCCTGCA
AATG

>A05_ATS7-2

TGCCCGCCGCTCGCATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGTCTCTTTCTGTGTCTCAGCTTCCTCATCCTTAGGATGCCAGTATATTTCAGA
ATTCAAATGATTTAAGAATTACAAGTGTTTAGACTATTTCAGTAAATTTTGGAAGAGGTAGGAAAAAAATAGGAGGTGAAGATGAGCAAAATGCATTGTATCATATATGAAATCCTCAAT
TAATGGAATAGTATGTATTACCTGCTACATGCTAGGCCCTAATGGTAAGTATCAGATGCTATTGCAATTGCTATAATTTCTATCATTTGCATGAGTATATGACTTGGTCAAGAGAACTCC
ATTGCACTAAATCTGAAAAATTAAAATCTAGGCTTAATTCCAAAATTTGAAATTTTAGCATTACTAAAGGGGAGACAGCAGAGAAAATGTTTGAAACTGACCAGCTTTTTGCACCAAGT
TACAAGTAGTAGCCCACTAATACTAATAATTATATTAAATTAAATAAACAGAGGTCATAATCAACCATATTAAGTTGACTAATTTGATTAATGTGCGTTTATTAATATTAAGTTAATTAA
ATTGGTTAATATGATTGAATATGATTAATTAACAATAATAAGCTATTCTTTTCTCATTATAAATGATACCATTGAATGCCAAATGCAGGGCATATAAATTCATTAAGGACATGTAAAAGG
CCTCTAAAGTTGGAAACATACATACTAATAATAGCTTTTGATAACCTTCCAATGCTAATCCATACATAGGACTGGTCACTAAATCATACCTCATGTGTCAAATCTACTTTTTTTTACAGT
TTGAAGTTTAGAATATTTTATAGTCTTAAATCATTTAAAAAATTATATTTTGTTTCCATGTGTGAAAATTGCATAAAACTTAAATGTCCTATGTCCATAATTTCTAAATGTAACATGGTT
AAATTTCCAAGTCCTTGTATATTTCCCTGTATTGTTCCTCTATCCATAAATGGTATCTTAAGCCTAAAAGTAAAAATGCCTTCGTCCACCTTTTCAA

>B05_ATS7-3

TTTCCCGCCCGCCTCGATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGCAATTCGCCCTTGTCTCTTTCTGTGTCTCAGCTTCCTCATCCTTAGGATGCCAGTATATTTC
AGAATTCAAATGATTTAAGAATTACAAGTGTTTAGACTATTTCAGTAAATTTTGGAAGAGGTAGGAAAAAAATAGGAGGCGAAGATGAGCAAAATGCATTGTATCATATATGAAATCCTCA
ATTAATGGAATAGTATGTATTACCTGCTACATGCTAGGCCCTAATGGTAAGTATCAGATGCTATTGCAATTGCTATAATTTCTATCATTTGCATGAGTATATGACTTGGTCAAGAGAACT
CCATTGCACTAAATCTGAAAAATTAAAATCTAGGCATAATTCCAAAATTTGAAATTTTAGCATTACTAAAGGGGAGACAGCAGAGAAAATGTTTGAAACTGACCAGCTTTTTTGCACCAA
GTTACAAGTAGTAGCCCACTAATACTAATAATTATACTTCCATTAAGAGAAAAAAATAGATTGAACAAAATCAATCATGAGGCAAAGTAGGAAATTTGAAGAATTTTCAAAATATAGAAC
TATTTTTAAAATACTAGATGCATTGGGGACATAGCTTAATCGTGAGTGCTTATTTTTCACTCATGAGGCTATCCTCAGTACCAGGTATCAGAGTATAAAACTTTTAATTTCATTACTGGG
GAAATAGAAGTGTGAGTATTAAAAGTACAAGAATAGTGCCACTTAACTTGCACAAAACAAAAATGAAACGAAACCCCCCACAAGAGCTGAGTAAAGACCTATTACTGGTAAGTTTACGTC
CCTTAAGCAGTATGAAAAATATCTTAATTCCTTGACGTAGAGAGGGGCAGAATGGTAAGAAGGAAAAAAATGTTATTTTTTCAGGATAACTATTTGCAAAATTTTTACCCTGGCTTTCTT
TAAAAAAAGAAAAAAAAAAATAAAAAGTTCTTGTCGATCGCCGTGGCCCAAAGGGCCGAATCTCGCCGGCCCGCCTAAAATTCAACTTCCGCCCCTACTAGTGAGGGTTCGTATTAC

>E05_ATS8-3

TGTTTGCCCGCTCGCATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGTCTCTTTCTGTGTCTCAGCTTCCTCATCCTTAGGATGCCAGTATATTTCAG
AATTCAAATGATTTAAGAATTACAAGTGTTTAGACTATTTCAGTAAATTTTGGAAGAGGTAGGAAAAAAATAGGAGGTGAAGATGAGCAAAATGCATTGTATCATATATGAAATCCTCAA
TTAATGGAATAGTATGTATTACCTGCTACATGCTAGGCCCTAATGGTAAGTATCAGATGCTATTGCAATTGCTATAATTTCTATCATTTGCATGAGTATATGACTTGGTCAAGAGAACTC
CATTGCACTAAATCTGAAAAATTAAAATCTAGGCTTAATTCCAAAATTTGAAATTTTAGCATTACTAAAGGGGAGACAGCAGAGAAAATGTTTGAAACTGACCAGCTTTTTTGCACCAAG
TTACAAGTAGTAGCCCACTAATACTAATAATTATATTAAATTAAATAAACAGAGGTCATAATCAACCATATTAAGTTGACTAATTTGATTAATGTGCGTTTATTAATATTAAGTTAATTA
AATTGGTTAATATGATTGAATATGATTAATTAACAATAATAAGCTATTCTTTTCTCATTATAAATGATACCATTGAATGCCAAATGCAGGGCATATAAATTCATTAAGGACATGTAAAAG
GCCTCTAAAGTTGGAAACATACATACTAATAAATAGCTTTTGATAACCTTCCAATGCTAATCCATACATAGGACTGGTCACTAAATCATACCTCATGTGTCAAATCTACTTTTTTTTTACA

GTTGAAGTTTAGAATATTTTATAGTCTTTAAATCATTAAAAAAATTATATTTTGTTTCATGTTGTGAAAATTGCATAAAACCTTAAATTGTCTATGTCCATAATTTCTAATGTAACATGG
TAAATTTTCAAGTTACTTGTATATTTTCCTGTTGGTTCTCTATTCATAATTGTATCTTGGCCCAAAAAGTAAAATGCTTTTGTCCACCTTTT

B.2.2 Human

>A05_Hu-111F

CGGACGCCTCCATATAGGGCGAATATGGGCCCATCAAGATGCACGCATCGAGCCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTT
TTTTTTTACCTTCATATTTACCATTTTTATTTATTTTTCTTTGTGGAGATTCCTTTTTGTAGATCCAAGTTTGTTGCTGGTGTTATTTTCCTTTTACTTGAAAAAATTTCTTAAACATTT
CTAAAGTATAGGTCTACTGGCTACAAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGATGCATAGCTTGAGTATTCTATAGTGTCACCTA
AATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTG
AGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGC
TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAG
AACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGG
TGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCCGCTTACCGGATACCTGTCG

>E06_HuAS-111F

TTTCCTGTGAACGACTACTATAGGGCGAATTGGGCCCTCTAGATGCGGAGCACGAGCGGTCGGCAGTGTGATGGATATCTGCAGGGACGCCCTTGGGAGAGCGACTAGTACTTTTTTTTT
TTTTTTACCTTCATATTTACCATTGCAGGGGCTTCTTTGTGGAGAAGGTTTTTGTATTATCCAAGTTTGGTGCTGGTGTTATTTTCCTTTTACTTGAAAAAATTTTAATAACATTTCTAA
AGTATAGGTCCCCTGGCTATAAAGGGCGAATTCCAGCACACTGTCGGCCGTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGATGCATTTTATAGATTATTCTATTTTGTAATATTAAT
AGCCCCCCCCGTTTTTTTTTTTTCATAGCTGTTTCCTGTGTGAAATTGTTATCTTTTTTTTTTTTCCACACAACATACTTTTTCGGAAGCATAAAGTGTTTTTTTTGTGGTGCCTAAT
GAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTT
GGGCGCTCTTCCGCTTTCCTCGCTCACTGACTCGCTTGCGCTCGGTCGTTCGGCTTGCGGCGAGCGGTATCAGCTCACTTAAAGGCGGGTAATTACGGGTTATCCACAGAACCAGGGGGA
TAACGGC

>F06_HuAS-112F

TTGATACGCTCTATAGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTT
TTTTTTTATATTTACCTTCATATTTACCATTTTTATTTATTTTTCTTTGTGGAGATTCCTTTTTGTAGATCCAAGTTTGTTGCTGGTGTTATTTTCCTTTTACTTGAAAAAATTTCTTAA
ACATTTCTAAAGTATAGGTCTACTGGCTACAAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGATGCATAGCTTGAGTATTCTATAGTGT
CACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAA
TGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATT
GGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCCACAGAATCAGGGGATAACCG
CAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAAACCCGTAAAAAGGCCGCGTTGCTGGCGTTTTTTCCATAGGCTTCCGCCCCCCCCTTGACCAGCATCACCAAAATTTG
ACGCTCAAGGTCAGAGGTGGCCGAACCCCCGACAGGACTATAAAGATTCCAGGCCGTTTTCCCCCCTGGAAGCTCCCCTCCTGCGCTCTCCTCGTTCCGACCCCTGGCCGCTTACCGGAT
TACACTGTCCGCCCTTTTCTCCTTCTCGGGAAGGCCGTGGC

>G06_HuAS-121F

CGGTGATCGCCTCCTATAGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTT
TTTTCTAACACTTGGTAATTTCTGTTTTTTATTTAAAATAATTGCCATTCTAATAGGCGTAAAAAGTCATTATGGTCTTAATTTACATTTTCTTCATGACTGCTGTTGAGAAGTTTTCTT
TTGTGTGTCTGTTCAAATATTTTCACATTCTTTTCAGAACATATGTTCTTTGTCAGATAGATAAATGTATTTCTTCTGTATCCTTACTGATTTTCTGTCCACTTACTAATTACCGAGGGA
AGAATGTTGAAATCTTCAACTATAATTTTGAGTTTATCTTCCTTCTGCTCTGTTCATTTTTTCTTCATATGCTTTGAAGCAAAGTCTTTAAAGTTGTGTCCTTAGGTGTGTTCACATTTA
GGACTGTTAGATCTCCTTGATTAATTGATGTCCTTATCATTATGAAATGTCTCTGTTTATCGCTGGTAATGTTCCTTAT

>H06_HuAS-122F

TTTTTTACCCTCTATAGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTAACGCTGCTGTGGAATCTATAACACATATAGA
AGTAAAATGTTAAGATCATAATACCAAAAAGAATGAGGGAGAGTAAATGGATGTATACTATTCCAAAATTCTTACATTATACATTAAGAGGAATAATATTTTTGAAGGTAGATTGAGGTA
AGTTAAAGGTGCATATTGTAAATCCTAGAAGTATCACTGGGGGGAAAAAAAAACAGAGGGACAAGCTAATAATGGAGACAATACAAAATTCCAAAAATGCTCCATTAACCGAAAAGAAGGC
AGAAAAGGAAGAAAAGATTAGACAAATAGAAAACAAATAGTGTCTTAAACACAATTATATTGCCAACTGCAATAAATGTAAATGGACTTAATCAAAAAGCAACGATCGTAAAAAGTCAAT
CTGAAAGCAAGATGCAACTATATGCTATCTAAAAGAAAGACATTAAATGACCAATAAGTTGAAAATAAAAACACAGCAAAATGATATAATGTGCAAAAATTAATCATAAGAAAGGAATGG
ATACACTAATATCAAAGTAGACTTTTAGATAAGGAACATTACCAGCGATAAACAGAGACATTTCATAATTGATAAAGACATCAATTTAATCAAGGAGTATCTAAACAGTCCCTTAAATG

>B05_Hu-211F

CAGACGCCTCTATAGGGCGAATCTGGGCCCTCAAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTCAGTGAACGCCAAACAAGTAAGATTAGAATTAC

AATTAGAATCAGAAATGACAAAAAAGAAACATGAACTGTGAGATGCAATATAAAAACTAATGGTGAACAAAGCAAAGAATAAAAAATAAAAACGGAAAGATTAAAAGGAAAATGTGACTG
TAAGTAGAGTGACCTTTATATTATACAGAAGGTAAAGCAGAATGAACATGTGAGGAATCTAAAAAACAGATATAAAATAGAAACAACTGAAGGCTTATATAATTAGTCTATAAAGCTAGG
AACCACTGCTTAGTAATTTCTCGTAAGTGGAGTCTCTTTGTTCTGAGACGTTGCAAAACTCCATTCCATCTCAAAGTTATTTTTAATGTTAAAATTTGTAAGAGTGAGCAACGGTATCCT
AAATAGTTCTAGGCCTTGTAGAAATAGACTGTAAGAGACATCTGTGACACAACTATATTCAGTATGATGTAGGGTCTGTTACAATGGGATGAAGTGAATTTAGATTATTTCAAGACTAAA
CCTACAATTGTGAACTCATTAGTACCATGAACCGAACAAAAACCTATTAAATATAGAACAAGACGATTTTCATGAAAAAATAAATATGAACAAATCAGTATTTTCAGTTGTTTGAAATAT
AGGTTAACCTTCAATTACTGTTTTATGGCCGGGTGCAGTGGCTCACGCCACGCCTCTAATCCCAGCACTTTGGGAGGCCGAGGCAGGCGGATCATGAGGTCAGGAGATCGAGACCATCCT
GGCTAACATGGTGAAACTACGTCCCTACTACAAATACAAAAAAAAAAATTAGCTGGGCGCCGTGGGTGGGGCGCCTATAGTCCCAAGCT

>C05_Hu_212F .

AGAGTGACCTTTATATTATACAGAAGGTAAAGCAGAATGAACATGTGAGGAATCTAAAAAACAGATATAAAATAGAAACAACTGAAGGCTTATATAATTAGTCTATAAAGCTAGGAACCA
CTGCCTAGTAATTTCTCGTAAGTGGAGTCTCTTTGTTCTGAGACGTTGCAAAACTCCATTCCATCTCAAAGTTATTTTTAATGTTAAAATTTGTAAGAGTGAGCAACGGTATCCTAAATA
GTTCTAGGCCTTGTAGAAATAGACTGTAAGAGACATCTGTGACACAACTATATTCAGTATGATGTAGGGTCTGTTACAATGGGATGAAGTGAATTTAGATTATTTCAAGACTAAACCTAC
AATTGTGAACTCATTAGTACCATGAACCGAACAAAAACCTATTAAATATAGAACAAGACGACTTTCATGAAAAAATAAATATGAACAAATCAGTATTTTCAGTTGTTTGAAATATAGGTT
AACCTTCAATTACTGTTTTATGGCCGGGTGCAGTGGCTCACGCCACGCCTCTAATCCCAGCACTTTGGGAGGCCGAGGCAGGCGGATCATGAGGTCAGGAGATCGAGACCATCCTGGCTA
ACATGGTGAAACTACGTCCCTACTAAAAATACAAAAAAAATTAGCTGGGCGCCGTGGTGGGCGCCTATAGTCCCAGCTACTTGGGAGGCTGATGCAGAAGAATGGCATGAACCCGGGAGG
CGGAGCTTGCAGTGAGCCGAGACAGAGCCACTGCACTGTGGCCTGGGCAAAAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAGTACTAGTCGACGCGTGGCCAAGGGCGAATTCTGCAG
ATATCCATCACACTGGCGGCCGCTCGAGCATGCATCTTAAGGGCCCAGATTCGCCCTATAGTAGCGACAG

>D05_Hu-212R

TCGACGCCAGCTATTTAGGTGACACTATAGAATACTCAAGTCTATGCATCAAGCTTGGTACCGAGCTCGGCTCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTCAGTGAACG
CCAAACAAGTAAGATTAGAATTACAATTAGAATCAGAAATGACAAAAAAGAAACATGAACTGTGAGATGCAATATAAAAACTAATGGTGAACAAAGCAAAGAATAAAAAATAAAAACGGA
AAGATTAAAAGGAAAATGTGACTGTAAGTAGAGTGACCTTTATATTATACAGAAGGTAAAGCAGAATGAACATGTGAGGAATCTAAAAAACAGATATAAAATAGAAACAACTGAAGGCTT
ATATAATTAGTCTATAAAGCTAGGAACCACTGCCTAGTAATTTCTCGTAAGTGGAGTCTCTTTGTTCTGAGACGTTGCAAAACTCCATTCCATCTCAAAGTTATTTTTAATGTTAAAATT
TGTAAGAGTGAGCAACGGTATCCTAAATAGTTCTAGGCCTTGTAGAAATAGACTGTAAGAGACATCTGTGACACAACTATATTCAGTATGATGTAGGGTCTGTTACAATGGGATGAAGTG
AATTTAGATTATTTCAAGACTAAACCTACAATTGTGAACTCATTAGTACCATGAACCGAACAAAAACCTATTAAATATAGAACAAGACGACTTTCATGAAAAAATAAATATGAACAAATC
AGTATTTTCAGTTGTTTGAAATATAGGTTAACCTTCAATTACTGTTTTATGGCCGGGTGCAGTGGCTCACGCCACGCCTCTAATCCCAGCACTTTGGGAGGCCGAGGCAGGCGGATCATG
AGTCAGGAGATCGAGACCATCCTGGCTAACATGGTGAACTACGTCCTACTAAAAATACAAAAAAAAAATTAGCTGGGCGCCGTGGTGGGCGCCTATAGTCCAGCTACTTTG

>E05_Hu-221F

ATGTCGCTATATAGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTCAGTGAACGCCAAACAAGTAAGATTAGAATTACAA
TTAGAATCAGAAATGACAAAAAAGAAACATGAACTGTGAGATGCAATATAAAAACTAATGGTGAACAAAGCAAAGAATAAAAAATAAAAACGGAAAGATTAAAAGGAAAATGTGACTGTA
AGTAGAGTGACCTTTATATTATACAGAAGGTAAAGCAGAATGAACATGTGAGGAATCTAAAAAACAGATATAAAATAGAAACAACTGAAGGCTTATATAATTAGTCTATAAAGCTAGGAA
CCACTGCTTAGTAATTTCTCGTAAGTGGAGTCTCTTTGTTCTGAGACGTTGCAAAACTCCATTCCATCTCAAAGTTATTTTTAATGTTAAAATTTGTAAGAGTGAGCAACGGTATCCTAA
ATAGTTCTAGGCCTTGTAGAAATAGACTGTAAGAGACATCTGTGACACAACTATATTCAGTATGATGTAGGGTCTGTTACAATGGGATGAAGTGAATTTAGATTATTTCAAGACTAAACC
TACAATTGTGAACTCATTAGTACCATGAACCGAACAAAAACCTATTAAATATAGAACAAGACGATTTTCATGAAAAAATAAATATGAACAAATCAGTATTTTCAGTTGTTTGAAATATAG
GTTAACCTTCAATTACTGTTTTATGGCCGGGTGCAGTGGCTCACGCCACGCCTCTAATCCCAGCACTTTGGGAGGCCGAGGCACGCGGATCATGGGGTCAGGAGATCGAGACCATCCTGG
CTAACATGGTGAAACTACGTCCCTACTAAAATACAAAAAAAATTAGCTGGGCGCCGTGGTGGGCGCCCATAGTCCCAGC

>F05_Hu-222F

ATGACGACTATATAGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTT
TTTGAGACAGAGTCTCGCTCTTTTGCCCAGGCCACAGTGCAGTGGCTCTGTCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCATGCCATTCTTCTGCATCAGCCTCCCAAGTAGCTGG
GACTATAGGCGCCCACCACGGCGCCCAGCTAATTTTTTTTTGTATTTTTAGTAGGGACGTAGTTTCACCATGTTAGCCAGGATGGTCTCGATCTCCTGACCTCATGATCCGCCTGCCTCG
GCCTCCCAAAGTGCTGGGATTAGAGGCGTGGCGTGAGCCACTGCACCCGGCCATAAAACAGTAATTGAAGGTTAACCTATATTTCAAACAACTGAAAATACTGATTTGTTCATATTTATT
TTTTCATGAAAATCGTCTTGTTCTATATTTAATAGGTTTTTGTTCGGTTCATGGTACTAATGAGTTCACAATTGTAGGTTTAGTCTTGAAATAATCTAAATTCACTTCATCCCATTGTAA
CAGACCCTACATCATACTGAATATAGTTGTGTCACAGATGTCTCTTACAGTCTATTTCTACAAGGCCTAGAACTATTTAGGATACCGTTGCTCACTCTTACAAATTTTAACATTAAAAAT
AACTTTGAGATGGAATGGAGTTTTGCAACGTCTCAGAACAAAGAGACTCCACTTACGAGAAATTACTAAGCAGTGGTTCCTAGCTTTATAGACTAATTATATAAGCCTTCAGTTGTTTCT
ATTTTATATCTGTTTTTTAGATTCCTCACATGTTCATTCTGCTTTACCTTCTGTATAATATAAAGGTCACTCTACTTACAGTCACATTTCCCTTTTAATCTTTCCGTTT

>G05_Hu-222R

TGTCGCCAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTCAGTGAACGCC
AAACAAGTAAGATTAGAATTACAATTAGAATCAGAAATGACAAAAAAGAAACATGAACTGTGAGATGCAATATAAAAACTAATGGTGAACAAAGCAAAGAATAAAAAATAAAAACGGAAA
GATTAAAAGGAAAATGTGACTGTAAGTAGAGTGACCTTTATATTATACAGAAGGTAAAGCAGAATGAACATGTGAGGAATCTAAAAAACAGATATAAAATAGAAACAACTGAAGGCTTAT

ATAATTAGTCTATAAAGCTAGGAACCACTGCTTAGTAATTTCTCGTAAGTGGAGTCTCTTTGTTCTGAGACGTTGCAAAACTCCATTCCATCTCAAAGTTATTTTTAATGTTAAAATTTG
TAAGAGTGAGCAACGGTATCCTAAATAGTTCTAGGCCTTGTAGAAATAGACTGTAAGAGACATCTGTGACACAACTATATTCAGTATGATGTAGGGTCTGTTACAATGGGATGAAGTGAA
TTTAGATTATTTCAAGACTAAACCTACAATTGTGAACTCATTAGTACCATGAACCGAACAAAAACCTATTAAATATAGAACAAGACGATTTTCATGAAAAAATAAATATGAACAAATCAG
TATTTTCAGTTGTTTGAAATATAGGTTAACCTTCAATTACTGTTTTATGGCCGGGTGCAGTGGCTCACGCCACGCCTCTAATCCCAGCACTTTGGGAGGCCGAGGCAGCGGATCATGAGG
TCAGGAGATCGAGACCATCCTGGCTAACATGGTGAAACTACGTCCCTACTAAAAATACAAAAAAAAATTAGCTGGGCGCCGTGGTGGGCGCCTATAGTCCCAGCTACTTGGGAGGCTGAT
GCAGAAAGA

>G05_Hu-222R

TGTCGCCAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTCAGTGAACGCC
AAACAAGTAAGATTAGAATTACAATTAGAATCAGAAATGACAAAAAAGAAACATGAACTGTGAGATGCAATATAAAAACTAATGGTGAACAAAGCAAAGAATAAAAAATAAAAACGGAAA
GATTAAAAGGAAAATGTGACTGTAAGTAGAGTGACCTTTATATTATACAGAAGGTAAAGCAGAATGAACATGTGAGGAATCTAAAAAACAGATATAAAATAGAAACAACTGAAGGCTTAT
ATAATTAGTCTATAAAGCTAGGAACCACTGCTTAGTAATTTCTCGTAAGTGGAGTCTCTTTGTTCTGAGACGTTGCAAAACTCCATTCCATCTCAAAGTTATTTTTAATGTTAAAATTTG
TAAGAGTGAGCAACGGTATCCTAAATAGTTCTAGGCCTTGTAGAAATAGACTGTAAGAGACATCTGTGACACAACTATATTCAGTATGATGTAGGGTCTGTTACAATGGGATGAAGTGAA
TTTAGATTATTTCAAGACTAAACCTACAATTGTGAACTCATTAGTACCATGAACCGAACAAAAACCTATTAAATATAGAACAAGACGATTTTCATGAAAAAATAAATATGAACAAATCAG
TATTTTCAGTTGTTTGAAATATAGGTTAACCTTCAATTACTGTTTTATGGCCGGGTGCAGTGGCTCACGCCACGCCTCTAATCCCAGCACTTTGGGAGGCCGAGGCAGCGGATCATGAGG
TCAGGAGATCGAGACCATCCTGGCTAACATGGTGAAACTACGTCCCTACTAAAAATACAAAAAAAAATTAGCTGGGCGCCGTGGTGGGCGCCTATAGTCCCAGCTACTTGGGAGGCTGAT
GCAGAAAGA

>F05_Hu222F

ATGACGACTATATAGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTT
TTTGAGACAGAGTCTCGCTCTTTTGCCCAGGCCACAGTGCAGTGGCTCTGTCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCATGCCATTCTTCTGCATCAGCCTCCCAAGTAGCTGG
GACTATAGGCGCCCACCACGGCGCCCAGCTAATTTTTTTTTGTATTTTTAGTAGGGACGTAGTTTCACCATGTTAGCCAGGATGGTCTCGATCTCCTGACCTCATGATCCGCCTGCCTCG
GCCTCCCAAAGTGCTGGGATTAGAGGCGTGGCGTGAGCCACTGCACCCGGCCATAAAACAGTAATTGAAGGTTAACCTATATTTCAAACAACTGAAAATACTGATTTGTTCATATTTATT
TTTTCATGAAAATCGTCTTGTTCTATATTTAATAGGTTTTTGTTCGGTTCATGGTACTAATGAGTTCACAATTGTAGGTTTAGTCTTGAAATAATCTAAATTCACTTCATCCCATTGTAA
CAGACCCTACATCATACTGAATATAGTTGTGTCACAGATGTCTCTTACAGTCTATTTCTACAAGGCCTAGAACTATTTAGGATACCGTTGCTCACTCTTACAAATTTTAACATTAAAAAT
AACTTTGAGATGGAATGGAGTTTTGCAACGTCTCAGAACAAAGAGACTCCACTTACGAGAAATTACTAAGCAGTGGTTCCTAGCTTTATAGACTAATTATATAAGCCTTCAGTTGTTTCT
ATTTTATATCTGTTTTTTAGATTCCTCACATGTTCATTCTGCTTTACCTTCTGTATAATATAAAGGTCACTCTACTTACAGTCACATTTCCCTTTTAATCTTTCCGTTT

>E05_Hu-221F

ATGTCGCTATATAGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTCAGTGAACGCCAAACAAGTAAGATTAGAATTACAA
TTAGAATCAGAAATGACAAAAAAGAAACATGAACTGTGAGATGCAATATAAAAACTAATGGTGAACAAAGCAAAGAATAAAAAATAAAAACGGAAAGATTAAAAGGAAAATGTGACTGTA
AGTAGAGTGACCTTTATATTATACAGAAGGTAAAGCAGAATGAACATGTGAGGAATCTAAAAAACAGATATAAAATAGAAACAACTGAAGGCTTATATAATTAGTCTATAAAGCTAGGAA
CCACTGCTTAGTAATTTCTCGTAAGTGGAGTCTCTTTGTTCTGAGACGTTGCAAAACTCCATTCCATCTCAAAGTTATTTTTAATGTTAAAATTTGTAAGAGTGAGCAACGGTATCCTAA
ATAGTTCTAGGCCTTGTAGAAATAGACTGTAAGAGACATCTGTGACACAACTATATTCAGTATGATGTAGGGTCTGTTACAATGGGATGAAGTGAATTTAGATTATTTCAAGACTAAACC
TACAATTGTGAACTCATTAGTACCATGAACCGAACAAAAACCTATTAAATATAGAACAAGACGATTTTCATGAAAAAATAAATATGAACAAATCAGTATTTTCAGTTGTTTGAAATATAG
GTTAACCTTCAATTACTGTTTTATGGCCGGGTGCAGTGGCTCACGCCACGCCTCTAATCCCAGCACTTTGGGAGGCCGAGGCACGCGGATCATGGGGTCAGGAGATCGAGACCATCCTGG
CTAACATGGTGAAACTACGTCCCTACTAAAATACAAAAAAAATTAGCTGGGCGCCGTGGTGGGCGCCCATAGTCCCAGC

>D05_Hu-212R

TCGACGCCAGCTATTTAGGTGACACTATAGAATACTCAAGTCTATGCATCAAGCTTGGTACCGAGCTCGGCTCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTCAGTGAACG
CCAAACAAGTAAGATTAGAATTACAATTAGAATCAGAAATGACAAAAAAGAAACATGAACTGTGAGATGCAATATAAAAACTAATGGTGAACAAAGCAAAGAATAAAAAATAAAAACGGA
AAGATTAAAAGGAAAATGTGACTGTAAGTAGAGTGACCTTTATATTATACAGAAGGTAAAGCAGAATGAACATGTGAGGAATCTAAAAAACAGATATAAAATAGAAACAACTGAAGGCTT
ATATAATTAGTCTATAAAGCTAGGAACCACTGCCTAGTAATTTCTCGTAAGTGGAGTCTCTTTGTTCTGAGACGTTGCAAAACTCCATTCCATCTCAAAGTTATTTTTAATGTTAAAATT
TGTAAGAGTGAGCAACGGTATCCTAAATAGTTCTAGGCCTTGTAGAAATAGACTGTAAGAGACATCTGTGACACAACTATATTCAGTATGATGTAGGGTCTGTTACAATGGGATGAAGTG
AATTTAGATTATTTCAAGACTAAACCTACAATTGTGAACTCATTAGTACCATGAACCGAACAAAAACCTATTAAATATAGAACAAGACGACTTTCATGAAAAAATAAATATGAACAAATC
AGTATTTTCAGTTGTTTGAAATATAGGTTAACCTTCAATTACTGTTTTATGGCCGGGTGCAGTGGCTCACGCCACGCCTCTAATCCCAGCACTTTGGGAGGCCGAGGCAGGCGGATCATG
AGTCAGGAGATCGAGACCATCCTGGCTAACATGGTGAACTACGTCCTACTAAAAATACAAAAAAAAATTAGCTGGGCGCCGTGGTGGGCGCCTATAGTCCAGCTACTTTG

>C05_Hu_212F

AGAGTGACCTTTATATTATACAGAAGGTAAAGCAGAATGAACATGTGAGGAATCTAAAAAACAGATATAAAATAGAAACAACTGAAGGCTTATATAATTAGTCTATAAAGCTAGGAACCA
CTGCCTAGTAATTTCTCGTAAGTGGAGTCTCTTTGTTCTGAGACGTTGCAAAACTCCATTCCATCTCAAAGTTATTTTTAATGTTAAAATTTGTAAGAGTGAGCAACGGTATCCTAAATA
GTTCTAGGCCTTGTAGAAATAGACTGTAAGAGACATCTGTGACACAACTATATTCAGTATGATGTAGGGTCTGTTACAATGGGATGAAGTGAATTTAGATTATTTCAAGACTAAACCTAC

AATTGTGAACTCATTAGTACCATGAACCGAACAAAAACCTATTAAATATAGAACAAGACGACTTTCATGAAAAAATAAATATGAACAAATCAGTATTTTCAGTTGTTTGAAATATAGGTT
AACCTTCAATTACTGTTTTATGGCCGGGTGCAGTGGCTCACGCCACGCCTCTAATCCCAGCACTTTGGGAGGCCGAGGCAGGCGGATCATGAGGTCAGGAGATCGAGACCATCCTGGCTA
ACATGGTGAAACTACGTCCCTACTAAAAATACAAAAAAAATTAGCTGGGCGCCGTGGTGGGCGCCTATAGTCCCAGCTACTTGGGAGGCTGATGCAGAAGAATGGCATGAACCCGGGAGG
CGGAGCTTGCAGTGAGCCGAGACAGAGCCACTGCACTGTGGCCTGGGCAAAAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAGTACTAGTCGACGCGTGGCCAAGGGCGAATTCTGCAG
ATATCCATCACACTGGCGGCCGCTCGAGCATGCATCTTAAGGGCCCAGATTCGCCCTATAGTAGCGACAG

>B05_Hu-211F

CAGACGCCTCTATAGGGCGAATCTGGGCCCTCAAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTCAGTGAACGCCAAACAAGTAAGATTAGAATTAC
AATTAGAATCAGAAATGACAAAAAAGAAACATGAACTGTGAGATGCAATATAAAAACTAATGGTGAACAAAGCAAAGAATAAAAAATAAAAACGGAAAGATTAAAAGGAAAATGTGACTG
TAAGTAGAGTGACCTTTATATTATACAGAAGGTAAAGCAGAATGAACATGTGAGGAATCTAAAAAACAGATATAAAATAGAAACAACTGAAGGCTTATATAATTAGTCTATAAAGCTAGG
AACCACTGCTTAGTAATTTCTCGTAAGTGGAGTCTCTTTGTTCTGAGACGTTGCAAAACTCCATTCCATCTCAAAGTTATTTTTAATGTTAAAATTTGTAAGAGTGAGCAACGGTATCCT
AAATAGTTCTAGGCCTTGTAGAAATAGACTGTAAGAGACATCTGTGACACAACTATATTCAGTATGATGTAGGGTCTGTTACAATGGGATGAAGTGAATTTAGATTATTTCAAGACTAAA
CCTACAATTGTGAACTCATTAGTACCATGAACCGAACAAAAACCTATTAAATATAGAACAAGACGATTTTCATGAAAAAATAAATATGAACAAATCAGTATTTTCAGTTGTTTGAAATAT
AGGTTAACCTTCAATTACTGTTTTATGGCCGGGTGCAGTGGCTCACGCCACGCCTCTAATCCCAGCACTTTGGGAGGCCGAGGCAGGCGGATCATGAGGTCAGGAGATCGAGACCATCCT
GGCTAACATGGTGAAACTACGTCCCTACTACAAATACAAAAAAAAAAATTAGCTGGGCGCCGTGGGTGGGGCGCCTATAGTCCCAAGCT

>H05_Hu-311F

ATGTCGCTATATAGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTCTTTT
GTGGCATGAGTTGTTTTTGTTTTTAATTTGTTGTGCTGTTACTAAAGTTCTGAGGGCTGCAGTTAAAACATTCCAATTTCTCCCTTCCTTCCATCTTTCTTTATTGATTGATTCTCAAGA
TTTTGCACAGAAAACTCCATGTGGGCTGGAACAGCAGATTTCAGCTACATATATGAACAAATCCTTTATTATTATTATAATTATTTTTTGCGTGAAAGTGTTACATATTCTTTCACTTGT
ATGTACAGAGAGGTTTTTCTGAATATTTATTTTAAGGGTTAAATCACTTTTGCTTGTGTTTATTACTGCTTGAGGTTGAGCCAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTG
GATCCGAGCTCGGTACCAAGCTTGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAAAT
TCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTG
CCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCATCCTCGCTC

>A06_Hu-311R

TCGAGCCAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGGCTCAACCTC
AAGCAGTAATAAACACAAGCAAAAGTGATTTAACCCTTAAAATAAATATTCAGAAAAACCTCTCTGTACATACAAGTGAAAGAATATGTAACACTTTCACGCAAAAAATAATTATAATAA
TAATAAAGGATTTGTTCATATATGTAGCTGAAATCTGCTGTTCCAGCCCACATGGAGTTTTCTGTGCAAAATCTTGAGAATCAATCAATAAAGAAAGATGGAAGGAAGGGAGAAATTGGA
ATGTTTTAACTGCAGCCCTCAGAACTTTAGTAACAGCACAACAAATTAAAAACAAAAACAACTCATGCCACAAAAGAAAAAAAAAAGTACTAGTCGACGCGTGGCCAAGGGCGAATTCTG
CAGATATCCATCACACTGGCGGCCGCTCGAGCATGCATCTAGAGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCC
TGGCGTTACCCAAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGA
CGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGAGGAGGGT

>B06_Hu-312F

ATGACGCTACTATAGGGCGAATTGGGCCCTCTAGTATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTGGCTCAACCTCAAGCAGTAATAAACACAAGCAA
AAGTGATTTAACCCTTAAAATAAATATTCAGAAAAACCTCTCTGTACATACAAGTGGAAGAATATGTAACACTTTCACGCAAAAAAATAATTATAATAATAATAAAGGATTTGTTCATAT
ATGTAGCTGAAATCTGCTGTTCCAGCCCACATGGAGTTTTCTGTGCAAAATCTTGAGAATCAATCAATAAAGAAAGATGGAAGGAAGGGAGAAATTGGAATGTTTTAACTGCAGCCCTCA
GAACTTTAGTAACAGCACAACAAATTAAAAACAAAAACAACTCATGCCACAAAAAAAAAAAAAAAGTACTAGTCGACCGCGTGGCCAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTA
GTGGATCCGAGCTCGGTACCAAGCTTGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCAT

>C06_Hu-312R

AGTCGCCAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGGCCACGCGGT
CGACTAGTACTTTTTTTTTTTTTTGTGGCATGAGTTGTTTTTGTTTTTAATTTGTTGTGCTGTTACTAAAGTTCTGAGGGCTGCAGTTAAAACATTCCAATTTCTCCCTTCCTTCCATCT
TTCTTTATTGATTGATTCTCAAGATTTTGCACAGAAAACTCCATGTGGGCTGGAACAGCAGATTTCAGCTACATATATGAACAAATCCTTTATTATTATTATAATTATTTTTTTGCGTGA
AAGTGTTACATATTCTTCCACTTGTATGTACAGAGAGGTTTTTCTGAATATTTATTTTAAGGGTTAAATCACTTTTGCTTGTGTTTATTACTGCTTGAGGTTGAGCCAAGGGCGAATTCT
GCAGATATCCATCACACTGGCGGCCGCTCGAGCATGCATCTAGAGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACC
CTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGA
CGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCAC
AATACGCCAGACTATTCACCGTCAAGCTCTAAATCGGGGGCTCCCTTTAAGGGTTCCGATTTAGTGCTTTACCGGCACCTCGACCCCAAAAAAACTTTGATTAAGGGTGATGGTTC

>D06_Hu-321F

CTGGACGACTACTATAGGGCGAATTGGGCCCTCTAGATGCATAGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTT
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTACTTTTGGTTGGGTTTATTACTGGTTGGGGGTGAGCCAAGGGGGAATTCCCCCCCACTGGGGGCC
GTTTCTTGTGGATCCCAGCTCGGTACCCAACTTGGTGCATAACTTGGGTTTTTTTTTGTGGCCCCTAAATAGCTTGGGGTAATCCTGGGCATAGCTTTTTCCCGGGGGAAAATTTTATCC
CCTTACAAATCCCCACAACATACGAGCCGGAAGCATAAAGTGTAAAACCCGGGGGGCCTAATGGGGGAGCCAACTCCCATTTATTGGGTTGGGCCTACCGGCCCCTTTTCCAGCGGGAAA
ACCGGCGGGCCCCCTGCTTTAAAGAATTCGCCCACCCCCGGGGGGGGGGGGTTTGCGTTTTGGGGGCCTTTCCCCTTTCCCCGTTATGGACCCCTTGCGCCTGGTCGTTCGGGTGGGGG
GAGGGGTATTAGCTCCCTCAAAGGGGGGGAATCCCGGTTTTCCCCCAAATCGGGGTTCCCCCAGGAAAAAAATTTGACCAAAGGCCCCCAAAAGGCCCGGACCCTAAAAGGGCCCCTTCT
TGGGTTTTTTCCTAGGGTCTCCCCCCCTGACAGGTATACAAAAATCACCCTCCAGTCTAGGGGGGGAACCCCGCAGGGCTATATAGATCCCACGGGTTTTCCCCTGGAAATCCCCTTGCG

>E06_Hu-321R

TGTCGCCAGCTATTTAGGTGACACTATAGAATACTCAAGCTATGCATCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGGCTCAACCTC
AAGCAGTAATAAACACAAGCAAAAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGTACTAGTCGACGCGTGGCCAAGGGCGA
ATTCTGCAAATATCCATCACACTGGCGGCCGCTCGAGCATGCATCTAGAGGGCCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGA
AAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCAAATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAAAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCACCCTGAA

>F06_Hu-322F

TTGACGCTATATAGGGCGAATTGGGCCCTCTAGTATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTT
TTGCGTGAAAGTGTTACATATTCTTTCACTTGTTTGTACAGAGAGGTTTTTCTGAATATTTATTTTAAGGGTTAAATAACTTTTGCTTGTGTTTATTACTGCTTGAGGTTGAGCCAAGGG
CGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTG
TTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCA
CTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTG
CGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCC
GGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAAGTGGCGAAACCCCGACAGGACTATAAAGATA
CCACGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCAGATACCTGTCCGCCTTTCTTCCTTCTGCAGC

>A07_HuAS-322R

TGCATGCCACTTTTAGGTGACACTATAGAATACTCAAGCTATGCAACAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTGGCTCAACCTCA
AGCAGTAATAAACACAAGCAAAAGTGATAAAACCCTTAAAATAAATATTCAGAAAAACCTTTTTGTACATACAAGTGAAAGAATATGTAACACTTTCACGCAAAAAAAAAAAAAAAAAATA
CTAGTCGACGCGTGGCCAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCATGCATCTAGAGGAGCCAATTCGCCCTATAGTGAGTCGTATTACAATTTTTTTTTTT
TTTTTTGTTACAACGTCGTGACTGGTAAAAACACGGCGTTACCCAACTTAATCGCCTTTTTTTTTTTTTTTCCTTTCGCCAGCTGGTGTAATAACGAAAAAGAACCCCCCCCTCGCCCT
TCCTAACAATTGCGCAGCCTGAATGGCGAATGGACGCCCCCTGTAGCGGCGCATTAATCGCGGCGGGTGTGGTGGTTACTCCCCAGCATGACCACTACACTTGCTAGCGCCCTAGTGCCC
GCTCCTTTAGATTTATTCCCTTCCTATATTCGCCACGATTCGACGGCTTTCCCCTCAAAGCACTAAAACGGGGGCGCCC

B.3 SNP analysis

Table B.5: SNP information from DBAxC57 hybrid mice

SNP	Location	Allelic Freq	DP4[2]
T/G	Chr7:66507227	-1.0	0,0,7,6
C/T	Chr7:66508429	-0.93	1,2,13,25
C/A	Chr7:66508469	-0.96	1,1,30,25
G/A	Chr7:66509079	-1.0	0,0,13,18
A/T	Chr7:66509131	-1.0	0,0,17,13

[2]Position: reference forward, reference reverse, alternative forward, alternative reverse

B.4 Exon usage

Table B.6: Exon genomic locations for *Ube3a-AS/UBE3A-AS*

Isoform	Chromosome	Start	End	Strand
AS Iso1	Chr7	66,452,265	66,453,513	-
AS Iso2	Chr7	66,465,881	66,485,019	-
AS Iso3	Chr7	66,499,737	66,530,634	-
AS Isoform	Chr15	25,663,935	25,667,541	+

B.5 Ballgown analysis

The custom annotation generated from stranded, high-depth reads (SRP01204) was used for guided StringTie assembly for Ballgown ready analysis for downstream analysis with Ballgown. Ballgown [151, 328] (version 2.6.0), a Bioconductor package, was used to analyze the highly expressed transcripts between tissues. A CSV file with phenotype information was generated and imported into R to generate ballgown objects. Ballgown objects were filtered by transcription FPKM variance (≥ 1). The average transcript expression profile was plotted with `plotMeans`.

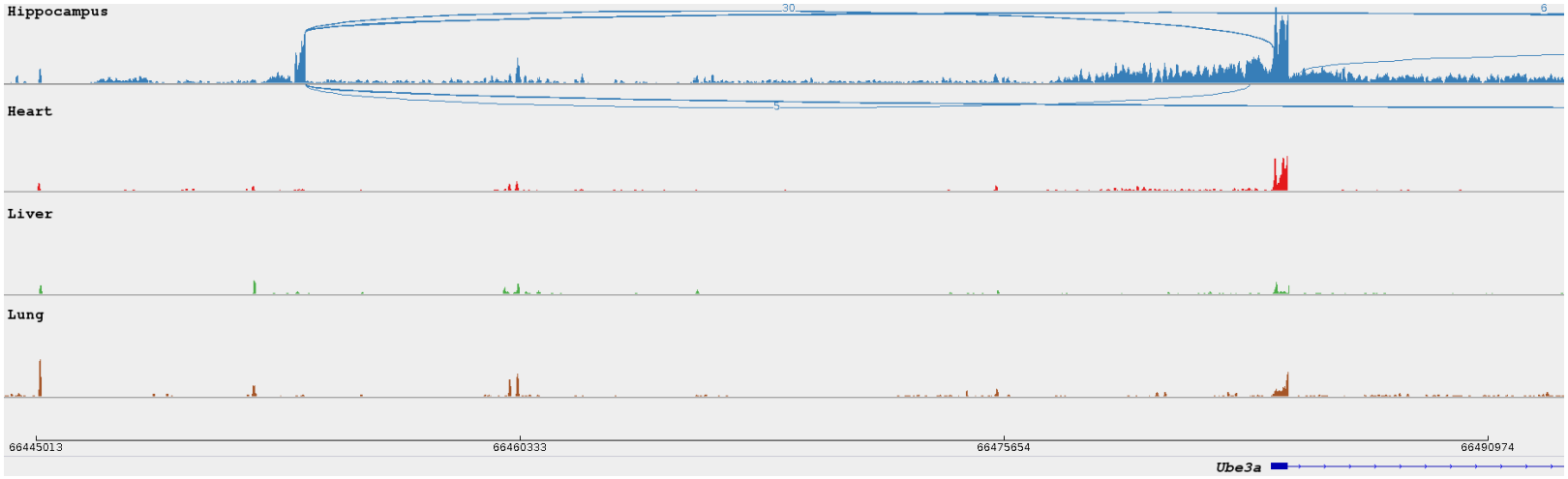


Figure B.1: Sashimi plots demonstrating tissue-specific splicing and alternative splicing in mouse at the 5' end of *Ube3a* (n = 6). Minimum junction coverage = 5. Data generated from Sanger Institute.

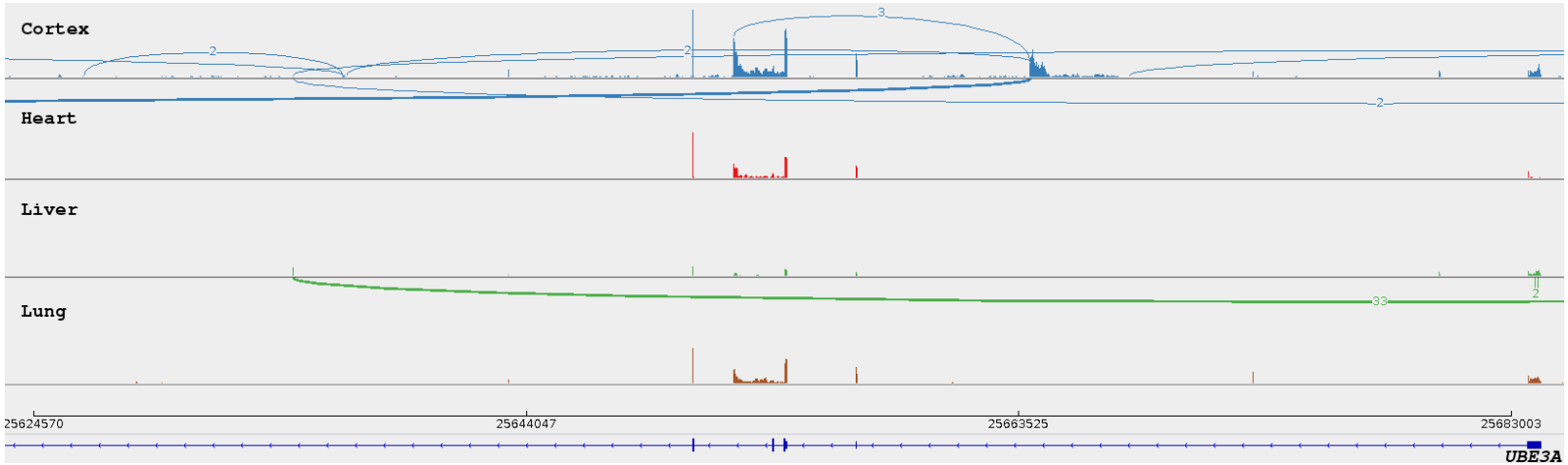


Figure B.2: Sashimi plots demonstrating tissue-specific splicing and alternative splicing in human at the 3' end of *UBE3A* (n = 3). Minimum junction coverage = 2. Data generated from Uhlen *et al.* (2015).

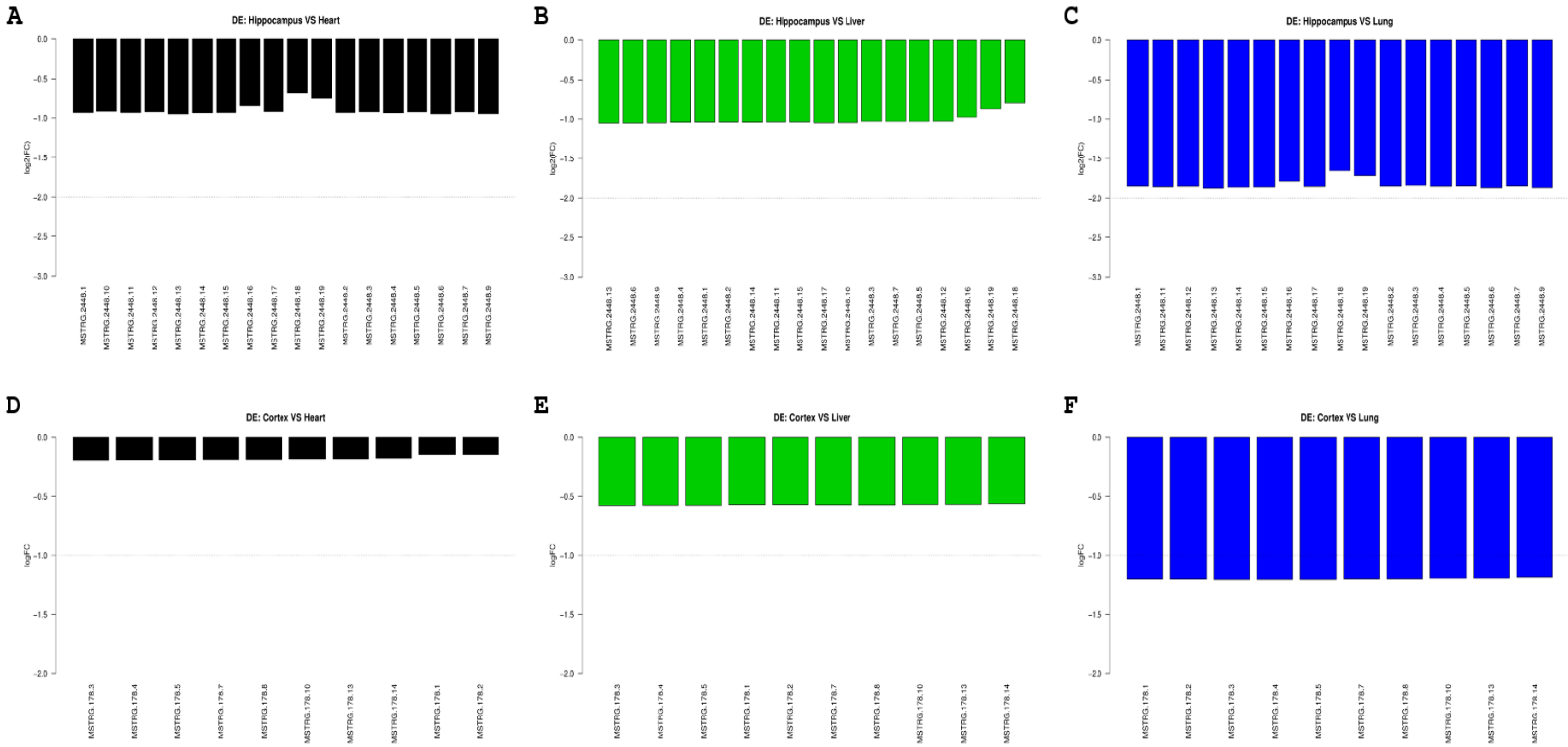


Figure B.3: The antisense transcripts are downregulated in non-brain tissues. Comparison of mouse hippocampus to **A.** heart, **B.** liver, and **C.** lung. Data generated from Sanger Institute hybrid mice. Comparison of human cortex to **D.** heart, **E.** liver, and **F.** lung. Data generated from Uhlen *et al.* (2015).

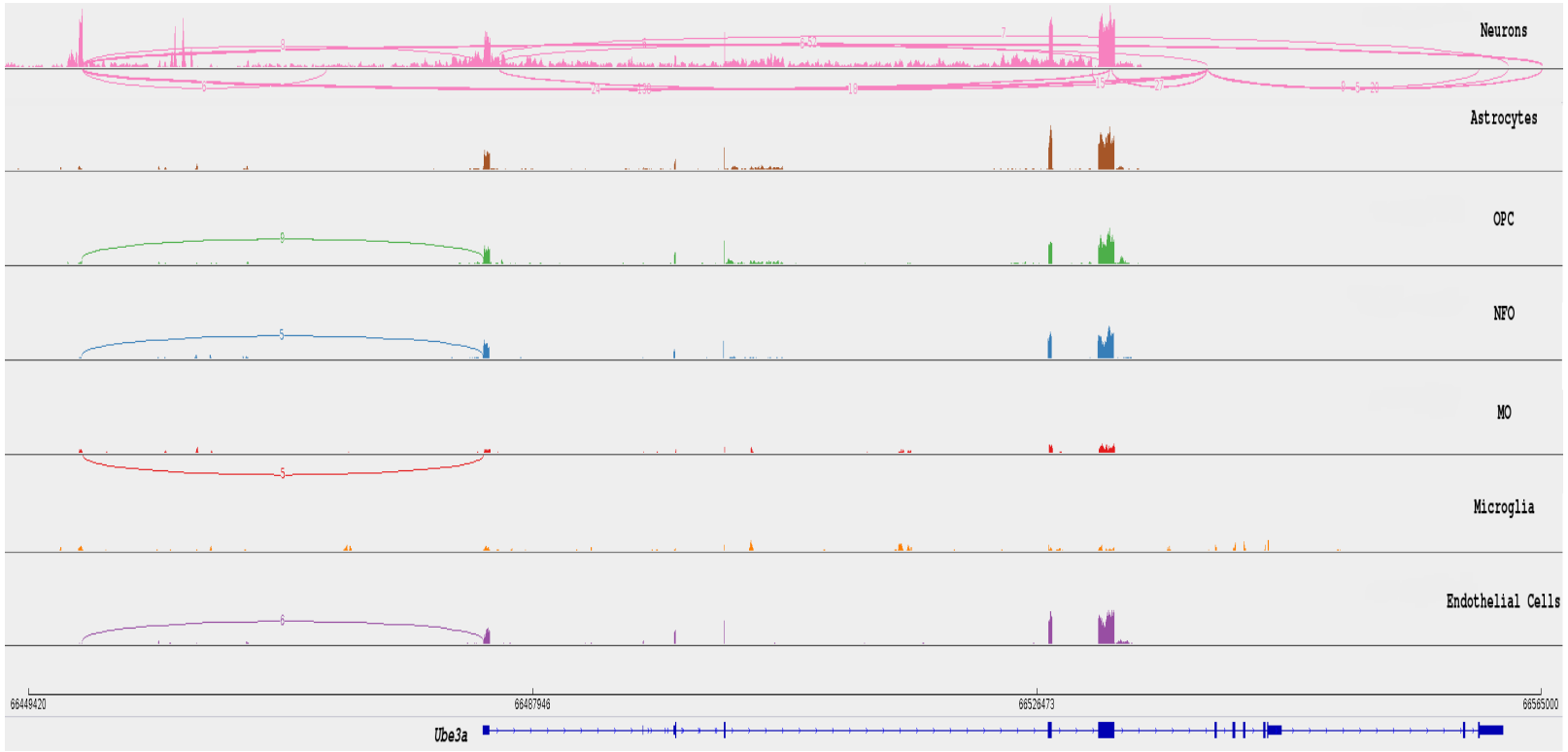


Figure B.4: Sashimi plots demonstrating splicing and alternative splicing events present in neurons, but not other cell-types of the cerebral cortex. Minimal junction coverage = 5. Data generated from Zhang *et al.* (2014). Abbreviations: OPC - oligodendrocytes precursor cells, NFO - newly formed oligodendrocytes, and MO - myelinating oligodendrocytes.

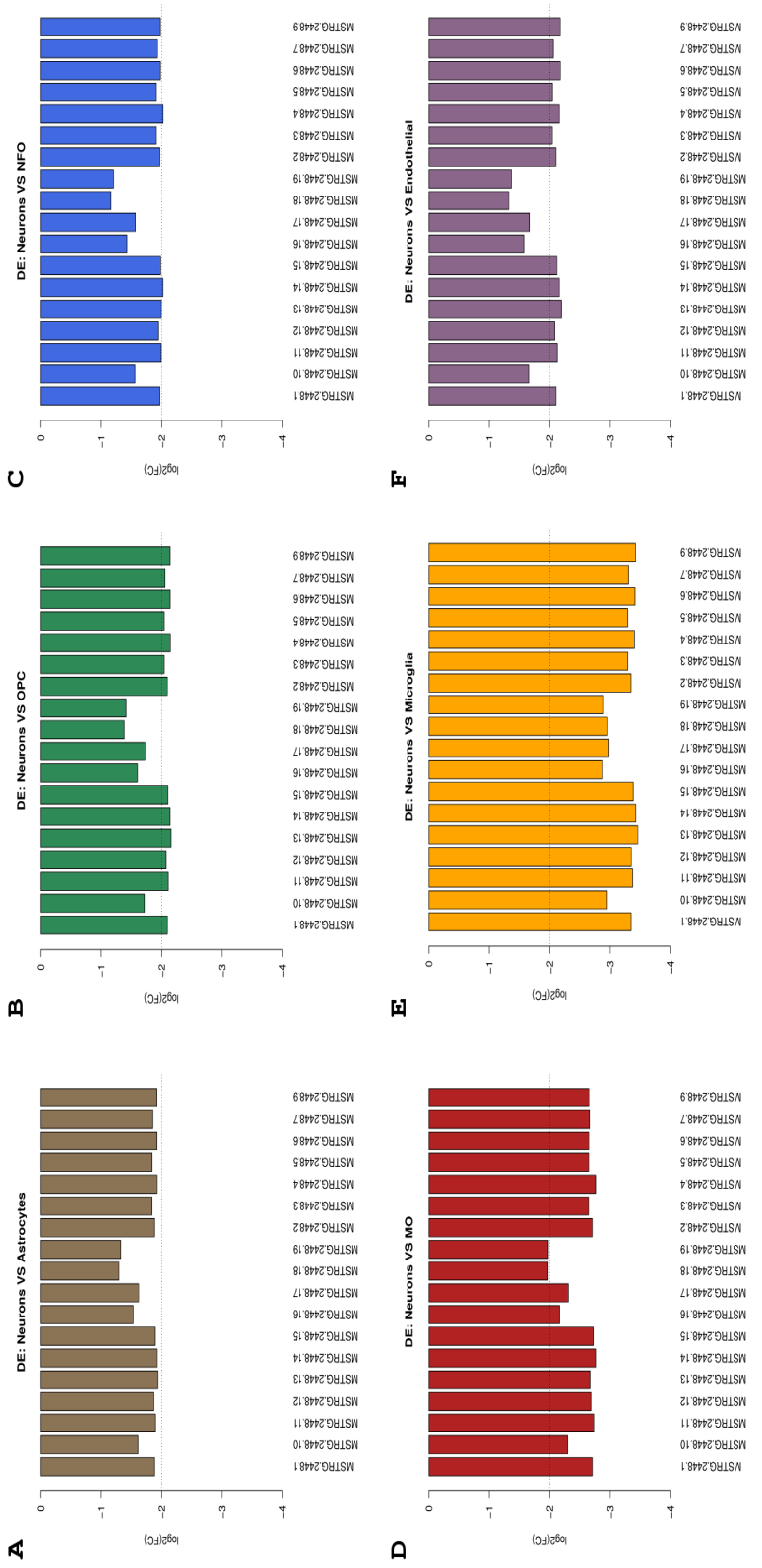


Figure B.5: *Ube3a-AS* is downregulated in non-neuronal cell-types. Comparison of neurons to **A.** astrocytes, **B.** OPC, **C.** NFO, **D.** MO, **E.** microglia, and **F.** endothelial cells. Data generated from Zhang *et al.* (2014). Abbreviations: OPC - oligodendrocytes precursor cells, NFO - newly formed oligodendrocytes, and MO - myelinating oligodendrocytes.

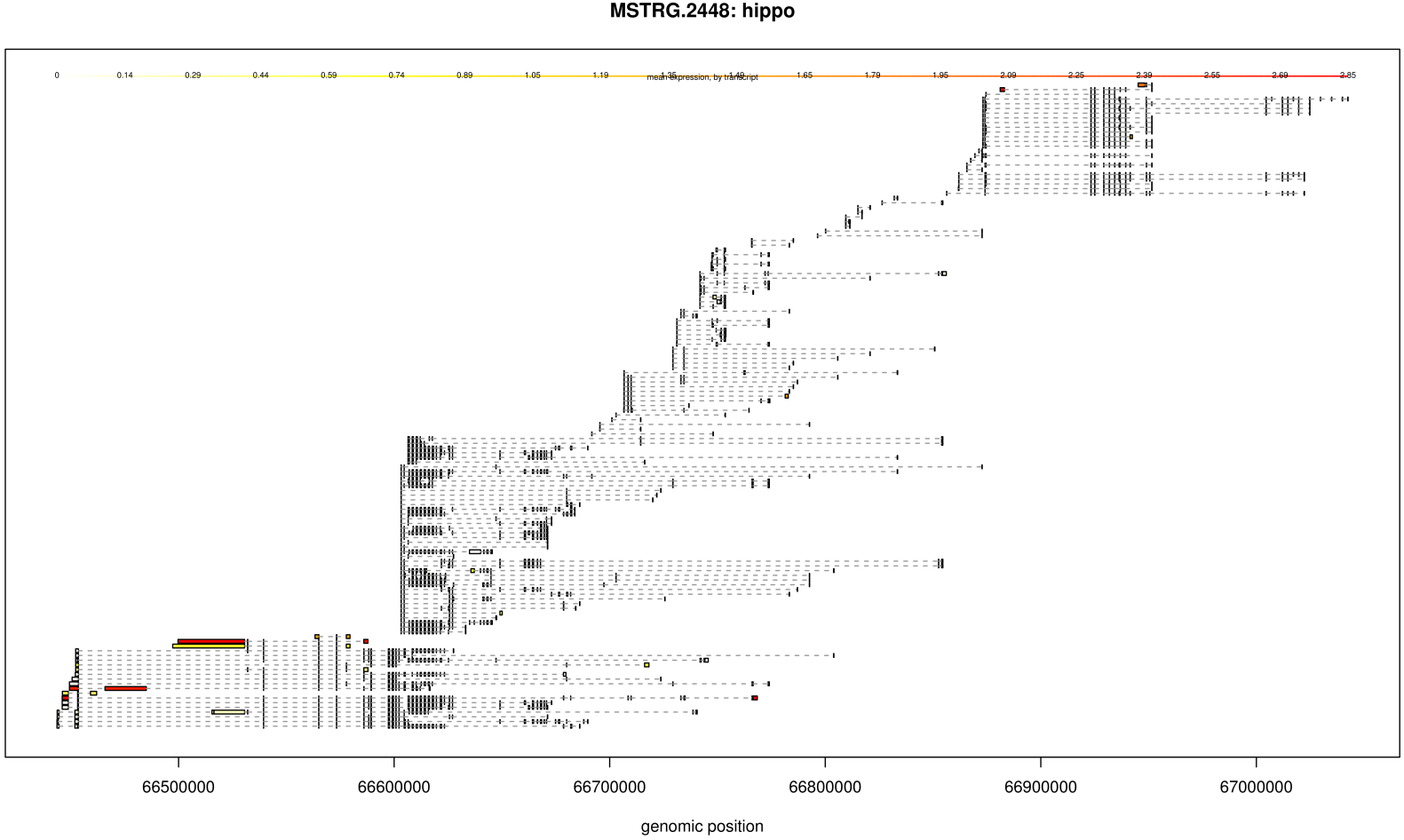


Figure B.6: Average transcript expression of the assemblied polycistronic transcriptional unit in mouse hippocampus. Lower left-hand is location of *Ube3a* antisense transcripts.

APPENDIX C

SUPPLEMENTAL DATA - CHAPTER THREE

C.1 Methods extended

Table C.1: Mouse tissue information by study

Study	Strain	Tissue	Accession	Read count (million)
ERP000591	C57xDBA	heart	ERR032227	31.0
			ERR032228	8.6
			ERR032229	30.1
			ERR032238	29.5
			ERR032230	27.8
			ERR032231	31.2
		hippocampus	ERR032232	18.8
			ERR032233	33.6
			ERR032234	23.8
			ERR032235	23.2
			ERR032236	22.6
			ERR032237	32.7
		liver	ERR032203	28.6
			ERR032204	30.0
			ERR032205	29.3
			ERR032206	30.8

Table C.3: *Continued...*

Study	Strain	Tissue	Accession	Read count (million)
			ERR032207	31.3
			ERR032208	31.4
		lung	ERR032221	11.7
			ERR032222	25.8
			ERR032223	23.3
			ERR032224	12.1
			ERR032225	28.4
			ERR032226	18.3
SRP012040	C57BL/6J	cerebellum	SRR567488	151
			SRR567489	145
		cortex	SRR567480	156
			SRR032481	166
		frontal lobe	SRR567478	186
			SRR567479	159
SRP017966	C57xCASTEi	vehicle	SRR649455	94.4
			SRR649456	97.4
			SRR649457	12.9
			SRR649458	10.6
			SRR649459	18.8
		topotecan	SRR649460	102
			SRR649461	113

Table C.3: *Continued...*

Study	Strain	Tissue	Accession	Read count (million)
			SRR649462	16.9
			SRR649463	8.8
			SRR649464	14.6
SRP033200	Aldh1l1-EGFP	astrocytes	SRR1033783	29.6
			SRR1033784	32.0
	NA	neurons	SRR1033785	37.9
			SRR1033786	33.9
		oligodendrocyte precursor cells	SRR1033787	32.2
			SRR1033788	32.5
		newly formed oligodendrocytes	SRR1033789	32.1
			SRR1033790	30.5
		myelinating oligodendrocytes	SRR1033791	33.4
			SRR1033792	29.7
		microglia	SRR1033793	29.2
			SRR1033794	30.0
	Tie2-EGFP	endothelial cells	SRR1033795	36.5
			SRR1033796	33.8

Table C.2: *Ube3a* Mechanism Primer List

Primer Name	**Sequence**	**Reference**
ActB Fwd	GGCTGTATTCCCCTCCATCG	[35]
ActB Rev	CCAGTTGGTAACAATGCCATGT	[35]
Map2 Fwd	GCCAGCCTCAGAACAAACAG	
Map2 Rev	AAGGTCTTGGGAGGGAAGAAC	
Ube3a-AS 1 Fwd	GGCTCTACGAGAAGCTGACTG	
Ube3a-AS 1 Rev	GTTGCCATCACCTTCAGTTC	
Ube3a-AS 3 Fwd	GCTACATGCTAGGCCCTAATG	
Ube3a-AS 3 Fwd	ATGGAGTTCTCTTGACCAAGTC	
Ube3a^{YFP} Fwd	GGTGACTAATGAATCGCCCTTA	
Ube3a^{YFP} Rev	GTTTACGTCGCCGTCCAG	
Iso4 3'RACE Fwd	CAAGGCTGACTTCAAACTCAGATA	
Iso4 3'RACE nxt[1]Fwd	TCTCCTGTTTCTGCTTTCTGAG	
Ube3a Exon 4 Fwd	ACCAGGAGAATCCCAGTCTGA	
Ube3a Exon 4.1 Rev	ATTTGATGCTGGTCATGGTG	
Ube3a Exon 5 Rev	TCATTCGTGCAGGCCTCATT	

C.2 Gene prediction

C.2.1 Splicing into Ube3a exon 4.1

[1]nxt: nested

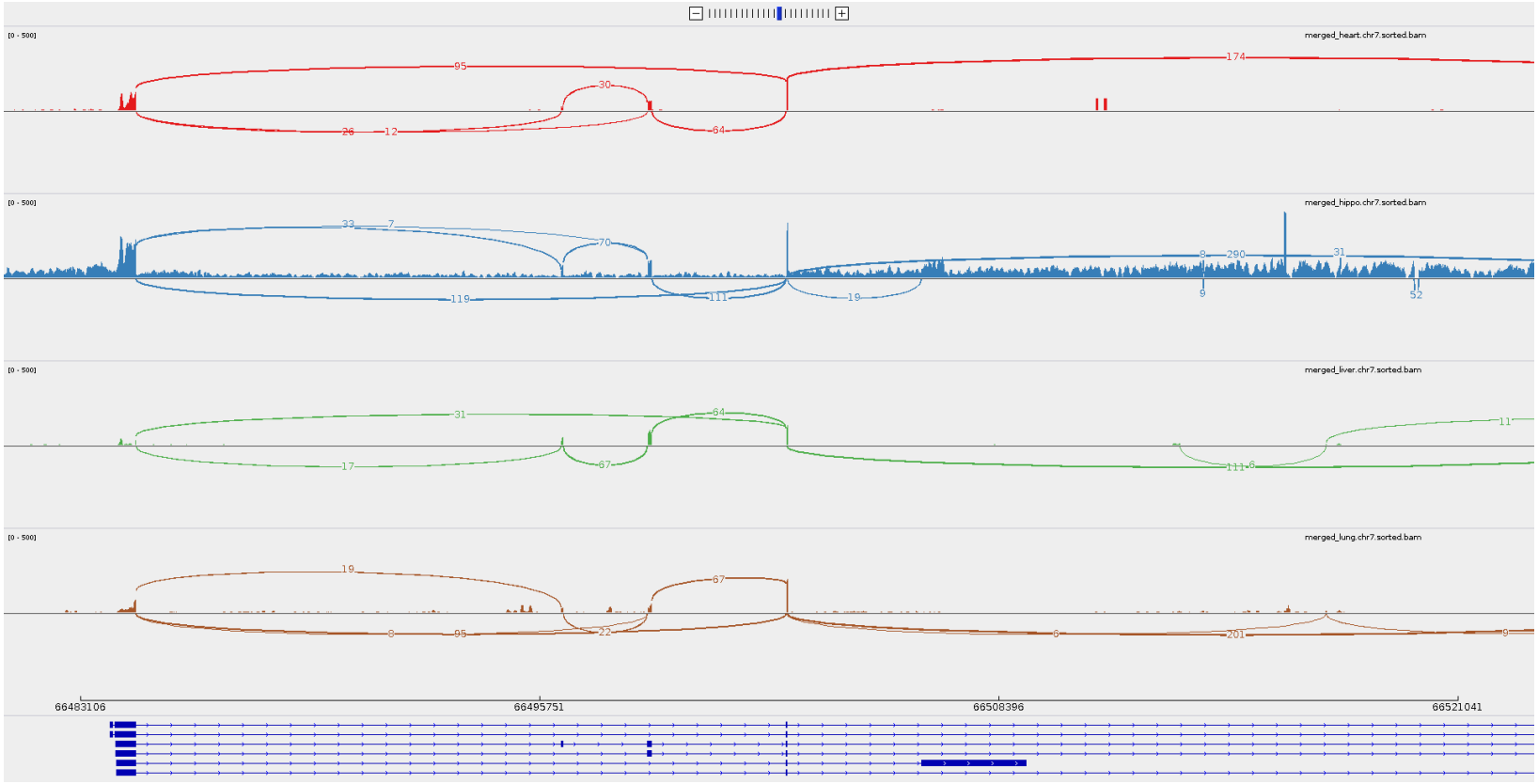


Figure C.1: Sashimi plot comparing heart (red), hippocampus (blue), liver (red), and lung (brown) forward strand demonstrating splicing into novel exon 4.1 only in the brain. Exon coverage of 0 to 500, minimal junction coverage = 5, max junction coverage = 1000.

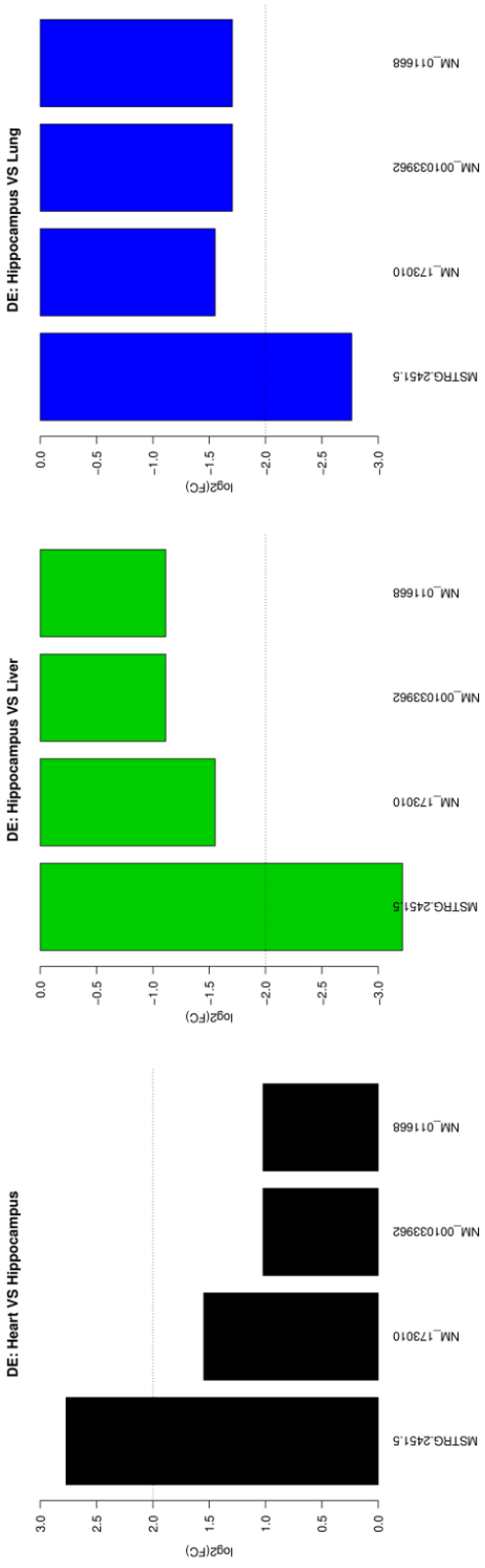


Figure C.2: Log2 fold-change plots of the four potential isoforms comparing heart, liver, and lung to hippocamus.

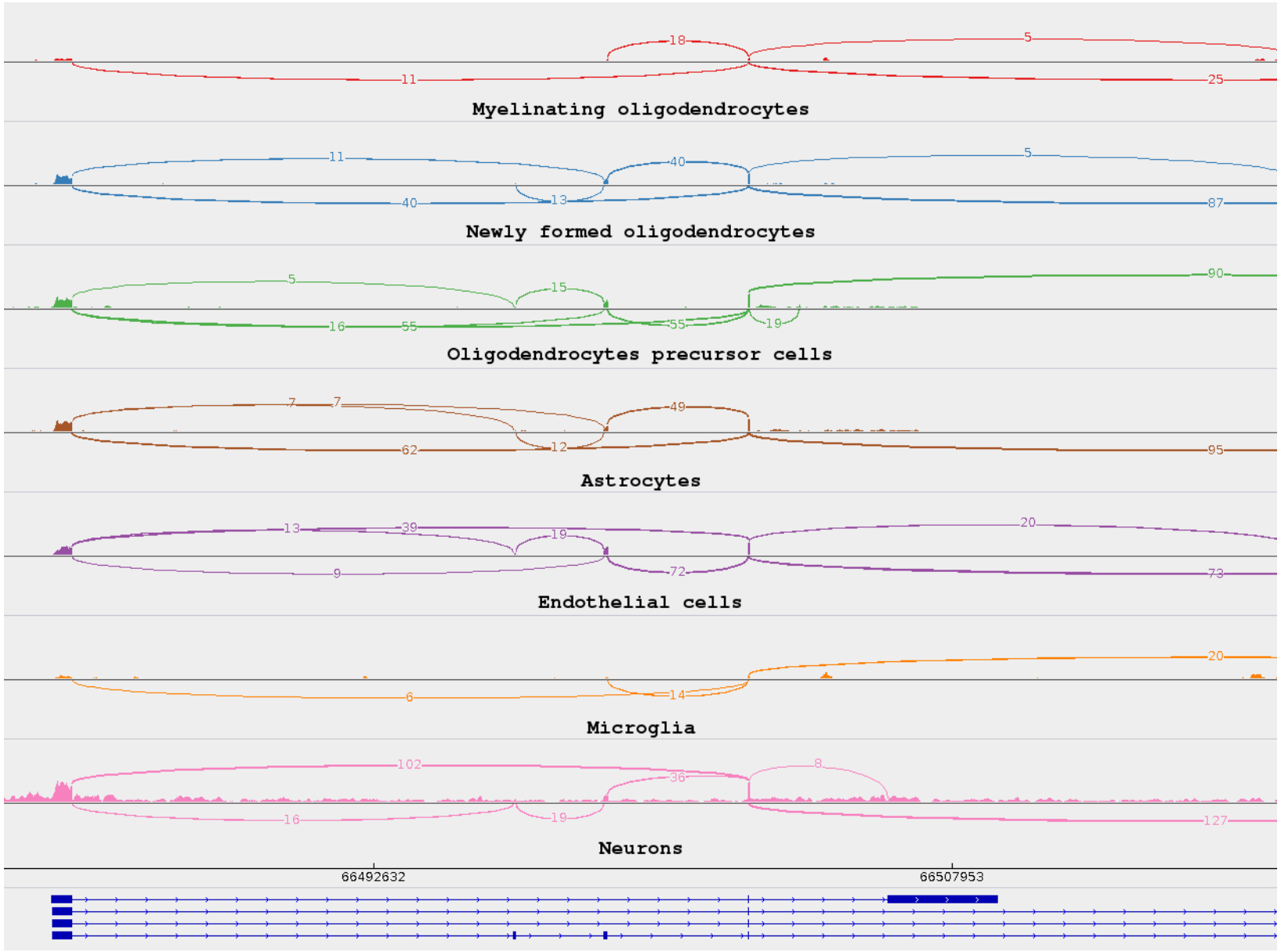


Figure C.3: Sashimi plot comparing different cell populations in the mouse cerebral cortex. Forward strand only, with exon coverage = 0 to 500, minimal junction coverage = 5, max junction coverage = 1000.

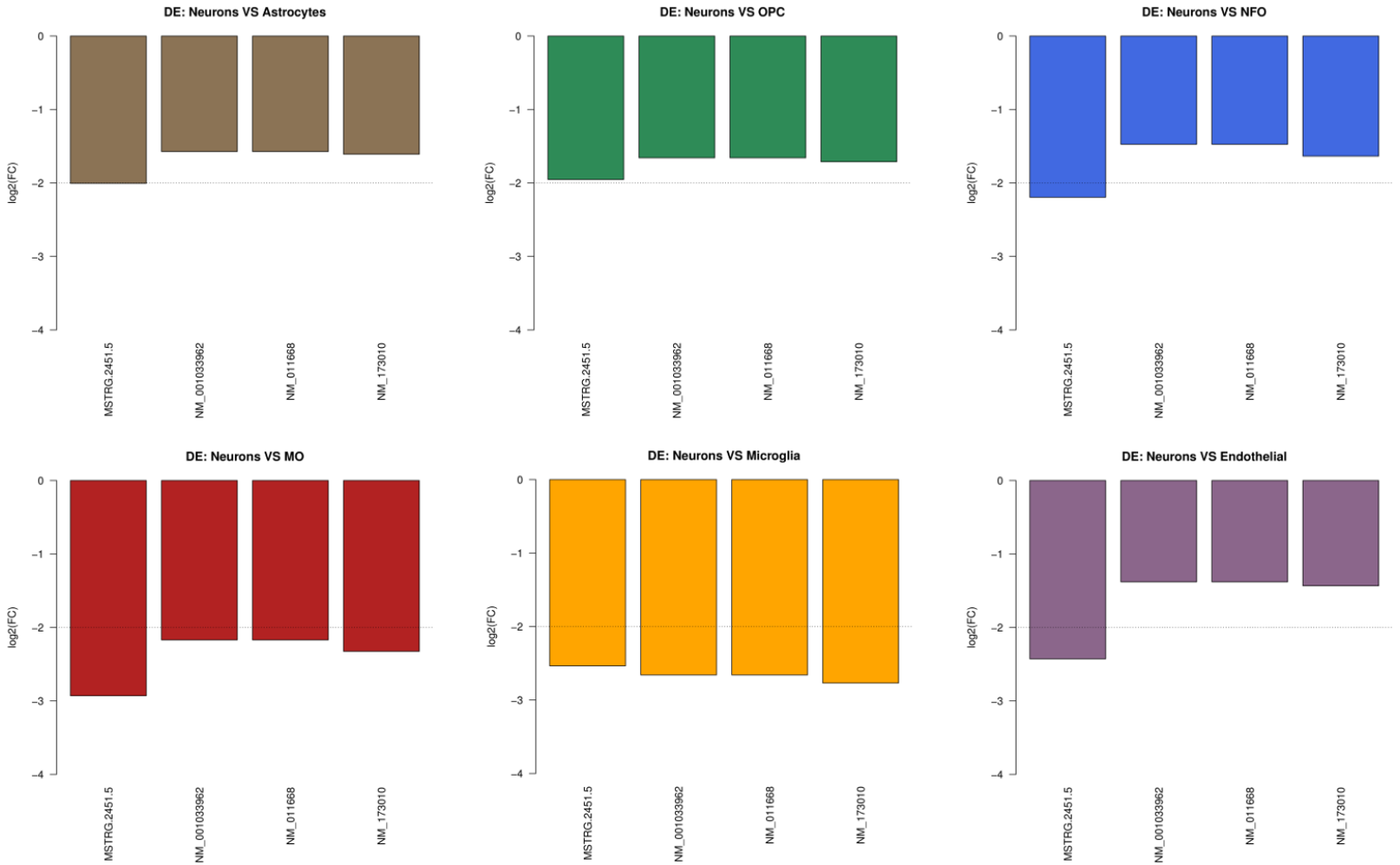


Figure C.4: Log2 fold-change plots of the four potential isoforms comparing other cell population in the mouse cerebral cortex to neurons.

C.3 Gene structure analysis

```
Sequence    5:   MSTRG.2451.5, from 1 to 3512.
************************************************************************
Query protein sequence    2 (File: NP_766598.1)
  1 MKRAAAKHLI ERYYHQLTEG CGNEACTNEF CASCPTFLRM DNNAAAIKAL ELYKINAKLC
 61 DPHPSKKGAS SAYLENSKGA SNNSEIKMNK KEGKDFKDVI YLTEEKVYEI YEFCRESEDY
121 SPLIRVIGRI FSSAEALVLS FRKVKQHTKE ELKSLQEKDE DKDEDEKEKA ACSAAAMEED
181 SEASSSRMGD SSQGDNNVQK LGPDDVTVDI DAIRRVYSSL LANEKLETAF LNALVYLSPN
241 VECDLTYHNV YTRDPNYLNL FIIVMENSNL HSPEYLEMAL PLFCKAMCKL PLEAQGKLIR
301 LWSKYSADQI RRMMETFQQL ITYKVISNEF NSRNLVNDDD AIVAASKCLK MVYYANVVGG
361 DVDTNHNEED DEEPIPESSE LTLQELLGDE RRNKKGPRVD PLETELGVKT LDCRKPLISF
421 EEFINEPLND VLEMDKDYTF FKVETENKFS FMTCPFILNA VTKNLGLYYD NRIRMYSERR
481 ITVLYSLVQG QQLNPYLRLK VRRDHIIDDA LVRLEMIAME NPADLKKQLY VEFEGEQGVD
541 EGGVSKEFFQ LVVEEIFNPD IGMFTYDEAT KLFWFNPSSF ETEGQFTLIG IVLGLAIYNN
601 CILDVHFPMV VYRKLMGKKG TFRDLGDSHP VLYQSLKDLL EYEGSVEDDM MITFQISQTD
661 LFGNPMMYDL KENGDKIPIT NENRKEFVNL YSDYILNKSV EKQFKAFRRG FHMVTNESPL
721 KYLFRPEEIE LLICGSRNLD FQALEETTEY DGGYTRESVV IR-

Predicted gene structure (within gDNA segment 1 to 3512):
Exon 1, 3: 64 (62 n); Protein 745, 762 (18 aa); score: 0.027

MATCH MSTRG.2451.5+ NP_766598.1 0.027 62 0.027 P
PGS_MSTRG.2451.5+_NP_766598.1 (3  64)

Alignment:
GAGCCCTCGC CCGGCAGGGT TGGCGCGCGC TGCCTGTCGG GATACTCGGT CCGCC-CACC 61
 E  P  S   P  G  R  V   G  A  R   C  L  S   G  Y  S  V   R     T
 |     +               |  .           .   .        +   |
 E  E  T   T  E  Y  D   G  G  Y   T  R  E   S  V  V  I   R  -  -  762
```

TAG 64

*

* 763

C.4 3'RACE sequences

>A01_4b-4.seq

TGCACGCAAGCTCGATAACCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTTCTCCTGTTTCTGCTTTCTGAGTGTGAGATTAAAGGTGTTGTACCACCATGACCA
GCATCAAATTCATGAGAAAAAAAAATTCTTTACCTTTCTAATCATCTAAGAAATGAAACAACAGTGAGATACTACTTCCTACCTGCTAGAATATATGAATGTTCATGTTGGTGAGGTTGTA
AGCAAACAGAAATCTTGTACACAGTTGTTTGGAATGTAAATTAGATCGATTATGGAAAATAACAGGTTCCTAAAAATTGCAATGACCTAGCAATTGTATTATAGAAATATATAGTGTGTT
ATAAGAGATACCTGCAGTTTAATTTTGATTGCAGCATTAATTAAACAAAATGCAGAAATAAACCAGTGTTTGTCAGTG

>B01_4b-5.seq

TGACGCCGCTCGCATAACCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTTCTCCTGTTTCTGCTTTCTGAGTGTGAGATTAAAGGTGTTGTACCACCATGACCA
GCATCAAATTCATGAGAAAAAAAAATTCTTTACCTTTCTAATCATCTAAGAAATGAAACAACAGTGAGATACTACTTCCTACCTGCTAGAATATATGAATGTTCATGTTGGTGAGGTTGT
AAGCAAACAGAAATCTTGTACACAGTTGTTTGGAATGTAAATTAGATCGATTATGGAAAATAACAGGTTCCTAAAAATTGCAATGACCTAGCAATTGTATTATAGAAATATATAGTGTG
TTATAAGAGATACCTGCAGTTTAATTTTGATTGCAGCATTAATTAAACAAAATGCAGAAATAAACCAGTGTTTGTCAGTGGCTAAACAAATAAAAATTGACATATATGATGATGTAATT
TTTCGCCATTAAAATTTTTTTCTGAAATCTTATGTCGGTGACAACATGGATGGAAATGATCCTTTATATTTTCAAAGAAATAAATCAGCACAAAAACAGGTAGTGCATCATCTTCCTGT
ACATAGACTGTAAAAACATTCTAAAAATTGTTTGTTACTTAGAGTAGAATAAATGATAGACATAAAAAATGCAAAATGAATAGAATGTAGTTCAGGACCAAAATTGTAGAGCATGGTGA
CTTCTTTCTGTATATTTTCTTGAATATCACTAAAAGTCAACTATTAAGTATATGTGCAAAAAGTGCTATGTTAGAATTTATAAACTCATTTGATGTATCACTTTTGCATATGATAATAT
ATGTAGTTCTTATTTGTCAATGAAAATAAAAACTTTGGGAAAAACAAAAAAAAAAAAAAAAAAAAAGATTAAGAATAAAAAAAAAAAAAAAAAAAGTCTAGTCGACGCGTGGCACAAAGGGC
GAATTTCGCGGCCGCTAAATTTCAATTCGCCCTATAGTGAGTCGTATAACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGAAAAACCTGCCGTTACCAACTTAATCGCCTTGAC
AGCACATCCCCCATTACGCAAC

>C01_4b-6.seq

TCCACGCCCGCCTTAGATAACCCCTCACTAAAGGGACTAGTTCCTGCAGGTTTAAACCGAATTCGCCCTTTCTCCTGTTTCTGCTTTCTGAGTGTGAGATTAAAGGTGTTGTACCACCA
TGACCAGCATCAAATTCATGAGAAAAAAAATTCTTTACCTTTCTAATCATCTAAGAAATGAAACAACAGTGAGATACTACTTCCTACCTGCTAGAATATATGAATGTTCATGTTGGTGA
GGTTGTAAGCAAACAGAAATCTTGTACACAGTTGTTTGGAATGTAAATTAGATCGATTATGGAAAATAACAGGTTCCTAAAAATTGCAATGACCTATCAATTGTATTATCTAGAAATAT
ATACTGTGT

>F02_4b-1.seq

TGAGCCGCTCGTATTAACCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTTTCCAAAGTTTTTATT
TTCATTGACAAATAAGAACTACATATATTATCATATGCAAAAGTGATACATCAAATGAGTTTATAAATTCTAACATAGCACTTTTGCACATATACTTAATAGTTGACTTTTAGTGATAT
TCAAGAAAATATACAGAAAGAAGTCACCATGCTCTACAATTTTGGTCCTGAACTACATTCTATTCATTTTGCATTTTTTATGTCTATCATTTATTCTACTCTAAGTAACAAACAATTTT
TAGAATGTTTTTACAGTCTATGTACAGGAAGATGATGCACTACCTGTTTTTGTGCTGATTTATTTCATTGAAAATATAAAGGATCATTTCCATCCATGTTGTCACCGACATAAGATTTC
AGAAAAAAATTTTAATGGCGAAAAATTACATCATCATATATGTCAATTTTTATTTGTTTAGCCACTGACAAACACTGGTTTATTTCTGCATTTTGTTTAATTAATGCTGCAATCAAAAT
TAAACTGCAGGTATCTCTTATAACACACTATATATTTCTATAATACAATTGCTAGGTCATTGCAATTTTTAGGAACCTGTTATTTTCCATAATCGATCTAATTTACATTCCAAACAACT
GTGTACAAGACTTCTGTTTGCTTACAACCTCACCAACATGAACATTCATATATTCTAGCAGGTAGGAAGTAGTATCTCACTGTTGTTTCATTTCTTAGATGATTAGAAAGGTAAAGAAT
TTTTTTTCTCATGAATTTGATGCTGGTCATGGTGGTACAACACCTTTAATCTCACACTCAGAAAGCAGAAACAGGAGAAAGGGCGAATTCGCGGCCCGCTAAATTCAATTCGCCCTATA
GTGAGTCGTATTACAATTTCACTGGCCCGTCGTTTTACACGTCGTGACTGGAAACCCTGCGTACCACTATTCGCTGCAGCACATCCCCATTCGCAGCTGCGTATAGCGAGAGCCCGCAC
GATCGCCTCACAGGTTGCCAGCCTATACGTACGGCAGTAGTTACCTTAGAGAAGCCGTTATCGACTGTTTATGA

>G02_4b-2.seq

TGAGCCGCTCTATTAACCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTTCTCCTGTTTCTGCTTTCTGAGTGTGAGATTAAAGGTGTTGTACCACCATGACCAG
CATCAAATTCATGAGAAAAAAAATTCTTTACCTTTCTAATCATCTAAGAAATGAAACAACAGTGAGATACTACTTCCTACCTGCTAGAATATATGAATGTTCATGTTGGTGAGGTTGTA
AGCAAACAGAAATCTTGTACACAGTTGTTTGGAATGTAAATTAGATCGATTATGGAAAATAACAGGTTCCTAAAAATTGCAATGACCTAGCAATTGTATTATAGAAATATATAGTGTGT
TATAAGAGATACCTGCAGTTTAATTTTGATTGCAGCATTAATTAAACAAAATGCAGAAATAAACCAGTGTTTGTCAGTGGCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAT
AAAAAAAAAAAAAAAAAAAAAGGTACTAGTCGACCCGTGGCCAAGGGCAAATTCGCGGCCGCTAAATTCAATTCCCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACAA
CGTCGGGACTGGGAAAACCCTGGCGTTCCCAAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAAAGGCCCGCACCGTTCGCCCCTTCCCAACATTTGC

GCAGCCTTTTCGTACGGGAGTTTAAGGGTTTACACCCTATAAAAAAAGAGAGCCGTTTTCTTTTTGTTTGGGGATGTACAGAATTGATTTTTTTTGACCCCCCGGGGCGACGGGATGGG
GGAATCCCCCTGGCCAGTGCCCGTTTGCTGTCAAATAAAGTTCTCCCGTGAACTTTACCCCGGGGGGCATATCCGGGATTGAAAGCTGGCCCATTATGTCCCCCCATTTGGCCATGGTG
CCGGTCTCCCTTTATCGGGGAAGAAAATGGCTGATCTCAGCCCACCGCGAAATGACATTCAAACGCCTTAACCTGATGTTTCTGGGGAAATATTA

>H02_4b-3.seq

TGCAGCCGCTCGTATTAACCCCTCACTAAAGGGACTAGTCCTGCAGGTTTAAACGAATTCGCCCTTTCTCCTGTTTCTGCTTTCTGAGTGTGAGATTAAAGGTGTTGTACCACCATGAC
CAGCATCAAATTCATGAGAAAAAAAATTCTTTACCTTTCTAATCATCTAAGAAATGAAACAACAGTGAGATACTACTTCCTACCTGCTAGAATATATGAATGTTCATGCTGGTGAGGTT
GTAAGCAAACAGAAATCTTGTACACAGTTGTTTGGAATGTAAATTAGATCGATTATGGAAAATAACAGGTTCCTAAAAATTGCAATGACCTAGCAATTGTATTATAGAAATATATAGTG
TGCTATAAGAGATACCTGCAGTTTAATTTTGATTGCAGCATTAATTAAACAAAATGCAGAAATAAACCAGTGTTTGTCAGTGGCTAAACAAATAAAAATTGACATATATGATGATGTAA
TTTTTCGCCATTAAAATTTTTTTCTGAAATCTTATGTCGGTGACAACATGGATGGAAATGATCCTTTATATTTTCAATGAAATAAAATCAGCACAAAAAACAGGTAGTGCATCATCTTC
CTGTACATAGACTGTAAAAAACATTCTAAAAAATTGTTTGTTACTTAGAGTAGAATAAATGATAGACATAAAAATGCAAATGAAAT

C.5 SNP analysis

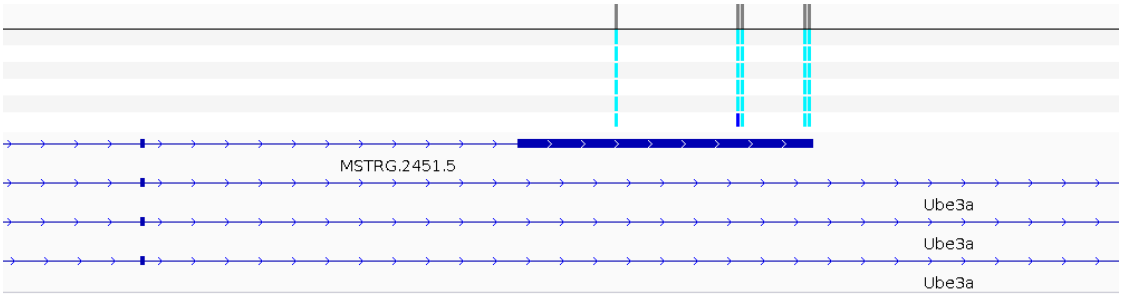


Figure C.5: Five informative snps located within exon 4.1. Paternal allele specific expression.

C.6 Temporal regulation of isoform 4

Table C.3: Mouse tissue information by study

Study	Strain	Tissue	Accession	Read count (million)
SRP048593	C57BL/6J	E18-hippo	SRR1772425	41.3
			SRR1772429	34.9
		P1-hippo	SRR1772426	38.5
			SRR1772430	34.6
		P10-hippo	SRR1772427	34.1
			SRR1772431	41.1
		P30-hippo	SRR1772428	43.4
			SRR1772432	41.8

Temporal hippocampal RNA-seq datasets were extracted from E18, P1, P10 and P30 mice [141]. RNA-seq data analyzed with Illumina HiSeq 2500 platform as single-end, stranded reads.

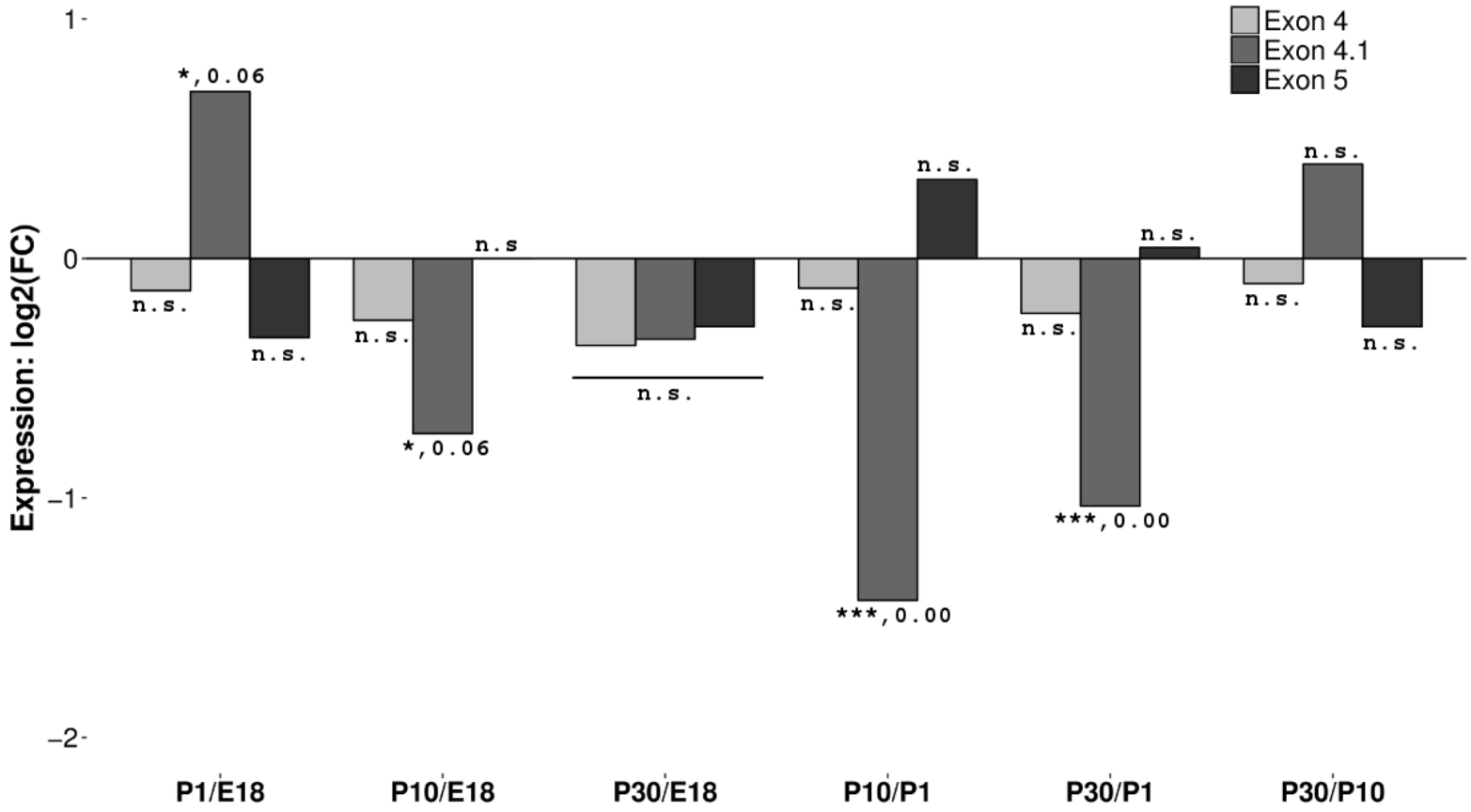


Figure C.6: Neuron-specific *Ube3a* isoform 4 is temporally regulated.

APPENDIX D

SUPPLEMENTAL DATA - CHAPTER FOUR

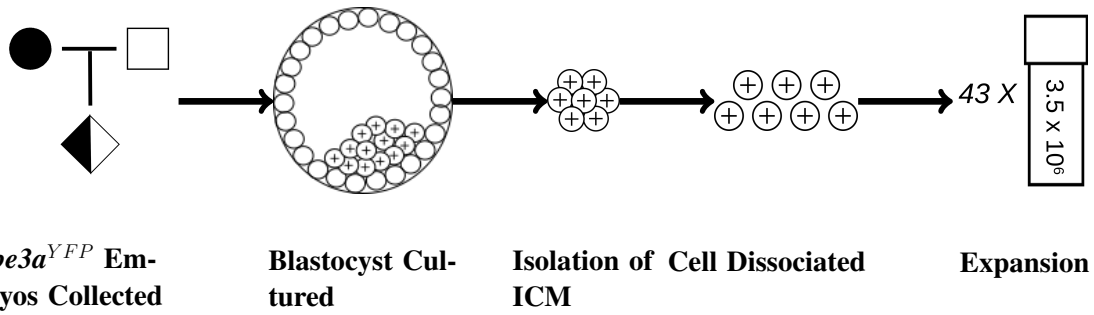


Figure D.1: Example schematic of *Ube3a*^{YFP} embryonic stem cell generation and expansion.

Table D.1: Composition of Medium

Medium	**Components**	**Company**	**Cat. #**	**Notes**
ES Medium	KnockOut DMEM	Life Technologies	10829018	N.A.
	FBS		1600044	10%
	GlutaMAX		35050061	2 mM
	Non-essential amino acids		111400050	1%
	Penicillin / streptomycin		15140122	1%
	β-mercapto-ethanol	Sigma-Aldrich	M6250-100ML	0.1 mM
CA Medium	DMEM	Life Technologies	11995073	N.A.
	KnockOut Serum Replacement		10828028	15%
	GlutaMAX		35050061	2 mM
	Non-essential amino acids		111400050	1%
	Penicillin / streptomycin		15140122	1%
	β-mercapto-ethanol	Sigma-Aldrich	M6250-100ML	0.1 mM

Table D.1: *Continued...*

Medium	**Components**	**Company**	**Cat. #**	**Notes**
	Retinoic Acid		R2625-50MG	0.5 mM
N2 Medium	DMEM/F-12	Life Technologies	11330057	1:1
	Neurobasal Medium		21103049	1:1
	GlutaMAX		35050061	2 mM
	B27		17504044	1%
	BSA	Sigma-Aldrich	A7906-100G	50 μg/ml
	Progesterone		P8783-1G	20 nM
	Putrescence		P5780-5G	100 nM
	ITS Supplement	Roche Life Science	11074547001	1%
Complete Medium	Advanced DMEM/F-12	Life Technologies	12634028	1:1
	Neurobasal Medium		21103049	1:1
	GlutaMAX		35050061	2 mM
	B27		17504044	1%
	BDNF		PHC7074	50 ng/ml

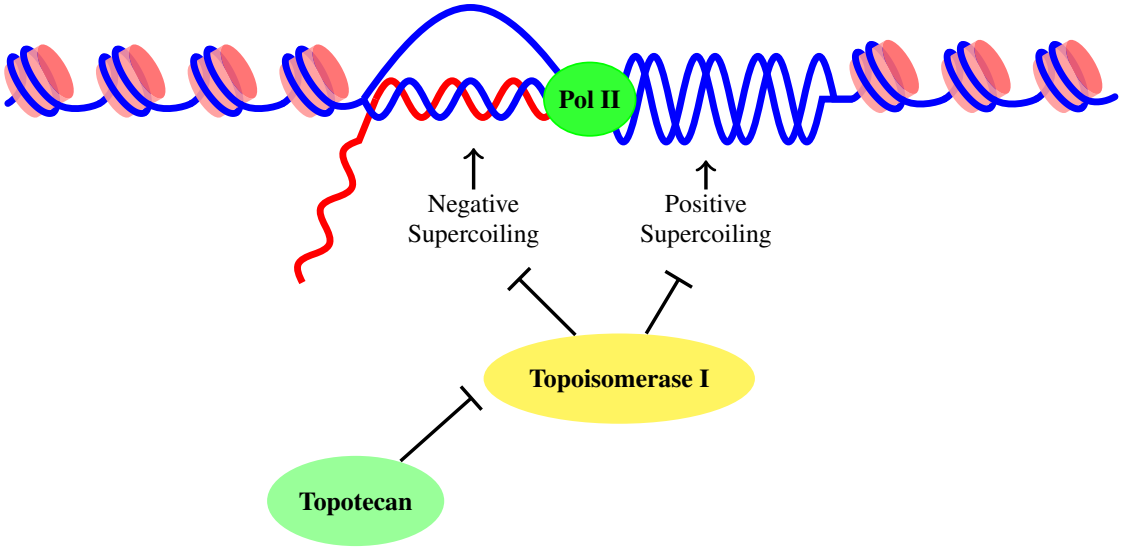


Figure D.2: A model of Topotecan inhibition of topoisomerase I actions.

Table D.2: Shapiro-Wilk Results

Plate	**Well**	**Treatment**	**PValue**	***N***
Plate 1	D-24	Vehicle	5.68e-15	289
	P-24		7.72e-12	111
	G-1		1.28e-07	65
	P-1		9.10e-11	138
	H-24		1.22e-09	131
	H-23	Topotecan	1.98e-07	239
	D-2		0.001	122
	K-2		0.442	17
	I-2		0.008	291
	O-23		1.09e-13	343

Table D.2: *Continued...*

Plate	**Well**	**Treatment**	**PValue**	***N***
Plate 2	G-24	Vehicle	4.16e-09	30
	P-1		9.79e-18	183
	A-24		5.67e-26	250
	G-1		0.442	23
	M-24		6.39e-14	243
	O-2	Topotecan	3.58e-04	249
	E-2		7.32e-06	314
	J-23		1.13e-06	209
	I-2		0.033	311
	D-2		8.02e-06	372

Table D.2: *Continued...*

Plate	**Well**	**Treatment**	**PValue**	***N***
Plate 3	I-24	Vehicle	0.572	9
	J-1		6.55e-13	158
	M-1		7.13e-14	186
	O-1		3.96e-12	70
	B-24		2.08e-07	50
	L-23	Topotecan	NA	2
	K-2		0.588	173
	E-23		0.014	132
	O-2		6.38e-07	228
	M-2		3.48e-05	233
Plate 4	C-24	Vehicle	2.01e-13	228
	A-1		6.75e-08	90
	G-24		1.01e-10	197
	M-1		9.33e-09	73
	H-1		1.90e-15	201
	A-23	Topotecan	2.35e-06	243
	K-2		0.116	202
	I-23		3.13e-07	135
	D-23		3.24e-05	224
	B-2		6.32e-07	227

Table D.2: *Continued...*

Plate	**Well**	**Treatment**	**PValue**	***N***
Plate 5	H-1	Vehicle	8.16e-17	154
	C-24		1.60e-08	130
	A-24		0.628	5
	L-24		0.002	62
	B-1		1.06e-14	221
	D-23	Topotecan	0.027	153
	P-2		0.136	111
	E-2		5.46e-05	190
	K-2		0.870	13
	C-2		0.002	161
Plate 6	H-24	Vehicle	3.01e-13	124
	J-24		0.040	21
	M-1		2.30e-12	94
	L-24		8.61e-11	111
	B-1		1.04e-07	36
	L-2	Topotecan	1.02e-07	148
	O-23		0.617	7
	H-2		1.46e-21	98
	M-23		7.57e-07	136
	M-2		7.22e-06	136

Table D.2: *Continued...*

Plate	**Well**	**Treatment**	**PValue**	***N***
Plate 7	J-1	Vehicle	4.78e-08	101
	J-24		8.82e-06	37
	D-24		1.59e-15	87
	F-1		4.72e-22	100
	P-1		3.53e-06	81
	M-2	Topotecan	0.005	184
	P-23		8.24e-06	116
	C-23		0.002	122
	H-2		0.018	115
	E-23		0.051	146
Plate 8	P-24	Vehicle	0.0001	40
	C-1		1.85e-10	203
	N-1		2.38e-09	129
	A-24		9.40e-07	82
	E-24		1.28e-14	119
	J-2	Topotecan	0.058	149
	K-23		0.005	158
	A-2		0.001	61
	E-2		0.295	171
	G-23		NA	2

Table D.2: *Continued...*

Plate	**Well**	**Treatment**	**PValue**	***N***
Plate 9	F-1	Vehicle	0.181	3
	B-24		0.490	4
	A-1		0.053	3
	N-24		9.29e-06	18
	F-24		8.47e-09	115
	K-23	Topotecan	0.086	76
	P-2		0.441	59
	L-23		0.003	58
	H-2		0.007	97
	F-2		0.001	68

```
require(tikzDevice)
tikz('time-course2.tex', standAlone = TRUE, width = 6, height = 5)

data <- read.table("YFP-timecourse-data.txt")

dpd2  <- subset(data, DPD == 2)[,2]
dpd4  <- subset(data, DPD == 4)[,2]
dpd6  <- subset(data, DPD == 6)[,2]
dpd8  <- subset(data, DPD == 8)[,2]
dpd12 <- subset(data, DPD == 12)[,2]
dpd13 <- subset(data, DPD == 13)[,2]

mydata  <- data.frame(DPD.02 = dpd2, DPD.04 = dpd4, DPD.06 = dpd6, DPD.08 = dpd8, DPD.12 = dpd12, DPD.13
= dpd13)
mydata2 <- stack(mydata)

xcoord  <- rep(0, length(mydata2$ind))
xcoord[mydata2$ind=="DPD.02"]<- 1
xcoord[mydata2$ind=="DPD.04"]<- 2
xcoord[mydata2$ind=="DPD.06"]<- 3
xcoord[mydata2$ind=="DPD.08"]<- 4
xcoord[mydata2$ind=="DPD.12"]<- 5
xcoord[mydata2$ind=="DPD.13"]<- 6

boxplot(mydata, names = c("2","4","6","8","12","13"), par(cex.axis=1.5,mar=c(4,4.5,0.4,0.4)))
par(new=T)
plot(xcoord, mydata2$values, xlim=c(0.5,6.5), axes=F, ylab="YFP Intensity", xlab="DPD",cex.lab=1.5)

dev.off()
tools::texi2dvi('time-course2.tex',pdf=T)
```

Figure D.3: Sample script for generating boxplots for embryonic stem cell-derived neurons time course

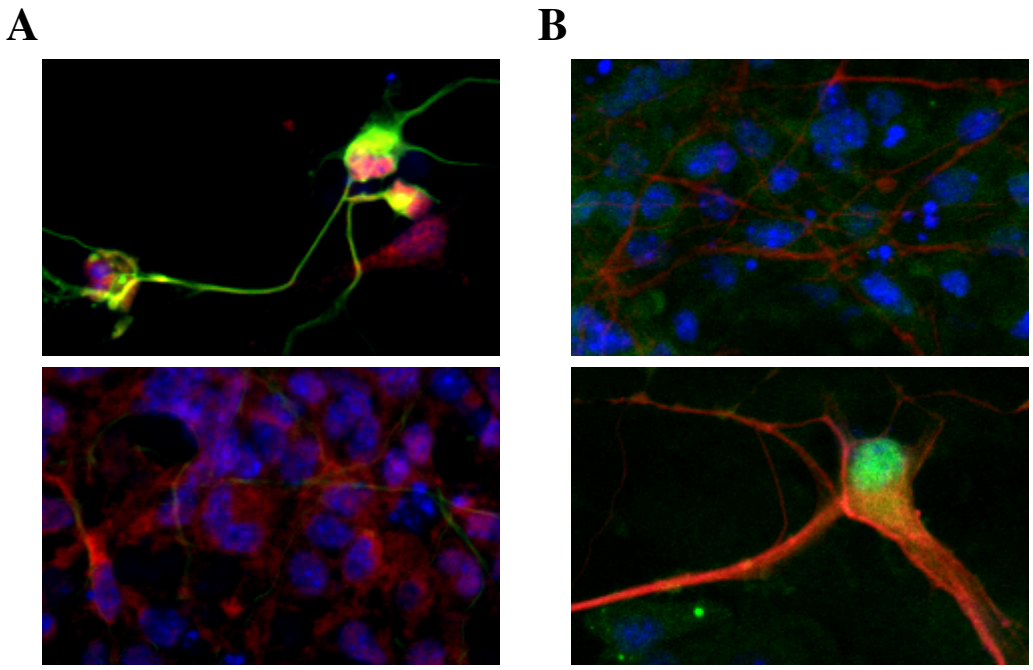


Figure D.4: Topotecan induces reactivation of paternal *Ube3a* allele in ES cell-derived neurons. **A)** Confocal image (40X magnification) of *Ube3a*^{YFP} ES cell-derived neurons at 2 and 13 days post dissociation (DPD) demonstrating the imprinting of paternal *Ube3a*. Nuclei marker TO-PRO-3 (blue), GFP (red), and βIII Tub (green). **B)** Confocal image (40X magnification) of ES cell-derived neurons at 13 DPD with vehicle (water) or Topotecan (300 nM) treatment demonstrating the reactivation of paternal *Ube3a*. Nuclei marker TO-PRO-3 (blue), GFP (green), and βIII Tub (red).

```
## Input files from a directory
NeuN   <- "raw/NeuN-Overlap"; Nuclei <- "raw/NucleiOverlap"
UpperQ <- "raw/UpperQ"; GFP    <- "raw/GFPonly"

## Output directory
output_dir <- "analysis/NeuN/"

## List files using full file name
files_neun   <- list.files(NeuN, full.names = T); files_nuclei <- list.files(Nuclei, full.names = T)
files_upperq <- list.files(UpperQ, full.names = T); files_gfp    <- list.files(GFP, full.names = T)

stat1 <- vector(); stat2 <- vector(); stat3 <- vector()

## Define functions in script
upper <- function(data){
  data  <- sort(data, na.last = NA)
  h1    <- ((length(data)-1)*0.99)+1
  upper <- data[floor(h1)]+((h1-floor(h1))*(data[floor(h1)+1]- data[floor(h1)]))
  return(upper)
}
lower <- function(data){
  data  <- sort(data, na.last = NA)
  h2    <- ((length(data)-1)*0.75)+1
  lower <- data[floor(h2)]+((h2-floor(h2))*(data[floor(h2)+1]- data[floor(h2)]))
  return(lower)
}
zfactor <- function(data1,data2){
  z <- 1 - ((3*(sd(data1, na.rm = T) + sd(data2, na.rm = T)))/(abs(mean(data1, na.rm = T) - mean(data2,
na.rm = T))))
  return(z)
}
ssmd <- function(data1,data2){
  beta <- (median(data1, na.rm = T) - median(data2, na.rm = T))/(sqrt(mad(data1, constant = 1, na.rm = T)
+ mad(data2, constant = 1, na.rm = T)))
  return(beta)
}

## Beginning of script
for (i in seq_along(files_neun)){
  file <- read.csv(files_neun[i])
  colnames(file)[2] <- "Well.R"
  txt  <- subset(file, Well.R == "2" | Well.R == "23")[,3]
  ntxt <- subset(file, Well.R == "1" | Well.R == "24")[,3]
  nam  <- paste("Plate", i, "_NeuN", sep="")
  data <- list(ntxt, txt)
  for (j in 1:2){
    treatment <- c("ntxt","txt")
    output1   <- mean(data[[j]], na.rm = T)
    output2   <- sd(data[[j]], na.rm = T)
    output3   <- lower(data[[j]])
    output4   <- median(data[[j]], na.rm = T)
    output5   <- upper(data[[j]])
    plate     <- paste(nam, treatment[j], sep="_")
    analysis  <- data.frame(Mean = output1, Std = output2, Lower = output3, Median = output4, Upper =
output5)
    write.table(analysis, file = paste(output_dir, plate, ".csv", sep = ""), sep = ",", append = F,
row.names = F, col.names = T)
  }
  ## Statistical analysis
  stat1 <- c(stat1, t.test(txt,ntxt)$p.value); stat2 <- c(stat2, zfactor(txt,ntxt)); stat3 <- c(stat3,
ssmd(txt,ntxt))
}

pvalues <- data.frame(Ttest = stat1, Zfactor = stat2, SSMD = stat3)
write.table(pvalues, file = paste(output_dir, "pvalues_NeuN_analysis.csv", sep = ""), sep = ",", append =
F, row.names = F, col.names = T)
```

Figure D.5: Sample script for plate analysis

```
## Input files from a directory
NeuN   <- "raw/NeuN-Overlap"
Nuclei <- "raw/NucleiOverlap"
UpperQ <- "raw/UpperQ"
GFP    <- "raw/GFPonly"

## Output directory
output_dir <- "analysis/NeuN/"

## List files using full file name
files_neun   <- list.files(NeuN, full.names = T)
files_nuclei <- list.files(Nuclei, full.names = T)
files_upperq <- list.files(UpperQ, full.names = T)
files_gfp    <- list.files(GFP, full.names = T)

## Plate analysis will use only NeuN-Overlap data
stat.ntxt <- vector()
stat.txt  <- vector()

## Beginning of script
for (i in seq_along(files_neun)){
  file <- read.csv(files_neun[i])
  colnames(file)[2] <- "Well.R"
  txt1  <- subset(file, Well.R == "2")[,3]
  txt2  <- subset(file, Well.R == "23")[,3]
  ntxt1 <- subset(file, Well.R == "1")[,3]
  ntxt2 <- subset(file, Well.R == "24")[,3]
  nam   <- paste("Plate", i, "_NeuN", sep="")
  data  <- list(ntxt1, ntxt2, txt1, txt2)

  for (j in 1:4){
    treatment <- c("ntxt1","ntxt2", "txt1", "txt2")
    output1   <- mean(data[[j]], na.rm = T)
    output2   <- sd(data[[j]], na.rm = T)

    plate     <- paste(nam, "well-comparison", treatment[j], sep="_")
    analysis  <- data.frame(Mean = output1, Std = output2)
    write.table(analysis, file = paste(output_dir, plate, ".csv", sep = ""), sep = ",", append = F,
row.names = F, col.names = T)
  }

  ## Statistical analysis
  stat.ntxt <- c(stat.ntxt, t.test(ntxt1,ntxt2)$p.value)
  stat.txt  <- c(stat.txt, t.test(txt1,txt2)$p.value)
}

pvalues <- data.frame(Vehicle = stat.ntxt, Topotecan = stat.txt)
write.table(pvalues, file = paste(output_dir, "pvalues_well-comparison_NeuN_analysis.csv", sep = ""), sep
= ",", append = F, row.names = F, col.names = T)
```

Figure D.6: Sample script for well-effect analysis

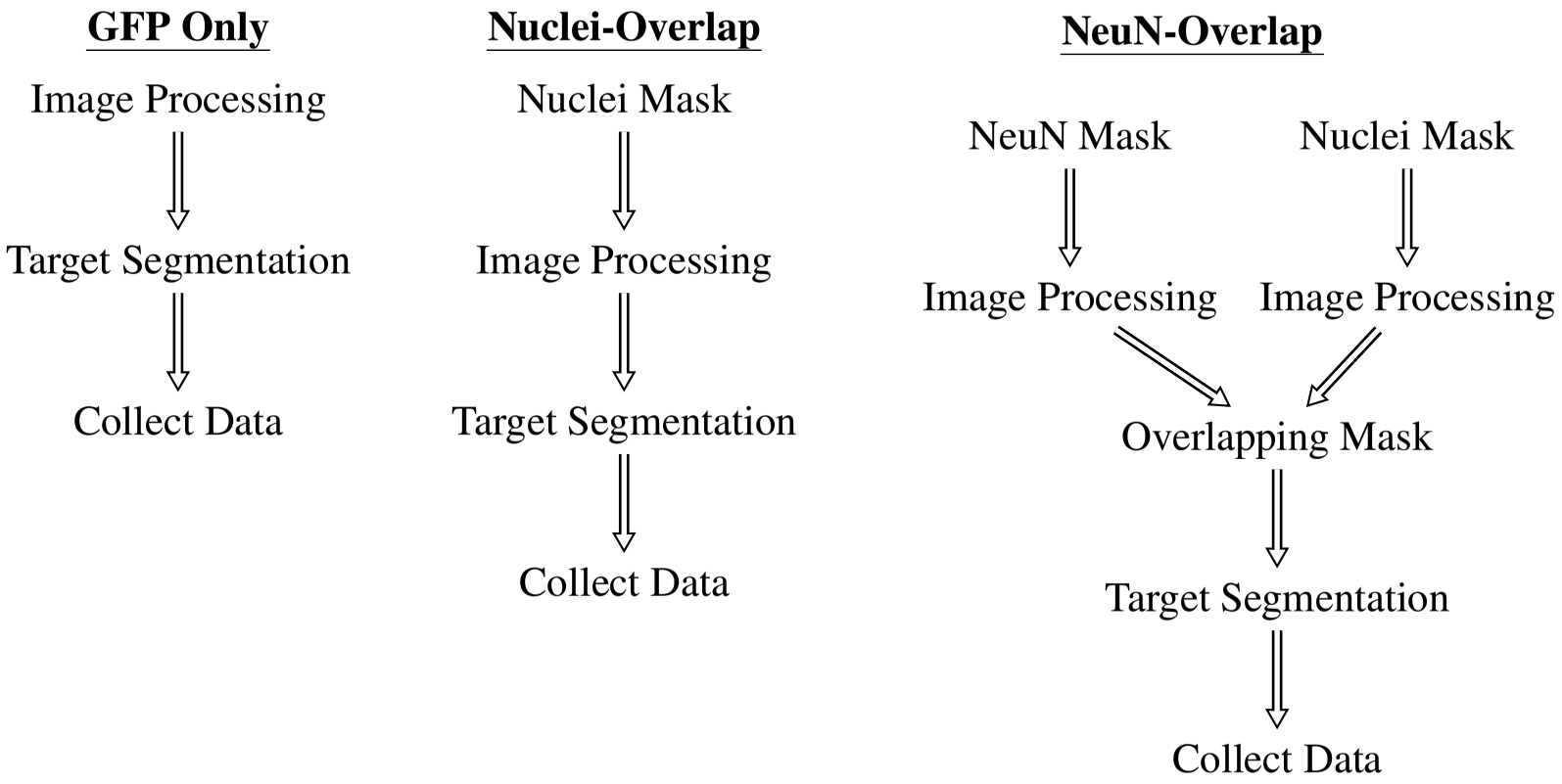


Figure D.7: Flow chart of analysis methods. Note that UpperQ method uses the GFP only data set and collects data from the 75th to 99th percentile.

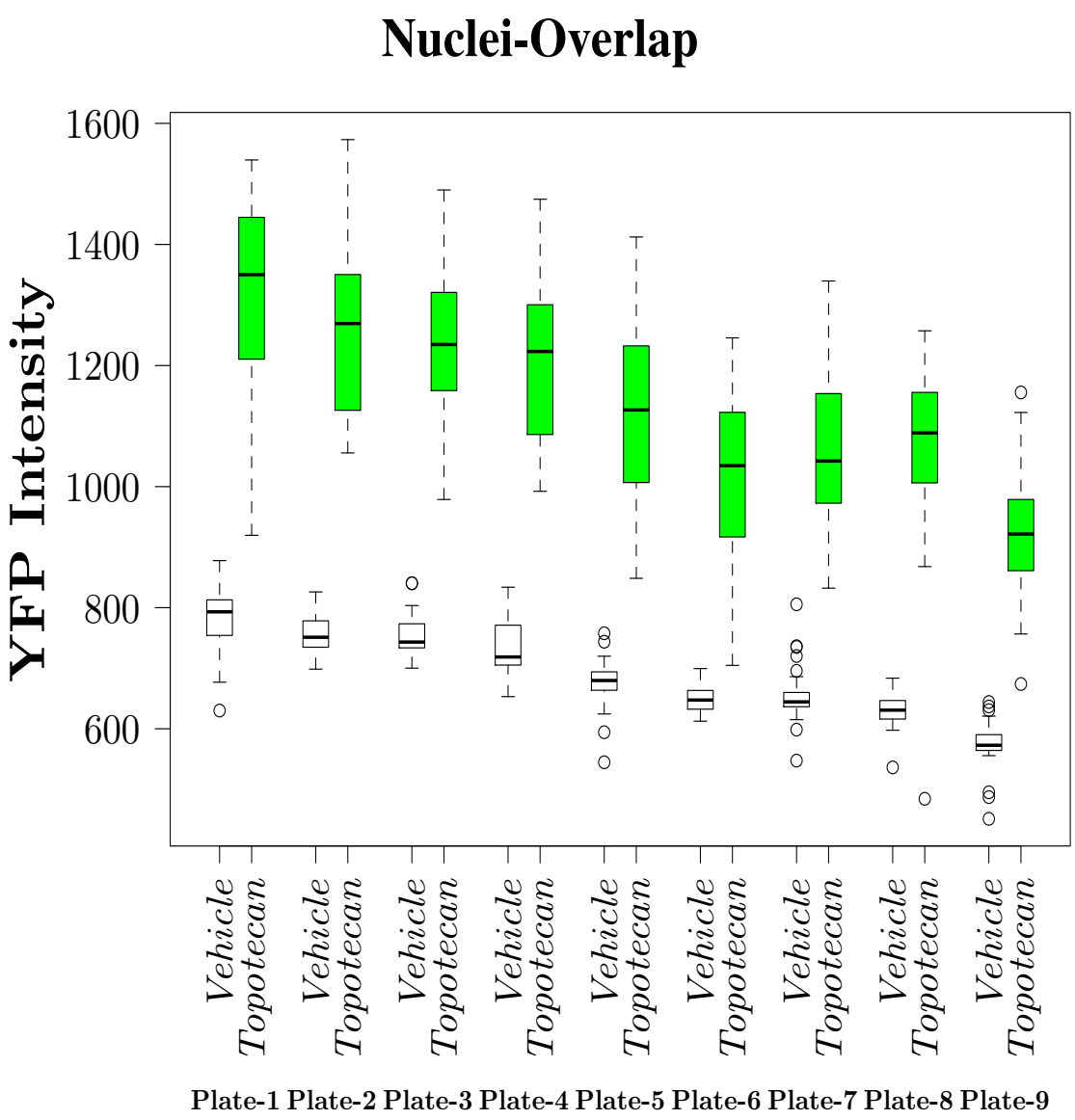


Figure D.8: The decrease in Ube3a^{YFP} intensity as a function of time, **Nuclei-Overlap** method.

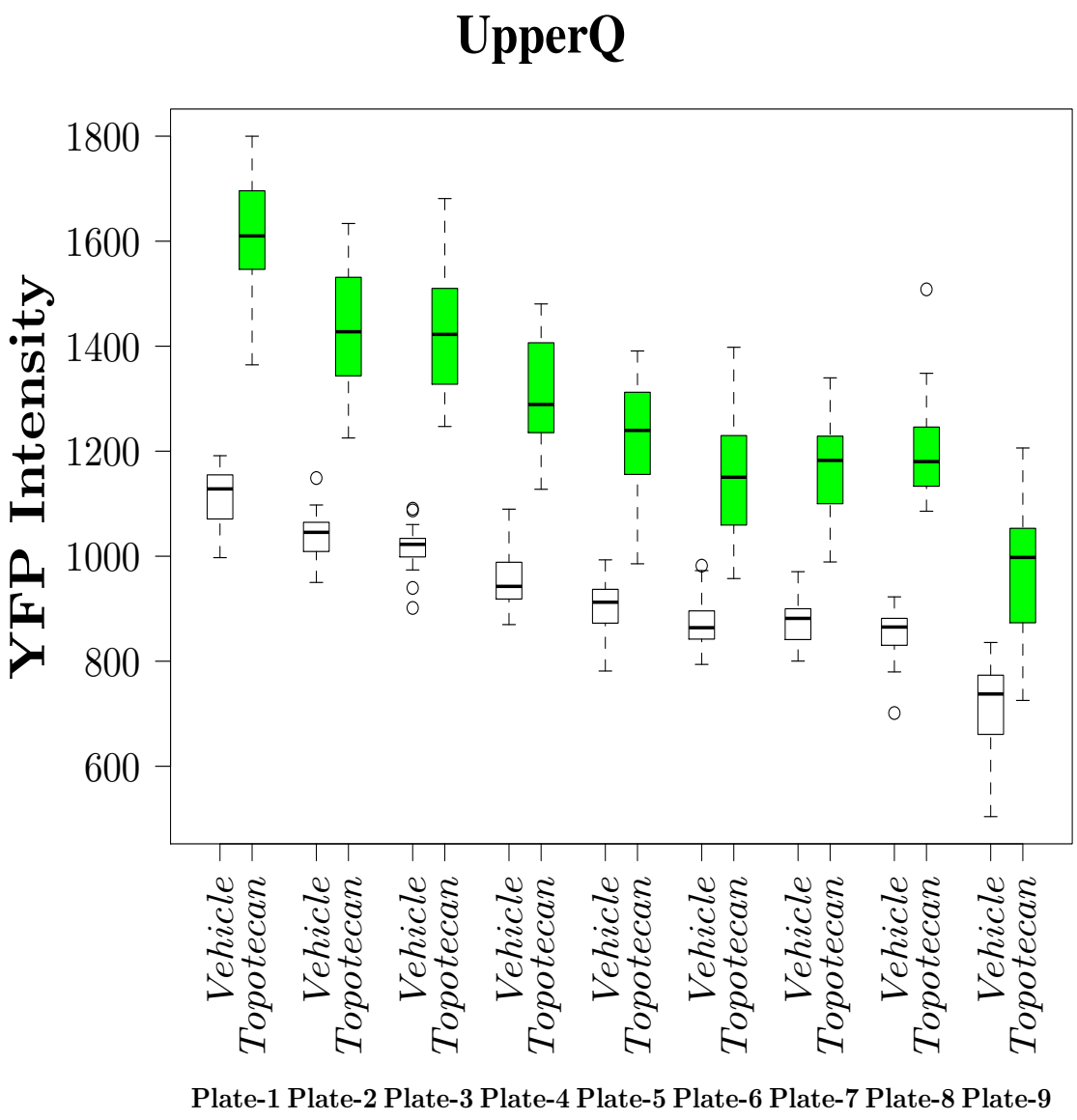


Figure D.9: The decrease in Ube3a^{YFP} intensity as a function of time, **UpperQ** method.

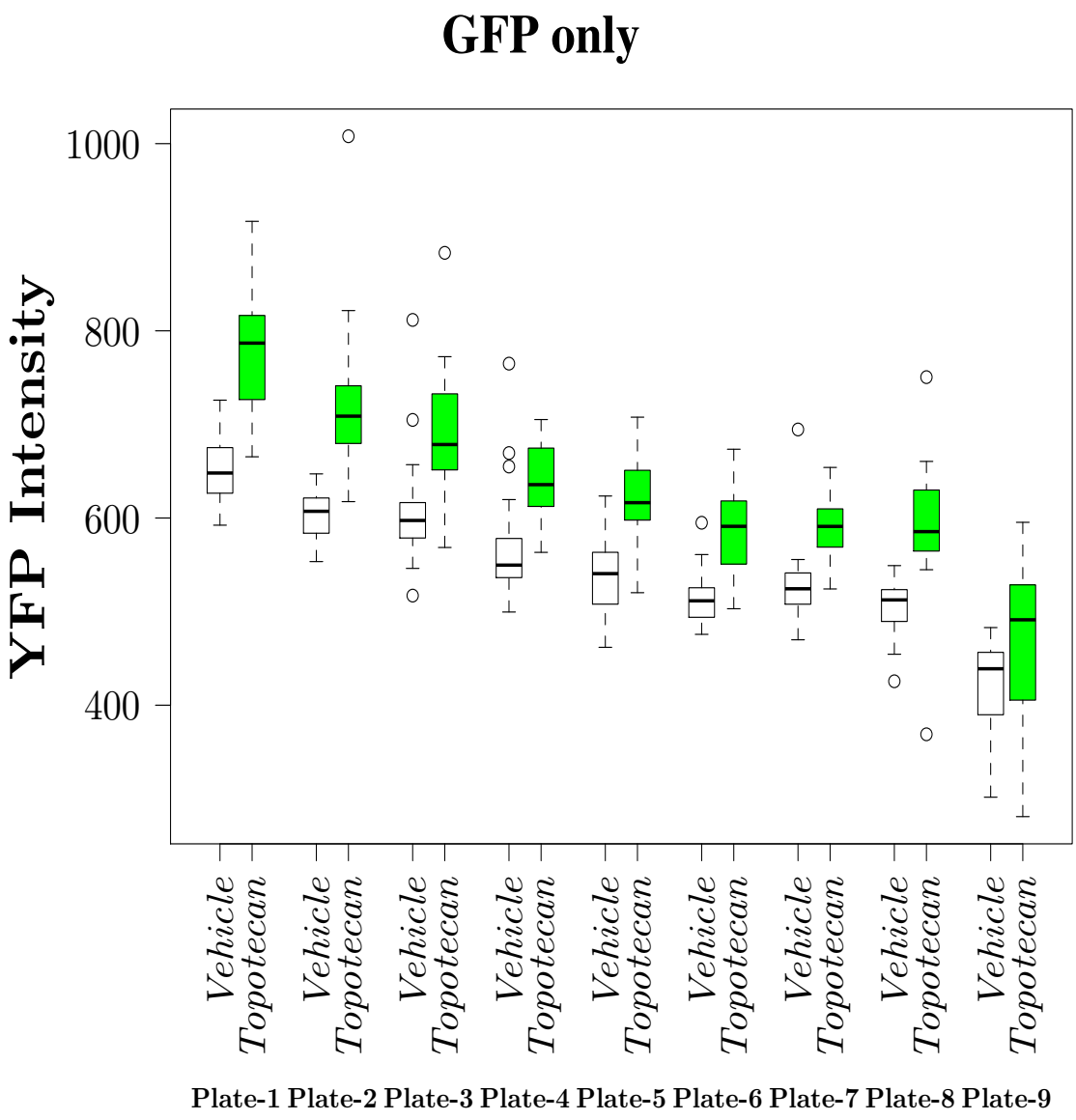


Figure D.10: The decrease in Ube3a^{YFP} intensity as a function of time, **GFP only** method.

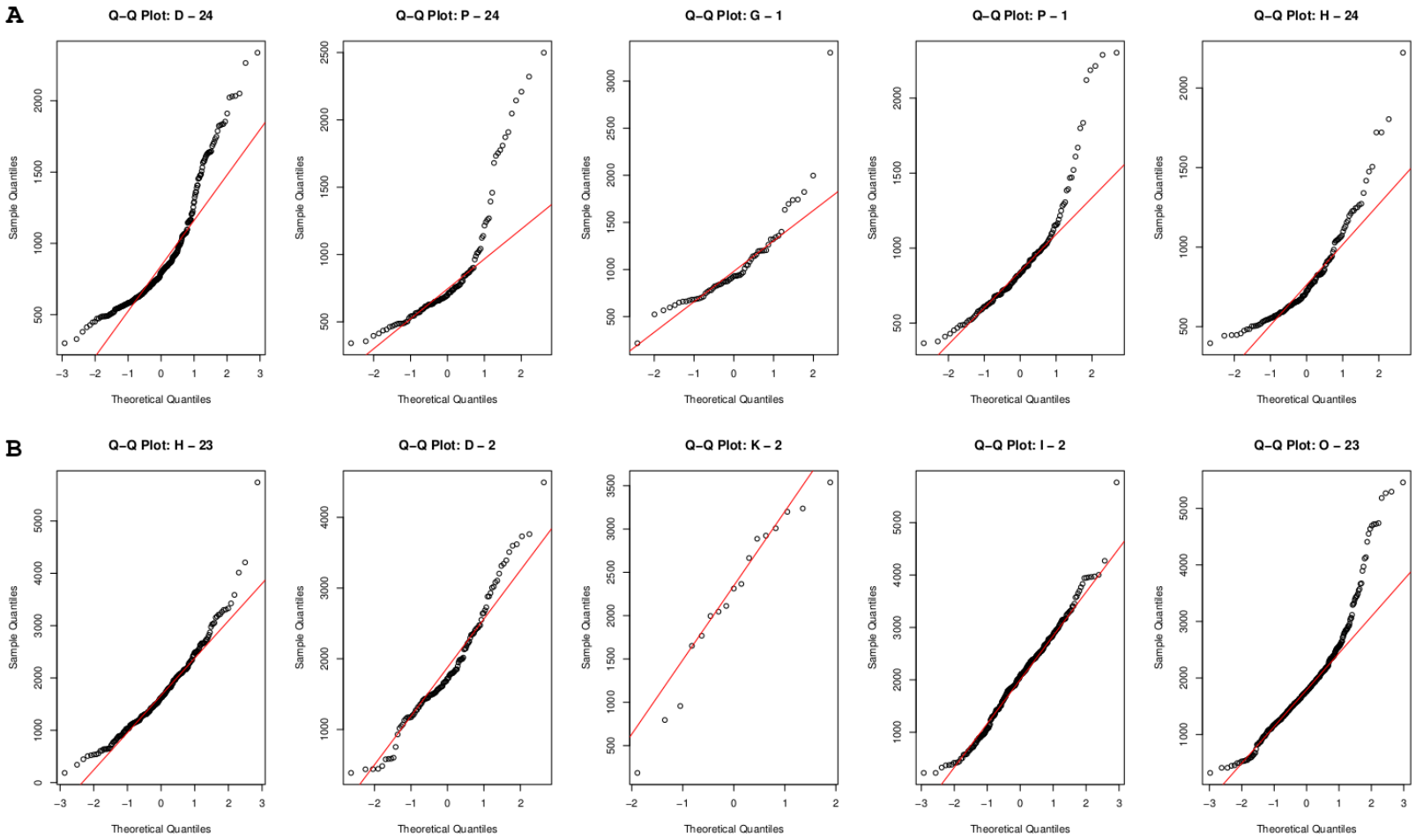


Figure D.11: Normal Q-Q plot for randomly selected wells for Plate 1, NeuN-Overlap method. **A.** Vehicle treated wells, and **B.** Topotecan treated wells.

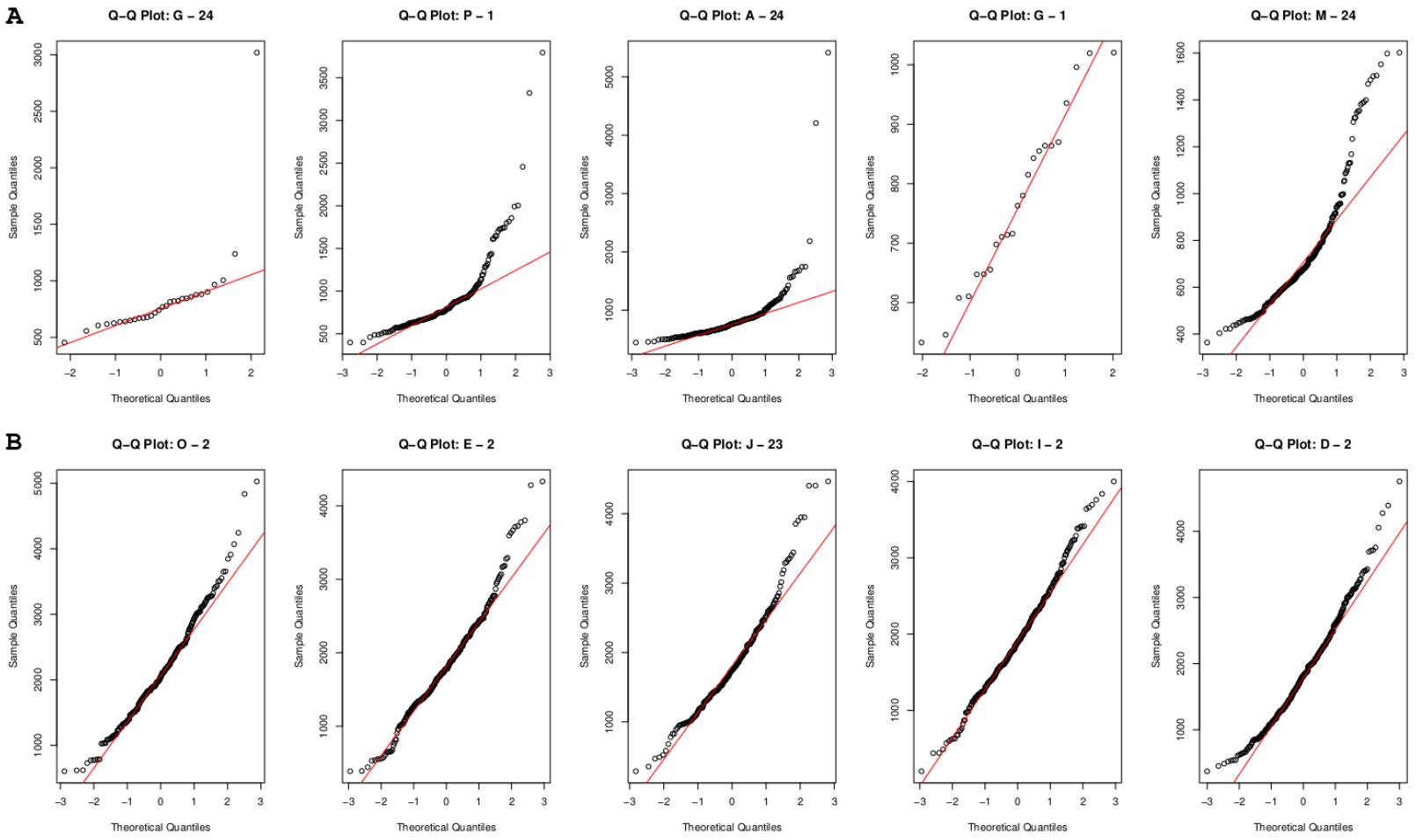


Figure D.12: Normal Q-Q plot for randomly selected wells for Plate 2, NeuN-Overlap method. **A.** Vehicle treated wells, and **B.** Topotecan treated wells.

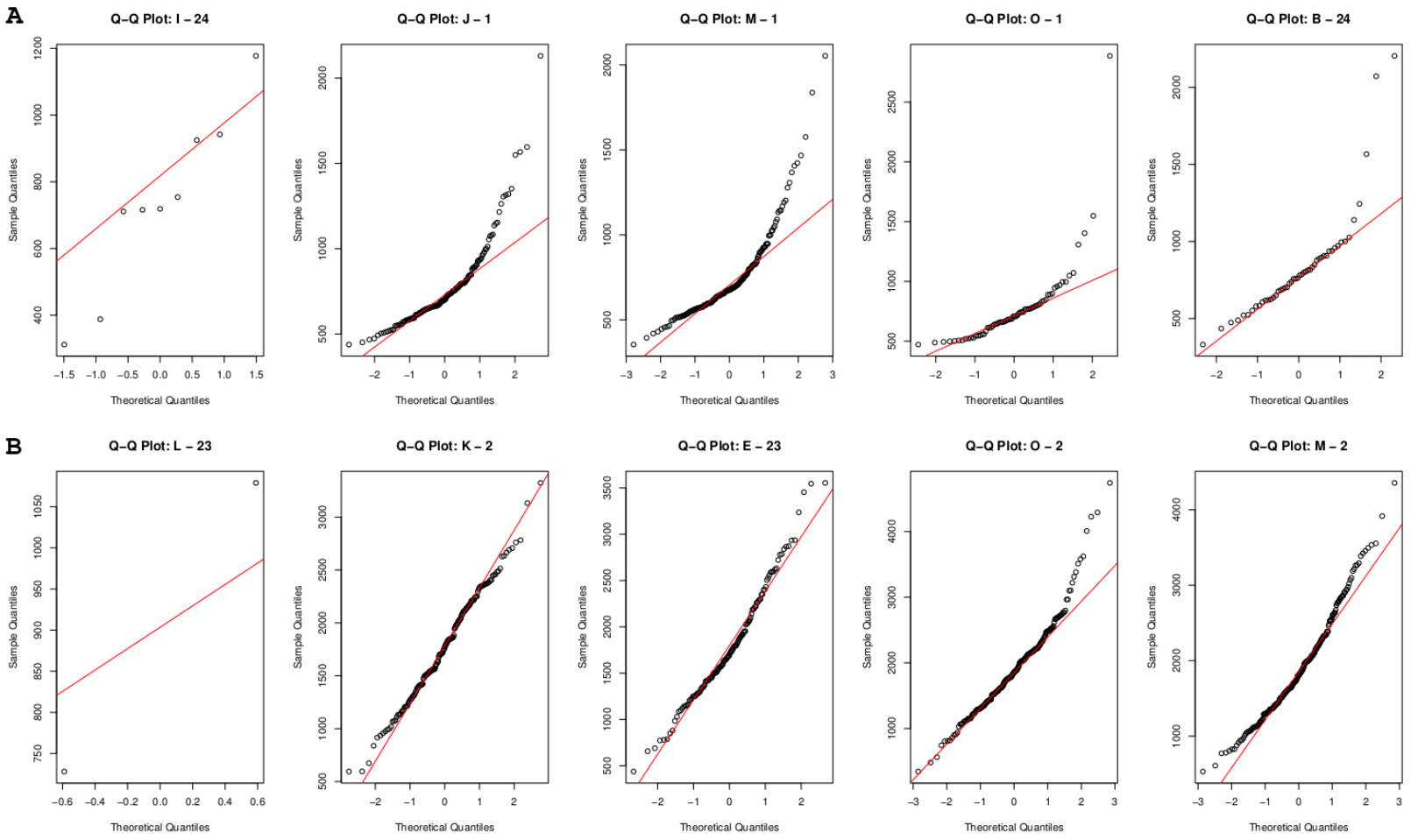


Figure D.13: Normal Q-Q plot for randomly selected wells for Plate 3, NeuN-Overlap method. **A.** Vehicle treated wells, and **B.** Topotecan treated wells.

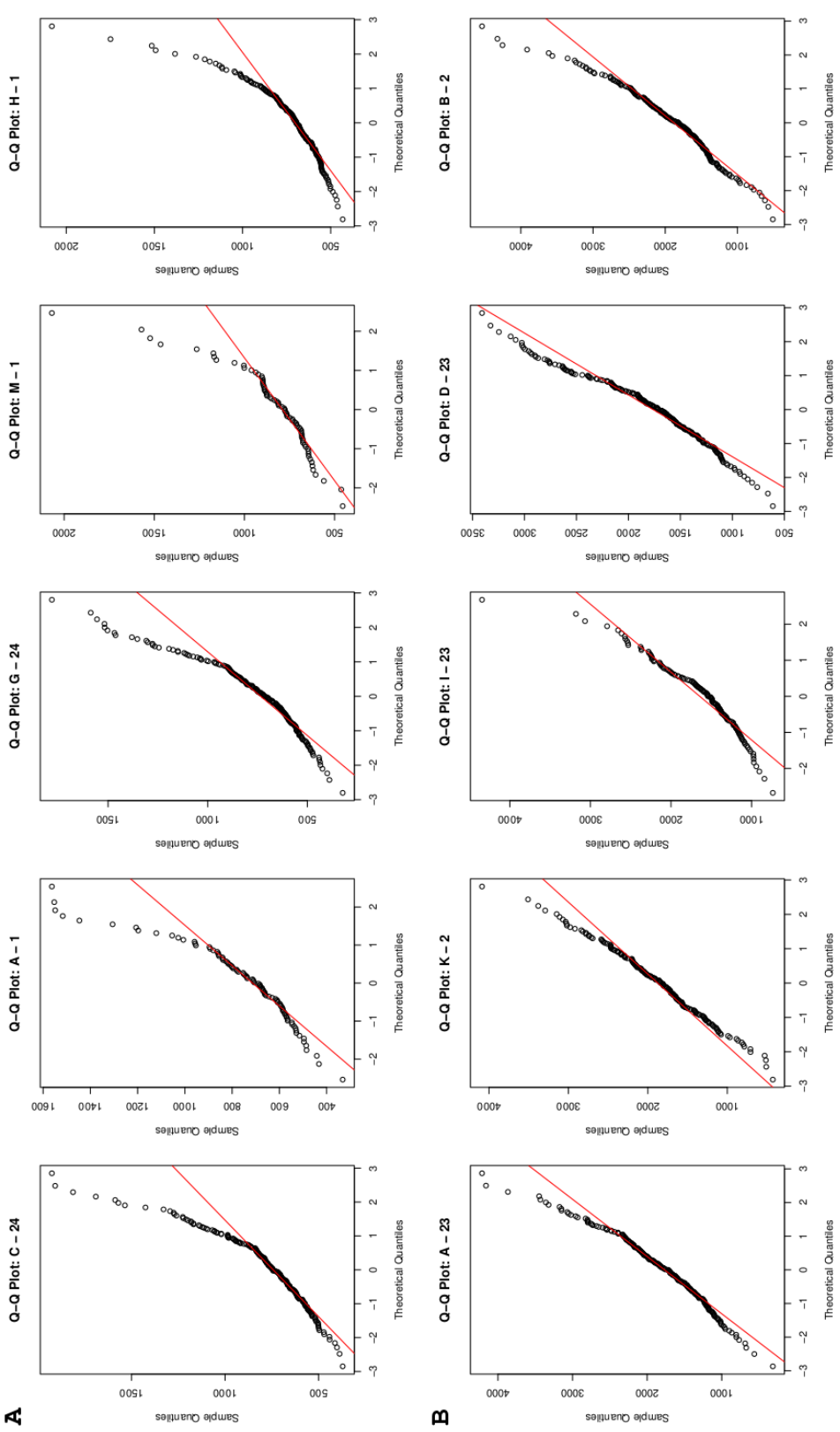


Figure D.14: Normal Q-Q plot for randomly selected wells for Plate 4, NeuN-Overlap method. **A.** Vehicle treated wells, and **B.** Topotecan treated wells.

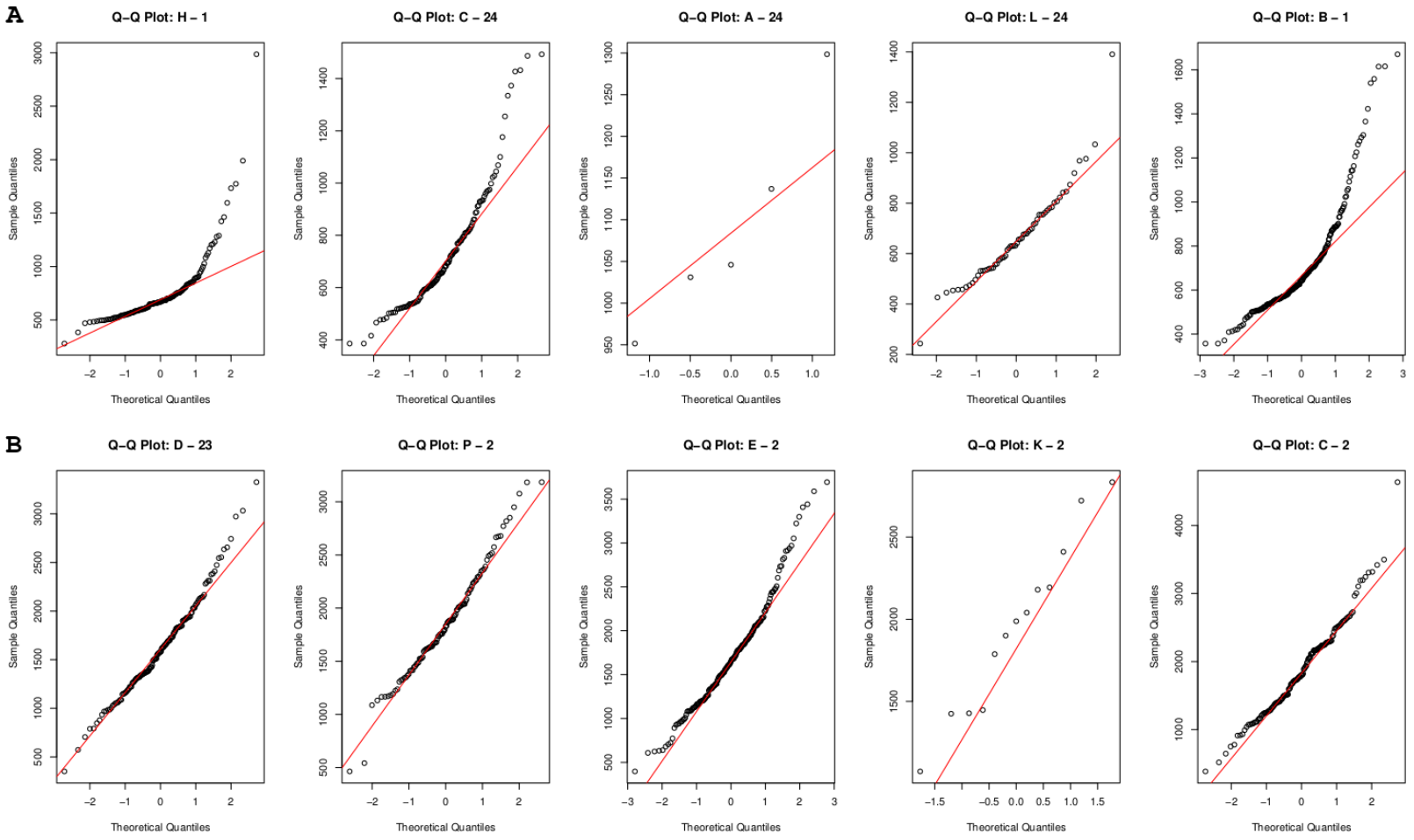


Figure D.15: Normal Q-Q plot for randomly selected wells for Plate 5, NeuN-Overlap method. **A.** Vehicle treated wells, and **B.** Topotecan treated wells.

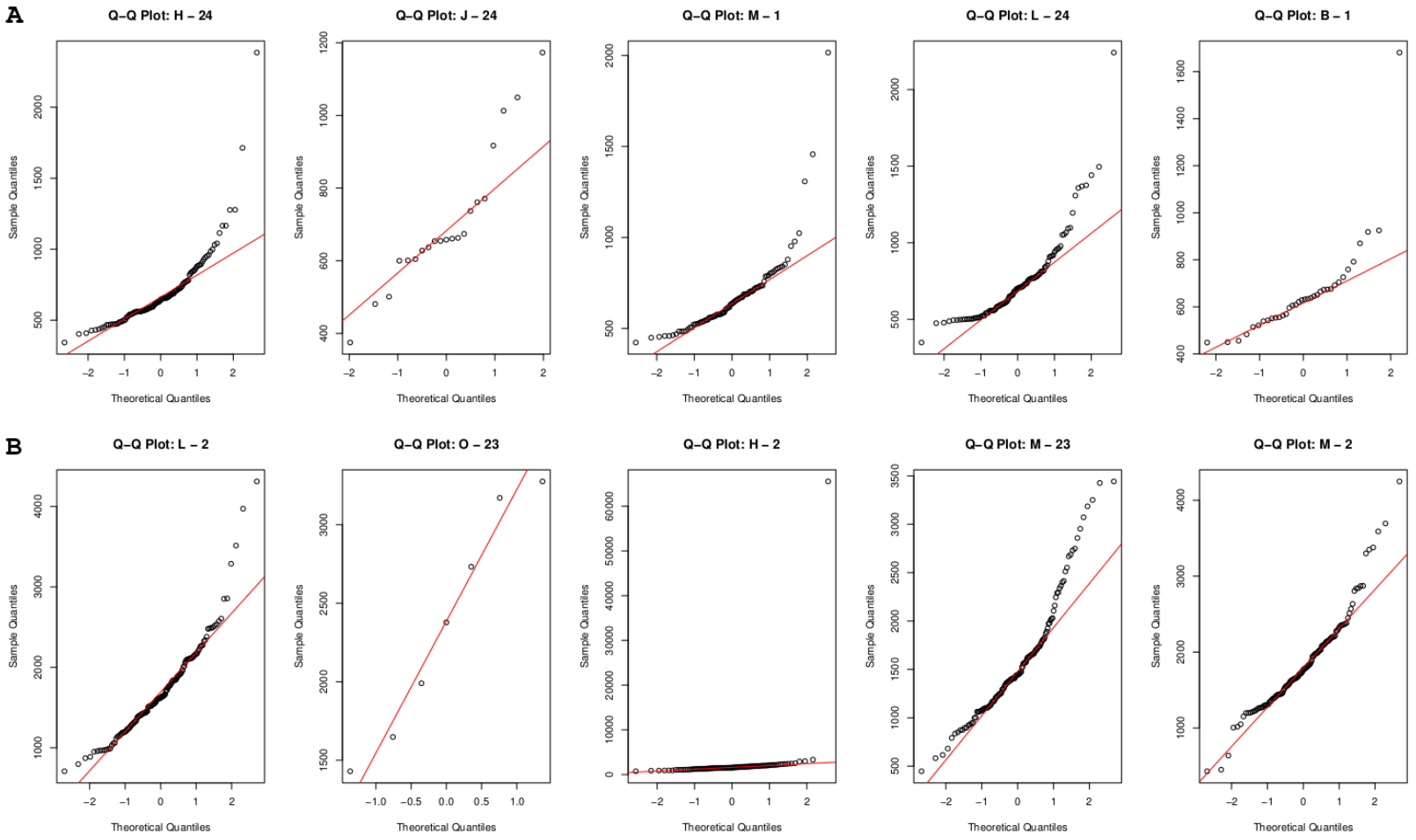


Figure D.16: Normal Q-Q plot for randomly selected wells for Plate 6, NeuN-Overlap method. **A.** Vehicle treated wells, and **B.** Topotecan treated wells.

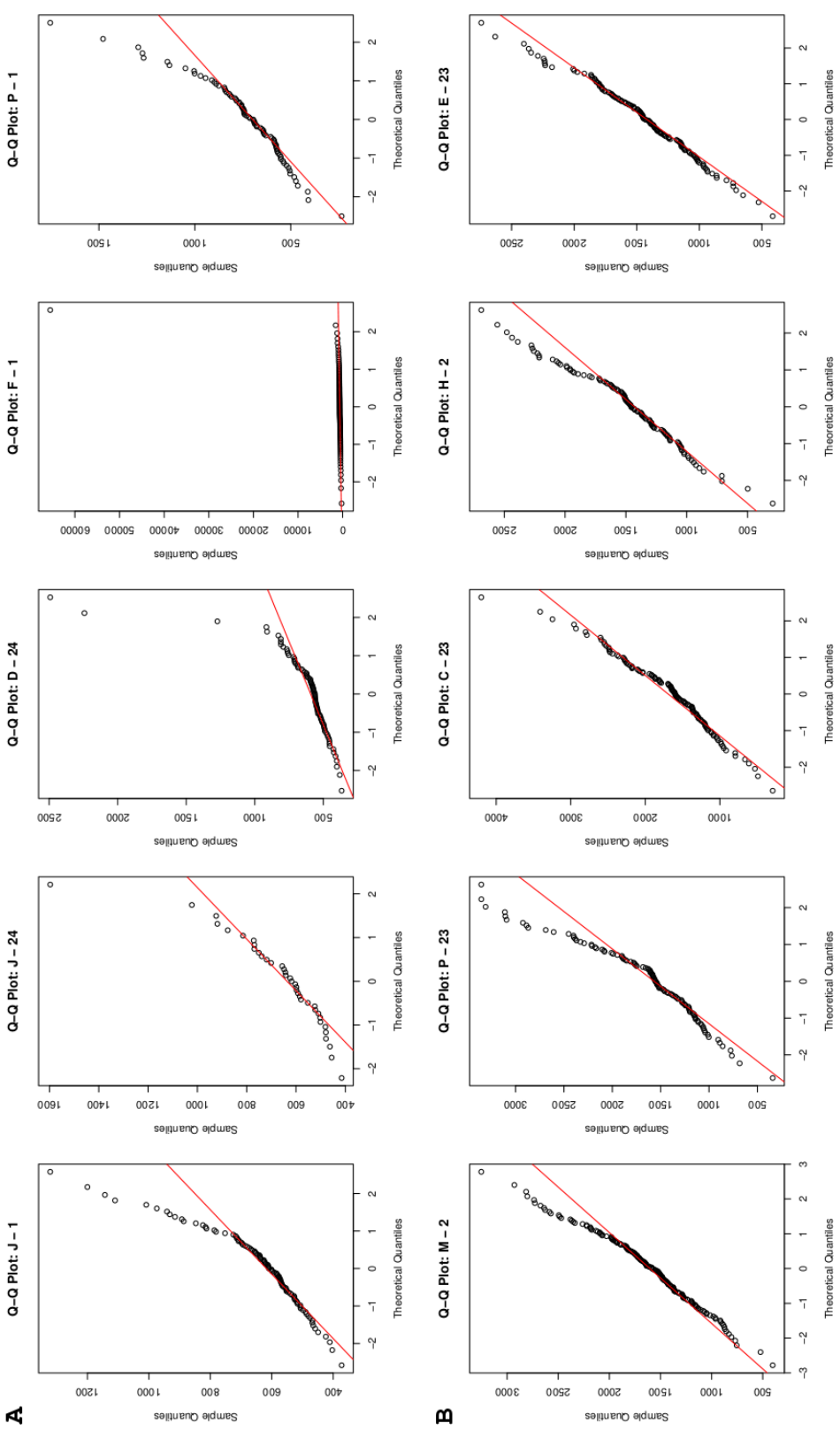


Figure D.17: Normal Q-Q plot for randomly selected wells for Plate 7, NeuN-Overlap method. **A.** Vehicle treated wells, and **B.** Topotecan treated wells.

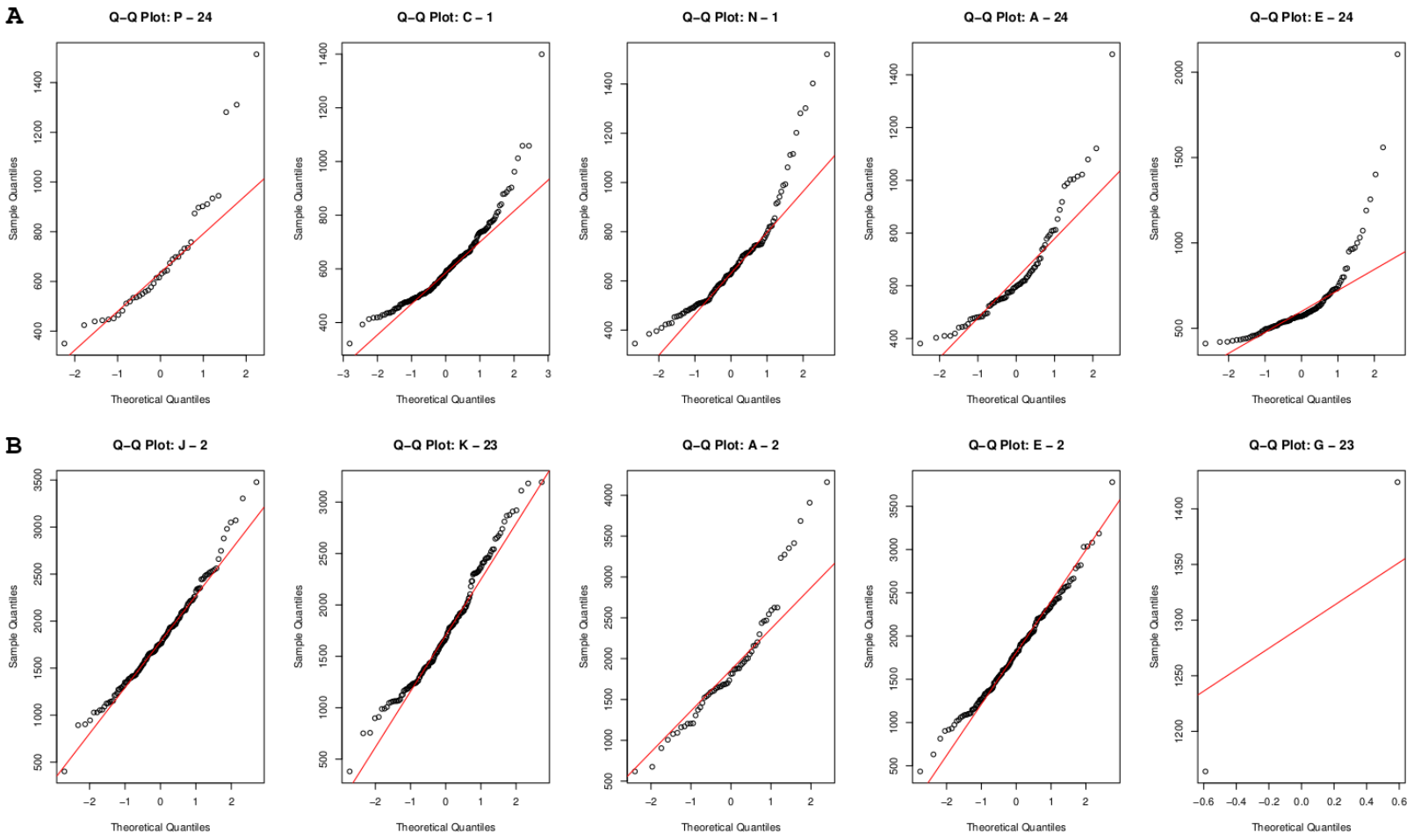


Figure D.18: Normal Q-Q plot for randomly selected wells for Plate 8, NeuN-Overlap method. **A.** Vehicle treated wells, and **B.** Topotecan treated wells.

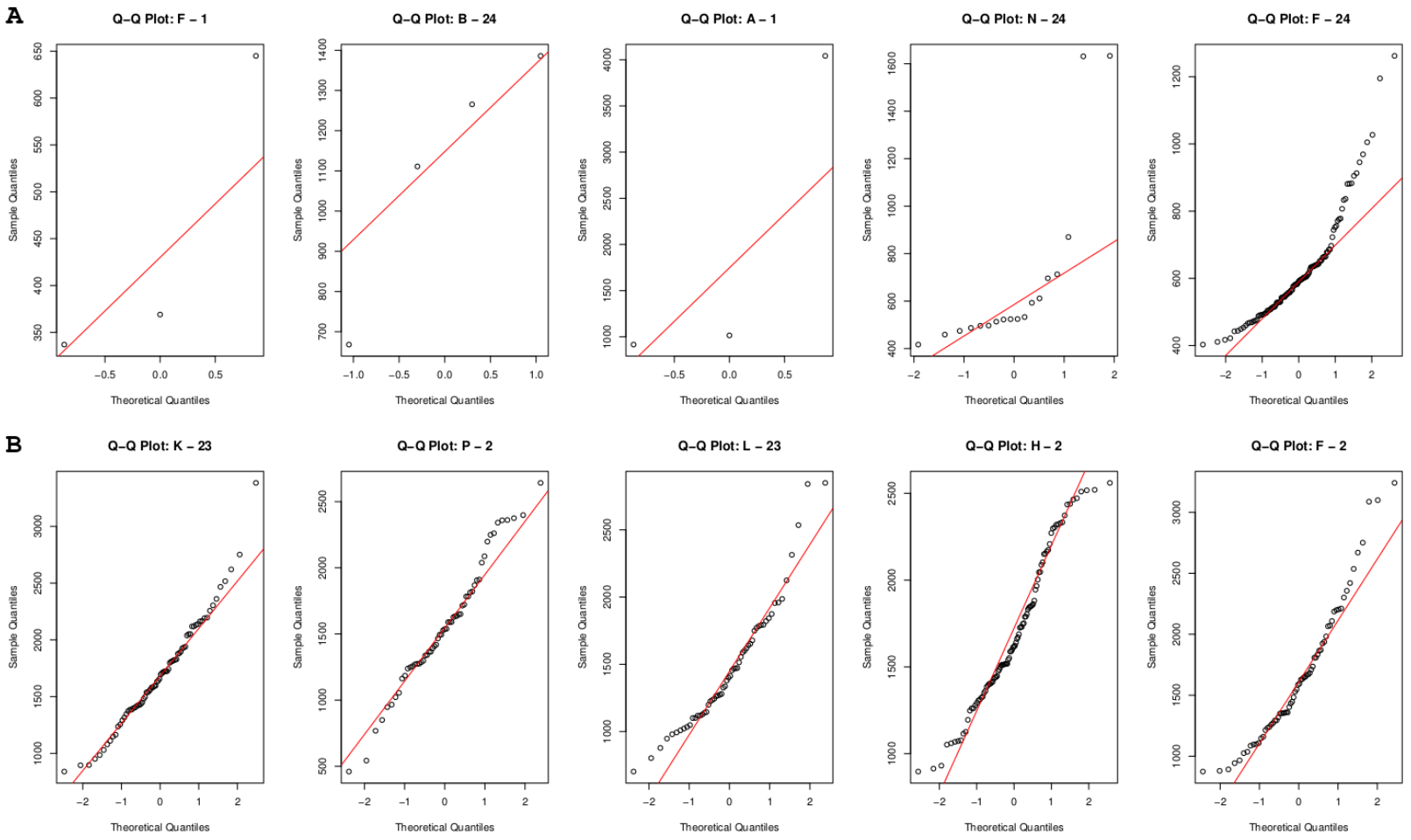


Figure D.19: Normal Q-Q plot for randomly selected wells for Plate 9, NeuN-Overlap method. **A.** Vehicle treated wells, and **B.** Topotecan treated wells.

```r
##This script tests for normality and returns results as figure and table
rm(list=ls())##Clear workspace
library(plyr);library(dplyr);library(devtools)
split.files <- function(file){
    file$Section <- as.character(file$Section)
    well.split <- ldply(strsplit(file$Section, " - "))
    names(well.split) <- c("Well.Letter", "Well.Number")
    new.file <- cbind(well.split, file)
    return(new.file)
}
outdir <- "/PATH/TO/OUTPUT"
files_full <- list.files("/PATH/TO/FILES/", full.names = T)
pdf(file=paste0(outdir,Sys.Date(),"_HTS_Random-QQPlots.pdf"),width=15,height=9
)
distnorm_full<- list()
for (i in seq_along(files_full)){
    file <- read.csv(files_full[i]);new.file <- split.files(file)
    txt.wells  <- subset(new.file, Well.Number == "2" | Well.Number == "23")
    ntxt.wells <- subset(new.file, Well.Number == "1" | Well.Number == "24")
    wellT.unique <- unique(txt.wells$Section);wellN.unique <- unique(ntxt.well
s$Section)
    par(mfrow = c(2, 5))
    ntxt.test <- sample(wellN.unique, 5)    ## Unseeded
    st_crt.pvalue  <- vector();count_crt <- vector()
    for (j in 1:5){
        crt  <- subset(ntxt.wells, Section == ntxt.test[j])
        if (length(crt$YFPintensity) >= 3){
            st_crt.pvalue  <- c(st_crt.pvalue, shapiro.test(crt$YFPintensity)$
p.value)
        } else {st_crt.pvalue  <- c(st_crt.pvalue, NA)}
        count_crt <- c(count_crt, length(crt$YFPintensity))
        qqnorm(crt$YFPintensity, main = paste0("Q-Q Plot: ",ntxt.test[j]));
        qqline(crt$YFPintensity, col = 2)
    }
    txt.test  <- sample(wellT.unique, 5)    ## Unseeded
    st_topo.pvalue <- vector();count_topo <- vector()
    for (j in 1:5){
        topo <- subset(txt.wells, Section == txt.test[j])
        if (length(topo$YFPintensity) >= 3){
            st_topo.pvalue <- c(st_topo.pvalue, shapiro.test(topo$YFPintensity
)$p.value)
        } else {st_topo.pvalue <- c(st_topo.pvalue, NA)}
        count_topo <- c(count_topo, length(topo$YFPintensity))
        qqnorm(topo$YFPintensity, main = paste0("Q-Q Plot: ",txt.test[j]));
        qqline(topo$YFPintensity, col = 2)
    }
    par(mfrow = c(1,1));st.pvalues <- c(st_crt.pvalue, st_topo.pvalue)
    test.wells <- c(ntxt.test, txt.test);counts <- c(count_crt, count_topo)
    tmp <- seq_along(files_full);plate_number <- paste0("Plate_", tmp[i])
    distnorm_full[[i]] <- data.frame(Plate=rep(plate_number,length(test.wells)
),Well=test.wells,PValue=st.pvalues,N= counts,stringsAsFactors=F)
}
shapiroWilk <- dplyr::bind_rows(distnorm_full)
write.table(shapiroWilk, file = paste0(outdir,Sys.Date(),"_ShapiroWilk_Normali
tyTest_pvalues.csv"),sep =",", append=F,row.names=F,col.names=T)
dev.off()
```

Figure D.20: Example script for Shapiro-Wilk analysis and Q-Q plot generation.

APPENDIX E

PROTOCOLS

The following is a list of detailed protocols used in this thesis.

E.1 Genotyping

E.1.1 DNA extraction

Two-day Genotyping

1. Move tails (or ear punches) to 1.5 ml tube that does not leak if necessary.
2. Add 400 Proteinase K mastermix to each tail.
   (a) 400 μl of NTES, **Table E.1**
   (b) 1 μl of Proteinase K
3. Incubate overnight at 50°C, constant agitation.
4. Next day, spin down for 5 min at 14,000 rpm.
5. Pour supernatant into new labeled tube.
6. Add 650 μl of isopropanol that has been cooled in -20°C for $>$ 1 h[1].
7. Shake by hand until a precipitate forms.
8. Spin down for $>$ 1 min at 14,000 rpm.
9. Pour supernatant off and set upside-down to dry briefly on a paper towel[2].
10. Add 1 ml of 70% EtOH to the tubes.
11. Gently free the pellet from the side of tube with a flick.
12. Set the sample to rock for $>$ 30 min at room temperature[3].

[1] If precipitate does not form immediately, place in -20°C for $>$ 2 h.

[2] Care should be taken when pouring off supernatant so as not to loss DNA at bottom of tube.

[3] The longer the wash, the cleaner the sample. Rock overnight for best results. For really dirty samples, 70% EtOH can be changed out.

13. Spin down sample for 1 min at 14,000 rpm.
14. Pour off 70% EtOH and blot dry on paper towel.
15. Set on heat-block with lid open for about 10 min (10 - 15 min) at 37°C to 65°C.
16. Add purified water (or elution buffer) based on the amount of DNA in the tube and close lid for overnight incubation[4].

KAPA Mouse Genotyping

1. Setup mastermix for DNA extraction reactions.
   (a) 22 μl of PCR-grade water
   (b) 2.5 μl of 10X KAPA Express Extract Buffer
   (c) 0.5 μl of 1 U/μl KAPA Express Extract Enzyme
2. Add 25 μl/sample in PCR 120 μl tubes.
3. Run lysis protocol on PCR machine[5].
   (a) Lysis step - 75°C for 10 min
   (b) Enzyme inactivation - 95°C for 5 min
4. Centrifuge samples briefly to pellet cellular debris.
5. Dilute DNA extract 10-fold with 10 mM Tris-HCl (pH 8.0 - 8.5)[6].

E.1.2 Ube3a^{YFP} Genotyping

Standard

1. Measure DNA with nanodrop before continuing[7].
2. For best results, dilute each sample to 50 ng/μl.
3. Make up mastermix.
   (a) 11.4 μl of water

[4] If uncertain about the amount of water to add, use low amounts like 50 μl and adjust after experience.
[5] KAPA Express Extract will not completely degrade the tissue like Proteinase K, so there will be intact tissue in the tube after lysis.
[6] This is necessary to dilute out the salts leftover by lysis.
[7] For Ube3a^{YFP} Genotyping, 100 ng/μl required.

(b) 5.0 μl of 5X Taq Buffer

(c) 2.5 μl of $MgCl_2$

(d) 1.0 μl of dNTP

(e) 1.0 μl of forward primer

(f) 0.15 μl of R_1 primer

(g) 0.5 μl of R_2 primer

(h) 1.25 μl of DMSO

(i) 0.2 μl of Taq Polymerase

4. Add 23 μl of mastermix to each PCR tube.
5. Add 2 μl of diluted DNA to each tube.
6. Run PCR protocol.
   (a) 1 cycle of 95°C for 5 min
   (b) 35 cycles of
      - 95°C for 40 s
      - 54°C for 40 s
      - 72°C for 60 s
   (c) 1 cycle of 72°C for 7 min
   (d) Hold at 4°C
7. Spin down and store at 4°C.

KAPA PCR Reaction

1. Make up master mix.
   (a) 7 μl of PCR-grade water
   (b) 10 μl of 2X KAPA2G Fast Genotyping Mix with dye
   (c) 1.0 μl of forward primer
   (d) 0.5 μl of R_1 primer

(e) 0.5 μl of R_2 primer

2. Add 19 μl of mastermix to each PCR tube.
3. Add 1 μl of diluted DNA to each tube.
4. Run PCR protocol.
   (a) 1 cycle of 95°C for 3 min
   (b) 35 cycles of
      - 95°C for 15 s
      - 58°C for 15 s
      - 72°C for 30 s
   (c) 1 cycle of 72°C for 2 min
   (d) Hold at 4°C
5. Spin down and store at 4°C.

E.1.3 Gel Electrophoresis

1. Make 1 to 1.5% agarose gel.
   (a) Measure out $\sim$1 to 1.5 g of agarose
   (b) Pour agarose powder into glass flask
   (c) Add 100 ml of 1xTAE buffer to glass flask, **Table E.1**
   (d) Microwave until agarose completely dissolves[8]
   (e) Let agarose solution cool down for 10 min[9]
   (f) Add 5 μl of ethidium bromide (EtBr) or other DNA imaging solution[10].
2. Setup gel tray if needed.
3. Pour the agarose solution into a gel tray with desired well comb in place and let solidify for 10 to 15 min.

[8]Be careful that it does not boil over. It is a good idea to microwave for 30 s and stop and swirl, then continue to boil.

[9]Use running water to speed the process up if desired.

[10]EtBr is a known mutagen, so use with caution (i.e. always wear nitrile gloves when working with it).

4. Place gel into gel electrophoresis unit with 1xTAE until gel is completely covered.
5. Load DNA molecular weight ladder into first lane of the gel and PCR samples after it.
6. Run at 100V for $>$ 1 h.
7. Use UV light to visualize DNA fragments.

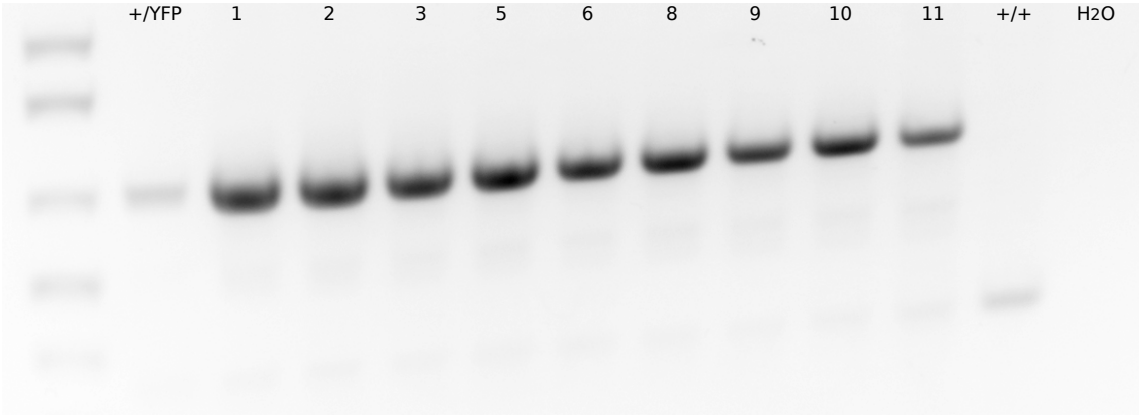


Figure E.1: Example of standard Ube3a^{YFP} Genotyping using Ube3a$^{+/YFP}$ ES cells gel image.

Table E.1: Media composition

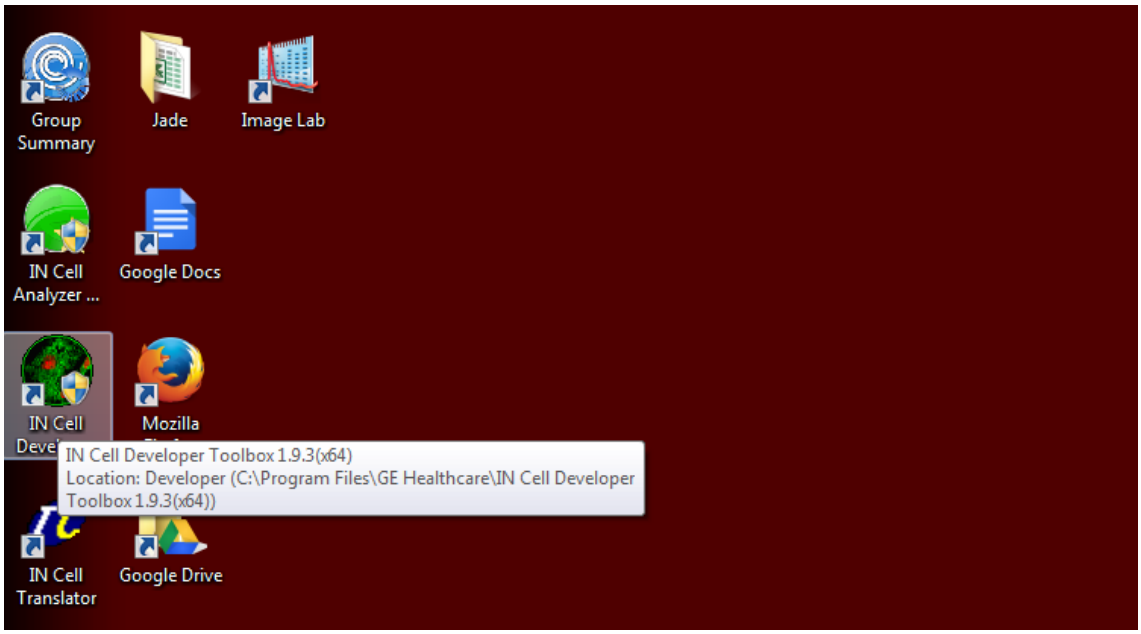
Media	**Components**	**Amount**
NTES Buffer	Sterile H_2O	315 ml
	SDS	1% (50 ml of 10%)
	NaCl	0.1 M (10 ml of 5 M)
	EDTA	0.1 M (100 ml of 0.5 M)
	Tris	0.05 M, pH 8 (25 ml of 0.5 M)
50X stock TAE	Tris-base	242.2 g
	Glacial acetic acid	57.1 ml
	EDTA	0.5 M, pH 8, 100 ml
	Sterile H_2O	up to 1 L

E.2 High-throughput Image Analysis

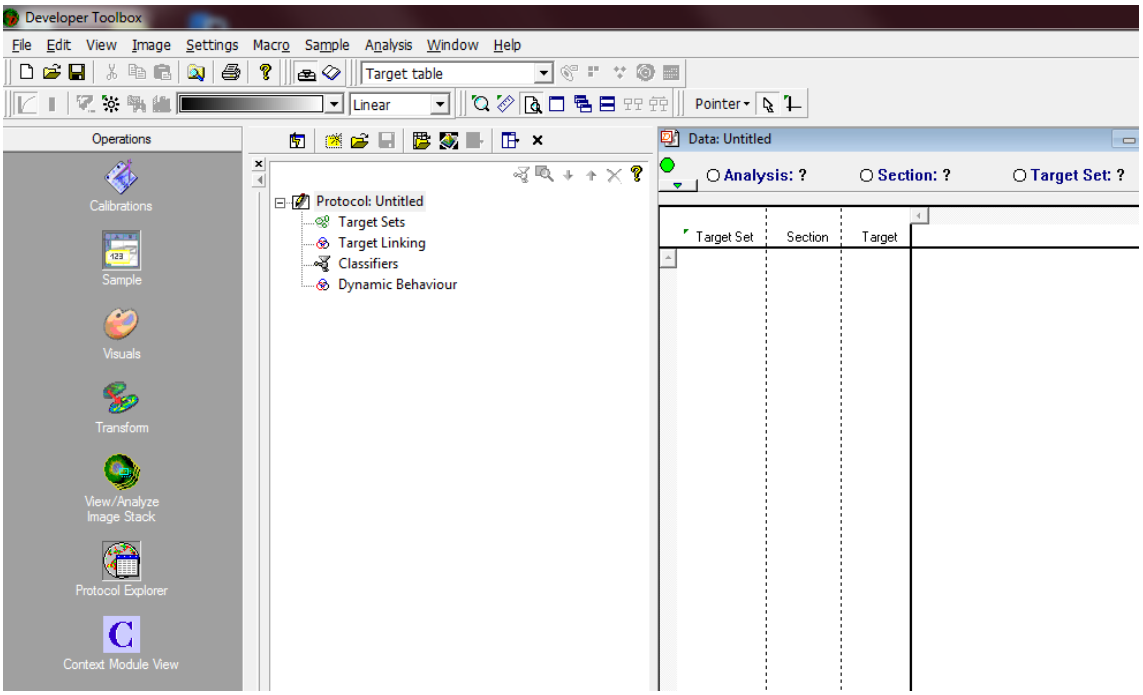
E.2.1 Image Processing

GFP only

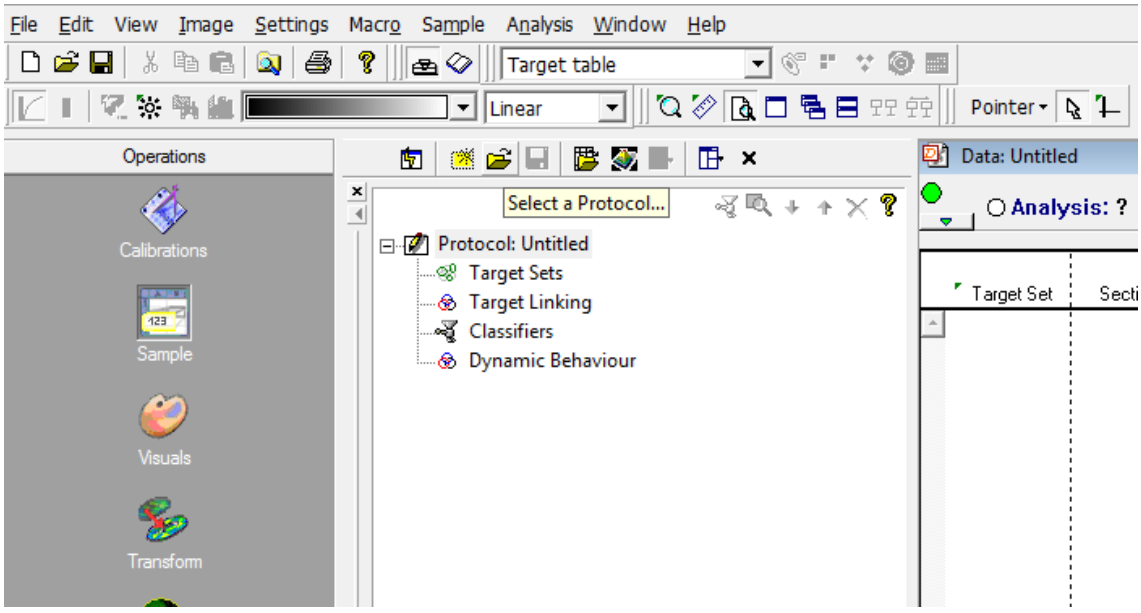
1. Launch Developer Toolbox



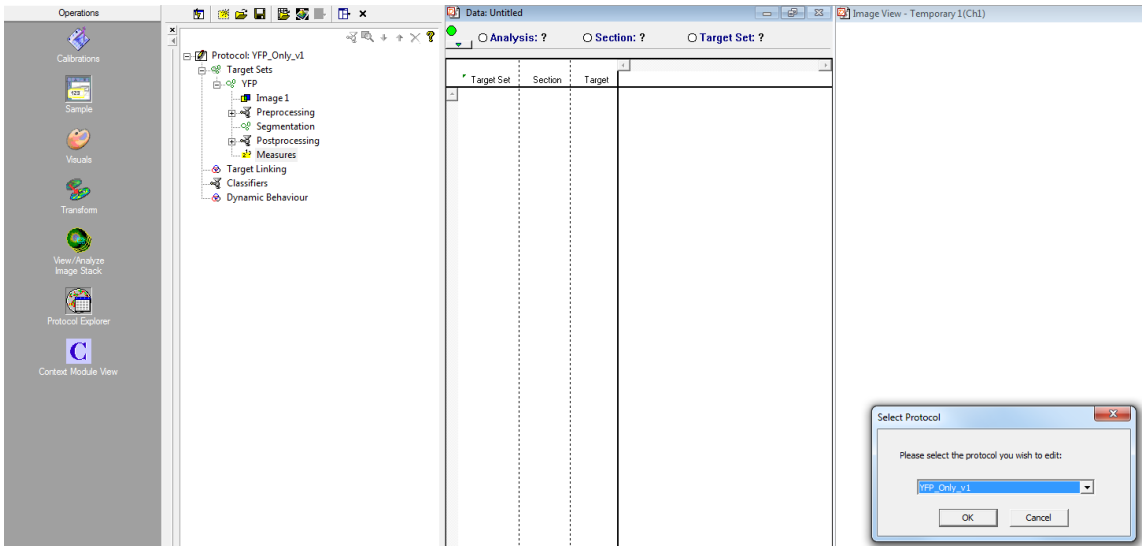
2. Enter Protocol Explorer

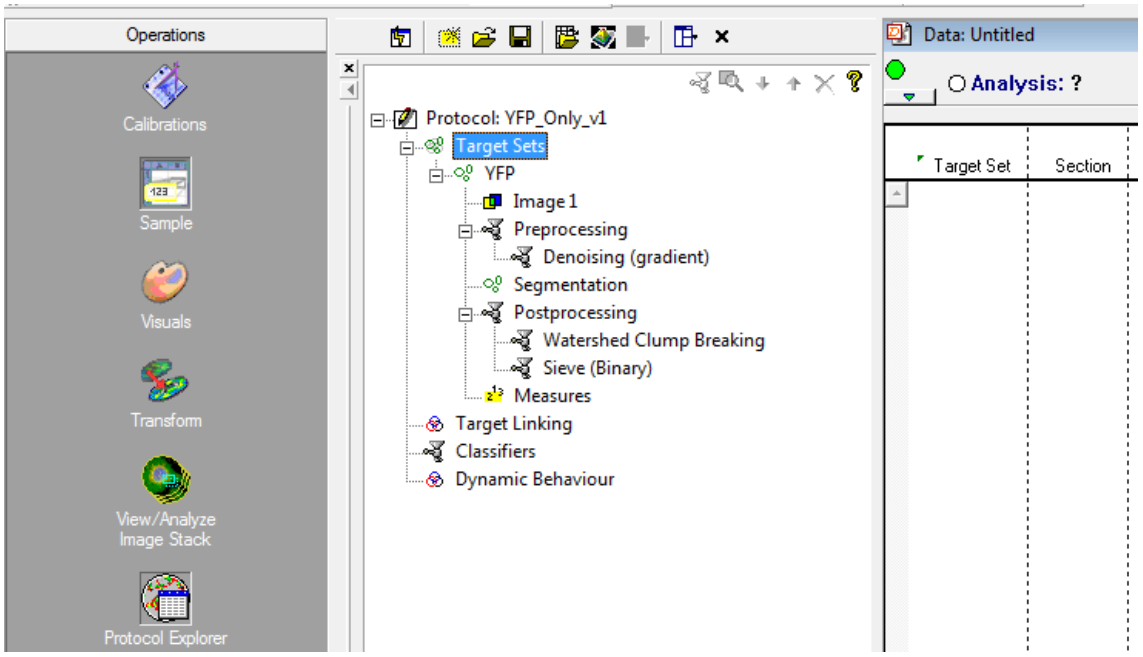


3. Open protocol

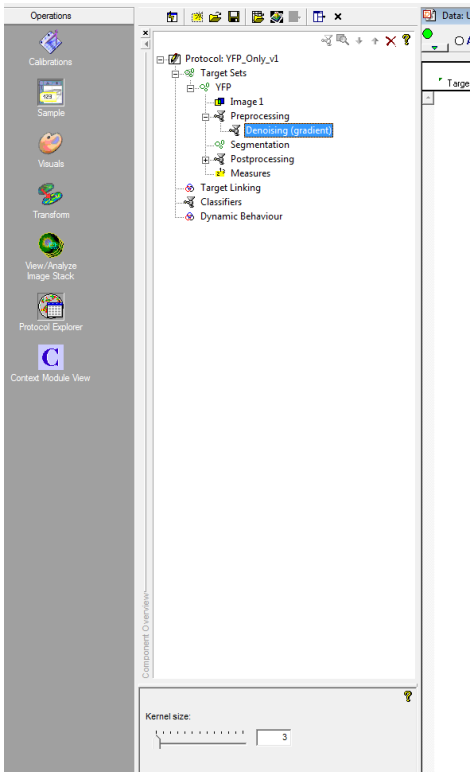


4. Select protocol: GFP only uses *YFP Only v1*

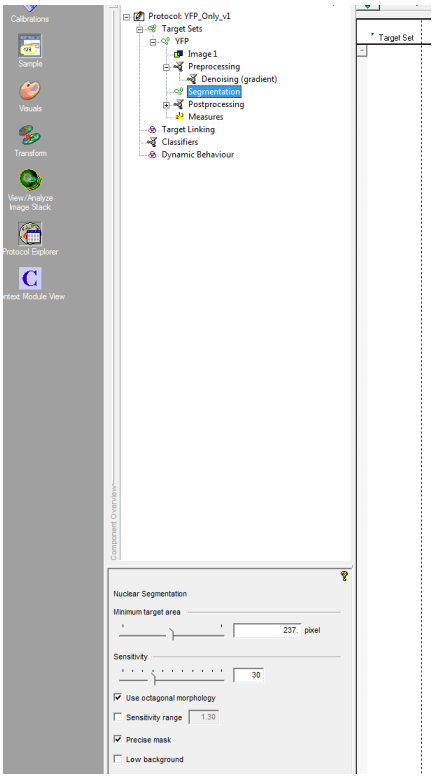




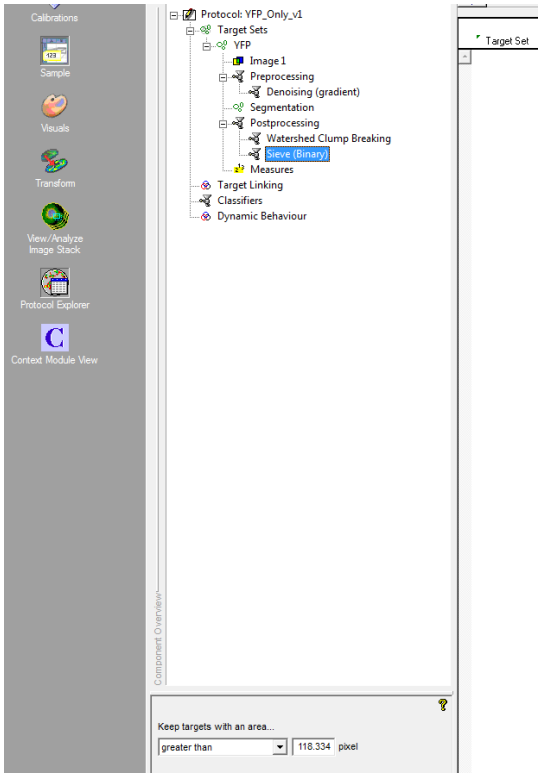
5. Within the protocol, targets are pre-processed using Denoising (gradient) at kernel size = 3.



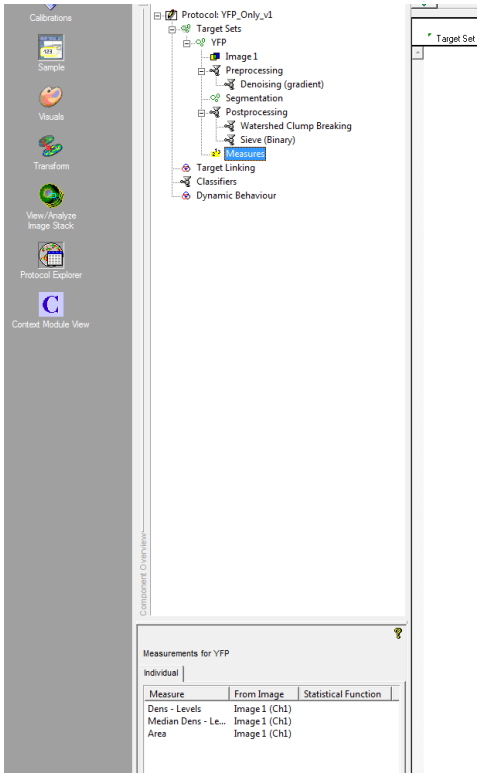
6. Within the protocol, targets are segmented via
    - Nuclear Segmentation
    - Minimum target area (237 pixel)
    - Sensitivity equal to 30
    - Octagonal morphology
    - Precise mask



7. Within the protocol, post-processing uses the following
    - Watershed clump breaking
    - Sieve (binary)



8. Within the protocol, targets can be directly measured using GFP channel.
   (a) Density Levels of target
   (b) Median Density Levels of target
   (c) Target Area



Nuclei-Overlap

Nuclei-Overlap image processing method uses the same protocol for GFP channel; however adds the Nuclei channel as well.

1. Process GFP channel
    (a) Pre-processing steps for GFP channel as above.
    (b) Segmentation for GFP channel as above.
    (c) Post-processing steps for GFP channel as above.
    (d) Collect Median Intensity (Density Levels) for GFP channel.
2. Process Nuclei channel
    (a) Pre-processing steps for Nuclei channel - Denoising, kernel size = 3.
    (b) Segmentation for Nuclei channel - optimize for maximum nuclei targeting.
    (c) Post-processing steps for Nuclei channel - Watershed & Sieve.
    (d) Collect cell count (*optional*).
3. Generate overlapping mask
    (a) Create linking target
    (b) Nuclei target 70% overlapping of GFP target
4. Collect data, Median Density & Area from GFP channel

NeuN-Overlap

NeuN-Overlap image processing method uses the same protocol for Nuclei-Overlap method; however with additional NeuN channel.

1. Process GFP channel
    (a) Pre-processing steps for GFP channel as above.
    (b) Segmentation for GFP channel as above.
    (c) Post-processing steps for GFP channel as above.
    (d) Collect Median Intensity (Density Levels) for GFP channel.
2. Process Nuclei channel
    (a) Pre-processing steps for Nuclei channel - Denoising, kernel size = 3.
    (b) Segmentation for Nuclei channel - optimize for maximum nuclei targeting.
    (c) Post-processing steps for Nuclei channel - Watershed & Sieve.
    (d) Collect cell count (*optional*).
3. Process NeuN channel
    (a) Pre-processing steps for NeuN channel - Denoising, kernel size = 3.
    (b) Segmentation for NeuN channel - optimized for maximum target selection.
    (c) Post-processing steps for NeuN channel - Watershed & Sieve.
    (d) Collect cell count (*optional*).
4. Generate overlapping mask
    (a) Create linking target
    (b) Nuclei target 70% overlapping of NeuN target
    (c) Use generated mask to measure on GFP channel.
5. Collect data, Median Density & Area from GFP channel

E.2.2 R programming analysis

GE Developer will produce text files (csv) that can be imported into programs such as R[11].

1. Edit text files by removing the top two lines (either by command-line or by hand) for easy of R programming import.

```
====================================>YFP<====================================
,,Ch 1,Ch 1,Ch 1,
Section,Target,Dens - Levels,Median Dens - Levels,Area - µm2,
A - 1,1,244.64,229.00,250.140,
A - 1,2,319.60,318.50,239.154,
A - 1,3,303.27,259.00,181.690,
A - 1,4,547.05,520.00,128.873,
```

[11]For Excel version 2016, data is truncated and should not be opened in this type of program.

2. The following R script was developed for high-throughput data processing.
    - Well mean
    - Well median
    - Well standard deviation
    - Lower quantile
    - Upper quantile
    - Cell count

```
##Edited Jan 29th, 2016 by Jade Benjamin.
##The following code is a user friendly script for analysis of high-throughput screening data analyzed
from GE INCell Developer Software.

library(gtools) #Libraries required: gtools
drug.library <- ask(msg = "What is the name of drug library?")
assay.name <- paste("_NucleiOverlap_Analysis_", drug.library, sep="")
#Check that input files have been edited.
checkfiles <- function(){
  c <- readline(prompt = "Have the top 2 lines of each input file been deleted? (Y/N): ")
  if(((c != "N") & (c != "n")) & ((c != "Y") & (c != "y"))){
    return(checkfiles())}
  return(c)}

answer <- checkfiles()#If input files have not been edited, then stop script.
if((answer == "N") | (answer =="n")){stop("Please edit files in Notepad!")}
# User chooses directory file with input data.
setwd(choose.dir(caption = "Select Input Folder"))
directory <- getwd()
output_dir <- paste(as.character(choose.dir()), "\\", sep = "")
files_full <- list.files(directory, full.names = T)

well_upper <- function(data){##Calculate well upper quantile
  h1 <- ((length(data)-1)*0.99)+1
  upper <- data[floor(h1)]+((h1-floor(h1))*(data[floor(h1)+1]- data[floor(h1)]))
  return(upper)}
well_lower <- function(data){##Calculate well lower quantile
  h2 <- ((length(data)-1)*0.75)+1
  lower <- data[floor(h2)]+((h2-floor(h2))*(data[floor(h2)+1]- data[floor(h2)]))
  return(lower)}
no.plate <- function(){##Plate correction for large libraries
  n <- readline(prompt = "Please enter the initial plate number: ")
  if(!grepl("^[0-9]+$", n)){
    return(no.plate())}
  return(as.integer(n)-1)}

num <- no.plate()##Start script, enter initial plate number
for (i in seq_along(files_full)){## Read all files into variable
  files <- read.csv(files_full[i])
  well_unique <- unique(files$Section)
  tmp <- seq_along(files_full) + num
  plate_number <- paste("Plate_", tmp[i], "_", sep="")
  output1 <- vector();output2 <- vector();output3 <- vector();output4 <- vector()
  output5 <- vector();output6 <- vector()
  for (j in seq_along(well_unique)){
    wells <- subset(files, Section==well_unique[j])##Seperates each well
    tmp1 <- sort(wells[, 3]) ##pulls out median density
    ##Calculate mean, standard deviation, median, upper and lower quantile, and cell count
    well_mean <- mean(tmp1, na.rm=T);well_sd <- sd(tmp1, na.rm=T)
    well_median <- median(tmp1, na.rm=T);well_upperq <- well_upper(tmp1)
    well_lowerq <- well_lower(tmp1);cell_count <- length(tmp1)
    ##Output each variable
    output1 <- c(output1, well_mean);output2 <- c(output2, well_sd)
    output3 <- c(output3, well_median);output4 <- c(output4, well_upperq)
    output5 <- c(output5, well_lowerq);output6 <- c(output6, cell_count)
  }
  ##plate analysis with background substracted
  plate <- data.frame(Well = well_unique, Mean = output1, Std = output2, Median = output3, Upper =
output4, Lower = output5, Count = output6)
  ##Need to output file with different names based on file name
  plate_name <- paste(output_dir, plate_number, sep="")
  write.table(plate, file = paste(plate_name, Sys.Date(), assay.name, ".csv", sep=""), sep = ",",
append=FALSE, row.names = FALSE, col.names=TRUE)
}
```

3. The R script outputs csv file for down-stream data analysis.
4. Another R script was developed for Well separation for by Well analysis in down-stream data analysis (csv).

```
## Get file list of all data within directory
setwd("E:/DrugHits/HTS1_Analysis/R/Preswesk/")
directory <- getwd()
output_dir <- "E:/DrugHits/HTS1_Analysis/Analysis/Preswesk/"
##MUST EDIT FILES SO THAT TOP TWO LINES ARE DELETED
##USE NOTEPAD SO AS NOT TO LOSS ANY DATA
files_full <- list.files(directory, full.names = T)

## Read all files into variable
for (i in seq_along(files_full)){
  files <- read.csv(files_full[i])
  well_unique <- unique(files$Section)
  tmp <- seq_along(files_full)
  plate_number <- paste("Plate_", tmp[i], "_", sep="")
  output1 <- vector();output2 <- vector();output3 <- vector()
  output4 <- vector();output5 <- vector();output6 <- vector()
  for (j in seq_along(well_unique)){
    wells <- subset(files, Section==well_unique[j])##Seperates each well
    well_density <- wells[, 3] ##pulls out mean density
    tmp1 <- sort(well_density)
    upper_h <- ((length(tmp1)-1)*0.75)+1
    lower_h <- ((length(tmp1)-1)*0.25)+1
    ##Calculate mean, standard deviation, median, upper and lower quantile, and cell count
    well_mean <- mean(tmp1); well_sd <- sd(tmp1)
    well_median <- median(tmp1,na.rm=T)
    well_upperq <- tmp1[floor(upper_h)]+((upper_h-floor(upper_h))*(tmp1[floor(upper_h)+1]- tmp1[floor
(upper_h)]))
    well_lowerq <- tmp1[floor(lower_h)]+((lower_h-floor(lower_h))*(tmp1[floor(lower_h)+1]- tmp1[floor
(lower_h)]))
    cell_count <- length(tmp1)
    output1 <- c(output1, well_mean);output2 <- c(output2, well_sd)##Output each variable
    output3 <- c(output3, well_median);output4 <- c(output4, well_upperq)
    output5 <- c(output5, well_lowerq);output6 <- c(output6, cell_count)
  }
  plate <- data.frame(Well = well_unique, Mean = output1, Std = output2, Median = output3, Upper =
output4, Lower = output5, Count = output6)
  plate_name <- paste(output_dir, plate_number, sep="")
  write.table(plate, file = paste(plate_name, Sys.Date(), "_Well_Analysis_Prestwick.csv", sep=""),
              sep = ",", append=FALSE, row.names = FALSE, col.names=TRUE)
}
```

5. UpperQ (Upper quantile analysis) R script separates out the 75-99 percentile targets per plate.

```
## Get file list of all data within directory
setwd("E:/DrugHits/HTS1_Analysis/R/Preswesk/")
directory <- getwd()
output_dir <- "E:/DrugHits/HTS1_Analysis/Analysis/Preswesk/"
##MUST EDIT FILES SO THAT TOP TWO LINES ARE DELETED
##USE NOTEPAD SO AS NOT TO LOSS ANY DATA
files_full <- list.files(directory, full.names = T)

quantile_mean75 <- function(data, q75, q99){
  results <- vector()
  for (k in floor(q75):floor(q99)){
    holder <- data[floor(k)]+((k-floor(k))*(data[floor(k)+1]- data[floor(k)]))
    results <- c(results, holder)
  }
  well_mean = mean(results)
  return (well_mean)}
quantile_std75 <- function(data, q75, q99){
  output <- vector()
  for (n in floor(q75):floor(q99)){
    holder2 <- data[floor(n)]+((n-floor(n))*(data[floor(n)+1]- data[floor(n)]))
    output <- c(output, holder2)
  }
  well_std = sd(output)
  return (well_std)}

for (i in seq_along(files_full)){## Read all files into variable
  files <- read.csv(files_full[i])
  well_unique <- unique(files$Section)
  tmp <- seq_along(files_full)
  plate_number <- paste("Plate_", tmp[i], "_", sep="")
  output <- vector()
  output2 <- vector()
  for (j in seq_along(well_unique)){
    wells <- subset(files, Section==well_unique[j])##Seperates each well
    well_density <- wells[, 3] ##pulls out mean density
    tmp1 <- sort(well_density)
    h1 <- ((length(tmp1)-1)*0.75)+1
    h2 <- ((length(tmp1)-1)*0.99)+1
    well_median <- median(tmp1,na.rm=T)
    ##Output vector of values between 75% and 99% percentile
    output <- c(output, quantile_mean75(tmp1, h1, h2))
    output2 <- c(output2, quantile_std75(tmp1, h1, h2))
  }
  plate <- data.frame(Well = well_unique, Mean_q75 = output, Std_q75 = output2)
  plate_name <- paste(output_dir, plate_number, sep="")
  write.table(plate, file = paste(plate_name, Sys.Date(), "_Upper_Quartile_Analysis_Prestwick.csv",
sep=""),
              sep = ",", append=FALSE, row.names = FALSE, col.names=TRUE)
}
```

6. Calculates well statistics and outputs csv file.

E.2.3 Shiny App

The app can be found here:
`https://kj-benjamin90.shinyapps.io/angelman-hts-app/`.

There are four main files and/or directories including: 1) ui.R, 2) server.R, 3) bootstrap.css, and 4) training datasets. The `bootstrap.css` file is used for webpage style. The `training datasets` are used to generate simulated plate data.

The app is broken into three major functions. The first checks that the text files have been edited. It throws a warning if it has not been ("No"), and if the files have been ("Yes") then prompts the user to upload the files to be analyzed. The user than has the option to check quality control of the plates be comparing topotecan treated wells with vehicle wells, or to directly download the processed data for downstream analysis.

Quality control of the plates is the second major function. The app looks at an unpaired one-sided Student's T Test (ttest), the z factor (Zfactor), and the strictly standardized mean difference (SSMD). This data can also be downloaded at the click of a button.

The final function is the generation of simulated plate data. This function simulates 9 plate randomly with each download. These files can than be used without outside manipulation for plate analysis and quality control.

The `ui.R` and `server.R` scripts are provided below:

E.2.3.1 ui.R

```
library(shiny)

shinyUI(fluidPage(theme = "bootstrap.css",
  headerPanel("High-throughput Drug Screen Analysis"),
  fluidRow(
    column(4,
           wellPanel(
```

```
          h4("Instructions:"),
          p("This app is for downstream analysis of GE Developer
          image analysis using the NeuN Overplay protocol, which
          outputs median YFP intensity in column 3. It does the
          following: 1) generate output summary statistics (mean,
          standard deviation, median, and cell count) for each well
          per plate imported; 2) quality control data assuming
          positive (col 2 & 23) and negative controls (col 1 & 24)
          in a 384 well plate."))),
column(4, align = "center",
       textInput("drug", label = h4("What is the name of drug
                                    library?"),
                 value = "Enter library name..."),
       numericInput("plates", "Please enter the initial plate
                               number: ",
                    1, min = 1, max = 100),
       h4("Have the top 2 lines of each input file been deleted?"),
          actionButton("yes", "Yes"),
          actionButton("no", "No"),
          conditionalPanel("input.yes",
                           br(),
                           fileInput("files",
                                     "Please choose files below:",
                                     multiple = TRUE,
                                     accept = c(".csv",".txt")),
                           downloadButton("multiDownload",
                                          "Download"))),
column(4,
       wellPanel(
          h4("Generate Simulation Data:"),
          checkboxInput("simulation", "Yes", FALSE),
```

```
            conditionalPanel(
              condition = "input.simulation == true",
              downloadButton("testDownload","Download"),
              helpText("Note: Data is randomly generated with each
                       download. The top 2 lines do not need to
                       be deleted to run analysis successfully."))))
          ),
    hr(),
  fluidRow(
    column(8, align = "center", offset = 2,
           h4("Plate Quality Control"),
           checkboxInput("quality","Yes", FALSE))
  ),
  fluidRow(
    conditionalPanel(
      condition = "input.quality==true",
        column(4,
               h5("Left side of Plate", align = "center"),
               tableOutput("left")),
        column(4,
               h5("Entire Plate", align = "center"),
               tableOutput("all")),
        column(4,
               h5("Right side of Plate", align = "center"),
               tableOutput("right")))),
  fluidRow(
    conditionalPanel(
      condition = "input.quality==true",
        column(8, align="center", offset = 2,
               downloadButton("qualDownload", "Download"))))
))
```

E.2.3.2 server.R

```
library(shiny)
library(plyr)
library(dplyr)

## Define functions in script
zfactor <- function(data1,data2){
    data1 <- as.numeric(unlist(data1))
    data2 <- as.numeric(unlist(data2))
    z <- 1 - ((3*(sd(data1, na.rm = T) + sd(data2, na.rm = T))) /
              (abs(mean(data1, na.rm = T) - mean(data2, na.rm = T))))
  return(z)
}
ssmd <- function(data1,data2){
    data1 <- as.numeric(unlist(data1))
    data2 <- as.numeric(unlist(data2))
    beta <- (median(data1, na.rm = T) - median(data2, na.rm = T)) /
        (sqrt((mad(data1, na.rm = T))^2 + (mad(data2, na.rm = T))^2))
  return(beta)
}

shinyServer(function(input, output){
#Input files
    library.name <- reactive({
        paste0("_NeuNOverlap_Analysis_", input$drug, "_")
    })
    observeEvent(input$no, {
        showModal(modalDialog(
            title = "Important message",
            "Please edit files in a text editor like Notepad!",
```

```
        easyClose = TRUE
    ))
})
num <- reactive({
    input$plates - 1
})
plate <- reactive({
    inFile <- input$files
    if(is.null(inFile)){
        return(NULL)
    }
    inFile <- inFile[order(inFile[,1]),]
    datalist <- list()
    tmp <- input$plates
    for (i in seq_len(nrow(inFile))){
        lst <- read.csv(inFile[[i, 'datapath']])
        unique.wells <- unique(lst$Section)
        out.mean <- vector()
        out.sd <- vector()
        out.med <- vector()
        out.count <- vector()
        for (j in seq_along(unique.wells)){
            wells <- subset(lst, Section==unique.wells[j])
            well.mean  <- mean(wells[,3], na.rm = TRUE)
            well.sd    <- sd(wells[,3], na.rm = TRUE)
            well.med   <- median(wells[,3], na.rm = TRUE)
            cell.count <- length(wells[,3])

            out.mean  <- c(out.mean, well.mean)
            out.sd    <- c(out.sd, well.sd)
            out.med   <- c(out.med, well.med)
```

```
            out.count <- c(out.count, cell.count)
        }
        plate.name <- rep(paste0("Plate_", tmp),
                          length(unique.wells))
        plate <- data.frame(PlateID = plate.name,
                            Well = unique.wells,
                            Mean = out.mean,
                            Std = out.sd,
                            Median = out.med,
                            Count = out.count)
        datalist[[i]] <- plate
        tmp <- tmp + 1
    }
    plate.data <- dplyr::bind_rows(datalist)
    return(plate.data)
})
left <- reactive({
    leftData <- plate()
    if(is.null(leftData)){
        return(NULL)
    }
    plateNames1 <- unique(leftData$PlateID)
    dlist1 <- plyr::dlply(leftData, "PlateID",
                          function(x) x[, c("Well", "Median")])
    stat1  <- vector()
    stat2  <- vector()
    stat3  <- vector()
    for (k in plateNames1){
        file <- dlist1[[k]]
        file$Well <- as.character(file$Well)
        well.split <- ldply(strsplit(file$Well, " - "))
```

```
        names(well.split) <- c("Well.Letter", "Well.Number")
        new.file <- cbind(well.split, file)
        txt   <- subset(new.file, Well.Number == "2",
                        select = Median)
        ntxt  <- subset(new.file, Well.Number == "1",
                        select = Median)
        stat1 <- c(stat1, t.test(txt,ntxt,
                                 alternative="greater")$p.value)
        stat2 <- c(stat2, zfactor(txt,ntxt))
        stat3 <- c(stat3, ssmd(txt,ntxt))
    }
    pvalues <- data.frame(ID = plateNames1, Ttest = stat1,
                          Zfactor = stat2, SSMD = stat3)
    return(pvalues)
})
all <- reactive({
    allData <- plate()
    if(is.null(allData)){
        return(NULL)
    }
    plateNames2 <- unique(allData$PlateID)
    dlist2 <- plyr::dlply(allData, "PlateID",
                          function(x) x[, c("Well", "Mean",
                                            "Std", "Median",
                                            "Count")])
    stat1 <- vector()
    stat2 <- vector()
    stat3 <- vector()
    for (j in plateNames2){
        file <- dlist2[[j]]
        file$Well <- as.character(file$Well)
```

```
        well.split <- ldply(strsplit(file$Well, " - "))
        names(well.split) <- c("Well.Letter", "Well.Number")
        new.file <- cbind(file, well.split)
        txt  <- subset(new.file, Well.Number == "2" |
                                 Well.Number == "23",
                       select = Median)
        ntxt <- subset(new.file, Well.Number == "1" |
                                 Well.Number == "24",
                       select = Median)
        stat1 <- c(stat1, t.test(txt, ntxt,
                                 alternative="greater")$p.value)
        stat2 <- c(stat2, zfactor(txt, ntxt))
        stat3 <- c(stat3, ssmd(txt, ntxt))
    }
    pvalues <- data.frame(ID = plateNames2, Ttest = stat1,
                          Zfactor = stat2, SSMD = stat3)
    return(pvalues)
})
right <- reactive({
    rightData <- plate()
    if(is.null(rightData)){
        return(NULL)
    }
    plateNames3 <- unique(rightData$PlateID)
    dlist1 <- plyr::dlply(rightData, "PlateID",
                          function(x) x[, c("Well", "Median")])
    stat1  <- vector()
    stat2  <- vector()
    stat3  <- vector()
    for (k in plateNames3){
        file <- dlist1[[k]]
```

```
        file$Well <- as.character(file$Well)
        well.split <- ldply(strsplit(file$Well, " - "))
        names(well.split) <- c("Well.Letter", "Well.Number")
        new.file <- cbind(well.split, file)
        txt   <- subset(new.file, Well.Number == "23",
                        select = Median)
        ntxt  <- subset(new.file, Well.Number == "24",
                        select = Median)
        stat1 <- c(stat1, t.test(txt, ntxt,
                                 alternative="greater")$p.value)
        stat2 <- c(stat2, zfactor(txt, ntxt))
        stat3 <- c(stat3, ssmd(txt, ntxt))
    }
    pvalues <- data.frame(ID = plateNames3, Ttest = stat1,
                          Zfactor = stat2, SSMD = stat3)
    return(pvalues)
})
output$multiDownload <- downloadHandler(
    filename = function(){
        paste0("Plate-DataTables_", library.name(),
               gsub("\\D", "_", Sys.time()), ".zip")
    },
    content = function(file){
        tempData <- plate()
        if(is.null(tempData)){
            return(NULL)
        }
        do.call(file.remove, list(list.files(tempdir(),
                                  full.names = TRUE)))
        tempdir <- tempdir()
        plateNames <- unique(tempData$PlateID)
```

```
        dlist <- plyr::dlply(tempData, "PlateID",
                             function(x) x[, c("Well", "Mean",
                                               "Std", "Median",
                                               "Count")])
        for(i in plateNames){
            write.csv(x = dlist[[i]],
                      file = paste0(tempdir, "/", i,
                                    library.name(),
                                    Sys.Date(),".csv"),
                      row.names = FALSE)
        }
        zip(zipfile = file, files = tempdir)
    }
)
output$left <- renderTable({
    if(input$quality==TRUE){
        left()
    }
})
output$all <- renderTable({
    if(input$quality==TRUE){
        all()
    }
})
output$right <- renderTable({
    if(input$quality==TRUE){
        right()
    }
})
output$qualDownload <- downloadHandler(
    filename = function(){
```

```
        paste0("Plate_QualityControl", library.name(),
               gsub("\\D", "_", Sys.time()), ".zip")
    },
    content = function(file){
        do.call(file.remove, list(list.files(
                                tempdir(), full.names = TRUE)))
        tempdir2 <- tempdir()
        write.csv(x = left(),
                  file = paste0(tempdir2, "/",
                                "Plate_left-side", ".csv"),
                  row.names = FALSE)
        write.csv(x = all(),
                  file = paste0(tempdir2, "/",
                                "Plate_entire", ".csv"),
                  row.names = FALSE)
        write.csv(x = right(),
                  file = paste0(tempdir2, "/",
                                "Plate_right-side", ".csv"),
                  row.names = FALSE)
        zip(zipfile = file, files = tempdir2)
    }
)
output$testDownload <- downloadHandler(
    filename = function(){
        paste0("TestData_NeuNOverlap_",
               gsub("\\D", "_", Sys.time()), ".zip")
    },
    content = function(file){
        do.call(file.remove, list(list.files(
                                tempdir(), full.names = TRUE)))
        tempdir3 <- tempdir()
```

```
files_full <- list.files("./train-data/",
                         full.names = TRUE)
for (k in seq_along(files_full)){
    files <- read.csv(files_full[k])
    files$Section <- as.character(files$Section)
    well.split <- ldply(strsplit(files$Section, " - "))
    names(well.split) <- c("Well.Letter", "Well.Number")
    new.file <- cbind(well.split, files)
    txt.left   <- subset(new.file, Well.Number == "2",
                         select = YFPintensity)
    ntxt.left  <- subset(new.file, Well.Number == "1",
                         select = YFPintensity)
    txt.right  <- subset(new.file, Well.Number == "23",
                         select = YFPintensity)
    ntxt.right <- subset(new.file, Well.Number == "24",
                         select = YFPintensity)
    rest       <- subset(new.file, Well.Number != "1" &
                                   Well.Number != "2" &
                                   Well.Number != "23" &
                                   Well.Number != "24",
                         select = c(Section, YFPintensity))
## This returns several data.frames with $YFPintensity for values.
    plate <- data.frame()
    ## Column 1 (ntxt.left)
    size1 <- sample(5:400, 16)
    row1  <- vector()
    for (j in 1:16){
      yfp1    <- sample(ntxt.left$YFPintensity,
                        size = size1[j], replace = TRUE)
      well1   <- rep(paste(LETTERS[j], "-", "1", sep=" "),
                     size1[j])
```

```
  df.tmp1 <- cbind(well1, 1:size1[j], yfp1)
  row1    <- rbind(row1, df.tmp1)
}
plate <- row1
## Column 2 (txt.left)
size2 <- sample(5:400, 16)
row2  <- vector()
for (j in 1:16){
  yfp2    <- sample(txt.left$YFPintensity,
                    size = size2[j], replace = TRUE)
  well2   <- rep(paste(LETTERS[j], "-", "2", sep=" "),
                 size2[j])
  df.tmp2 <- cbind(well2, 1:size2[j], yfp2)
  row2    <- rbind(row2, df.tmp2)
}
plate <- rbind(plate, row2)
## Columns 3 to 22 (rest)
for (i in 3:22){
  size <- sample(5:400, 16)
  num <- i
  row <- vector()
  for (j in 1:16){
    yfp    <- sample(rest$YFPintensity,
                     size = size[j], replace = TRUE)
    well   <- rep(paste(LETTERS[j], "-", num, sep=" "),
                  size[j])
    df.tmp <- cbind(well, 1:size[j], yfp)
    row    <- rbind(row, df.tmp)
  }
  plate <- rbind(plate, row)
}
```

```
## Column 23 (txt.right)
size23 <- sample(5:400, 16)
row23  <- vector()
for (j in 1:16){
  yfp23    <- sample(txt.right$YFPintensity,
                     size = size23[j], replace = TRUE)
  well23   <- rep(paste(LETTERS[j],"-","23",sep=" "),
                  size23[j])
  df.tmp23 <- cbind(well23, 1:size23[j], yfp23)
  row23    <- rbind(row23, df.tmp23)
}
plate <- rbind(plate, row23)
## Column 24 (ntxt.right)
size24 <- sample(5:400, 16)
row24  <- vector()
for (j in 1:16){
  yfp24    <- sample(ntxt.right$YFPintensity,
                     size = size24[j], replace = TRUE)
  well24   <- rep(paste(LETTERS[j],"-","24",sep=" "),
                  size24[j])
  df.tmp24 <- cbind(well24, 1:size24[j], yfp24)
  row24    <- rbind(row24, df.tmp24)
}
plate <- rbind(plate, row24)
plate.data <- as.data.frame(plate)
names(plate.data) <- c("Section", "Zipget",
                       "YFPintensity")
tmp <- seq_along(files_full)
plate_number <- paste0("Plate_", tmp[k], "_")
write.csv(plate.data,
          file = paste0(tempdir3, "/", plate_number,
```

```
                                        Sys.Date(),
                                        "_SimulatedData",".csv"),
                            row.names = FALSE)
                }
                zip(zipfile = file, files = tempdir3)
            }
        )
})
```

E.3 Immunocytochemistry

E.3.1 Fixation - Cell Culture

1. Wash media from slides gently with 1xPBS, twice.
2. Remove 1xPBS and add 4% PAF/4% Sucrose to each coverslip[12].
3. Agitate for 10 min at room temperature.
4. Rinse coverslips gently twice with 1xPBS.
5. Add 1xPBS to wells and cover with aluminum foil.
6. Store at 4°C.

E.3.2 Staining - Cell Culture

1. Make up 5% goat serum fresh in 0.3% Triton X-100 in 1xPBS.
2. Transfer coverslips to new plate, blotting off excess PBS.
3. Add 5% goat serum fresh to coverslips and incubate at room temperature with slight agitation for 1 h.
4. Prepare primary antibodies using 5% goat serum.
   - anti-GFP diluted at 1:1000 (anti-rabbit)
   - anti-NeuN diluted at 1:250 (anti-mouse)
5. Replace 5% blocking solution (goat serum) with primary antibodies[13].
6. Incubate primary antibodies for 1 h at room temperature with slight agitation.
7. Wash three times coverslips with 0.1% Tween20 in 1xPBS; 10 min at room temperature with slight agitation.
8. Prepare secondary antibodies using 5% goat serum during last wash at 1:200 dilution.
9. Replace last wash with secondary antibody solution.

[12] Use freshly made fixation solution ($\leq$ 7 days).
[13] Be careful of bubbles.

10. Incubate secondary antibodies for 1 h at room temperature with slight agitation covered with aluminum foil[14].
11. Wash twice with 0.1% Tween20 in 1xPBS for 10 min each at room temperature with slight agitation.
12. Replace wash with TO-PRO-3 (1:1000) in 0.1% Tween20 in 1xPBS for 10 min at room temperature with slight agitation.
13. Wash coverslips with 0.1% Tween20 in 1xPBS for 10 min with slight agitation at room temperature.
14. Store at 4°C until ready for imaging.

[14] Aluminum foil and dark conditions should be used for the remainder of staining protocol.

E.4 Rapid Amplification of cDNA Ends (3')

The following uses the 3' RACE System (Cat. No. 18373019) from ThermoFisher Scientific. Everything but RNA, Taq Polymerase, and gene specific primers provided within the kit, as such developing gene specific primers is the first step for 3' RACE.

E.4.1 First-Strand cDNA Synthesis

1. Combine DEPC-treated water and 2.5 - 5 μg of total RNA to a final volume of 11 μl to a PCR tube.
2. Add 1 μl of 10 μM AP solution and mix with a flick.
3. Spin down briefly and heat mixture in PCR thermocycler.
   (a) 70°C for 10 min
   (b) 4°C for $>$ 1 min
4. Spin down briefly and add the following
   (a) 2 μl of 10X PCR buffer
   (b) 2 μl of 25 mM $MgCl_2$
   (c) 1 μl of 10 mM dNTP mix
   (d) 2 μl of 0.1 M DTT
5. Mix gently and spin down before placing on PCR thermocycler held at 42°C.
6. Equilibrate sample for 2 - 5 min.
7. Add 1 μl of SuperScript^{TM} II RT and perform the following
   (a) 42°C for 50 min
   (b) 70°C for 15 min
   (c) Hold at 4°C
8. Spin down and add 1 μl of RNase H.
9. Mix and spin down before incubating for 20 min at 37°C.
10. Store at -20°C or proceed to Amplification of Target cDNA.

E.4.2 Amplification of Target cDNA

1. Combine the following in a PCR tube.
   (a) 14.6 μl of DEPC-treated water
   (b) 2 μl of 10X PCR buffer
   (c) 1.2 μl of $MgCl_2$
   (d) 0.4 μl of Gene Specific forward primer
   (e) 0.4 μl of AUAP (abridged universal amplification primer) reverse primer
   (f) 0.4 μl of dNTP
   (g) 0.8 μl of cDNA (from First Strand Synthesis)
   (h) 0.2 μl of Taq Polymerase
2. Place on thermocycler[15]
   (a) 1 cycle of 94°C for 3 min
   (b) 35 cycles of
      - 94°C for 30 s
      - (gradient) 65°C for 60 s
      - 72°C for 60 s
   (c) 1 cycle of 72°C for 5 min
   (d) Hold at 4°C
3. Preform second amplification in new PCR tube.
   (a) 36.5 μl of DEPC-treated water
   (b) 5.0 μl of 10X PCR buffer
   (c) 3.0 μl of $MgCl_2$
   (d) 1.0 μl of Nested-Gene Specific forward primer
   (e) 1.0 μl of AUAP reverse primer

[15]For optimal amplification a gradient PCR should be run.

(f) 1.0 μl of dNTP

(g) 2.0 μl of cDNA (from first amplification - optimized)

(h) 0.5 μl of Taq Polymerase

4. Run at optimal melting temperature on thermocycler.

(a) 1 cycle of 94°C for 3 min

(b) 35 cycles of

- 94°C for 30 s
- 65°C for 60 s
- 72°C for 60 s

(c) 1 cycle of 72°C for 5 min

(d) Hold at 4°C

5. Place 4 μl of PCR reaction into new tube and store at 4°C.

E.4.3 Electroporation Cloning & Insert Verification

1. Run PCR product ($\sim$46 μl) out on 1.5% agarose gel (100V, $>$ 1 h).
2. Cut out amplified band(s) for gel purification.
3. Use Gel Purification kit to purify PCR products.
4. Ligate PCR products into TOPO TA vector.

(a) Using TOPO TA Cloning kit, add the following

- 0.5 μl of 2.1 TOPO Vector
- 0.5 μl of RNase-free water
- 1.0 μl of salt solution
- 4.0 μl of PCR product (direct, and gel purified in different tubes)

(b) Incubate for 30 min at room temperature

(c) Dilute ligation reactions $1/4$ with RNase-free water (18 μl of H_2O)

(d) Add 2 μl to electro-competent cells for electroporation

(e) Add mixature to curvettes for electroporation

(f) After electroporation add SOC media to cells

(g) Shake at 37°C for 1 h

(h) Plate on Ampicilin+ and X-gal treated Luria broth (LB)-agar plates

- 50 μl plate
- High concentration plate

(i) Culture overnight at 37°C

5. Perform plasmid DNA purification and measure DNA with nanodrop.
6. Verify insertion with restriction enzyme (RE) digestion (EcoRI-HF, NEB).

(a) Combine the following in PCR tube

- 1 μg of plasmid DNA
- 5 μl of CutSmart Buffer
- 1 μl of RE
- Water to final volume of 50 μl

(b) On thermocycle

- 37°C for 40 min
- 65°C for 20 min
- Hold at 4°C

(c) Run out on 1 - 1.5% agarose gel (100V, $\sim$ 1 h)

(d) Image gel for insert verification

7. Send DNA for sequencing.

E.5 Reverse-transcription PCR/qPCR

The SuperScript® IV First-Strand Synthesis System from ThermoFisher Scientific is used for the following protocol.

E.5.1 First-Strand Synthesis

1. Using DNase treated RNA (2-5 μg total RNA), combine with 1 μl of 50 μM Oligo d$(T)_{20}$ primer, 1 μl of 10 mM dNTP mix, and DEPC-treated water to final volume of 13 μl in PCR tube.
2. Mix gently and spin down before heating on thermocycler at 65°C for 5 min.
3. Incubate at 4°C for $\leq$ 1 min.
4. Make a mastermix of the following components
   (a) 4 μl of 5X SSIV Buffer
   (b) 1 μl of 100 mM DTT
   (c) 1 μl of Ribonuclease Inhibitor
   (d) 1 μl of either SuperScript® IV Reverse Transcriptase or DEPC-treated water
5. Add 7 μl of mastermix to primer annealed RNA.
6. Incubate mixture for 10 min at 50°C followed by 10 min at 80°C.
7. (*optional*) Add 1 μl of RNase H to samples and incubate for 20 min at 37°C.
8. Store cDNA at -20°C.

For reverse-transcription PCR, dilute cDNA and proceed to running PCR reactions.

E.5.2 SYBR Green qPCR

1. Design qPCR primers and verify size and single band with PCR amplification.
2. Setup qPCR plate layout.
3. Dilute enough cDNA at 1:4 for experiment.

4. Create 1:4 serial dilution standards (≤ 5) of pooled cDNA[16].
5. Dilute samples 1:5 (1:20 dilution from original cDNA).
6. Make mastermix for each primer set
   (a) 5 μl of 2X SYBR Green Buffer
   (b) 500 nM forward primer (final)
   (c) 500 nM reverse primer (final)
   (d) DEPC-treated water to final volume of 8 μl
7. Add mastermix to plate following layout.
8. Add samples to plate.
9. Seal off plate, vortex, and spin down.
10. Run qPCR reaction
    (a) 1 cycle of 50°C for 2 min
    (b) 1 cycle of 95°C for 10 min
    (c) 40 cycles of
        - 95°C for 15 s
        - 60°C for 60 s
    (d) Dissociation Curve
        - 95°C for 15 s
        - 60°C for 20 s
        - 72°C for 15 s
11. Analysis data.

[16]Recommend using 4 technical replicas for experimentation.

E.6 RNA Extraction

For all protocols, work in laminar flow-hood if possible with work area and pipettes cleaned with RNA-zap if no RNA only pipettes designated. Additionally, all solutions not provided by the kit should be made with DEPC-treated water (i.e. sodium acetate and 75% EtOH).

E.6.1 Tissue RNA isolation

The following is a TRIzol method for RNA extraction and isolation.

TRIzol extraction

1. Set centrifuge for 4°C.
2. Prepare Sonicator for RNA extraction by cleaning with RNase-Zap and DEPC-treated water.
3. Homogenize samples (held on ice) in 1 ml of TRIzol for 60 s[17].
4. Incubate at room temperature for 5 min.
5. Add 200 μl of Chloroform and shake by hand vigorously for 15 s.
6. Incubate for 2 - 3 min at room temperature.
7. Centrifuge ($12,000 \times g$) for 15 min at 4°C[18].
8. Transfer aqueous phase to new RNase-free tube.
9. Add 500 μl of isopropanol and incubate for 10 min at room temperature.
10. Centrifuge ($12,000 \times g$) for 10 min at room temperature.
11. Aspirate and discard supernatant and resuspend pellet in 1 ml of 75% EtOH.
12. Vortex to mix and centrifuge at ($7,500 \times g$) for 5 min at 4°C.
13. Briefly dry RNA pellet before adding at least 100 μl of DEPC-treated water.
14. Incubate for 10 min at 60°C.

[17] Clean Sonicator after every use with RNase-Zap and DEPC-treated water.
[18] RNA will be in colorless upper phase.

15. Store at -80°C.

RNA clean-up & concentration[19]

1. Add $\frac{1}{10}$ volume of sodium acetate (i.e. 10 μl from above extraction) to sample.
2. Add 1 volume of isopropanol to sample.
3. Incubate overnight at -20°C.
4. Centrifuge for 15 min at $15,000 \times g$ and discard supernatant carefully.
5. Add 1 ml of 70% EtOH and centrifuge for 2 min at $20,000 \times g$.
6. Aspirate supernatant and dry pellet.
7. Re-suspend with RNase-free water.

E.6.2 Cell culture RNA isolation

The following protocol is for a monolayer of cells ($\leq 5 \times 10^6$) using the PureLink® RNA Mini Kit.

1. Prepare lysis buffer by adding 2-mercaptoethanol (βME) to Lysis Buffer (10 μl per 1 ml of Lysis solution).
2. Rinse cells gently twice with 1xPBS.
3. Add 300 μl of prepared lysis buffer to plated cells.
4. Incubate at room temperature for 3 - 5 min with back and forth motion.
5. Pipette up and down ten times to break up cells.
6. Transfer lysate to new RNase-free tube.
7. Add 1 volume of 70% EtOH to each cell lysate.
8. Vortex until no precipitate is visible.
9. Transfer up to 700 μl of sample into spin cartridge.
10. Centrifuge for 30 s at $12,000 \times g$ and discard flow-through[20].

[19] There are kits that can do this as well.

[20] The following centrifuge are done at room temperature.

11. Repeat until the entire sample has been processed.
12. Add 700 μl of Wash Buffer I to spin cartridge.
13. Centrifuge for 30 s at $12,000 \times g$.
14. Discard flow-through and collection tube. Place spin cartridge in new collection tube.
15. Add 500 μl of Wash Buffer II and centrifuge for 30 s at $12,000 \times g$. Discard flow-through and repeat wash.
16. Centrifuge spin cartridge at $12,000 \times g$ for 90 s to dry membrane.
17. Discard collection tube and insert spin cartridge into recovery tube.
18. Add 30 μl of RNase-free water to center of spin cartridge and incubate for 1 min at room temperature.
19. Centrifuge spin cartridge at $12,000 \times g$ for 2 min.
20. Store purified RNA at -80°C.

E.6.3 Cytoplasm & nuclear cell RNA isolation

The following protocol is from the Ambion® PARIS^{TM} Kit for RNA extraction.

1. Prepare reagents and setup equipment.
   (a) Set heat block to 95°C
   (b) Set centrifuge to 4°C
   (c) Add βME to 2X Lysis Buffer[21]
   (d) Add 100% EtOH to 2/3 Wash Buffer[22]
2. Place Cell Fractionation Buffer on ice before use.
3. Allow 2X Lysis Buffer and Wash Solution 1 to equilibrate to room temperature.
4. Heat an aliquot of Elution Solution to 95°C on heat block.
5. Rinse monolayer of cells ($\leq 10 \times 10^6$) with 1xPBS.

[21] Only add for first time use.
[22] Only add for first time use.

6. Add 250 μl of Cell Fractionation Buffer to each well[23].
7. Incubate at room temperature for 5 - 10 min with back and forth motion, or until cells completely detached from plate.
8. Transfer to RNase-free tube and incubate on ice for 10 min.
9. Centrifuge samples at 4°C for 5 min at $500 \times g$.
10. Carefully transfer supernatant to new RNase-free tube and place on ice. This is the cytoplasmic fraction.
11. Add 200 μl of Cell Fractionation Buffer to pelleted nuclear fraction, and gently re-suspend pellet.
12. Spin down at 4°C for 1 min, $500 \times g$.
13. Aspirate and discard supernatant.
14. Add 1 volume equal to cytoplasmic volume of Cell Fractionation Buffer.
15. Vortex vigorously to lyse nuclei, then place on ice.
16. Add 1:1 2X Lysis Buffer to cytoplasmic and nuclear fractions and invert tubes several time to mix well.
17. (*optional*) Cellular debris removal[24]
    - Pass lysate through syringe needle several times to reduce viscosity and remove debris
    - Spin down and place supernatant into new RNase-free tube
18. Add 1 volume of 100% EtOH to mixture[25] and invert several times.
19. Add $\leq$ 700 μl of the sample to the filter cartridge.
20. Centrifuge for 30 s ($12,000 \times g$) and discard flow-through.
21. Repeat until all the sample as been passed through the filter.
22. Add 700 μl of Wash Solution 1 to column.

[23] Assuming 6-well plate
[24] If there is any debris or the samples are too viscous, then they will not pass through the filter cartridges.
[25] This is the volume before Lysis Buffer added.

23. Centrifuge for 30 s ($12,000 \times g$) and discard flow-through.
24. Add 500 μl of Wash Solution 2/3 to column.
25. Centrifuge for 30 s ($12,000 \times g$) and discard flow-through. Repeat wash once.
26. Place filter cartridge into new collection tube and add 40 - 60 μl of Elution Buffer.
27. Centrifuge for 30 s ($12,000 \times g$).
28. Add 10 - 60 μl of Elution Buffer to filter cartridge and centrifuge for 30 s ($12,000 \times g$).
29. Store purified RNA at -80°C.

E.7 RNA-seq Analysis

All work performed on Centos 7: Linux, kernal release: 3.10.0-514.6.1.el7.x86_64, on 64-bit system. The computer has an AMD FX(tm)-8350 Eight-Core Processor, 32G of RAM, 4.6T of working disk space with an additional 2T of backup disk space.[26]

E.7.1 Software installation

Tools used in this dissertation include the following: FastQC, Trimmomatics, HISAT2, SAMtools, StringTie, Gffcompare, Gffread[27], BEDtools, BCFtools, GENESEQER, IGV, Python, R, ballgown, edgeR, and Rsubread.[28] I recommend installing most of these softwares locally and adding to PATH, as some of them may have library issues that can mess with the system.

Tools set up in one directory (tools) within RNA-seq directory (rna-seq).

SAMtools & BCFtools

```
cd ~/rna-seq/tools/
git clone https://github.com/samtools/samtools.git
git clone https://github.com/samtools/bcftools.git
cd samtools
make -j4
cd ../bcftools
make -j4
```

HISAT2 & StringTie

[29]

[26]This is a home built computer.

[27]This is part of Cufflinks, but can be downloaded separately.

[28]If working on a cluster, then you should check to see what software they have. If something is not present, put in a request. Additionally, do not expect to use the latest updates if working on a cluster.

[29]Version 2.0.4 of HISAT2 is not the lastest.

```
cd ~/rna-seq/tools/
#download with wget (hisat2-2.0.4-Linux_x86_64.zip)
unzip hisat2-2.0.4-Linux_x86_64.zip
git clone https://github.com/gpertea/stringtie
cd stringtie
make release
```

FastQC

```
cd ~/rna-seq/tools/
#download with wget (fastqc_v0.11.5.zip)
unzip fastqc_v0.11.5.zip
chmod 755 ./FastQC/fastqc
```

Trimmomatics

```
cd ~/rna-seq/tools/
#download with wget (Trimmomatic-0.36.zip)
unzip Trimmomatic-0.36.zip
mkdir -p ~/rna-seq/refs/adapters
cp -v ./Trimmomatic-0.36/adapters/* ~/rna-seq/refs/adapters/
```

Gffcompare & Gffread

```
cd ~/rna-seq/tools/
git clone https://github.com/gpertea/gffcompare.git
git clone https://github.com/gpertea/gffread.git
cd gffcompare
make release
cd ../gffread
```

```
make -j4
```

BEDtools

```
cd ~/rna-seq/tools/
git clone https://github.com/arq5x/bedtools2.git
cd bedtools2
make -j4
```

IGV

```
cd ~/rna-seq/tools/
#download with wget (IGV_2.3.91.zip)
unzip IGV_2.3.91.zip
chmod 755 ./IGV_2.3.91/igv.sh
```

GENESEQER

```
cd ~/rna-seq/tools/
#download with wget (GeneSeqer-2-26-2014.tar.gz)
tar -zxvf GeneSeqer-2-26-2014.tar.gz
cd GENESEQER
make -j4
```

R & Python

First check to see if R and Python are already installed with `which R` and `which python`. If they are there, then R will often prompt user to update to latest version when in application, which can be done as root. Note of caution: Python 2 is not the same as Python 3, so if you are using MISO and have Python 3 installed, then you'll need to download Python 2 for it to work as MISO was built with Python 2.

```
sudo yum install R.x86_64 -y
sudo yum install python2-pip.noarch python34u-pip.noarch -y
```

Note: Installing pip for python will also install python as a dependency.

Bioconductor & Packages

This can be done as a local install or global with root access. If you have root access enter R with `sudo` command.

```
R #Launches R
install.packages(c("devtools","dplyr","gplots","ggplot2"))
#Follow prompting
source("http://bioconductor.org/biocLite.R")
biocLite(c("genefilter","ballgown","edgeR","Rsubread"))
quit(save="no")
```

Clean up tar files with `rm -rf`. Add the tools location to PATH.

Add to PATH

```
emacs -nw ~/.bashrc

# Load_PATHS
export PATH=$HOME/bin:$PATH
export PATH=/usr/local/bin:$PATH
export RNA_HOME=$HOME/rna-seq
export RNA_REFS=$RNA_HOME/refs
export RNA_ALIGN=$RNA_HOME/alignments
export PATH=$RNA_HOME/tools/samtools:
       $RNA_HOME/tools/gffread:
```

```
        $RNA_HOME/tools/FastQC:
        $RNA_HOME/tools/hisat2-2.0.4:
        $RNA_HOME/tools/bedtools2/bin:
        $RNA_HOME/tools/stringtie:
        $RNA_HOME/tools/gffcompare:
        $RNA_HOME/tools/Trimmomatic-0.36:
        $RNA_HOME/tools/bcftools:
        $RNA_HOME/tools/IGV_2.3.90:
        $RNA_HOME/tools:GENESEQER/bin:
export LD_LIBRARY_PATH="/usr/local/lib:$LD_LIBRARY_PATH"
```

The `export PATH` should be all one line. The indentation is for viewing in this dissertation. Additionally, if you don't have `emacs`, any editor will can be used. Furthermore, if you don't already have `git` downloaded you can just use `wget` for everything.

E.7.2 Reference genomes and annotation

To get everything at once, Illumina's iGenome can be used, but there will need to be some cleaning up of unnecessary files to save disk space.

In a bash script, run with `bash script.name.sh`. Or change permissions and run without bash command.

```
#!/bin/bash
mkdir -p $RNA_REFS/human/hg19/fasta
mkdir -p $RNA_REFS/human/hg19/annotation
mkdir -p $RNA_REFS/mouse/mm9/fasta
mkdir -p $RNA_REFS/mouse/mm9/annotation


cd $RNA_REFS/
```

wget ftp://igenome:G3nom3s4u@ussd-ftp.illumina.com/Homo_sapiens/UCSC/hg19/Homo_sapiens_UCSC_hg19.tar.gz

wget ftp://igenome:G3nom3s4u@ussd-ftp.illumina.com/Mus_musculus/UCSC/mm9/Mus_musculus_UCSC_mm9.tar.gz

Copy over only Sequence and Annotation files to fasta and annotation directories: Whole genome fasta, individual chromosome fasta, and current annotation directory. Remove the rest with `rm -rf`. `gzip` all fasta files that will not be used to save disk space and remove TAR files.

Once the fasta files are downloaded, index the file you plan to use for alignment with SAMtools.

```
samtools faidx chr15.fa
```

E.7.3 Indexing with HISAT2

```
mkdir -p $RNA_REFS/human/hg19/indexes/splice-exons
mkdir -p $RNA_REFS/mouse/mm9/indexes/splice-exons
cd $RNA_REFS/human/hg19/indexes
hisat2_extract_splice_sites.py ../annotation/Genes/gene.gtf > \
                               ./splice-exons/splicesites.hg19.tsv
hisat2_extract_exons.py ../annotation/Genes/gene.gtf > \
                        ./splice-exons/exons.hg19.tsv
hisat2-build -p 8 --ss ./splice-exons/splicesites.hg19.tsv \
             --exon ./splice-exons/exons.hg19.tsv \
             ../fasta/chr15.fa chr15
cd $RNA_REFS/mouse/mm9/indexes
```

```
hisat2_extract_splice_sites.py ../annotation/Genes/gene.gtf > \
                                ./splice-exons/splicesites.mm9.tsv
hisat2_extract_exons.py ../annotation/Genes/gene.gtf > \
                         ./splice-exons/exons.mm9.tsv
hisat2-build -p 8 --ss ./splice-exons/splicesites.mm9.tsv \
             --exon ./splice-exons/exons.mm9.tsv \
             ../fasta/chr7.fa chr7
```

E.7.4 Downloading Data

This can be done in one of two ways. In a bash script with all of the locations listed, or with a for-loop.

```
#!/bin/bash
mkdir -p $RNA_HOME/datasets/human/hpa/reads/fastqc-report
cd $RNA_HOME/datasets/human/hpa/reads

wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR315/ERR315477/ERR315477_
1.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR315/ERR315477/ERR315477_
2.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR315/ERR315455/ERR315455_
1.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR315/ERR315455/ERR315455_
2.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR315/ERR315432/ERR315432_
1.fastq.gz
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR315/ERR315432/ERR315432_
2.fastq.gz
```

```
...
fastqc -t 8 *fastq.gz -o ./fastqc-report
```

```
#!/bin/bash
mkdir -p $RNA_HOME/datasets/human/hpa/reads/fastqc-report
cd $RNA_HOME/datasets/human/hpa/reads

for i in {432,455,477}
do
    wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR315/ERR315${i}/ERR315${i}_1.fastq.gz
    wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR315/ERR315${i}/ERR315${i}_2.fastq.gz
done
fastqc -t 8 *fastq.gz -o ./fastqc-report
```

If using a for-loop, can add all of the data at once for the same project. This is a simpliar script.

E.7.5 Quality Control

FastQC

FastQC functions to provide a way to view quality of raw sequence data working with BAM, SAM, or fastq files it exports data as pdfs (zipped) and HTML for easy viewing in web browser. It can be run interactively with it's JAVA online function, or as I use it in the above section offline as a part of a large-scale pipeline. To view the html files:

```
cd $RNA_HOME/datasets/human/hpa/reads/fastqc-report/
firefox *.html
```

Regardless of the quality, adapter sequences should be trimmed from the raw sequence.

Trimmomatics

In the same line as the download process, I prefer to use a for-loop in my bash scripts.

```
#!/bin/bash
```

```
mkdir -p $RNA_HOME/dataset/human/hpa/trimmed/SE
cd $RNA_HOME/datasets/human/hpa/trimmed

#Trimming Data
trim='/path/to/Trimmomatic-0.36/'
adapter_file='/path/to/adapter/file/TruSeq3-PE-2.fa'
reads='/path/to/raw/reads'

for i in {432,455,477}
do
    java -jar $trim/trimmomatic-0.36.jar PE -threads 8 \
        $reads/ERR315${i}_1.fastq.gz $reads/ERR315${i}_2.fastq.gz \
        ERR315${i}_pe1_trimmo.fastq.gz ERR315${i}_se1_trimmo.fastq.gz \
        ERR315${i}_pe2_trimmo.fastq.gz ERR315${i}_se2_trimmo.fastq.gz \
        ILLUMINACLIP:$adapter_file:2:30:10 LEADING:3 TRAILING:3 MINLEN:25
done

mv -v *se* ./SE
```

For single-end reads use the following:

```
java -jar $trim/trimmomatic-0.36.jar SE -threads 8 \
     $reads/filename.fastq.gz filename_trimmo.fastq.gz \
     ILLUMINACLIP:$adapter_file:2:30:10 LEADING:3 TRAILING:3 MINLEN:25
```

Here you can look at the quality again. `fastqc -t 8 *fastq.gz`. Make sure if you use the `-o` option for FastQC that there is already a directory with that filename. Otherwise, it will not work.

E.7.6 Alignment with HISAT2

HISAT2 [152] is a fast and sensititve alignment program for mapping next-gen sequencing reads from the same producers of TopHat2. Moreover, this is the improvement

on TopHat2. As there is a manual page `man hisat2`, I will add only a few of the important options below:

1. Main options
    (a) -x: Location of indexes (.ht2)
    (b) -1/-2: Read 1 and Read 2 for paired-end
    (c) -U: Unpaired reads (single-end)
    (d) --sra-acc: SRA file format
2. Input options
    (a) -q: fastq reads (default)
    (b) -f: fasta reads
3. Spliced alignment options
    (a) --min-intronlen: Default 20
    (b) --max-intronlen: Default 500000
    (c) --novel-splicesite-outfile: location for output
    (d) --rna-strandness: FR or RF for paired-end (RF for TruSeq)
    (e) --dta: alignments for StringTie analysis
    (f) --dta-cufflinks: alignments for Cufflinks analysis

Running with a for-loop:

```
#!/bin/bash


cd $RNA_ALIGN/hg19
mkdir -p unstranded
cd unstranded


#Run Hisat2
```

```
genome_index_prefix='/path/to/indexes/chr15'
output_SAM='/path/to/output/directory'
reads='/path/to/trimmed/fastq/reads'

for i in {432,455,477}
do
    hisat2 -p 8 -t --dta -x $genome_index_prefix \
           -1 $reads/ERR315${i}_pe1_trimmo.fastq.gz \
           -2 $reads/ERR315${i}_pe2_trimmo.fastq.gz \
           -q | \
           samtools view -@ 8 -Shb | \
           samtools sort -@ 8 \
           -o $output_SAM/ERR315${i}.chr15.sorted.bam
done
```

Merging & indexing

For viewing in IGV and a few other downstream programs, the sorted BAM files need to be merged by tissue and indexed.

```
samtools merge -@ 8 merged_cortex.sorted.bam \
               ERR315432.chr15.sorted.bam \
               ERR315455.chr15.sorted.bam \
               ERR315477.chr15.sorted.bam
samtools index merged_cortex.sorted.bam
```

By only indexing the merged files, this saves disk space. If disk space is not an issues, then each sorted BAM can be indexed within the hisat2 for-loop by adding:

```
samtools index $output_SAM/ERR315${i}.chr15.sorted.bam
```

E.7.7 Annotations with StringTie

StringTie [151, 154]: Transcript assembly and quantification for RNA-Seq, produced by Johns Hopkins University, Center for Computational Biology like Cufflinks.

Table E.2: StringTie Options

Flag Options	Description
-h	Prints help message and exits.
-v	Verbose mode, prints bundle processing details.
-o <path>	Set name for output GTF file.
-p <int>	Number of processing threads to use for assembly.
-G <gff>	Reference annotation file (gtf or gff3 format).
-l <label>	Sets label as prefix for the output transcripts.
-C <gtf>	Output file with names of fully covered transcripts. (-G)
-B	Output of Ballgown input table with coverage data for transcripts. (-G)
-b <path>	File path for Ballgown output. (-G)
-e	Only assemble reads that match reference. (-G, -B/-b)

For the annotation assembly, the *de novo* option was used. That is, no reference (-G) is used.

```
#!/bin/bash
mkdir -p $RNA_HOME/assemblies/human/hg19/unstranded
cd $RNA_HOME/assemblies/human/hg19/unstranded
sam_files='/path/to/sorted/BAM/files'
for i in {432,455,477}
```

```
do
     echo "Running sample: ERR315"$i
     stringtie -p 8 -l ERR315$i \
     -o ./ERR315${i}/transcripts.chr15.gtf \
     $sam_files/ERR315${i}.chr15.sorted.bam
done
ls -1 unstranded/*/transcripts.chr15.gtf > \
   unstranded/mergelist_gtf.txt
```

The next step is to use stringtie --merge (E.3) with all of the predicted transcriptomes to make one final transcriptome annotation for downstream analysis.

Table E.3: StringTie Merge Options

Flag Options	Description
-G	Reference annotation to include in the merging (gtf/gff3).
-o <path>	Set name for output GTF file.
-m <int>	Minimum input transcript length (default: 50).
-c <int>	Minimum input transcript coverage (default: 0).
-F <int>	Minimum input transcript FPKM (default: 0).
-T <int>	Minimum input transcript TPM (default: 0).
-f <float>	Minimum isoform fraction (default: 0.01).
-i	Keep merged transcripts with retained introns (default: no).
-l <label>	Name prefix for output transcripts.

Here, I use several different options to see which makes the best annotation.

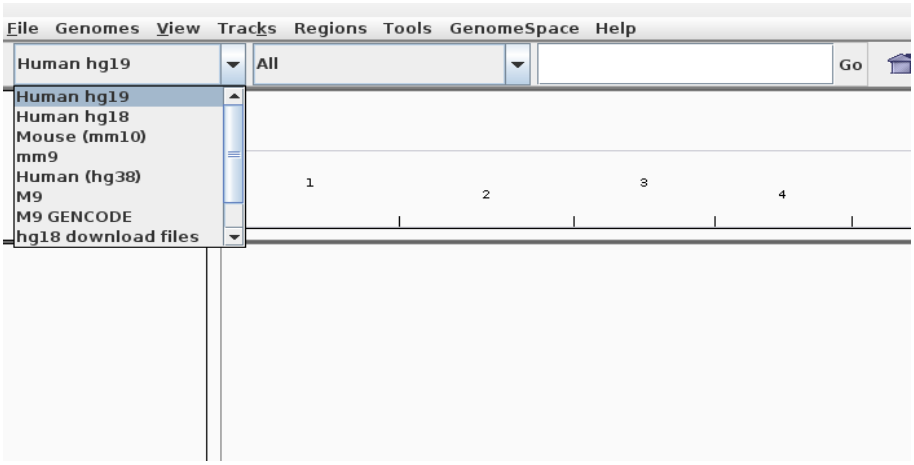
```
#!/bin/bash
#De novo assembly
annot='/path/to/annotation/file.gtf'
outputDIR='/path/to/output/directory'
mergeLIST='/path/to/merged/list/of/transcripts.gtf'

mkdir -p $RNA_HOME/assemblies/human/hg19/unstranded/strict
cd $RNA_HOME/assemblies/human/hg19/unstranded

stringtie --merge -p 8 -G $annot \
          -o $outputDIR/stringtie_merged.0.gtf $mergeLIST
stringtie --merge -p 8 -f 0.05 -G $annot \
          -o $outputDIR/stringtie_merged.1.gtf $mergeLIST
stringtie --merge -p 8 -F 5 -G $annot \
          -o $outputDIR/stringtie_merged.2.gtf $mergeLIST
stringtie --merge -p 8 -f 0.05 -F 5 -G $annot \
          -o $outputDIR/stringtie_merged.3.gtf $mergeLIST
stringtie --merge -p 8 -F 10 -G $annot \
          -o $outputDIR/stringtie_merged.4.gtf $mergeLIST
stringtie --merge -p 8 -F 5 -T 10 -G $annot \
          -o $outputDIR/stringtie_merged.5.gtf $mergeLIST
```

Once I have the correct parameters, I rename the file (`mv -v`) so that I know which is the annotation file to use.

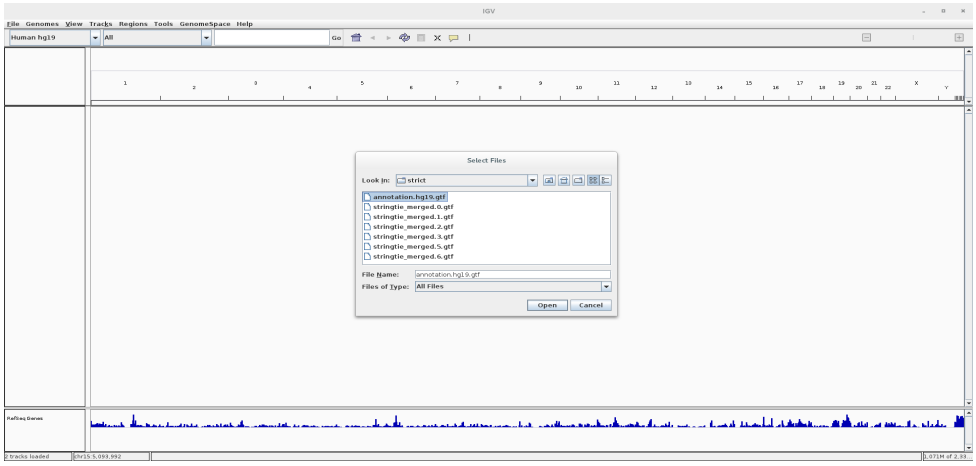
E.7.8 Working with IGV

1. Lauch IGV with `igv.sh`

2. Load Genome with the drop down menu

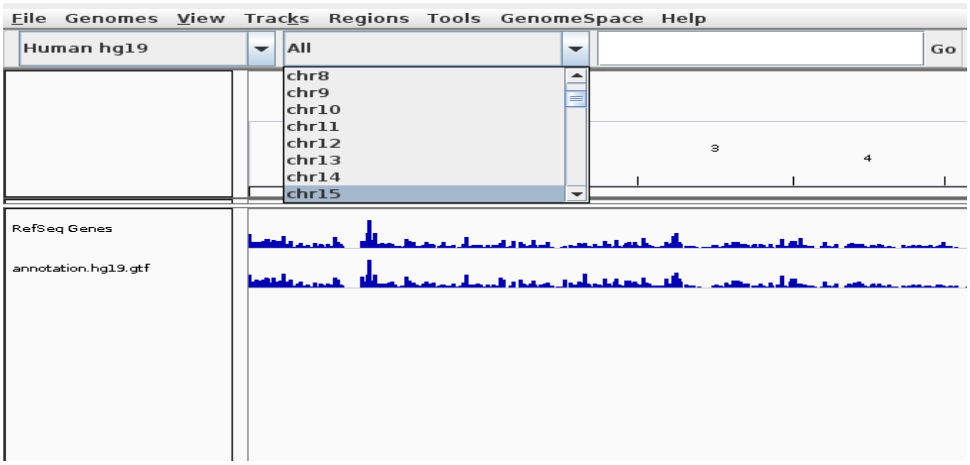


3. Load Genome annotation: File → Load from File...

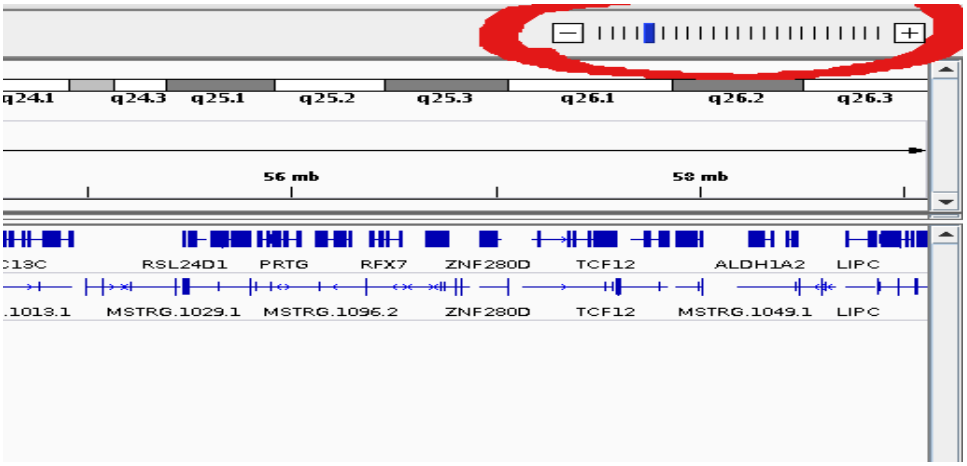


4. Zoom into region of interest by

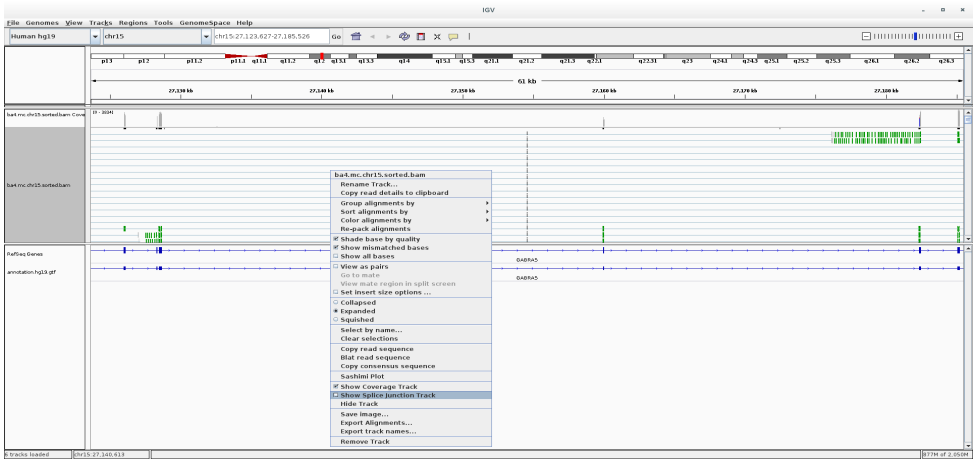
   (a) changing chromosomes:



   (b) zooming in with buttons:

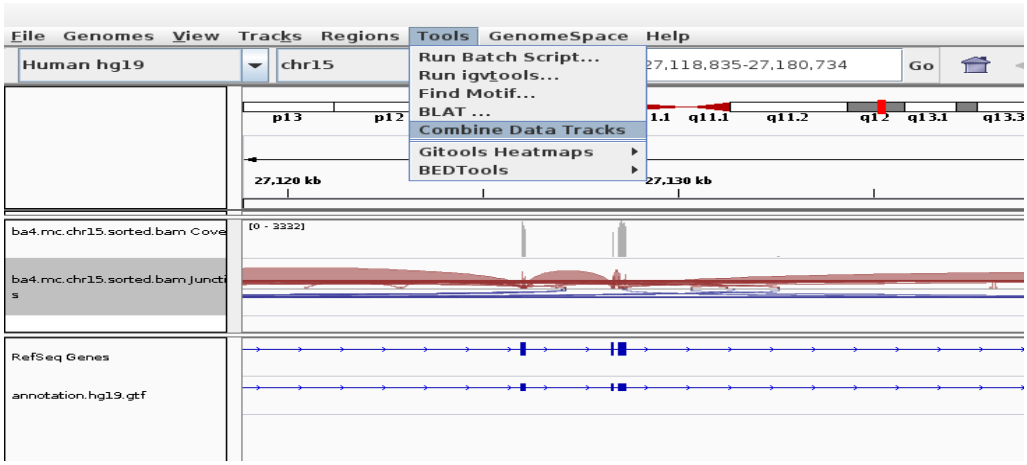


5. Load a sorted BAM file similar to loading a genome annotation
6. Show splicing junctions



7. Open a Sashimi plot

8. Save an image by right-clicking on screen
9. Use a tool with Tools
    (a) BEDtools options
    (b) BLAT options (no file input, but direct sequence)
    (c) And more...



BLAT

Generating BED files from Sanger sequence data like 3'RACE involves first using the tool `BLAT` within IGV and second exporting (right-click) it as a BED file. Once that has been done, file manipulation can be done with the following commands assuming all BED files have filename: iso*. Use `man` on any Bash function (i.e. paste) to read about it in more detail.

```
paste -s -d "\n" iso* > isoforms.hg19.bed
sortBed -i isoforms.hg19.bed > isoforms.hg19.sorted.bed
bedtools merge -i isoforms.hg19.sorted.bed > isoforms.hg19_merged.bed
```

Visualizing SNP data in IGV

```
awk '$1 == "7" { print $2 }' snp.file.tsv > snps.tsv
grep -f snps.tsv hippo.flt.vcf > informative.snps.vcf
```

In `emacs` copy over the header from the hippo.flt.vcf, so that I can view in IGV. Otherwise, cannot view this file in IGV.

E.7.9 SNP analysis

SNP information for region of interest downloaded from the Mouse Genomes Project - Query SNPs from Wellcome Trust Sanger Institute. To call variants, preform the following:

```
samtools mpileup -ugf chr7.fa -r chr7:region of interest \
                file1.sorted.bam file2.sorted.bam ... | \
                bcftools call -c - > var.raw.bcf
bcftools view var.raw.bcf | vcfutils.pl varFilter -D1333 > var.flt.vcf
```

The filter is based on mpileup output.

E.7.10 Gene prediction with GENESEQER

To analysis gene structures, the program GeneSeqer can be used in the commandline and as part of a pipeline if necessary. It requires FASTA files and an EST library. First step is to download protein or mRNA EST files from NCBI. Once that is complete, conversion of genome annotation and use of `GeneSeqer` as follows:

```
gffread -w annotation.file.fa -g /path/to/genome/fasta/chr7.fa \
        /path/to/annotation.file.gtf
GeneSeqer -s mouse -f -E est.files.fa -L comparison.files.fa > output
```

E.7.11 Manipulating files

Merging files uniquely (unique entries in a file):

```
awk '!a[$0]++' ./dir/* > ./new.unique.file
```

Separating out strands assuming strand:

```
#Find strand data column (stringtie gtf, col 7)
awk '$7 == "-"' annotation.file.gtf > minus.annot.gtf
awk '$7 == "+"' annotation.file.gtf > plus.annot.gtf
```

This varies on file layout. For instance, BED files normally have strand information in column 6. If you don't know where the strand information is at take a peek at it.

```
cat annotation.file.gtf | head
```

Use `man` or `-h/--help`[30] on any Bash function to learn about it's options.

```
man cat
head --help | less
awk -h | less
```

A slightly more complicated usage of `awk` to export only genes of interest.

```
awk -F "\t" '$7 == "+" { print $9 }' annotation.file.gtf | \
    awk -F "; " '$1 == "gene_id \"MSTRG.2451\""' > Ube3a.txt
grep -f Ube3a.txt annotation.file.gtf > Ube3a.gtf
```

The `-F` option for `awk` is used to tell the program to look for field separaters similar to how excel opens up CSV or TSV files. Here, I am telling `awk` to look for tab separation in the forward strand (+) and `print` the 9th column, which happens to have the gene information. In stead of dumping it to the screen, I pipe (|) the results into another `awk` command to use the ';' field separator to extract out all of the "gene_id" labeled as MSTRG.2451, which is the label StringTie gave *Ube3a* during assembly. The backslash (\) is used to comment out the (") as the `awk` command uses that as part of it's syntax.

[30] If using -h/--help, don't forget to pipe it into `less` so that you can view it easy. Exit with 'q'.

Once that is complete (taking only seconds), I direct the output into a new file Ube3a.txt. The file doesn't need to already be created to work. As such, it is very easy to overwrite files if you are not careful.

That all take seconds to run, afterwhich I use the `grep` command to use the pattern in the file, (`-f`), I just created to look in the annotation file and pull out all line-by-line the 'pattern' in the first file, directing the output to a new file Ube3a.gtf that now has only the *Ube3a* annotations. You can edit these smaller files in any text editor (like emacs), while trying to edit the original annotation.file.gtf would be a nightmare.

Converting GTF files with gffread

Bash isn't the only way to manipulate files. gffread can be used to convert GTF files into GFF file format, which is necessary to run MISO.

```
gffread annotation.file.gtf -o annotation.file.gff
```

gffread can also be used to convert GTF files into FASTA; however, to convert, an indexed genome (`samtools faidx`) is required.

```
gffread -w Ube3a.fa -g /path/to/chr7.fa Ube3a.gtf
```

E.7.12 Ballgown

To run Ballgown, StringTie needs to be run a second time for the samples with the `-e`, `-B`, and `-G` options.

```
stringtie -p 8 -G /path/to/annotation.gtf -eB \
          -o /path/to/output/sample.gtf \
          /path/to/sorted/BAM/file.bam
```

Once run for all the samples, a CSV file is need for importing the data when in R. This can be done by either editing a new file, in excel or Libre, or `printf`.

Running ballgown in R

I personally like using Ballgown [151, 328] (version 2.6.0) to look at the average transcript expression by tissue. Below is the script to so that. Unfortunately, for the antisense transcript, the snoRNAs are part of the assembly, so it's not as informative as one would like, but for *Ube3a* you get a pretty good idea of what's going on, on the expression level.

```
#Load libraries
library(ballgown)
library(ggplot2)
library(gplots)
library(genefilter)
library(devtools)

#Load phenotype data to generate ballgown object
pheno_unstranded <- read.csv("unstranded.mm9.csv")

##Generated ballgown object and save
bg_unstranded <- ballgown(samples=as.vector(pheno_unstranded$path),
                          pData=pheno_unstranded)
save(bg_unstranded, file='./bg_unstranded.rda')
load("bg_unstranded.rda")

pdf(file="file.name.pdf")

#Average expression of transcripts
plotMeans("MSTRG.2451",bg_unstranded, groupvar="type", legend=T)
plotMeans("MSTRG.2448",bg_unstranded, groupvar="type", legend=T)

dev.off()
```

Ballgown+

For more information on what can be done with ballgown, see the Nature protocol paper [151].

E.7.13 edgeR

For analysis with edgeR, I have provided a sample script. For more information, the edgeR manual can be of help. Briefly, the script loads libraries, runs featureCounts, normalizes the data, then calculates differential analysis based of the features counted via featureCounts.

```
##Clear workspace
rm(list=ls(all=T))

## load libraries
library(edgeR)
library(dplyr)
library(Rsubread)

##Importing BAM file location (unstranded)
pheno_data <- read.csv("bam_files.csv")
bam_files   <- pheno_data$path

fc_transcript <- featureCounts(bam_unstranded,
                               annot.ext="/path/to/annotation.gtf",
                               isGTFAnnotationFile=TRUE,
                               GTF.featureType="transcript",
                               GTF.attrType="transcript_id",
                               useMetaFeatures=FALSE,
                               allowMultiOverlap=TRUE,
                               strandSpecific=0,
                               isPairedEnd=TRUE,
```

```
                                nthreads=8)

fc_exons     <- featureCounts(bam_unstranded,
                              annot.ext="/path/to/annotation.gtf",
                              isGTFAnnotationFile=T,
                              GTF.featureType="exon",
                              GTF.attrType="gene_id",
                              useMetaFeatures=F,
                              allowMultiOverlap=T,
                              strandSpecific=0,
                              isPairedEnd=T,
                              nthreads=8)

save(fc_transcript, file='fc_transcript.rda')
save(fc_exons, file='fc_exons.rda')

##Load data
#load("fc_transcript.rda")
#load("fc_exons.rda")

transcripts <- DGEList(counts=fc_transcript$counts,
                    group=rep(1:4, each=3),##depend on samples
                    genes=fc_transcript$annotation)
keep.t      <- rowSums(cpm(transcripts) > 1) >= 3  #n=12, 25%
transcripts <- transcripts[keep.transcripts, ,keep.lib.sizes=F]
transcripts <- calcNormFactors(transcripts)

exons       <- DGEList(counts=fc_exons$counts,
                    group=rep(1:4, each=3),
                    genes=fc_exons$annotation)
keep.exons  <- rowSums(cpm(exons) > 1) >= 3
```

```
exons       <- exons[keep.exons, ,keep.lib.sizes=F]
exons       <- calcNormFactors(exons)

rownames(transcripts$samples) <- pheno_data$id #sample names
colnames(transcripts$counts)  <- pheno_data$id
rownames(exons$samples)       <- pheno_data$id
colnames(exons$counts)        <- pheno_data$id

##Obtain dispersion estimates
design <- model.matrix(~group, data=transcripts$sample)

transcripts <- estimateDisp(transcripts, design, robust=TRUE)
exons       <- estimateDisp(exons, design, robust=TRUE)

#Testing for DE (One-way Anova like Test)
fit1 <- glmFit(transcripts, design)
fit2 <- glmFit(exons, design)

lrt1 <- glmLRT(fit1, coef=2:4)##depended on design
lrt2 <- glmLRT(fit2, coef=2:4)

sink("DE_transcripts_ANOVA.txt")
print(topTags(lrt1, n = 10000))
sink()
sink("DE_exons_ANOVA.txt")
print(topTags(lrt2, n = 10000))
sink()

##Testing for DE (pairwise)
et1.1 <- exactTest(transcripts, pair=c("2","1"))
et1.2 <- exactTest(transcripts, pair=c("2","3"))
```

```
et1.3 <- exactTest(transcripts, pair=c("2","4"))

et2.1 <- exactTest(exons, pair=c("2","1"))
et2.2 <- exactTest(exons, pair=c("2","3"))
et2.3 <- exactTest(exons, pair=c("2","4"))

sink("DE_transcripts_tissue2VStissue1.csv")
print(topTags(et1.1, n = 10000))
sink()

sink("DE_exons_tissue2VStissue1.csv")
print(topTags(et2.1, n = 10000))
sink()

##Alternative to topTags & sink()
write.table(cbind(et2.1$genes,et2.1$table),
            "DE_exons_tissue2VStissue1_full.txt",
            sep="\t",row.names=FALSE)

##Differential Splicing
sp <- diffSpliceDGE(fit2, coef = 4,
                    geneid = "GeneID",
                    exonid = "Start")

sink("DE_splicing.txt")
print(topSpliceDGE(sp, n = 10000))
sink()
```

E.7.14 Graphic with ggplot2

`ggplot2` is a fantastic package for publication quality graphs in R. Everything can be manipulated for graphic. If you have trouble, there is an active community to help with

every problem. Below is a sample script although ggplots should be run in terminal.

```
##Clear workspace
rm(list=ls(all=T))

##Load libraries
library(ggplot2)
library(devtools)

pdf(file="filename.pdf", width=10, height=7)

data <- read.table("data.tsv",
                   sep="\t", header=TRUE)

data.plot <- ggplot(data=data,
                        aes(x=Sample, y=logFC, fill=GeneID)) +
    geom_bar(colour="black", width=0.8,
             stat="identity", position=position_dodge()) +
    geom_hline(aes(yintercept=0))

data.plot + scale_x_discrete(name="") +
    scale_y_continuous(name="Expression: log2(FC)",limits=c(-6,1)) +
    scale_fill_manual(values=c("grey","grey40","grey20")) +
    annotate("text",x=1,y=-4.45,label="***,0.00", size=5) +
    annotate("text",x=2,y=-3.74,label="***,0.00", size=5) +
    annotate("text",x=3,y=-5.23,label="***,0.00", size=5) +
    theme(legend.title=element_blank(),
          legend.position=c(0.88,0.92),
          legend.text=element_text(size=16),
          axis.text.x=element_text(face="bold", colour="black", size=16),
          axis.title.y=element_text(face="bold", size=18),
          axis.text.y=element_text(size=16,colour="black"),
```

```
          panel.grid.minor=element_blank(),
          panel.grid.major=element_blank(),
          panel.background=element_rect(fill="white"))

dev.off()
```

APPENDIX F

HIGH THROUGHPUT DRUG SCREENING OF MOUSE EMBRYONIC STEM CELL-DERIVED NEURONS

Summary

High-throughput screening (HTS) to identify new drugs for neurological disorders often rely on the use of mouse primary neuronal cultures; however, establishing primary cultures from mice is labor intensive and expensive. Moreover, most HTS facilities do not allow the use of primary cell lines because of the risks associated with contaminating other cell lines in the facility. In contrast, embryonic stem (ES) cells are permitted in most HTS facilities and can be reliable differentiated into neurons, generating an almost unlimited source of cells for large-scale studies. Thus, ES cell-derived neurons are an excellent model system for performing HTS to identify new therapies for neurological disorders. Here, we developed a high-throughput neuronal culture model via ES cells. Mouse C57BL/J6 ES cells were successfully differentiated into neurons on poly-d-lysine, and immunocytochemistry performed using high-throughput imaging system. These results are promising for the field of neurological disorders and drug discovery.

F.1 Mouse Embryonic Stem Cell Culture (Timing: 5 days)

Mouse embryonic stem (ES) cells are a powerful tool for the scientific discovery because of their ability of almost endless self-renew and potential to differentiate into multiple cell types, including neuronal cell types [329]. This pluripotency is a result of the cell type used to derive ES cells, inner mass of a developing blastocyst. As such, ES cells are often co-cultured with feeder cells. The pluripotency is facilitated by a complex pathway involving Wnt/β-catenin signaling and cross-talk between Wnt and LIF, leukemia

inhibitory factor [330, 331]. This replaces the need for inhibition of GSK3, which phosphorylates β-catenin marking it for ubiquitin-dependent degradation [332, 333], making LIF an essential factor in culturing self-renewing mouse ES cells [334–336].

F.1.1 Plating feeder cells for co-culture

As mentioned above, ES cells are often co-cultured with feeder cells when being maintained as self-renewing pluripotent cells. There are a number of feeder cell types to chose from; however, the most common fibroblasts used as feeder cells are mouse embryonic fibroblast (MEF) and SIM mouse embryoderived thioguanine and ouabain resistant (STO) [337, 338]. In this study, we use a genetically modified versus of the STO cell line, SNL 76/7, first established by Dr. Allen Bradley [226]. The SNL 76/7 is a unique STO line as it contains the murine LIF gene; thus, LIF does not need to be added to the culture media.

1. Coat six 100-mm tissue culture dishes with 0.1% gelatin for SNL adhesion. Add $\sim$7 ml of gelatin (StemCell Technologies) to each dish and incubate for 30 min at room temperature.
2. Aspirate the gelatin from the dishes and allow them to dry for $\sim$5 min.
3. Culture feeder cells for the ES cells co-culture by plating one vial of Mitomycin C-inactivated SNL (approx. 2.25×10^7 cells, 60x concentration) into 60 ml of STO media (see recipe in **Table F.1**) and plate onto fixed 100-mm tissue culture dishes.
   (a) Quick thaw vial at 37°C using a water bath for $\sim$3 min
   (b) Add to 10 ml of STO media
   (c) Centrifuge ($<$270g for 5 min)
   (d) Decant supernatant and resuspend in 60 ml of STO media
   (e) Add 10 ml to each plate
   (f) Disperse the cells with back and forth motion

4. Incubate overnight before using (37°C, 5% CO_2). SNL must be plated at least one day before adding ES cells. SNL feeder plates are generally good after plating up to 7 days.

F.1.2 Plating ES cells

Expansion of ES cell is important for downstream HTS studies. As such, ES cells expansion need to be optimized for individual cell-lines. For C57BL/6J ES cells, the following number of passages is sufficient.

1. Condition one SNL plate with 10 ml of ES media (see recipe in **Table F.1** for at least 2 h before plating ES cells. This allows for LIF expression from the SNL to be added to the ES media, which is required to maintain ES cell pluripotency and ability to self-renew.
2. Defrost 1 vial of ES cell, C57BL/6J, (approx. 3.5×10^6 cells) in 37°C water bath for 3 min. Add to 10 ml of ES media. Transfer to 10 ml of ES media in 15 ml tube and spin down (<270g for 5 min).
3. Decant supernatant and resuspend in 1 ml of ES media. Add cell suspension of SNL conditioned plates. Disperse the cells with back and forth motion. Incubate the ES cells overnight at 37°C, 5% CO_2.
4. 24 h after plating, ES cells form small colonies, **Figure F.1A**. Change media with 10 ml of ES media. 48 h after plating, ES cells should be 70-80% confluent, be careful not to let the ES cells over grow. Condition remaining five SNL plates with ES media for 2 h. Passage cells.
    (a) Aspirate media, and rinse twice with room temperature sterile 1xPBS. Add 2 ml of TrypLE Express (Life Technologies) and incubate for 5 min. Add 3 ml of ES media to neutralize the trypsin and break the colonies into single cell suspension. Transfer to sterile 15 ml tube and spin down (<270g, 5 min).

(b) Aspirate supernatant and gently resuspend in 5 ml of ES media. Add 1 ml of cell suspension to each conditioned SNL plate. Disperse the cells with back and forth motion and incubate overnight. Change media next day.

5. 48 h after passaging, ES cells should be 70-80% confluent. These cells are ready for differentiation.

F.2 Neural Induction (Timing: 7 days)

To induce differentiation of ES cells into neurons, one method is to separate ES cells from SNL and cultured them in suspension as demonstrated in **Figure F.1B**. This can be done one of two ways. The first is several feeder free passages, and the second is by using gelatin coated flasks to separate the SNL feeder cells from ES cells. For HTS purposes, the gelatin technique is used. This saves time and resources.

1. Once ES colonies are ready for neural induction (day 6), coat one T175 with 20 ml of 0.1% gelatin for 30 min at room temperature. Aspirate gelatin and allow it to dry for $\sim$5 min.
2. Aspirate ES media from ES cells, and rinse twice with room temperature sterile 1xPBS. Add 2 ml of TrypLE Express and incubate for 5 min. Add 3 ml of CA media (see recipe in **Table F.1**) to neutralize the trypsin and break the colonies into single cell suspension.
3. Transfer cells suspension to T175 (approx. 25 ml, 5 ml/plate), and incubate for 30 min at 37°C, 5% CO_2. After incubation, SNL should have attached to the surface of the gelatin coated flask while the majority of the ES cells should remain floating. Collect the floating ES cells into 50 ml tube and spin down ($<$270g, 5 min). Aspirate supernatant and gently resuspend in 5 ml of CA media.
4. Count the cells using a hemocytometer, or automated cell counter, and plate 4×10^6 ES cells per 100-mm bacteriological Greiner Petri dishes.

5. Next day change CA media and split CAs 1:2 by carefully transferring the CA suspension into 50 ml tube (1 plate/tube)[1]. Let CAs settle for at least 3 min. Remove supernatant and resuspend CAs gently with 20 ml of CA media. Mix suspension gently with a 25 ml pipette and plate cells on new dishes.
6. After 48 h in suspension, change CA media as in step 5, however add RA (retinoic acid) at final concentration of 0.5 mM to CA media. This is another 1:2 split of the CAs, which may be optional depending on CAs density in suspension. For next four day change CA media without splitting (2 plates/tube) by resuspending each tube with 20 ml CA media with 0.5 mM RA.

F.3 Neuron Elongation & Maturation (Timing: 2+ days)

ES cell-derived neurons are traditionally plated on poly-dl-orithine/laminin co-coating [230–232]. This co-coat is feasible for 24-well format; however, for high-throughput screening (HTS) purposes poly-d-lysine coated 96- and 384-well plates are more readily available and less expensive, so we conducted experiments to determine if poly-d-lysine pre-coated 96-well plates (VWR) could be used for neuron elongation and maturation.

1. (Days Post Differentiation, DPD 0) To dissociate the CAs into single cell suspension, transfer CAs to 50 ml tubes (2 plates/tube), and wash CAs twice with PBS before trypsinizing.
   (a) Let CAs settle for $\sim$3 min, then remove supernatant and resuspend CAs gently with 20 ml of 1x PBS. Let CAs settle again for $\sim$3 min, and remove supernatant. Resuspend CAs gently with 5 ml 1x PBS and let settle for $\sim$3 min.
   (b) Label and open 1.5/1.6 ml tubes, and carefully transfer CAs at the bottom of 50 ml tubes into them. Spin down quickly for about 5 s, and carefully

[1](1) Splitting CAs is highly depended on number of CAs in suspension, this may vary depending on type of serum/serum replacement. (2) Do not use narrow pipettes to mix suspension to avoid dissociating the CAs. (3) By using new plates, this will further remove any lingering SNL feeder cells.

remove supernatant with pipet tip. Add 1 ml of 0.5% Trypsin-EDTA (Life Technologies). Vortex and place on heat block for 4 min at 37°C rotation and 450 rpm. Vortex the tubes and spin down quickly.

(c) Remove supernatant and resuspend CAs in CA media to neutralize trypsin. Dissociate CAs by pipetting up and down (approx. 10 times), and spin down quickly for 5 s. Carefully remove as much supernatant as possible and resuspend in N2 media (see recipe in **Table F.1**), which should be made fresh.

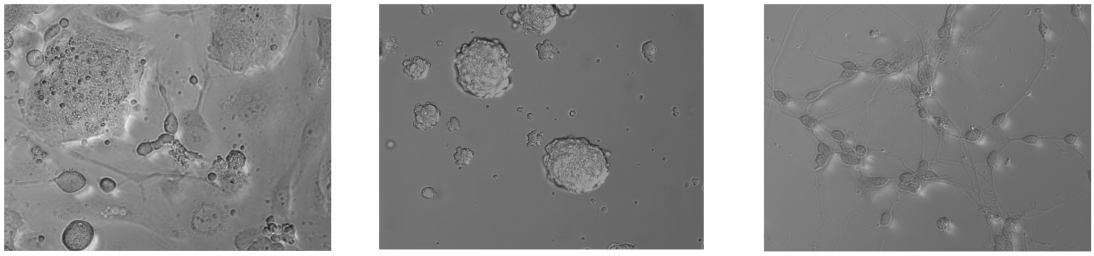
(d) Filter cell suspension through 40 μm cell strainer by applying drop by drop[2], and count cells using automated cell counter, or hemocytometer.

2. Plate 9.8×10^4 cells/well for 96-well plate using sterile hydrospeed in N2 media[3].
3. 24 h after plating cells should have attached to the plates as shown in **Figure F.1C**. (*optional*) Change the N2 media to further remove trace amounts of FBS.
4. 48 h after plating N2 media should be changed to Complete media (see recipe in **Table F.1**). Add the BDNF, brain derived neurotrophic factor, fresh for long term cultures of more than three days. Change Complete media every two to three days.

F.4 Representative Results

Figure F.2 shows an outline of the protocol. The first 5 days involves ES cell culture on SNL feeder cells, which is highly depended on cell lines growth. Following ES cell culture and expansion, the ES cells are separated from the SNL feeder cells expressing LIF to initiate the differentiation process. For high-throughput purposes, cells were grown in suspension and split to avoid overcrowding and obtain optimal numbers. In this case, one vial of ES cells generated $\sim$15-20 plates worth of 4×10^6 cells, which are split twice generating four times the initial number of cells. After neuronal induction, the cellular aggregates are dissociated and placed into serum free N2 media for 48 h. After which, the

[2]Avoid applying pressure.
[3]Plating density should be optimized for ES cell lines.



A. Co-culture of ES cells and Feeder cells

B. Embryoid bodies in suspension culture

C. Elongated ES cell derived neurons

Figure F.1: Differentiation of ES cells into neurons. Light microscope images of **A)** co-cultures of ES cells and SNL feeder cells at 40x magnification, **B)** embryoid boides in suspension at 20x magnification, and **C)** elongated neurons after three days of culture at 20x magnification.

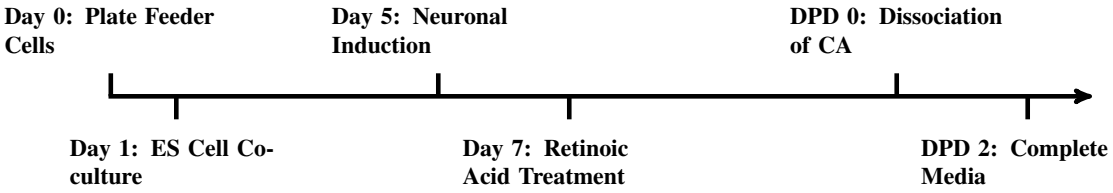
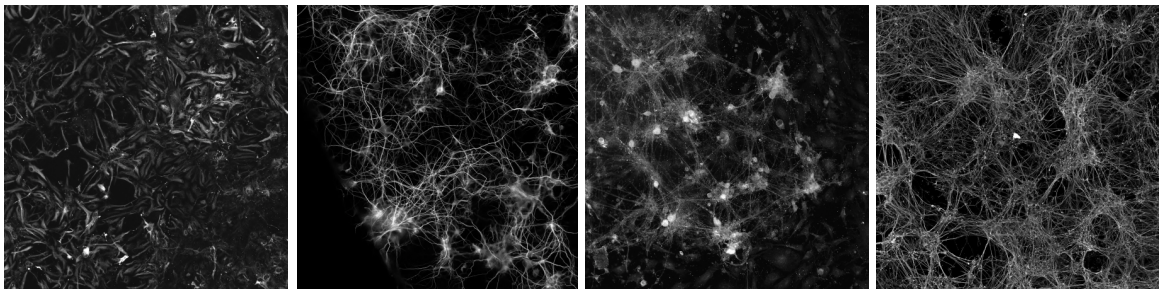


Figure F.2: Timeline for the differentiation of ES cells into neurons.

cells are changed to the Complete media for long term maintenance, with BDNF added to support neuronal growth for cultures lasting more than five days. **Figure F.3** shows ES cell-derived neurons that have been cultured for twelve days.

F.5 Discussion

Here, a high-throughput screening method optimized for drug discovery in neuronal cultures is described. For HTS purposes, a large initial number of ES cells is essential; therefore, each individual ES cell-line must be optimized for cell growth. In this method, the use of SNL 76/7 feeder cells is recommended as these cells secrete LIF - supplement required to maintain undifferentiated ES cell state - into the media eliminating the additional purchase of LIF.



Nestin **Map2** **Doublecortin** **βTubulin III**

Figure F.3: High-throughput screening immunofluorescence characterization of differentiated ES cells in 96-well format at day 12, 10x magnification.

For neuronal cultures including ES cell-derived neurons, neural connectivity patterns are crucial for proper function and development. For HTS assays, this adds an additional requirement for an extracellular matrix coating at the bottom of the wells for proper cell attachment and growth [339–341]. As such, extra cellular signaling proteins like laminin are needed. The majority of ES cell-derived neuron protocols utilize a co-coat with laminin [230, 232]; however, in these protocols the co-coating is applied by hand, which is not conductive to HTS. Furthermore, ordering optical-bottom, sterile plates with a laminin co-coat for 96- or 384-well plates is an expensive special order process that can take upwards to three months.

For those reasons, an alternative solution of using a single coating of poly-d-lysine was choice for this protocol. Poly-d-lysine promotes cell adhesion through ionic interactions [339]. Moreover, poly-d-lysine is a common substrate choice for culturing primary neurons [180]. In this protocol, we successfully differentiate ES cells into neurons on poly-d-lysine, to the best of our knowledge, for the first time.

Finally, this protocol provides an efficient approach for large-scale differentiation of ES cells into neurons. More importantly, the methods offers a platform for drug discovery for single gene neurological disorders that have mouse models currently available like

Angelman syndrome - a severe neurodevelopmental disorder. The power of using ES cells to derive neuron cultures extend beyond the world of drug discovery, especially with the advances in gene manipulation technology (i.e. CRISPR/Cas9 systems), where ES cells can be manipulated in culture before being expanded and differentiated into neurons to answer basic questions in the field of neuroscience. Altogether, this method has the potential to bridge the gap between basic and translational research.

F.6 Materials

Equipment[4]

1. Centrifuge
2. Water bath
3. Incubator
4. Microscope
5. Vortex
6. Mini Centrifuge
7. 100-mm Petri dishes
8. 15 ml & 50 ml conical tubes
9. 1.6 ml tubes
10. Bacteriological Petri dishes
11. 40 μm nylon cell strainer
12. Sterile filter 0.2 μm
13. Poly-d-lysine 96-well pre-coated plates

[4] All reagents and materials used must be sterile.

Table F.1: Media composition - Full

Media	**Components**	**Company**	**Cat. #**	**Notes**
	SNL 76/7 (untreated)	Applied StemCell, Inc.	ASF-1305	N.A.
STO Media	KnockOut DMEM	Life Technologies	10829018	N.A.
	FBS		1600044	10%
	GlutaMAX		35050061	2 mM
	Penicillin / streptomycin		15140122	1%
ES Media	KnockOut DMEM	Life Technologies	10829018	N.A.
	FBS		1600044	10%
	GlutaMAX		35050061	2 mM
	Non-essential amino acids		111400050	1%
	Penicillin / streptomycin		15140122	1%
	β-mercapto-ethanol	Sigma-Aldrich	M6250-100ML	0.1 mM

Table F.1: *Continued...*

Media	**Components**	**Company**	**Cat. #**	**Notes**
CA Media	DMEM	Life Technologies	11995073	N.A.
	KnockOut Serum Replacement		10828028	15%
	GlutaMAX		35050061	2 mM
	Non-essential amino acids		111400050	1%
	Penicillin / streptomycin		15140122	1%
	β-mercapto-ethanol	Sigma-Aldrich	M6250-100ML	0.1 mM
	Retinoic Acid		R2625-50MG	0.5 mM
N2 Media	DMEM/F-12	Life Technologies	11330057	1:1
	Neurobasal Medium		21103049	1:1
	GlutaMAX		35050061	2 mM
	B27		17504044	1%
	BSA	Sigma-Aldrich	A7906-100G	50 μg/ml
	Progesterone		P8783-1G	20 nM
	Putrescence		P5780-5G	100 nM

Table F.1: *Continued...*

Media	**Components**	**Company**	**Cat. #**	**Notes**
	ITS Supplement	Roche Life Science	11074547001	1%
Complete Media	Advanced DMEM/F-12	Life Technologies	12634028	1:1
	Neurobasal Medium		21103049	1:1
	GlutaMAX		35050061	2 mM
	B27		17504044	1%
	BDNF		PHC7074	50 ng/ml
MISC	0.1% Gelatin	Stem Cell Technologies	07903	N.A.
	TrypLE Express	Life Technologies	12604-021	N.A.
	0.5% Trypsin / EDTA		15400054	N.A.
	PBS		10010049	N.A.
	100-mm plates	Greiner Bio-One	633161	N.A.
	Poly-d-lysine 96-well plates	Thermo Scientific	152037	N.A.

Table F.2: List of Antibodies

Antibody	**Company**	**Cat. #**	**Dilution**
anti-Nestin	EMD Millipore	AB5922	1:200
anti-Map2	Santa Cruz	sc-20172	1:250
anti-DCx	Biotechnology	sc-8066	1:200
anti-mouse (555, Cy3)	Jackson Immuno	115-165-146	1:200
anti-rabbit (488, Cy2)		111-545-144	1:200
anti-TOPRO-3	Life Technologies	T3605	1:1000
anti-βTubulin III	Sigma-Aldrich	T5076	1:200
anti-Goat Serum		G9023-10ML	5%