# DYNAMIC REGULATION OF RNA EDITING IN HUMAN BRAIN DEVELOPMENT AND DISEASE

by
Taeyoung Hwang

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

March, 2016

# Abstract

RNA editing is increasingly recognized as a molecular mechanism regulating RNA activity and recoding proteins. Here, I surveyed the global landscape of RNA editing in human brain tissues and identify three unique patterns of A-to-I RNA editing during cortical development: stable high, stable low and increasing. RNA secondary structure and the temporal expression of adenosine deaminase acting on RNA (ADAR) contribute to cis- and trans- regulatory mechanisms of these RNA editing patterns, respectively. Interestingly, the increasing pattern in development is most apparent in brain and conserved in mouse brain development. The increasing pattern associates with the growth of cortical layers and neuronal maturation, correlates with mRNA abundance, and influences miRNA binding energy. Gene ontology analyses implicate the increasing pattern in vesicle or organelle membrane-related genes and glutamate signaling pathways. I also show that the increasing pattern is selectively perturbed in spinal cord injury and glioblastoma. These findings reveal dynamic and functional aspects of RNA editing in brain, providing new insight into epigenetic regulation of sequence diversity.

Readers:

Daniel R. Weinberger, M.D. (Thesis advisor)

Anthony K. L. Leung, Ph. D.

# Acknowledgments

It has been almost six years since I first enrolled in Ph. D. program and I am about to finish my journey to a Ph. D. very soon. It was not easy and not like as what I first expected. I had to go through many hurdles in studying biomedical sciences since I didn't take any related courses in undergraduate, as well as in orienting myself toward a career as a researcher in the field of biomedical engineering. Reflecting on the past time in this journey to a Ph. D., I realized that this chapter of life would not be successful without people who support and encourage me. I would like to acknowledge a couple of them here although many others are in my mind.

First of all, I would like to express my appreciation to my thesis advisor, Dr. Daniel R. Weinberger for his support and encouragement. Despite his busy schedule as C.E.O. of the Lieber Institute for Brain Development, Danny always finds time to take care of me as well as to advise me. His encouraging words have been motivation that I rely on up to the moment of defending this dissertation.

I thank my committee members, Dr. Steven Salzberg and Dr. Anthony K. L. Leung for their insightful suggestions and comments. Steven and Anthony have given their valuable advices throughout this dissertation. I am so happy to have them as my committee advisors. I am also deeply grateful to Dr. Chul-Kee Park for his expert views on brain tumors and his generosity to provide glioblastoma samples for this study.

Another appreciation is to my lab mates, including Dr. Jooheon Shin, Amanda Price, Anandita Rajpurohit, Carrie Wright and Dewey Kim. Without their cheers and daily chats, my graduate life would have no joy. In particular, I would like to express my appreciation for Jooheon who has given many life lessons.

Finally but most importantly, I want to shout out my gratitude to my family. My parents are truly the energy of my life. I can't thank my younger brother, JuYoung,enough, who plays a role as a stout pillar in Korea while I have been in the U.S. for my study. And my wife, Hyo Ju is the one healing me whenever I was tired, exhausted and disappointed.

I will remember all these supports in my heart.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

## Introduction

RNA editing is a molecular process perturbing RNA sequences in a post-transcriptional manner. In this chapter, I will review the RNA editing in general, focusing on A-to-I editing in messenger RNA.

## 1.1 RNA editing

RNA editing is a post-transcriptional modification that alters RNA sequences from their original DNA templates. These processes affect most cellular RNAs including messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA), as well as small RNA such as miRNA. RNA editing expands the repertoire of RNA transcripts, contributing to the complexity of genetic information through recoding amino acids and affecting regulatory roles of RNA.

**Overview**

In 1986, Benne et al. coined the term "RNA-editing" to describe molecular phenomena in which uridines were inserted or deleted in RNAs of trypanosome mitochondria (Benne et al. 1986). Since then, the term RNA editing has been used to describe various molecular processes resulting in the modification of RNA sequences differing from original DNA templates (Sansam & Emerson 2005).

Three types of editing (Figure 1.1) - insertion, deletion and substitution - have been identified in all major types of RNA (messenger, ribosomal, and transfer RNAs) as well as in miRNA in various species of eukaryotes, as shown in Table 1.1. In particular, eukaryotic organelles such as chloroplasts and mitochondria harbor the greatest variety of RNA editing (Gray 2012). In this study, I specifically focused on mRNA editing.

In mammals, two substitutional types of RNA editing have been identified mainly so far: cytidine is converted to uridine ("C-to-U editing") and adenosine is modified to inosine ("A-to-I editing"). Both are generated by a hydrolytic deamination process catalyzed by specific enzymes: apoB mRNA editing complex (APOBEC) for C-to-U editing and Adenosine Deaminase acting on RNA (ADAR) for A-to-I editing.

**C-to-U editing**

The most well characterized C-to-U editing in mammals is an event found in the apolipoprotein B (*apoB*) gene, which is also the first reported instance of C-to-U editing in human (Powell et al. 1987). C-to-U editing in the *apoB* gene converts a CAA (glutamine) codon to a UAA (stop) codon, resulting in a truncated *apoB* protein due to premature termination of translation (Figure 1.2). Importantly, this editing event is responsible for tissue-specific modification of the *apoB* transcript. While a full-length *apoB* protein, denoted by apoB100 (512 kilo daltons), is expressed in the liver, a truncated version of *apoB*, known as apoB48 (241 kilo daltons) is synthesized in intestine through C-to-U editing. Functionally, this tissue-specific modification of the *apoB* transcript has important consequences because apoB100 comprises a low density lipoprotein (LDL) complex that works for the transport of dietary cholesterol in liver, but apoB48, which lacks the C-terminal domain that in apoB100 binds to LDL receptors on cell membranes, is a component of the triglyceride-rich chylomicrons that deliver dietary lipids to intestine (Davidson & Shelness 2000).

C-to-U editing of *apoB* mRNA is mediated by a multicomponent protein complex, APOBEC that consists of a catalytic subunit for cytidine deamination, APOBEC-1, and several auxiliary proteins. One cofactor is an APOBEC-1 complementation factor (ACF), which serves to bind a target RNA with RNA recognition motifs (Blanc & Davidson 2010). In terms of *cis*-regulation, three sequence elements around the editing position, referred to as the mooring sequence, enhancer and spacer region, are known to be necessary factors for efficient editing (Backus & Smith 1992).

Other than C-to-U editing in the *apoB* gene, 32 additional mRNA targets of APOBEC-1 for C-to-U editing were identified by a recent genome-wide screen (Rosenberg et al. 2011). Here,

all C-to-U editing sites were found to be located in AU-rich segments of 3' untranslated regions (3' UTR) of transcripts. However, functional roles of these editing sites are not characterized yet.

**A-to-I editing**

A-to-I editing is the most prevalent form of RNA editing in the animal kingdom. It is catalyzed by a family of enzymes known as ADARs, which deaminate adenosines in double-stranded structures of RNAs (Savva et al. 2012). The effect of A-to-I editing can be diverse depending on the location of the edited nucleotide, because inosine is recognized as guanosine by the cellular machinery. For example, A-to-I editing at protein-coding regions can change a codon (Pullirsch & Jantsch 2010), while it also has potential to modulate splice site usage especially when residing in an intron (Rueter et al. 1999; Schoft et al. 2007).

In mammals, A-to-I editing was first identified in transcripts encoding glutamate-gated ion channels (Sommer et al. 1991). Here, A-to-I editing converts a glutamine (Q) to an arginine (R) in a gene encoding a subunit of glutamate receptors, affecting the ion permeability of the receptors. While originally believed to be a rare event, A-to-I editing event is now recognized as a widespread process in mammals. In particular, recent genome-wide studies identified huge number of A-to-I editing sites in human tissues (Li et al. 2011; Park et al. 2012; Ramaswami et al. 2013). Because of its importance, A-to-I editing is discussed in detail in the next section of this chapter (see 1.2 A-to-I editing).

**Evolutionary perspectives**

RNA editing can be considered as a mechanism for adaptive evolution (Gommans et al. 2009). It confers phenotypic variation at relatively low evolutionary cost because RNA editing is

not a hard-wired process - in other words, it can retain predominant production of the wild-type protein. If the RNA editing produces a beneficial variant, the genome may maintain the locus that produces the RNA-edited transcript for its novel function (Knisbacher & Levanon 2015). A recent example with octopus describes the utilization of RNA editing for adaptation to environment (Garrett & Rosenthal 2012). While an Antartic and a tropical octopus share the same potassium channel gene in their genomes, they have distinct editing status at a site in the channel's pore to accommodate different temperatures. Specifically, the site modulating gating kinetics is extensively edited, recoding an isoleucine (I) to a valine (V) in the Antarctic species while it is mostly unedited in the tropical species. This editing dramatically accelerates deactivation kinetics of potassium channels, suggestive of candidate mechanism for cold adaptation. This result suggests that RNA editing can contribute to the kind of adaptation to the physical environment achievable by genetic variation in general. Regarding this, it is hypothesized that RNA editing increases the evolvability of species by providing a molecular mechanism to express diverse phenotypic variations in response to changing environments, and selection may favor such a system with higher levels of genetic flexibility (Gommans et al. 2009).

RNA editing is also involved in a process known exonization of Alu elements, where Alu elements can become new exons (Pandey & Mukerji 2011). Alu elements are retro-transposable repeat sequences in a genome, found only in primates. RNA editing can affect exonization through the creation of a functional 3' splice site or "AG" and the alteration of functional exonic splicing enhancers within the exon (Lev-Maor et al. 2007). These observations suggest the possibility that RNA editing is involved in primate evolution.

## 1.2 A-to-I editing

**ADAR**

A-to-I editing is a deamination process of adenosines to inosine, which is mediated by the enzyme ADAR targeting double-stranded RNA (dsRNA). ADAR was initially characterized as a protein with double-stranded RNA (dsRNA)- unwinding activity (Bass & Weintraub 1987; Rebagliati & Melton 1987). Soon after, it was recognized as an enzyme responsible for A-to-I editing (Bass & Weintraub 1988; Wagnert et al. 1989). It turns out that ADAR targets duplex regions in RNA, switching A-U pairs to less stable I-U pairs.

There are three ADAR genes that have so far been identified in mammals: *ADAR1*, *ADAR2* and *ADAR3* (Nishikura 2010). *ADAR1* has two isoforms of a long "*ADAR1p150*" and a short "*ADAR1p110*", which are determined by alternative promoters and different start codons. *ADAR2* and *ADAR1p110* are relatively ubiquitously expressed while *ADAR1p150* is induced by interferon. *ADAR3* is mainly detected in central nervous system but its editing capacity is known to be inactive, at least *in-vitro*. ADARs share a common structure (Figure 1.3), consisting of a highly conserved, C-terminal deaminase (catalytic) domain and a variable number of double-stranded RNA-binding domains (dsRBDs) in the N-terminal half of the protein.

The targets of *ADAR1* and *ADAR2* differ but also overlap (I. X. Wang et al. 2013). For example, in *GluA2*, a key subunit of the AMPA receptor, the editing site converting an amino acid from glutamate (Q) to arginine (R), called as 'Q/R site', is only edited by *ADAR2*, while the site changing arginine (R) to glycine (G), denoted as 'R/G site', can be edited by both *ADAR1* and *ADAR2* (Higuchi et al. 2000).

ADARs are essential in mammals although ADAR-null invertebrates are viable with behavioral defects. Specifically, *ADAR1*-knockout mice are embryonic lethal at embryonic day

12.5 (E12.5) owing to defects in erythropoiesis, stress-induced apoptosis, liver disintegration, and overproduction of type I interferon (Wang et al. 2004; Hartner et al. 2004). Mice that lack *ADAR2* died within 3 weeks after birth as a result of seizures (Brusa et al. 1985; Higuchi et al. 2000). In the case of *C. elegans* with homozygous deletions of both *ADAR1* and *ADAR2*, defective chemotaxis is observed (Tonkin et al. 2002). Drosophila without the ADAR locus exhibits uncoordinated locomotion and age-dependent neurodegeneration (Palladino et al. 2000). Interestingly, the impaired phenotype caused by knockout of *ADAR2* in mice can be rescued by introducing an arginine (R) codon at the Q/R site in *GRIA2* (Higuchi et al. 2000), implying that Q/R editing is essential for survival in *ADAR2*-mediated A-to-I editing. However, the rescued *ADAR2* knockout mouse still showed significant changes in behavior, hearing ability, allergy parameters and transcript profiles of brain, suggesting that *ADAR2* affects broad physiology in mice (Horsch et al. 2011).

**Site-selective editing and hyper-editing**

ADARs can deaminate various numbers of adenosines in a target mRNA. Because of this characteristic, A-to-I editing can have two different modes: site-selective and nonselective, called hyper-editing. Typically, in site-selective editing, one or a few A-to-I editing sites are observed. In the case of hyper-editing, multiple adenosines are subjected to deamination (Wahlstedt & Öhman 2011). Coding regions of mRNAs usually undergo site-selective deamination while noncoding regions of mRNA such as introns and untranslated regions (UTRs) tend to be a target of hyper-editing.

Previous studies demonstrate that the number of A-to-I editing sites usually increases with the length of RNA duplexes and is affected by structure of RNA duplexes. Site-selective editing is often found in short duplexes, between 15 and 40 bp in length, interrupted by bulges

and internal loops. Hyper-editing, on the other hand, is usually observed in long (greater than 50 bp) and almost completely base-paired dsRNA (Hundley & Bass 2010).

Hyper-editing is enriched in Alu elements because their genomic structure is favorable to form a long and perfect double-stranded structure when expressed. The Alu repeat element with about 300 nucleotides in length is the most abundant primate-specific retroelement, and makes up more than 10% of the human genome. Most Alu repeats are located within genes (usually in introns or 3' UTRs), and are hence transcribed as part of the pre-mRNA transcript of the gene. Owing to the abundance of Alus, it is very frequent to find mRNA transcripts containing two nearby Alus in opposite orientation. As the mRNA molecule folds, these two Alus may form secondary RNA structures that are targeted by ADAR, resulting in a substrate for hyper-editing (Knisbacher & Levanon 2015). Recent genome-wide studies of the human transcriptome (Blow et al. 2004; Sakurai et al. 2014) confirm large numbers of A-to-I editing associated with Alu elements.

**Molecular function of A-to-I editing**

Site-selective editing in coding sequence (CDS) can alter amino acid sequences, diversifying protein isoforms. A limited number of cases have been identified so far, especially in genes involved in the central nervous system, such as ligand- and voltage-gated ion channels as well as G-protein-coupled receptors. In most cases, A-to-I editing creates multiple isoforms of proteins essential for balanced neuronal kinetics (Rosenthal & Seeburg 2012). Several examples are reviewed in a detail in the next section (1.3 A-to-I editing in brain and disease).

Site-selective editing is also found in intronic regions and has the potential to modulate splice site usage (Rueter et al. 1999; Schoft et al. 2007). The exemplary case is an A-to-I editing found in *ADAR2* (Figure 1.4). Here, A-to-I editing in an intron between exon3/4 and exon 5 of

*ADAR2* generates 3' splicing signal ('AG'), producing a longer transcript containing the 47-nucleotides insert. Insertion of the 47-nucleotides induces decreased *ADAR2* protein expression, representing a negative autoregulatory mechanism by which *ADAR2* can prevent its own expression (Rueter et al. 1999).

Moreover, site-specific editing in 3' UTRs evades or creates the binding of miRNA (Liang & Landweber 2007; Q. Wang et al. 2013). Specifically, Wang et al. showed that A-to-I editing in ARHGAP26, a negative regulator of the Rho family essential for muscle development, disrupts the binding of miR-30b-ep and miR-573, resulting in the loss of repression by these miRNAs (Q. Wang et al. 2013).

Functional roles of hyper-editing are not well characterized yet but several studies have suggested their functions. First, previous studies showed that synthetic inosine-containing RNAs are specifically retained in the nucleus, which is mediated by RNA-binding protein, p54nrb (Zhang & Carmichael 2001; DeCerbo & Carmichael 2005). They proposed the possibility that nuclear retention of inosine-containing RNA works as a cellular mechanism to suppress gene expression. In fact, an endogenous inosine-containing RNA in mouse, Cat2 transcribed nuclear-RNA (*Ctn RNA*) was shown to be retained in the nucleus and released upon cleavage (Prasanth et al. 2005). It should be noted here that a nuclear structure, the paraspeckle was involved in this nuclear retention, and p54nrb was found in the paraspeckle.

Second, hyper-edited RNA is known to be a target of tudor staphylococcal nuclease (Tudor-SN), an eukaryotic RISC (RNA-induced silencing complex) subunit. A previous *in-vitro* study showed that Tudor-SN specifically interacts with and promotes cleavage of synthesized hyper-edited dsRNA substrates containing I-U and U-I pairs (Scadden 2005). However, no endogenous substrate of Tudor-SN has been reported so far. Some researchers hypothesized that

9

endogenous RNA is targeted by Tudor-SN only in response to certain environmental conditions (Hundley & Bass 2010).

Lastly, there were reports that hyperedited dsRNA can exert its functions *in trans*. For example, inosine-containing dsRNA lead to global downregulation of gene expression (Scadden 2007), which is mediated by stress granules (SGs) interacting with edited transcripts. Also, the same group showed that hyperedited dsRNA with multiple IU pairs suppresses interferon induction and apoptosis (Vitali & Scadden 2010).

All of these studies demonstrated direct or indirect roles of hyper-editing, especially for sites in 3' UTR. However, their functional roles are still under debate. For example, endogenous mRNA with multiple inosines in their 3' UTRs have been found in mammalian and *C. elegans* cell cytoplasm, arguing that nuclear retention of inosine-containing RNA is not a general phenomenon  (Hundley & Bass 2010). The roles of hyper-editing might be diverse depending on the cellular context and further studies will give better understanding on the functionalities of hyper-editing.

## 1.3 A-to-I editing in brain and diseases

**A-to-I editing in brain**

It is believed that A-to-I editing is most abundant in brain among all mammalian tissues (Table 1.2) (Paul & Bass 1998). In fact, some important A-to-I editing events are found in the nervous system. These A-to-I editing events are often involved in amino acid changes regulating neurotransmission. The followings are some important examples of A-to-I editing.

The first example is A-to-I editing in glutamate receptors. In fact, this is the first A-to-I editing found in mammals (Sommer et al. 1991). It was found serendipitously while researchers studied the excitatory glutamate receptors. Cloning of the mammalian AMPA receptor subunit GRIA2/GluA2 showed a discrepancy between the mRNA and the corresponding DNA template, which indicated A-to-I editing. The function of the editing is significant because it recodes amino acids from a glutamine (Q) to an arginine (R) around the channel pore, affecting the permeability of calcium ions. In normal physiological situations, full editing prevents calcium ions from moving through the assembled receptors. The failure of this editing causes epileptic seizures and death in mice (Higuchi et al. 2000). In addition to AMPA glutamate receptors, kainate glutamate receptors including GRIK1/GluR5 and GRIK2/GluR6 also have Q/R editing (Barbon & Barlati 2011). In particular, GRIK2 Q/R editing is involved in synaptic plasticity via long-term potentiation (Vissel et al. 2001). In glutamate receptors, there is another important A-to-I editing site, namely an R/G editing site where an arginine (R) is converted to glycine (G) in the extracellular ligand binding domain of the receptors (Barbon & Barlati 2011). This is mainly found in subunits of AMPA receptors such as GRIA2/GluA2, GRIA3/GluA3 and GRIA4/GluA4. Editing at this site increases the recovery rate from desensitization of the ion channels, allowing faster response to repeated impulses (Kappler et al. 2002).

The second A-to-I editing example involves the serotonin receptor. Serotonin or 5-HT is a neurotransmitter that modulates a wide array of physiological processes including mood, appetite, pain perception, locomotion, memory and sexual behavior (Roth 2006; Berger et al. 2009). RNA editing is found in one of the G-protein-coupled serotonin receptor, 5-HT2cR. The A-to-I editing in 5-HT2cR, which alters the amino acid sequences of the second intracellular loop of the receptor, lead to a 10~15 fold reduction in the efficacy of the interaction between receptors and their G proteins  (Burns et al. 1997). Because 5-HT2cR is involved in many human

psychiatric and behavioral disorders including depression and schizophrenia, the A-to-I editing in 5-HT2cR draws much attention in terms of clinical implications (O'Neil & Emeson 2012).

The last example is A-to-I editing found in a voltage –gated potassium channel, Kv1.1. The Kv1.1 is one of the most widely expressed voltage-gated potassium channel playing an important role in excitability by repolarizing membranes and shaping the firing properties (Jan & Jan 2012). A-to-I editing in the Kv1.1transcript substitutes an isoleucine (I) for a valine (V) in the highly conserved ion-conducting pore of the channel, resulting in an increased recovery rate of the channel (Bhalla et al. 2004). This is an example showing that A-to-I editing shapes the action potential, a fundamental element of the neuronal system.

**A-to-I RNA editing in brain development**

Interestingly, some A-to-I editing events are differentially regulated in brain development and these differential regulations of A-to-I editing are expected to have functional significance in this respect. For example, several sites including the R/G site in glutamate receptors (e.g. GRIA2, GRI3, GRIA4) show a gradual increase in editing during brain development (Wahlstedt et al. 2009). Considering the effects of R/G editing on functions of glutamate receptors, it can be hypothesized that unedited status of R/G editing sites of glutamate receptors in fetal brain makes the recovery of the receptors from desensitization slow, which avoids their hyper-activation under elevated glutamate conditions.

Another instance can be found in GABAA receptors, a ligand-gated chloride channel. The GABAA receptor produces a hyperpolarizing influx of chloride ions when active in the mature brain. However, it generates depolarizing currents creating an excitatory response to GABA in fetal brain, which is crucial for a number of developmental processes including proliferation and synaptogenesis (Ben-Ari 2014). This functional switching is largely due to different subunit

composition of the receptors. The alpha subunit is particularly important as alpha3 (GABRA3) is found in the receptor in fetal brain while alpha1 (GABRA1) is expressed in adult brain (Rula et al. 2008). A-to-I editing in GABRA3 in part plays a role here because an edited GABRA3 transcript tends to be inefficiently transported to neurites and falls under enhanced lysosomal degradation, contributing to a component change of GABAA receptors from GABRA3 to GABRA1 (Daniel et al. 2011).

A recent study discovered a more striking case regarding A-to-I editing change in neuronal differentiation (Pachernegg et al. 2015). In general, GRIA2 Q/R site is known to be almost exclusively edited across brain development. However, Pachernegg et al. showed that editing levels of GRIA2 Q/R site increases rapidly during the very early stages of *in vitro* neural differentiation. They also demonstrated that neural progenitor cells (NPCs) express glutamate receptors permeable to calcium ion, when GRIA2 remains unedited by Q/R editing. Consistent with these results, delivery of ADAR2 into NPCs prevents neuronal differentiation (Whitney et al. 2008). Along with a gradual increase of R/G editing in brain development, a rapid increase of Q/R editing during early neurogenesis highlights the importance of A-to-I editing in neuronal development.

**A-to-I RNA editing in diseases**

Dysregulation of A-to-I editing is observed in many diseases, primarily neurological or psychiatric diseases (Gallo & Locatelli 2012). For example, aberrant A-to-I editing in GRIA2 is implicated in amyotrophic lateral sclerosis (ALS), malignant glioma and ischemia (Yamaguchi et al. 1999; Maas et al. 2001; Kawahara & Kwak 2004). Also, imbalance of A-to-I editing in the serotonin receptor, 5-HT2cR was found in psychiatric diseases such as suicide and schizophrenia (Niswender et al. 2001; Sodhi et al. 2001). However, psychiatric associations of A-to-I editing in

13

general are still debates (Lyddon et al. 2012; Eran et al. 2013). Moreover, functional roles of A-to-I editing in pathological conditions are largely unclear although deficiency of A-to-I editing at GRIA2 Q/R site is related to death of neuronal cells in some pathological conditions such as glioma and ALS (Slotkin & Nishikura 2013).

```
DNA:              AATTCGGCTACGCTCGAACGTACGTCA
RNA:              AAUUCGGCUAC|GCUCGAACGUACGUCA
```

Edited RNA:       AA–UCGGCUACAAGCUCGAACGUACGUUA

      Deletion      Insertion      Substitution
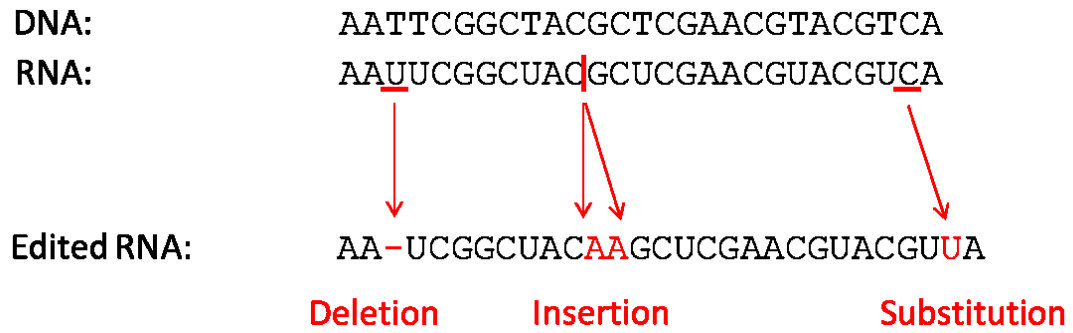
**Figure 1.1 Three types of RNA editing.** Deletion, Insertion and Substitution are depicted.
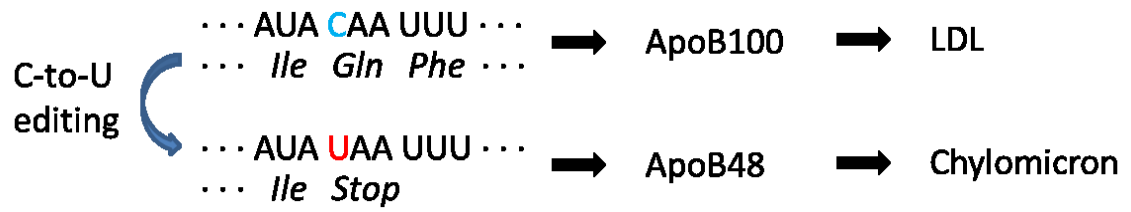
**Figure 1.2 RNA editing in apolipoprotein B (apoB) mRNA.** C-to-U editing at cytidine 6666 in apoB mRNA generates a protein isoform functionally different from the unedited one. LDL: low-density lipoprotein.
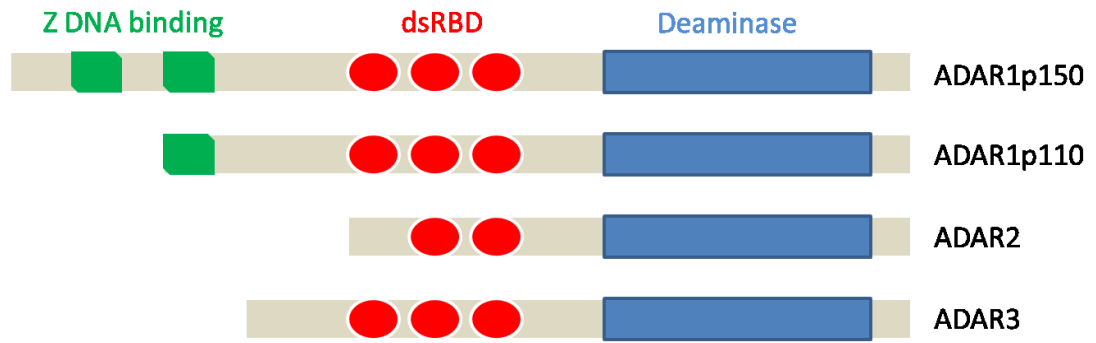
**Figure 1.3 Schematics of ADAR family**. dsRBD: double-stranded RNA binding domain.
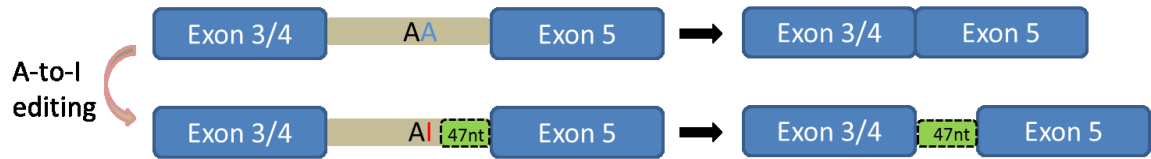
**Figure 1.4 A-to-I editing in ADAR2.** A-to-I editing in an intron between exon 3/4 and exon 5 extends the length of transcript by adding 47 additional bases before exon5.

| Type of editing | edited RNAs | Organisms | Genetic system |
|---|---|---|---|
| U insertion/deletion | mRNA | kinetoplastid protozoa | mitochondrial |
| N insertion | mRNA, rRNA, tRNA | Physarum polycephalum, other myxomycete slime molds | mitochondrial |
| C-to-U substitution | apolipoprotein B mRNA | mammals | nuclear |
| | mRNA, tRNA | land plants | mitochondrial, chloroplast |
| | tRNA$^{asp}$ (anticodon) | marsupials | nuclear |
| | tRNA$^{trp}$ (anticodon) | kinetoplastid protozoa | nuclear/mitochondrial |
| | cox1 mRNA | P. polycephalum | mitochondrial |
| | cox1, cox3 mRNAs | Naegleria gruberi | mitochondrial |
| U-to-C substitution | mRNA | land plants | mitochondrial, chloroplast |
| | cox1 mRNA | placazoan (Trichoplax adhaerens) | mitochondrial |
| | WT1 mRNA | mammals | nuclear |
| A-to-I substitution | mRNA, miRNA, viral RNA | metazoan animals | nuclear |

**Table 1.1 Examples of RNA editing in Eukaryotes.** *N insertion* means insertion of C, U, dinucleotide. 'nuclear/mitochondrial' indicates that the substrate is a nucleus-encoded tRNA that is imported into mitochondria where it undergoes editing. This table is modified from Gary 2012.

| Tissue | IMP (pmol) | One IMP for every: |
|--------|-----------|--------------------|
| Brain | 1.8 ± 0.9 | 17,000 nt |
| Lung | 0.9 ± 0.7 | 33,000 nt |
| Heart | 0.9 ± 0.6 | 33,000 nt |
| Thymus | 0.5 ± 0.3 | 60,000 nt |
| Muscle | 0.2 ± 0.1 | 150,000 nt |

**Table 1.2 Amounts of IMP in different rat tissues.** IMP: inosine monophosphate. This table is modified from Paul & Bass 1998.

# Chapter 2

# Identification of RNA editing events from genome-wide sequencing of

# RNA (RNA-seq)

In order to accurately identify RNA editing sites using RNA-seq, a computational pipeline is necessary to reliably identify RNA editing sites. In this chapter, I will review issues in identifying RNA editing sites from RNA-seq and describe the computational tool developed in this study to identify RNA editing sites at a genome-wide level from RNA-seq.

## 2.1. Overview

In principle, detecting RNA editing events is straight forward because it just requires comparisons of RNA sequences or complementary DNA (cDNA) sequences with template DNA sequences. In particular, A-to-I editing is easily identified as an adenosine to guanosine (A-to-G) discrepancy in the comparison because an inosine base pairs with cytosine in cDNA synthesis and Sanger sequencing. In fact, several important A-to-I editing sites were found serendipitously through comparisons of cDNA with the genomic sequence.

With the recent advent of next generation sequencing (NGS) technology that generates a huge number of short sequences ("sequencing reads") in a parallel way, it is now possible to have sequence information from DNA or RNA at a genome-wide level, where the genome-wide sequencing of DNA and RNA are referred to as "whole-genome sequencing" and "RNA-seq", respectively. Now, RNA editing can be found in a systematic way at a genome-wide level by comparing sequences between whole-genome sequencing and RNA-seq obtained from the same sample.

However, because whole-genome sequencing is still inefficient in terms of cost, an efficient way to identify RNA editing using only RNA-seq has been of special interest in the RNA editing field. In particular, it is still daunting to perform both whole-genome sequencing and RNA-seq for all the samples in a study necessary for a large sample size. A typical approach to identify RNA editing from the only RNA-seq is to compare sequencing data from RNA-seq with a reference genome, and controlling for false-positive calls at a reasonable rate. In this chapter, I will review several issues in the discovery of RNA-editing from RNA-seq and present the computational pipeline to detect RNA editing sites, which was developed for this study.

## 2.2. Issues in identifying RNA editing sites from RNA-seq data only

There are three main issues in calling RNA editing sites from RNA-seq: contamination of genomic variants, technical errors originating from inherent RNA-seq technology, and errors introduced by an alignment step during processing of RNA-seq data.

**Contamination of genomic variants**

When RNA editing is called from RNA-seq without the corresponding DNA template information, a typical approach is to compare RNA sequence determined by RNA-seq with a reference genome and detect sequence discrepancies as putative RNA editing sites. But significant proportions of these candidates are likely genomic variation, especially due to single nucleotide polymorphisms (SNPs) (Figure 2.1). Therefore, the false positive rate in the initial call obtained from a simple comparison between RNA-seq with reference genome is unacceptably high.

**Technical errors inherent to sequencing technology**

A typical sequencing read from NGS has non-uniform sequencing qualities from base to base. Generally, 3' ends of sequencing reads have poor sequencing quality (Figure 2.2). Therefore, one should be cautious if a called RNA editing site is enriched by sequencing reads whose bases at the editing site have poor sequencing qualities. For a given base of a sequencing read, sequencing quality is quantified by the Phred score, Q defined as follows:

$$Q = -10 \log_{10} P$$

23

where *P* is the probability that the base is called wrong. A Phred score smaller than 20 indicates that the accuracy of the base call is less than 99%. In addition to non-uniform sequencing qualities, 5' ends of sequencing reads can have random sequence regardless of biological sources due to a PCR step with random primers necessary for the preparation of a sequencing library.

**Alignment errors in RNA-seq**

Although most of sequencing reads from RNA-seq are accurately aligned to the reference genome when aligned by commonly-used software such as TopHat (Trapnell et al. 2009) or STAR (Dobin et al. 2013), there are still sequencing reads that are aligned ambiguously, mainly around the regions with RNA splicing and repeat sequence. Although these ambiguously-mapped reads may not be a critical problem when using RNA-seq as a tool to estimate mRNA abundances, currently a main reason for performing RNA-seq, they should be critically considered for an RNA editing study as they can be a source of false positive calls of RNA editing sites (Kleinman & Majewski 2012).

## 2.3. Development of computational tools to identify RNA-editing from RNA-seq data

In order to accurately identify RNA editing sites using RNA-seq, I developed a computational pipeline to make reliable calls of RNA editing sites, considering the issues discussed in the previous section. This pipeline consists of several steps including alignment, variant-calling, and filtering of false positives. Figure 2.3 summarizes the computational pipeline.

**Alignment**

RNA-seq reads were aligned by BWA (Li & Durbin 2009) to the reference genome (UCSC hg19) as well as the annotated transcriptome obtained from several public databases including NCBI RNA reference sequence collection (RefSeq), UCSC and Ensembl. Using the combined transcript information from multiple databases and not attempting to detect novel isoforms from RNA-seq data help to avoid false positives due to alignment errors.

**Variant calling**

After alignment is done, putative RNA editing sites were identified by a variant-calling software, "Genome Analysis Toolkit" (GATK) (DePristo et al. 2011), using uniquely mapped reads after PCR duplicates were removed by the Picard tool (Auwera et al. 2013). Note that A-to-I editing events appear as A-to-G discrepancies in variant-calling.

**Computational Filters**

For initial list of RNA editing sites, computational filters were applied to deal with the major issues in calling RNA editing sites from RNA-seq: (1) contamination of genomic variants, and (2) false positives due to sequencing or alignment errors. In our pipeline, potential genomic variants were filtered out through three steps: (i) removing all known SNPs in dbSNP version 137 (Sherry et al. 2001) except for SNPs of molecular type "cDNA"; (ii) taking out sites with variant-supporting reads only under the assumption that 100% editing efficiency is unrealistic; (iii) keeping variants detected in at least two individuals because they are unlikely to be rare variants. Possible false positive RNA editing sites due to sequencing or alignment errors were removed by taking advantage of their tendency to have bias in terms of positions on a sequencing read and

strands (Kleinman & Majewski 2012). After categorizing sequencing reads spanning a putative

RNA editing site into two groups, a reference-supporting group and a variant-supporting group

according to the alleles in the reads, we applied a two sample t-test and a Fisher-exact test to see

if the two groups were statistically different in terms of position bias (Figure 2.4) and strand bias

(Figure 2.5), respectively. For a given site, the position on a sequencing read was defined as the

smaller distance from either end of the read. If the p-value from the test was smaller than a given

threshold, the site was declared as a false positive. In this study, 0.01 was chosen as the p-value

threshold and used Bonferroni-correction for multiple testing correction.

**Final call of RNA editing sites**

Finally, reliable RNA editing sites were identified if there were enough sequencing reads

at the sites, such that at least one individual had more than or equal to 5 high-quality (PHRED

score $\geq$ 20) reads and more than two high-quality and variant-supporting reads were found in at

least one person. Sites with two or more variants were excluded under the assumption that RNA

editing is generated only by transition from purine to purine or pyrimidine to pyrimidine.

## 2.4. Validation of the computational tool

In order to evaluate the developed computational pipeline, we obtained both DNA and

RNA sequence information from two samples. Specifically, two pairs of DNA exome-sequencing

(exome-seq) and RNA-seq were generated and used to estimate the false discovery rate (FDR) of

the developed pipeline, under the assumption that observed DNA and RNA sequence differences are mainly caused by RNA editing.

**Exome-sequencing**

Exome-sequencing is a variant of whole genome sequencing. Here, instead of sequencing a genome, exon regions are enriched and sequenced. For a given sequencing capacity, exome-sequencing gives much higher number of sequencing reads for exonic regions, compared to whole genome sequencing, which makes it possible to call accurate genotype. I used the *Agilent Sure Select Capture* system to enrich exonic regions from DNA samples.

**Validation**

First, RNA editing sites were identified from individual RNA-seq through the above pipeline, except for the step requiring the multiple samples. Second, genotypes for RNA editing sites were called from matched exome-seq using uniquely-mapped reads as determined by BWA and GATK pruning tools for realignment and recalibration (Auwera et al. 2013). A given genomic site was declared as homozygous if reads supporting major alleles comprise a proportion greater than 90% of the sequencing depth. Finally, the false discovery rate (FDR) was estimated with the sites whose sequencing depths in exome-seq were greater than or equal to 20 to ensure reliable genotype calls. Here, false positives are RNA editing sites where corresponding genomic sites do not have homozygous genotypes with the reference allele. FDR of two matched sets are 5.1% and 6.0%, respectively (Table 2.1). It should be noted that the additional filters in the pipeline, which take advantage of the multitude of samples, were not used for this evaluation but would be expected to decrease FDR further in the application of pipelines to multiple samples.
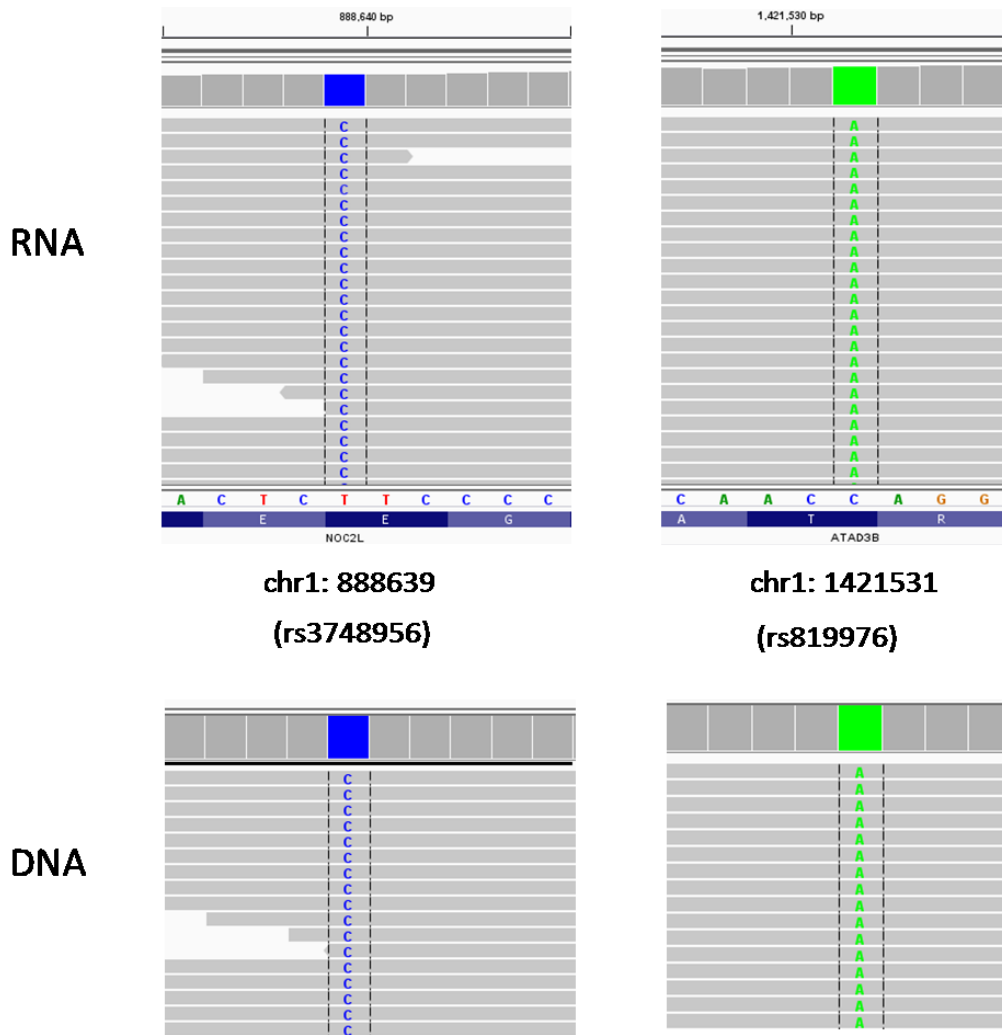
**Figure 2.1 False positive call of RNA editing at SNP sites.** Two sites are shown where RNA editing sites are initially called from RNA-seq (top) as reference sequences are T and C but sequencing reads have C and A at the sites, respectively. However, they are previously-annotated SNP sites (rs3748956 and rs819976). In fact, they turned out to be false positives, based on DNA sequencing (bottom) as DNA and RNA have same sequences.
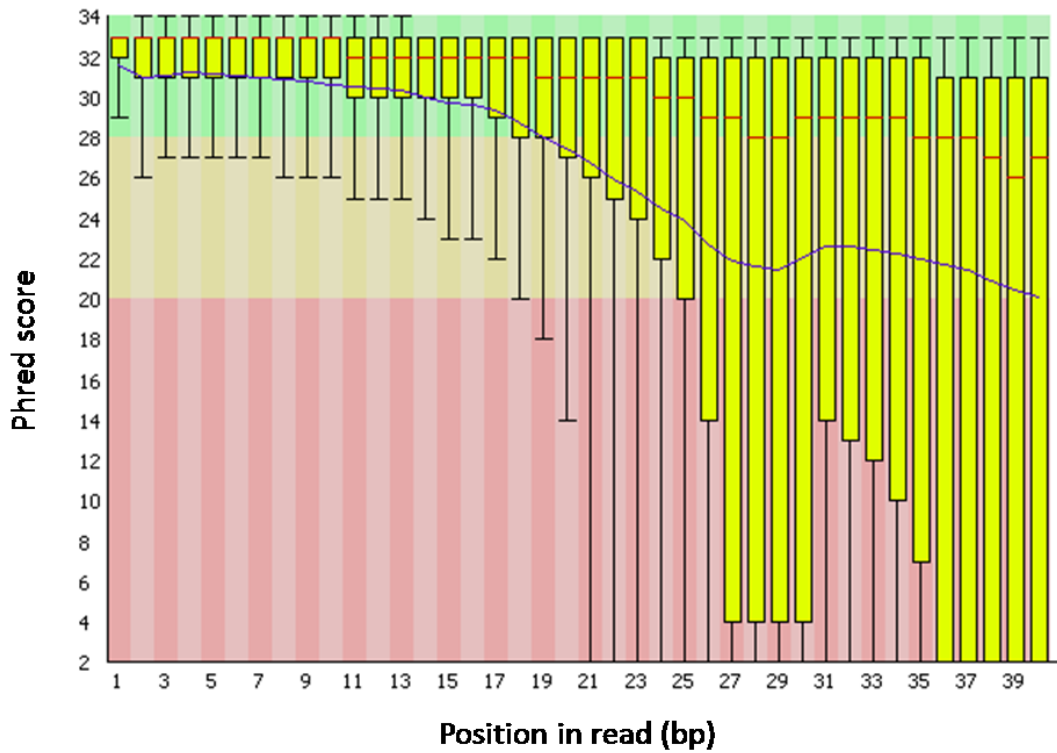
**Figure 2.2 Varying sequencing qualities in a sequencing read.** Histogram shows the distribution of Phred score across sequencing reads in a single run according to the position in a read. Here, the size of a sequencing read is 36 base. The average sequence quality, denoted by the line, is decreasing from 5' end to 3' end. The figure is adapted from the following website:

'https://en.wikibooks.org/wiki/Next_Generation_Sequencing_(NGS)/Pre-processing'.

**RNA-seq**
- polyA pull down library
- paired-end 100 bp

**Alignment**
- Using BWA
- Using multiple annotations:
  RefSeq, UCSC, Ensembl, Encode/Gencode
- Selecting uniquely-mapped reads
- Removal of suspected duplicates

**Variant call**
- Using GATK
- Recalibration of base quality scores
- Variant call against a reference genome

**Genomic variant filter**
- Remove genomic SNPs previously reported.
- Remove sites if edit rate=1 (Q20 reads>=30).
- Variants should be shared by at least two ppls.

**Bias filter**
- Strand bias filter (Fisher-exact test)
- Position bias filter (Two sample t-test)

**Figure 2.3 Computational pipeline to identify RNA editing sites from RNA-seq.** Step-by-step procedures are described with a brief description.
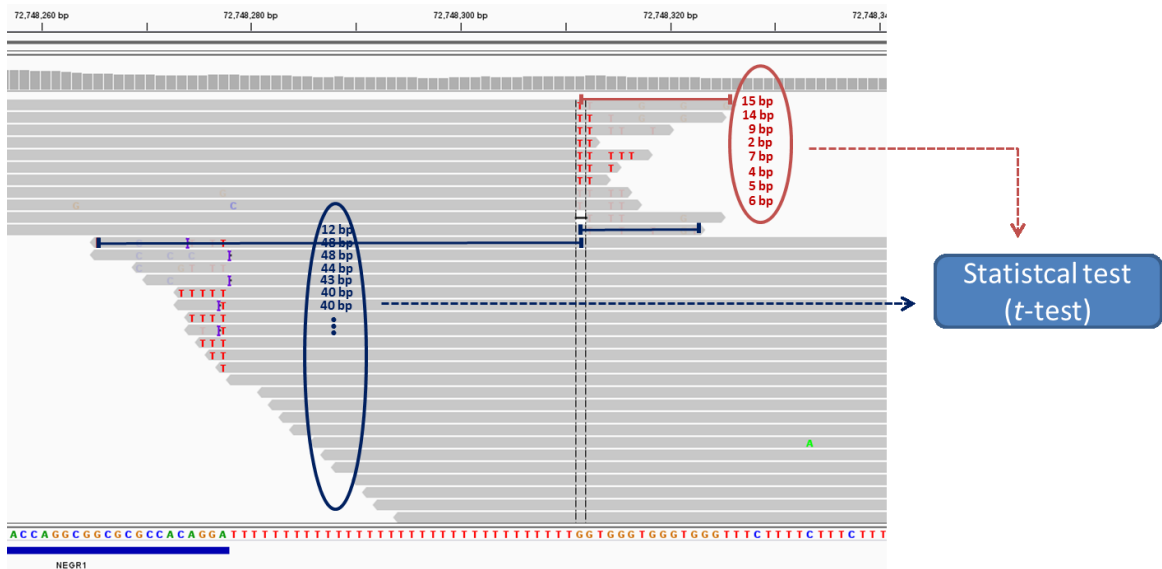
**Figure 2.4 Position bias filter.** For a given site, reference-supporting reads and variant-supporting reads are identified. They are statistically compared in terms of distance from an end of the sequencing reads. In the figure, aligned sequencing reads denoted by gray horizontal bars are described. Vertical dotted line indicates a site to be tested, where the reference is G (with no letter indication) and the variant is T (marked by red letter). Blue circle and red circle describe two groups consisting of reference-supporting reads and variant supporting reads, respectively. Numbers in the circles are distance from the tested site to an end of sequencing reads.
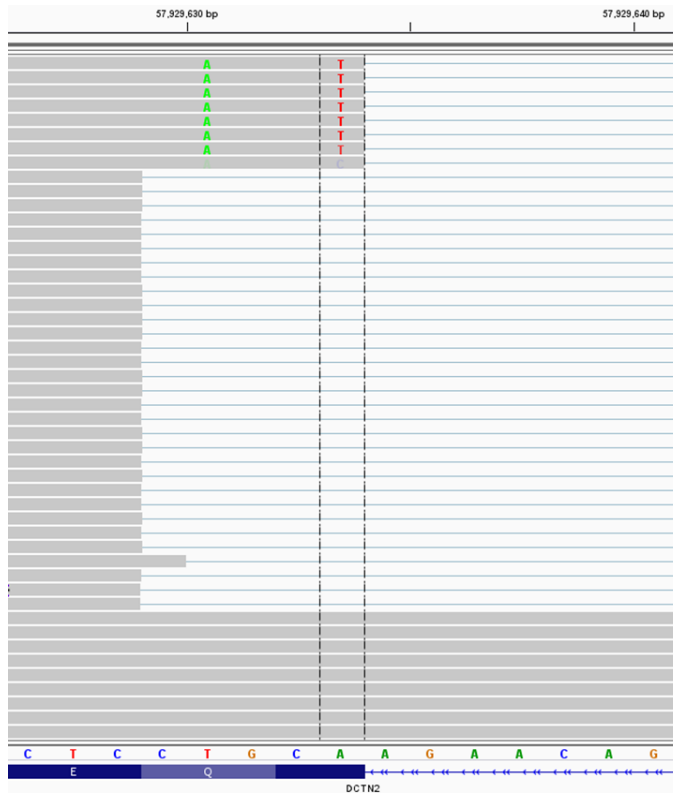
**Figure 2.5 Strand bias filter.** For a given site, reference-supporting reads and variant-supporting reads are identified. They are statistically compared in terms of strand of sequencing reads. In the figure, aligned sequencing reads denoted by gray horizontal bars are described. Vertical dotted line indicates a site to be tested, where the reference is A (with no letter indication) and the variant is T (marked by red letter). Two by two tables are constructed, as shown in the figure, to apply the Fisher exact test.

| Brain ID | DNA ID | RNA ID | Total number of reads in RNA-seq | The number of RNA editing sites | The number of testable sites (exome-sequencing depth >= 20) | The number of false positives | FDR (%) |
|---|---|---|---|---|---|---|---|
| B925 | D2136 | R3763 | 83,059,552 | 55,271 | 4,833 | 250 | 5.1 |
| B1016 | D2278 | R3670 | 112,386,874 | 214,400 | 7,119 | 428 | 6.0 |

**Table 2.1 Evaluation of computational pipeline for identifying RNA editing from RNA-seq.**
Two matched sets of DNA exome-sequencing (exome-seq) and RNA-seq were generated to evaluate the performance of our computational pipeline to identify RNA editing sites from RNA-seq. RNA editing sites were identified from RNA-seq through the pipeline described in methods, except for the step requiring the multiple samples. The false discovery rate (FDR) was estimated with the sites whose sequencing depths in exome-seq were greater than or equal to 20 (testable sites). The false positives are RNA editing sites whose genotypes are not homozygous with the reference alleles.

# Chapter 3

# Landscape of RNA editing in human brain development

Here, I analyzed RNA-seq in post-mortem human brain tissues obtained from a total of 33 individuals spanning fetal to old age in order to generate comprehensive RNA editing profiles in human cortical development. In addition to confirming previous results about RNA editing, I found that the genome-wide landscape of A-to-I editing reflects a set of uniquely regulated RNA editing sites, comprising three distinct patterns: stable high, stable low and increasing across cortical development.

## 3.1. Significance of RNA editing study in human brain development

Despite the apparent frequency and consequence of RNA editing in brain as discussed in the Chapter 1, its broader regulation and functional roles are unclear (Li & Church 2013). In particular, the contribution of RNA editing is largely unexplored in human cortical development where diverse molecular processes must be orchestrated in a timely and precise manner. Previous studies investigating A-to-I editing in human brain have offered a very limited picture of editing sites and of their developmental variation (Blow et al. 2004; Li & Church 2009; Li et al. 2013; Ramaswami et al. 2013; Sakurai et al. 2014). For examples, prior studies have looked at only known editing sites (Li et al. 2013) or only in adult brain samples (Blow et al. 2004; Sakurai et al. 2014). This limited profile of RNA editing makes it impossible to elucidate mechanisms or functions in the development of the nervous system. A single previous study that explored global A-to-I editing profiles across brain development in the mouse (Dillman et al. 2013) did not address mechanistic or functional implications. More importantly, the results from mouse brain tissues are limited in their translation into RNA editing in human brain, particularly as A-to-I editing is enriched in Alu repeats, a primate-specific DNA element (Athanasiadis et al. 2004; Kim et al. 2004). Therefore, it is valuable to investigate comprehensive RNA editing profiles in human cortical development, which can provides insights to understand the function and regulation of A-to-I editing.

## 3.2. Post-mortem human brain samples

Post-mortem brain tissues from 33 individuals without neurological or psychiatric illnesses were obtained as previously described (Jaffe et al. 2015) and dissected at the Lieber

Institute for Brain Development (LIBD), USA. Specifically, the samples are of prefrontal cortex (DLPFC) grey matter (BA 9/46) spanning from birth to the eight decade of life. Fetal tissue was taken from the prefrontal region, over the dorsal convexity of the frontal lobe just anterior to the temporal pole. The samples were categorized into six distinct life stages: fetal, infant (< 12 months), child ($1 \leq$ age <10), teen ($10 \leq$ age <20), middle ($20 \leq$ age <50) and old life (age $\geq$ 50). In addition, 20 brain samples were collected from 5 additional healthy old individuals, each generating four samples for different brain regions including dorsolateral prefrontal cortex (DLPFC), cerebellum (CB), hippocampus (HIPPO), and entorhinal cortex (ERC). In order to reduce heterogeneity in the samples, only brain tissue from Caucasian individuals were collected. The demographic details of the samples are presented in the Table 3.1.

## 3.3. Identification of RNA editing sites from RNA-seq across human brain development

**RNA-seq**

The poly-A enriched RNA-seq libraries were prepared according to the standard Illumina protocol. RNA-seq was performed to generate paired-end reads with 100-bp read length. Approximately 120M reads per sample were acquired (Tables 3.2).

**Identification of RNA editing sites**

As a discovery set to identify RNA editing sites, 33 DLPFC samples were interrogated by the computational pipeline described in Chapter 2. This results in the genome-wide identification

of RNA editing sites across human brain development. RNA editing types were assigned based on strand information from RefSeq.

## 3.4. Landscape of RNA-editing in human brain development

**Global characteristics of RNA editing in human brain development**

A total of 267,766 RNA variant sites were identified in our 33 fully sequenced samples across brain development. On average, over 66,000 sites were identified in an individual sample (Table 3.2). Four major variant types were identified in genic regions, including A-to-G, T-to-C, G-to-A and C-to-T, each comprising a proportion greater than 1% (Fig. 3.1 and Table 3.3). While T-to-C and G-to-A variants are not canonical RNA editing types, they can be understood as possible A-to-I editing and C-to-U editing, respectively, if we consider incomplete strand annotation or antisense transcription. With this consideration, the proportion of the two known RNA editing types (A-to-I and C-to-U) accounts for most of the RNA variants in the list (94%). In particular, A-to-I editing sites are disproportionately enriched, as expected. I also confirmed that 85.5% of the identified A-to-I editing sites have been found in previous studies of various human tissues (Ramaswami & Li 2014) (115,336 of 134,914 sites). These results collectively indicate a successful identification of RNA editing sites in our study.

Next, A-to-I editing was focused, which is the most enriched editing type in human brain development. It is found mainly in introns and in 3' UTRs as well as in Alu repeat regions (Table 3.4), which is consistent with previous studies using human brain tissues (Blow et al. 2004; Sakurai et al. 2014). Interestingly, as shown in figure 3.2, a large proportion of RNA editing sites in coding sequence (CDS) regions and 5' UTRs are not within Alu repeat regions that frequently

form double-stranded RNA secondary structures required for ADAR's target recognition. This indicates that A-to-I editing in CDS regions and in 5' UTRs is not simply a reflection of expansion of Alu repeats in primate evolution.

**Landscape of A-to-I editing rates in human brain development**

To explore A-to-I editing quantitatively on a genome-wide scale, the A-to-I editing rates were estimated across all samples. For a given RNA editing site in a given sample, the A-to-I editing rate was estimated as a ratio of the number of reads supporting RNA editing to the total number of reads covering that site (Figure 3.3). Only aligned reads with high sequencing quality (PHRED score $\geq$ 20) on the sites were used. A-to-I editing rates was compared across the six age groups: namely, fetal, infant, child, teen, middle and old age. Specifically, in order to identify differentially edited A-to-I editing sites across developmental stages, the sites on mRNA structure (5'UTR, CDS and 3' UTR) whose median number of reads across samples was greater or equal to 20 were evaluated by ANOVA among the six age groups followed by multiple test correction with FDR. 0.01 was chosen as a significance cutoff of FDR adjusted p-values. 748 sites were identified as ones with significant editing rate differences among the six age groups. Among these sites, 742 sites have a developmentally-increasing editing pattern from fetal to adult samples (Figure 3.4a, top in Figure 3.4b and Table 3.5). These increasing sites were found in at least 10 samples (Figure 3.5) and comprise five clusters based principally on their relative baseline editing rates during fetal life (Figure 3.6). In contrast, most of the other editing sites show stable editing rates across samples from fetal age to late life. Specifically, the majority of these stable sites showing low editing rates have a mean rate of about 0.1, but there are a few sites with stably high editing rates around 0.9 (bottom in Figure 3.4b). These global editing rate profiles during brain development were highlighted with three unique labels: 'stable high editing', 'stable low editing'

38

and 'increasing editing', which are hereafter collectively called 'developmental A-to-I editing patterns' (Figure 3.4b).

## 3.5. Comments on the increasing pattern

Because the increasing editing profile might also reflect cell composition change towards relatively more glia postnatally, I considered the possibility that increasing populations of glia account for these regional and developmental differences. However, in an available dataset with purified mouse cortical cells (Zhang et al. 2014), CDS-residing editing sites in the increasing pattern generally have higher editing rates in neurons than in non-neuronal cell types (Table 3.6). Also, in a comparison of mRNA levels of ADARs among different cell types using a previous human brain single-cell RNA-seq dataset (Darmanis et al. 2015), neurons have higher ADAR expression levels than various glial cell types (Figure 3.7). Thus, the possibility that the increasing pattern is related to relative glial enrichment in postnatal brain samples is highly unlikely. This is further confirmed by an analysis of editing rates after normalization to neuronal and/or glial proportions. After estimating relative neuronal and glial composition using a previously-proposed computational method (Jaffe et al. 2015) with available genome-wide DNA methylation data (Jaffe et al. 2016) from the same postnatal brain tissues in the 33 discovery set, we find that the increasing pattern in brain development is found to be present after the normalization of editing rates (Figure 3.8). Here, for fetal samples, two independent fetal tissues were used, which have both DNA methylation and RNA-seq datasets (Table 3.7). Finally, it should be noted that cerebellum, which has the highest proportion of neuronal cells among brain tissues, has both higher and lower editing rates, compared to neocortical regions, at the sites showing the increasing editing pattern in brain development (Figure 3.9 with Table 3.8). Because neurons have higher ADAR expression levels compared to other cell types as described, the

simple proportional differences of neuronal or glial population do not explain the lower editing rates at some sites in cerebellum. In this analysis, 20 brain samples (Table 3.7) were collected from 5 additional healthy old individuals, each generating four samples for different brain regions including dorsolateral prefrontal cortex (DLPFC), cerebellum (CB), hippocampus (HIPPO), entorhinal cortex (ERC). When editing rates were compared among brain regions, repeated measures ANOVA followed by FDR was performed to identify the editing sites showing differences across brain regions. These results collectively suggest that spatiotemporal changes of A-to-I editing rates in brain development do not simply reflect proportional differences in cellular compositions.
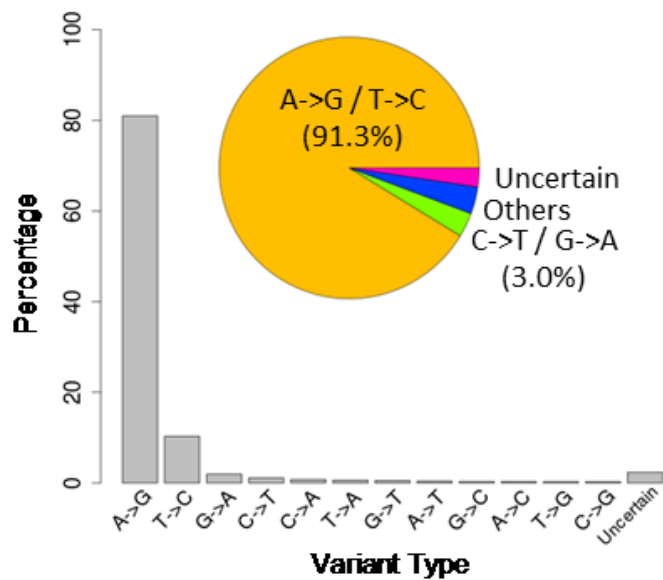
**Figure 3.1 RNA editing types in human brain development.** A-to-I editing is disproportionately enriched among RNA variants. A-to-I editing is described as A-to-G variant. Sites with ambiguous strand information are marked as 'Uncertain'.
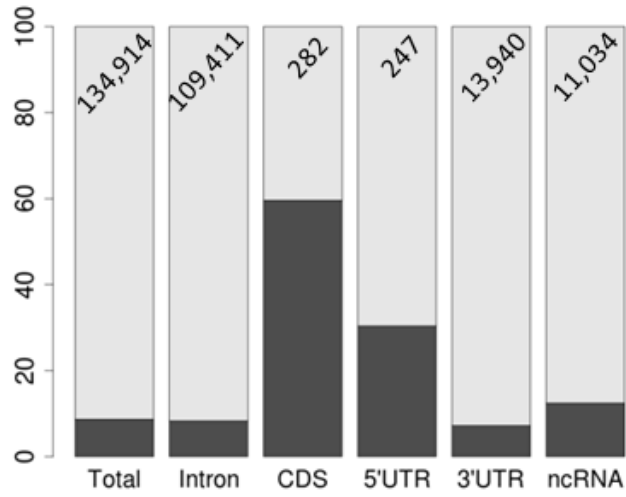
**Figure 3.2 Distribution of A-to-I editing sites according to gene regions.** The numbers of A-to-I editing sites across genic regions are shown at the top of the grey bar. The proportion of editing sites not residing in an Alu repeat is represented by black bar.
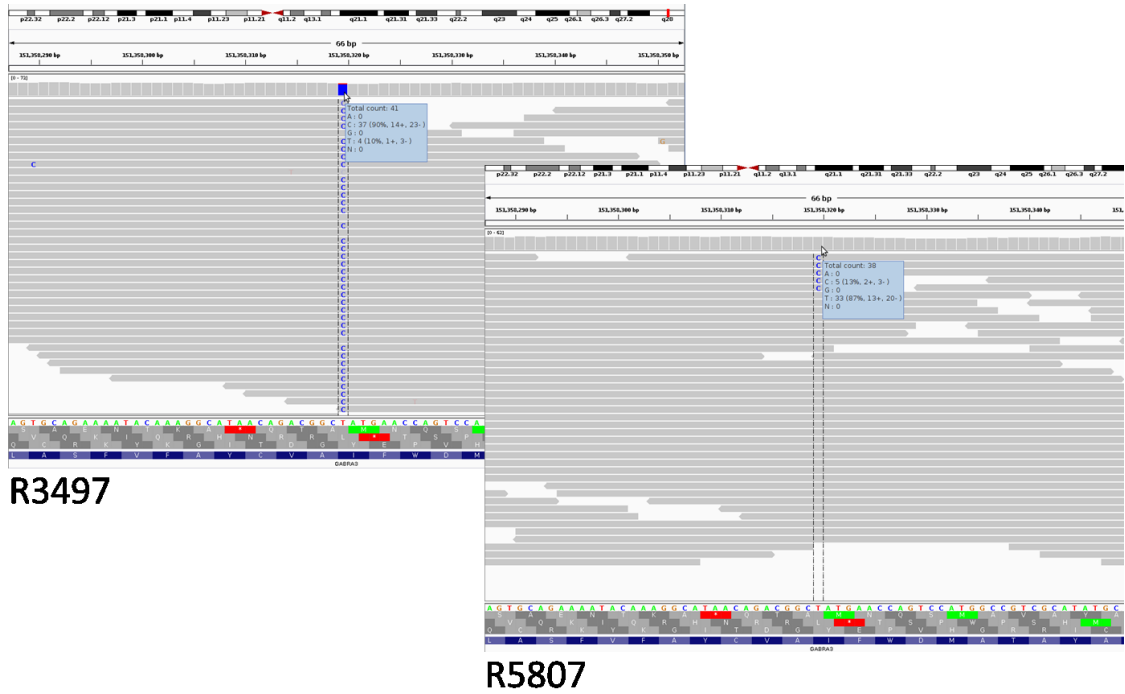
**Figure 3.3 Examples of RNA editing rates.** For a given site whose reference is 'T', editing rates are different between two samples: in R3497, 37 sequencing reads support the allele 'C' among a total of 41 reads, assigning 0.9 (≈37/41) as an editing rate. In R5807, only 5 reads in 38 reads have an allele 'C', resulting in an editing rates of 0.13 (≈5/38). Here, the sense strand is (-).
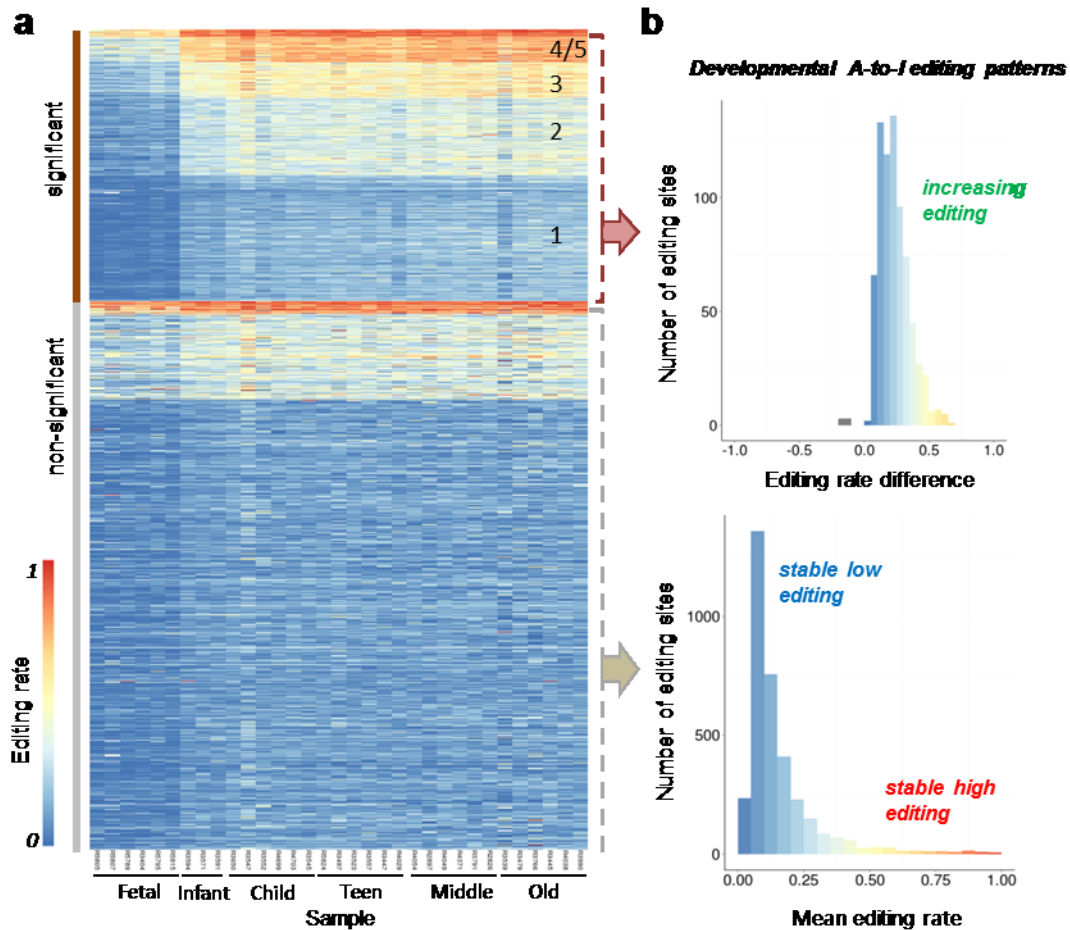
**Figure 3.4 Genome-wide profiles of A-to-I editing rates in human brain development.** (a) The editing rates for 4,282 A-to-I editing sites that have higher median numbers of sequencing reads in 33 brain samples. For a given RNA editing site in a given sample, the A-to-I editing rate was estimated as a ratio of the number of reads supporting RNA editing to the total number of reads covering that site. Color (from blue to red) indicates editing rate (from 0 to 1) for a given site (row) in a sample (column). Samples are ordered according to age from fetal (left) to old age (right). The differentially edited sites (marked as *significant*) across the six age groups are determined by ANOVA with multiple testing correction. The *significantly increasing* sites are clustered into five groups denoted by numbers in the figure, which vary principally by their absolute editing rates and not by the slope of developmental change (see figure 3.6). (b) The top

figure is the histogram of mean editing rate differences between prenatal age group and post-infant age groups for significant sites while the bottom histogram describes the mean editing rate across 33 samples for non-significant sites. Overall, the global profiles of A-to-I editing rates in human brain development are summarized by three patterns: 'stable low editing', 'stable high editing' and 'increasing editing' which collectively comprise a set of *'developmental A-to-I editing patterns'*.
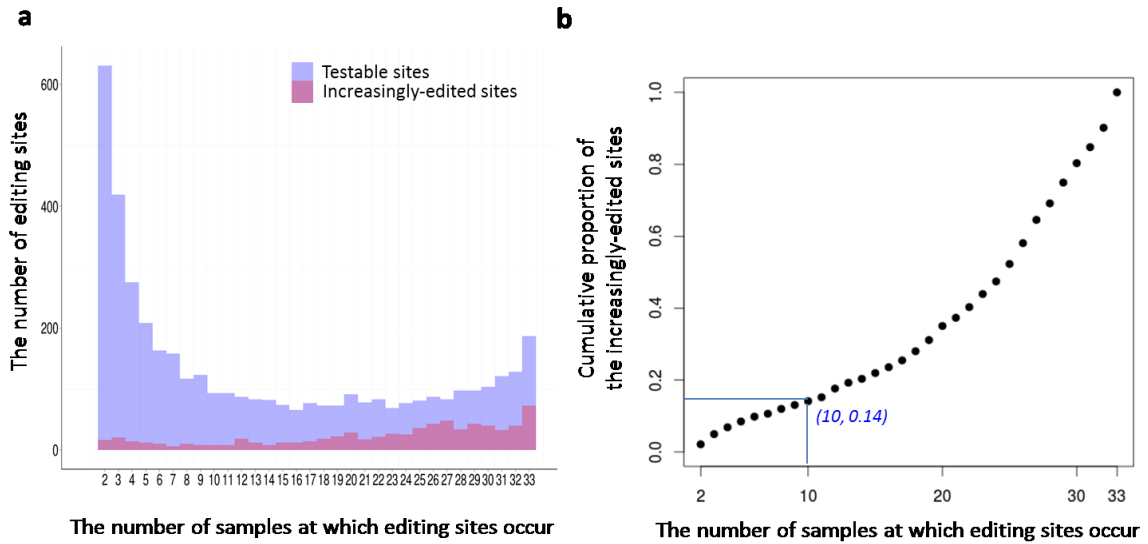
**Figure 3.5 The number of sites with increasing pattern in 33 human brain samples.** (a) Histogram of the number of editing sites according to the number of samples at which editing sites occur: Testable sites are sites whose median sequencing depth across samples is greater than 20 in mRNA regions (5' UTR, CDS and 3' UTR), which correspond to all sites in figure 3.4. The increasingly-edited sites indicate the sites with the increasing pattern in figure 3.4. (b) The cumulative proportion of the increasingly-edited sites according to the number of samples at which editing sites occur: only 14% of sites were found in 10 samples or less. In other words, 86% of the increasingly-edited sites were found in more than 10 samples.
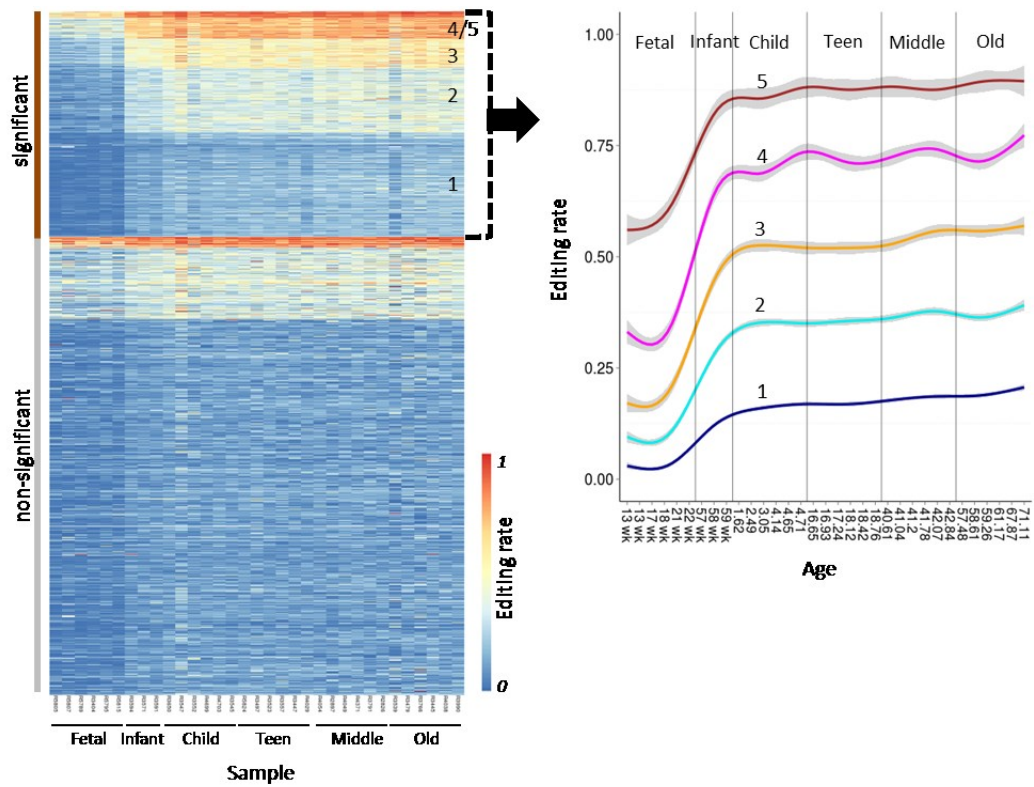
**Figure 3.6 Increasing editing patterns.** The *significant* sites determined by ANOVA followed by FDR-correction have an increasing editing pattern which is represented by five groups varying in their absolute editing rates but not in their slopes (noted by numbers) based on hierarchical clustering. Lines are generated by fitting generalized additive models (GAM) to data with integrated smoothness estimation. Shades indicate 95% confidence interval.
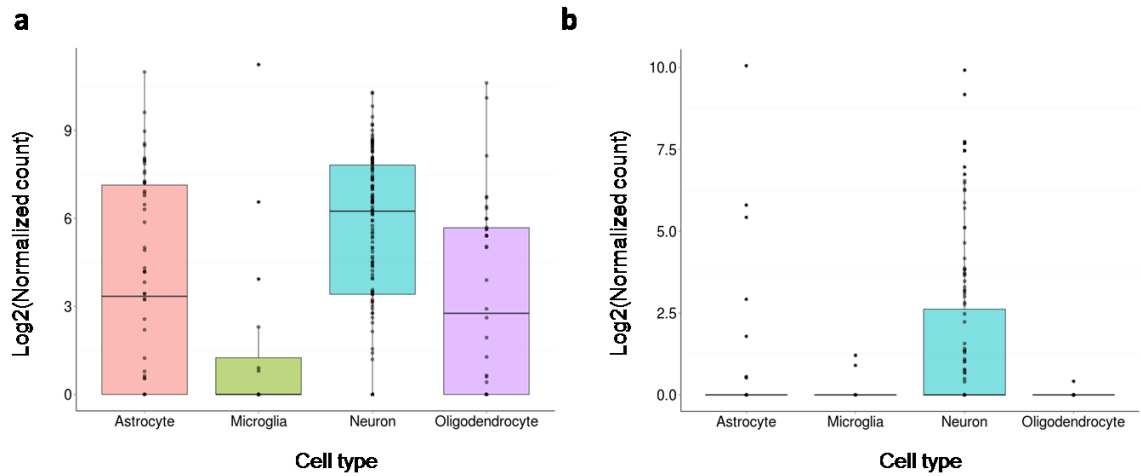
**Figure 3.7 Comparison of ADAR RNA level across cell types.** (A) ADAR1 (B) ADAR2. Normalized counts of sequencing reads from a previous single cell RNA-seq dataset (Darmanis et al. 2015) were used to compare RNA levels across cell types. Each dot represents a single cell. The number of cells are 62 (astrocyte), 16 (microglia), 131 (neuron), 38 (oligodendrocyte), respectively.
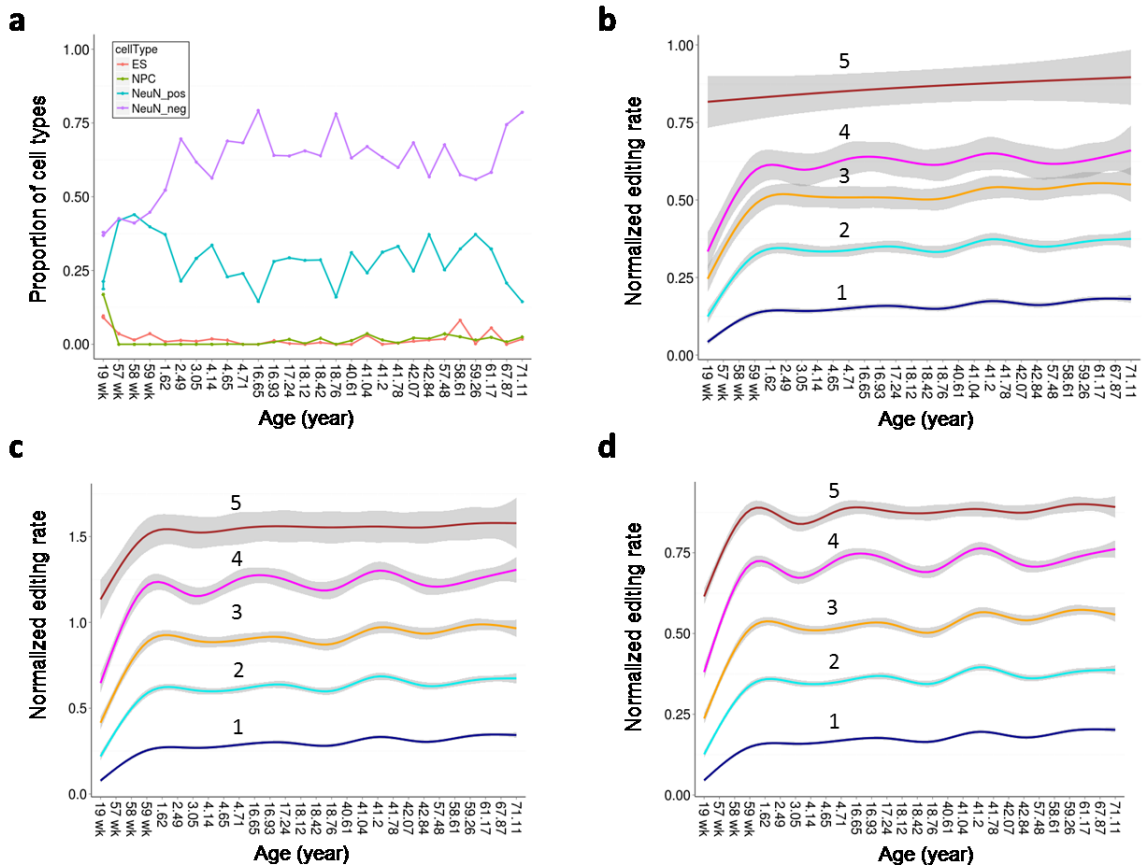
**Figure 3.8. Increasing pattern normalized by neuronal or glial proportions in human brain development.** (a) The proportions of four cell types including embryonic stem cells (ES), neural progenitor cells (NPC), neuronal cells (NeuN pos) and glial cells (NeuN Neg) were estimated from the same brain tissues in the 33 subject discovery set with the available genome-wide DNA methylation data using the previously-published algorithm (Jaffe et al. 2015). For fetal samples, we used two independent fetal tissues which have both DNA methylation and RNA-seq datasets. (b) Editing rates normalized by glial proportion (c) Editing rates normalized by the neuronal proportion (d) Editing rates normalized by the ratio of neuronal proportion to glial proportion. Normalization is simply done by dividing editing rates by the proportions. In b, c and d, lines indicate the five increasing clusters defined in figures 3.4 and 3.6 (The same colors and the same

numbers are used across the figures). Lines are generated by locally weighted scatterplot

smoothing (loess) regression with 95% confidence interval indicated by shades.
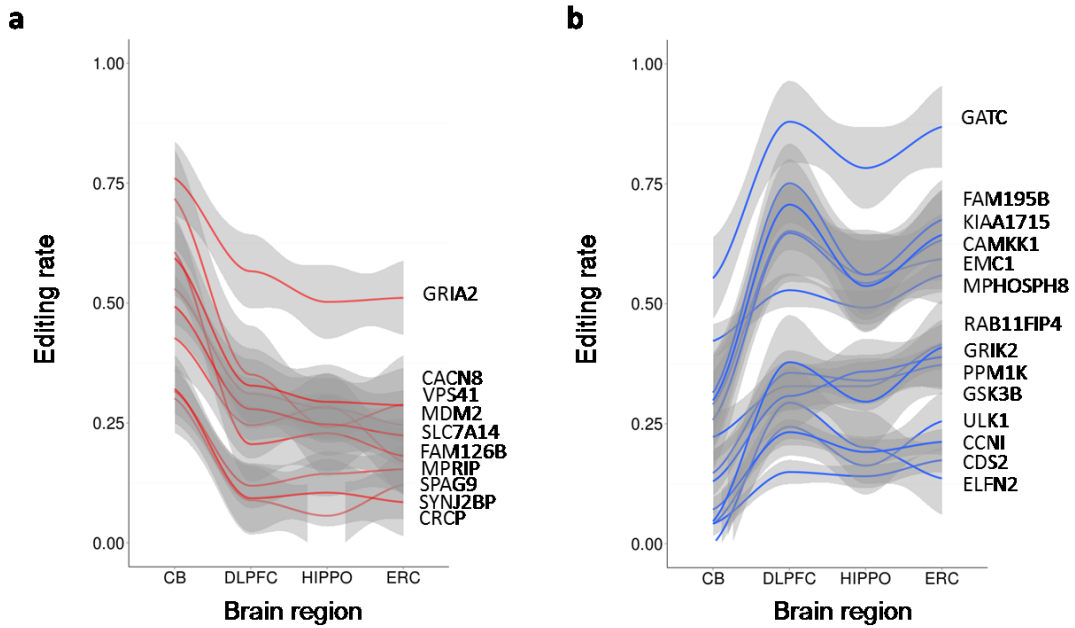
**Figure 3.9. Editing rate differences among brain regions.** A-to-I editing sites showing significant differences between cerebellum (CB) and other brain regions (DLPFC: dorsolateral prefrontal cortex, HIPPO: hippocampus, ERC: entorhinal cortex) are described. Each line corresponds to an editing site and the associated gene name is indicated. (a) Editing sites whose editing rates are higher in CB. (b) Editing sites with higher editing rates in neo cortical regions. Lines are generated by locally weighted scatterplot smoothing (loess) regression with 95% confidence interval indicated by shades.

| RNA ID | Region | RIN | Sex | Race | Age Group | Age (year) | Gestational week |
|---|---|---|---|---|---|---|---|
| R5805 | DLPFC | 9.4 | F | CAUC | Fetal | -0.52 | 13 |
| R5807 | DLPFC | 9.8 | M | CAUC | Fetal | -0.52 | 13 |
| R5789 | DLPFC | 8.9 | F | CAUC | Fetal | -0.44 | 17 |
| R3404 | DLPFC | 8.4 | M | CAUC | Fetal | -0.42 | 18 |
| R5795 | DLPFC | 9.6 | F | CAUC | Fetal | -0.36 | 21 |
| R5815 | DLPFC | 9.5 | M | CAUC | Fetal | -0.35 | 22 |
| R3594 | DLPFC | 8.5 | M | CAUC | Infant | 0.33 | 57 |
| R3571 | DLPFC | 8.8 | M | CAUC | Infant | 0.35 | 58 |
| R3591 | DLPFC | 8.8 | M | CAUC | Infant | 0.36 | 59 |
| R3650 | DLPFC | 8.2 | F | CAUC | Child | 1.62 | NA |
| R3547 | DLPFC | 8.1 | F | CAUC | Child | 2.49 | |
| R3552 | DLPFC | 7.1 | M | CAUC | Child | 3.05 | |
| R4699 | DLPFC | 8.7 | M | CAUC | Child | 4.14 | |
| R4703 | DLPFC | 7.9 | M | CAUC | Child | 4.65 | |
| R3545 | DLPFC | 7.9 | M | CAUC | Child | 4.71 | |
| R5824 | DLPFC | 8.1 | F | CAUC | Teen | 16.65 | |
| R3497 | DLPFC | 8.9 | M | CAUC | Teen | 16.93 | |
| R3523 | DLPFC | 9.7 | F | CAUC | Teen | 17.24 | |
| R3557 | DLPFC | 9 | M | CAUC | Teen | 18.12 | |
| R3447 | DLPFC | 8.3 | M | CAUC | Teen | 18.42 | |
| R4029 | DLPFC | 8.9 | F | CAUC | Teen | 18.76 | |
| R4054 | DLPFC | 8.6 | F | CAUC | Middle | 40.61 | |
| R2897 | DLPFC | 8.7 | M | CAUC | Middle | 41.04 | |
| R4049 | DLPFC | 8.6 | F | CAUC | Middle | 41.20 | |
| R4371 | DLPFC | 7.5 | M | CAUC | Middle | 41.78 | |
| R3791 | DLPFC | 8.5 | M | CAUC | Middle | 42.07 | |
| R2826 | DLPFC | 8.6 | M | CAUC | Middle | 42.84 | |
| R3539 | DLPFC | 7.7 | F | CAUC | Old | 57.48 | |
| R3479 | DLPFC | 8.1 | M | CAUC | Old | 58.61 | |
| R3766 | DLPFC | 8.4 | F | CAUC | Old | 59.26 | |
| R3445 | DLPFC | 8.2 | M | CAUC | Old | 61.17 | |
| R4038 | DLPFC | 8.3 | M | CAUC | Old | 67.87 | |
| R3990 | DLPFC | 8.5 | F | CAUC | Old | 71.11 | |

**Table 3.1 Demographic details of the human brain tissues.** RNA ID: sample identifier; Region: brain region, DLPFC (dorsolateral prefrontal cortex); RIN: RNA integrity number indicating RNA quality; Sex: M (male) or F (female); RACE: CAUC (Caucasian); Age Group:

six categories according to age, fetal (before birth), infant (birth to 12 months), child (1≤ age <10), teen (10≤ age <20), middle (20≤ age <50) and old life (age ≥ 50); Age: years old; Gestational week: age of fetal tissues represented by conventional week of pregnancy.

| RNA ID | Total number of reads | Number of RNA editing sties |
|---|---|---|
| R5805 | 80,337,568 | 25,567 |
| R5807 | 124,013,892 | 24,408 |
| R5789 | 115,578,312 | 49,525 |
| R3404 | 101,440,880 | 69,143 |
| R5795 | 135,185,718 | 72,255 |
| R5815 | 137,933,864 | 24,355 |
| R3594 | 115,856,986 | 93,553 |
| R3571 | 136,135,132 | 93,483 |
| R3591 | 155,835,950 | 86,731 |
| R3650 | 167,953,146 | 103,649 |
| R3547 | 93,909,690 | 61,348 |
| R3552 | 119,030,262 | 57,923 |
| R4699 | 95,116,394 | 69,460 |
| R4703 | 197,052,880 | 104,694 |
| R3545 | 122,085,850 | 76,074 |
| R5824 | 184,243,872 | 102,841 |
| R3497 | 107,834,650 | 65,739 |
| R3523 | 176,310,414 | 89,674 |
| R3557 | 100,916,134 | 62,852 |
| R3447 | 96,398,214 | 67,615 |
| R4029 | 186,789,930 | 91,690 |
| R4054 | 91,322,460 | 64,596 |
| R2897 | 93,746,042 | 47,099 |
| R4049 | 97,608,938 | 69,177 |
| R4371 | 105,055,680 | 70,288 |
| R3791 | 82,526,684 | 49,291 |
| R2826 | 118,781,082 | 73,717 |
| R3539 | 114,574,246 | 53,600 |
| R3479 | 87,319,298 | 43,724 |
| R3766 | 82,157,476 | 40,286 |
| R3445 | 93,543,730 | 58,576 |
| R4038 | 98,241,300 | 65,126 |
| R3990 | 149,259,434 | 65,025 |
| Average | 120,124,124 | 66,457 |

**Table 3.2 Number of RNA editing sites in individual samples with total number of sequencing reads**

| Type | Number | Proportion (%) |
|---|---|---|
| A->G | 134,914 | 81.0 |
| T->C | 17,133 | 10.3 |
| G->A | 3,192 | 1.9 |
| C->T | 1,803 | 1.1 |
| C->A | 1,219 | 0.7 |
| T->A | 991 | 0.6 |
| G->T | 824 | 0.5 |
| A->T | 647 | 0.4 |
| G->C | 528 | 0.3 |
| A->C | 465 | 0.3 |
| T->G | 457 | 0.3 |
| C->G | 437 | 0.3 |
| Uncertain | 3,915 | 2.3 |
| Total | 166,525 | 100.0 |

**Table 3.3 Number of RNA editing sites in genic regions according to types**

| Annotation | Number | Proportion (%) |
|---|---|---|
| Total | 134,914 | 100.0 |
| Intronic | 109,411 | 81.1 |
| 3' UTR | 13,940 | 10.3 |
| CDS | 282 | 0.2 |
| 5' UTR | 247 | 0.2 |
| ncRNA | 11,034 | 8.2 |

**Table 3.4 Number of A-to-I editing sites according to gene regions**

**Table 3.5 A-to-I editing sites with increasing pattern** See Appendix 1: Chromosome and Coordinate: location of A-to-I editing sites in UCSC human genome hg19; Strand, Gene and Gene region: gene annotation based on RefSeq; DB: whether a A-to-I editing site is found in RADAR(Ramaswami & Li 2014), a public A-to-I editing database; Prenatal mean, Post-infant mean, Mean difference (diff.), Cluster in Editing rate: summary of editing rates in 33 samples. Cluster indicates an index of cluster described in Figures 3.4 and 3.6.

| Site | Gene | Gene region | Neu. | Ast. | MO | NFO | OPC | MG | End. |
|------|------|-------------|------|------|-----|-----|-----|-----|------|
| chr1:172092348 | COPA | CDS | 0.05 | 0.22 | 0.07 | 0.28 | **0.49** | **0.55** | 0.48 |
| chr1:172074326 | NCSTN | CDS | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | **0.02** | **0.01** |
| chr19:40327452 | SORBS1 | CDS | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | NA | 0.00 |
| chr19:40327453 | SORBS1 | CDS | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | NA | 0.00 |
| chr7:130759165 | TACC2 | CDS | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| chr9:4456006 | GRIA4 | CDS | **0.65** | 0.33 | 0.01 | 0.09 | **0.16** | NA | 0.25 |
| chr14:50919694 | OSGEP | CDS | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| chr12:46700334 | NOVA1 | CDS | **0.08** | 0.00 | 0.00 | 0.00 | **0.02** | 0.00 | 0.00 |
| chr9:57144307 | NEIL1 | CDS | **0.03** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| chr11:102479070 | GPATCH8 | CDS | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| chr8:83661176 | GIPC1 | CDS | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| chr8:83661088 | GIPC1 | CDS | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| chr16:87940543 | GRIK1 | CDS | **0.53** | 0.00 | 0.00 | **0.12** | 0.06 | NA | NA |
| chr3:32561485 | MFN1 | CDS | **0.33** | **0.01** | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 |
| chr5:93189584 | CCNI | CDS | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 | **0.06** | **0.03** |
| chr3:80706908 | GRIA2 | CDS | **0.28** | **0.08** | 0.05 | 0.06 | 0.07 | 0.00 | 0.00 |
| chr3:80692286 | GRIA2 | CDS | **0.51** | 0.07 | 0.00 | 0.05 | **0.15** | 0.00 | 0.50 |
| chr17:45662949 | TMEM63B | CDS | **0.45** | 0.01 | 0.00 | 0.02 | **0.05** | 0.00 | 0.01 |
| chr10:49272776 | GRIK2 | CDS | **0.73** | 0.05 | 0.04 | 0.02 | **0.25** | 0.00 | 0.00 |
| chr10:49244347 | GRIK2 | CDS | **0.10** | 0.00 | 0.00 | 0.00 | 0.00 | NA | 0.00 |
| chr10:49244330 | GRIK2 | CDS | **0.88** | 0.05 | 0.00 | 0.04 | **0.19** | NA | 0.00 |
| chrX:150648527 | TRO | CDS | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | NA | 0.00 |
| chrX:41654252 | GRIA3 | CDS | **0.54** | 0.06 | 1.00 | 0.15 | **0.26** | NA | NA |
| chr9:57144308 | NEIL1 | CDS | **0.12** | 0.00 | 0.00 | **0.02** | 0.00 | 0.00 | 0.00 |
| chr3:80706912 | GRIA2 | CDS | **0.97** | 0.67 | 0.67 | 0.84 | **0.99** | 1.00 | 1.00 |
| chr14:75719719 | COG3 | CDS | 0.27 | 0.76 | 0.57 | 0.75 | **0.83** | 0.91 | **0.97** |

**Table 3.6 CDS-residing A-to-I editing sites in different mouse brain cell types.** Site: location of A-to-I editing sites in UCSC mouse genome mm10. The A-to-I editing sites whose editing rates are greater than 0 in at least one cell type are marked by red; Gene and Gene region: gene annotation of A-to-I editing sites based on RefSeq; Columns 4 to 10: editing rates in Neuron (Neu.), Astrocyte (Ast.), Myelinating Oligodendrocyte (MO), Newly Formed Oligodendrocyte (NFO), Oligodendrocyte Precursor Cell (OPC), Microglia (MG), Endothelial cells (End.). Gray

indicates detected A-to-I editing rate. Yellow and green describe the highest and second-highest editing rate, respectively.

| RNA ID | Region | RIN | Sex | Race | Age Group | Age | Gestational week | RNA-seq |
|---|---|---|---|---|---|---|---|---|
| R3715 | DLPFC | 7.3 | F | CAUC | Fetal | -0.40 | 19 | poly-A |
| R3390 | DLPFC | 9.9 | M | CAUC | Fetal | -0.40 | 19 | poly-A |
| R2855 | DLPFC | 8.5 | M | CAUC | Old | 56.09 | NA | Ribo-zero |
| R4936 | Hippocampus | 8.3 | | | | | | Ribo-zero |
| R10023 | CB | 7.9 | | | | | | Ribo-zero |
| R10065 | ERC | 7.4 | | | | | | Ribo-zero |
| R3395 | DLPFC | 8.9 | M | CAUC | Old | 57.63 | | Ribo-zero |
| R5455 | Hippocampus | 8.6 | | | | | | Ribo-zero |
| R10012 | CB | 7.4 | | | | | | Ribo-zero |
| R10055 | ERC | 6 | | | | | | Ribo-zero |
| R3766 | DLPFC | 8.4 | F | CAUC | Old | 59.26 | | Ribo-zero |
| R4869 | Hippocampus | 8.1 | | | | | | Ribo-zero |
| R10029 | CB | 7 | | | | | | Ribo-zero |
| R10070 | ERC | 6 | | | | | | Ribo-zero |
| R3052 | DLPFC | 7.3 | F | CAUC | Old | 75.57 | | Ribo-zero |
| R4757 | Hippocampus | 7.7 | | | | | | Ribo-zero |
| R10022 | CB | 7.3 | | | | | | Ribo-zero |
| R10092 | ERC | 6 | | | | | | Ribo-zero |
| R2839 | DLPFC | 8.3 | M | CAUC | Old | 77.99 | | Ribo-zero |
| R4992 | Hippocampus | 8.2 | | | | | | Ribo-zero |
| R10003 | CB | 7.6 | | | | | | Ribo-zero |
| R10052 | ERC | 7 | | | | | | Ribo-zero |

**Table 3.7 Additional brain samples.** RNA ID: sample identifier; Region: brain region, DLPFC (dorsolateral prefrontal cortex), CB (cerebellum), HIPPO (hippocampus), ERC (entorhinal cortex); RIN: RNA integrity number indicating RNA quality; Sex: M (male) or F (female); RACE: CAUC (Caucasian); Age Group: fetal (before birth) and old life (age ≥ 40); Age: years old; Gestational week: age of fetal tissues represented by conventional week of pregnancy. RNA-seq: RNA-seq library type, poly-A (poly-A enrichment), Ribo-zero (ribosomal RNA depletion).

| Site | Gene | Gene region | Mean editing rate | | | |
|------|------|-------------|-----|-------|-----|-------|
| | | | CB | DLPFC | ERC | HIPPO |
| chr1:19544002 | EMC1 | 3' UTR | 0.29 | 0.65 | 0.59 | 0.56 |
| chr12:69237056 | MDM2 | 3' UTR | 0.53 | 0.31 | 0.25 | 0.28 |
| chr12:120899075 | GATC | 3' UTR | 0.55 | 0.88 | 0.87 | 0.78 |
| chr12:132407137 | ULK1 | 3' UTR | 0.04 | 0.24 | 0.26 | 0.16 |
| chr13:20247023 | MPHOSPH8 | 3' UTR | 0.42 | 0.53 | 0.56 | 0.49 |
| chr14:70834087 | SYNJ2BP | 3' UTR | 0.3 | 0.09 | 0.12 | 0.06 |
| chr17:3763779 | CAMKK1 | 3' UTR | 0.26 | 0.65 | 0.63 | 0.54 |
| chr17:17092596 | MPRIP | 3' UTR | 0.43 | 0.25 | 0.17 | 0.28 |
| chr17:29861403 | RAB11FIP4 | 3' UTR | 0.22 | 0.33 | 0.42 | 0.33 |
| chr17:49042206 | SPAG9 | 3' UTR | 0.32 | 0.12 | 0.15 | 0.14 |
| chr17:79780692 | FAM195B | 3' UTR | 0.31 | 0.75 | 0.67 | 0.56 |
| chr19:54487949 | CACNG8 | 3' UTR | 0.72 | 0.35 | 0.29 | 0.24 |
| chr2:176791181 | KIAA1715 | 3' UTR | 0.3 | 0.71 | 0.64 | 0.54 |
| chr2:201842411 | FAM126B | 3' UTR | 0.61 | 0.21 | 0.18 | 0.23 |
| chr20:5175805 | CDS2 | 3' UTR | 0.04 | 0.15 | 0.17 | 0.14 |
| chr22:37765609 | ELFN2 | 3' UTR | 0 | 0.29 | 0.14 | 0.2 |
| chr3:119545199 | GSK3B | 3' UTR | 0.15 | 0.36 | 0.37 | 0.34 |
| chr3:170181366 | SLC7A14 | 3' UTR | 0.49 | 0.28 | 0.22 | 0.25 |
| chr4:77979680 | CCNI | exonic | 0.07 | 0.23 | 0.21 | 0.19 |
| chr4:89180382 | PPM1K | 3' UTR | 0.13 | 0.31 | 0.39 | 0.36 |
| chr4:158281294 | GRIA2 | exonic | 0.76 | 0.57 | 0.51 | 0.5 |
| chr6:102372572 | GRIK2 | exonic | 0.05 | 0.38 | 0.41 | 0.3 |
| chr7:38764438 | VPS41 | 3' UTR | 0.59 | 0.33 | 0.29 | 0.29 |
| chr7:65619303 | CRCP | 3' UTR | 0.32 | 0.09 | 0.09 | 0.1 |

**Table 3.8 A-to-I editing sites showing significant differences in editing rates among brain regions.** Mean editing rates in 5 individuals. Site: location of A-to-I editing sites in UCSC human genome hg19; Gene and gene region: gene annotation of A-to-I editing sites based on RefSeq; Editing rates in DLPFC (dorsolateral prefrontal cortex), CB (cerebellum), HIPPO (hippocampus), ERC (entorhinal cortex).

# Chapter 4

# Regulation of developmental A-to-I editing pattern

In Chapter 3, I reported "the developmental A-to-I editing patterns", comprising three distinct patterns across cortical development: "stable high", "stable low" and "increasing". In this chapter, I will demonstrate that these patterns are explained by the secondary structures of RNA and the temporal pattern of expression of ADAR enzymes and propose them as possible cis- and trans-regulatory mechanisms of the developmental A-to-I editing patterns, respectively.

## 4.1. Representative editing sites for developmental A-to-I editing pattern

To compare three distinct A-to-I editing patterns in brain development, the sites that explicitly represent each pattern were selected. Here, I considered that ADAR enzymes can edit multiple adenosines in a target transcript. This characteristic of editing can blur differences among the three editing patterns especially when sites from different patterns are close to each other in their genomic coordinates. For example, there are some cases of increasingly-edited sites that are within a few bases from stable high-edited sites. This is likely due to incomplete efficiency of editing at the sites around a stable high-edited site in fetal samples, which generates seemingly increasingly-edited sites but truly passive byproducts around stable high editing. Therefore, I selected the representative sites for each pattern, which are spatially isolated from other patterns. The following three parameters are used to formulate conditions to select representative sites for each pattern: i) magnitude of increasing of editing rates from fetal to adult samples (parameter 1), ii) maximum editing rates (parameter 3) at the neighboring editing sites within a given distance (parameter 2). Specific conditions are as follows. From now on, the following labels were used for groups of selected sites representing each pattern: *"Group I. Low"*, *"Group II. Increasing"*, and *"Group III. High"*.

**Editing sites in "Group II. Increasing": Representative sites for increasing pattern**

Sites representing the increasing pattern, which comprises *'Group II. Increasing'*, were chosen from the sites in figure 3.4 such that they have clear increasing pattern from fetal to adult ages and no stable high-edited sites are near increasingly-edited sites. These conditions were formalized by two quantitative requirements using the above three parameters: i) parameter 1

should be greater than a certain threshold, ii) parameter 3 should be less than a certain threshold for a given neighbor region specified by parameter 2. Currently we used 0.3, 800 bp, and 0.4 for the values of the parameters 1, 2 and 3, respectively. These values were determined systematically as follows: for parameter 2, I chose 800 bp to have an adequate number of sites for statistical tests (Figure 4.1a), considering that the number of selected sites decreases as parameter 2 increases. I assumed that 800 bp is enough to check effects of neighboring sites. Parameters 1 and 3 were chosen as conservatively as possible but keeping enough number of sites (Figure 4.1b). For parameter 1, 0.4 is too strict to identify enough sites, resulting in 0.3 being the choice. As for parameter 3, 0.4 is selected to maximize the number of selected sites. Using these conditions, a total of 58 sites were chosen for *'Group II: Increasing'* (denoted by green in Figure 4.2).

For all sites in group II except for 5 sites around DNA repeat regions, genotypes were checked in 12 selected samples covering fetal, infant and post-infant age groups in order to confirm that editing rates of about 50% after post-infant age are not affected by genomic variants. Specifically, a targeted DNA sequencing with extremely-high sequencing depth (mean: 5417 reads, minimum: 113 reads) was performed for accurate genotyping. Table 4.1 showed that in three samples, for example, all the amplified sites are homozygous with reference alleles, confirming that sites in 'Group II. Increasing' are not genomic variants.

**Editing sites in "Group I. Low": Representative sites for stable low patterns.**

For the stable low pattern, the sites were selected such that none of the stable high-edited and increasingly-edited sites are close to the stable low-edited sites. This was achieved by the conditions require that i) parameter 1 should be less than a certain threshold (0.1), ii) parameter 3 should be less than a certain threshold (0.3) for a given parameter 2 (800), iii) sites should have

average editing rates greater than 0.1 and less than 0.3. As a result, *'Group I: Low'* consists of 65 selected sites for the stable low pattern (denoted by blue in Figure 4.2).

**Editing sites in "Group III. High": Representative sites for stable high patterns.**

Here, the required conditions are i) parameter 1 should be less than a certain threshold (0.1), ii) sites should have average editing rates greater than 0.7. The 40 sites were selected for the stable high editing pattern and defined '*Group III. High*' (denoted by red in Figure 4.2).

To summarize, the selected sites comprise three groups each representing the three patterns respectively as follows: *'Group I: Low'* consisting of 65 selected sites for the stable low pattern, *'Group II: Increasing'* consisting of 58 selected sites for the increasing pattern and *'Group III: High'* consisting of 40 sites for the stable high pattern (Fig. 4.2 and Table 4.2).

## 4.2. *trans*-regulation

The expression levels of the known A-to-I editing enzymes across brain development were investigated as a potential *trans*-regulatory mechanism. There are two ADAR enzymes (ADAR1 and ADAR2) expressed in brain and known to be responsible for A-to-I RNA editing in humans. The expression level of ADARs was measured by RNA-seq with an RPKM (Reads Per Kilobase per Million) unit that is typically used for quantifying gene expression levels with RNA-seq. The figure 4.3 showed that changes of expression of ADARs are correlated selectively with the developmentally-increasing A-to-I editing pattern. Specifically, the mean Spearman

correlation coefficients between ADAR mRNA levels and editing rates of the "Group II. Increasing" are 0.56 and 0.51 for ADAR1 and ADAR2 respectively.

## 4.3. *cis*-regulation

For a potential *cis*-regulatory mechanism, the neighbor sequence preferences were first examined to compare the sensitivities to ADAR 1/2 among the three groups. There is no statistically significant difference in the frequency of 5' and 3' sequence preferences among the three groups, as all three groups share previously known 5' and 3' sequence preferences (Eggington et al. 2011) as T>A>C>G and G>C>A≈T respectively (Figure 4.4).

Next, RNA secondary structure was considered as another *cis*-regulatory mechanism. Although ADARs are known to recognize the double-stranded structure of target RNAs, the subtle differences of RNA secondary-structures around editing sites may affect the efficiency of A-to-I editing. I developed quantitative measures of the degree and the distance of double-stranded structure around A-to-I editing sites to compare the sensitivities of the three groups to ADAR 1/2 (see 4.4 methods). While the distance measure does not reveal significant differences (Figure 4.5), statistically-significant mean differences (ANOVA and post-hoc two-sample t-test) were observed in the degree of secondary structure among the three groups (Figure 4.6), where the group of increasingly-edited sites tend to have an intermediate degree of double stranded structure between the high and low editing sites. However, the high-edited sites have a broad range of degrees of double-stranded structure, which suggests additional or alternative regulatory mechanisms for these sites. In fact, evaluation of prior RNA-seq data from ADAR 1/2 knockdown by siRNA in B-cells (I. X. Wang et al. 2013) showed that the high-edited sites are not

affected by ADAR 1/2 knockdown as much as the increasingly-edited sites (Figure 4.7 and Table 4.3). Here, the effect of ADAR 1/2 knockdown for a given site is measured by the percent decrease of editing rate in ADAR siRNA-treated experiments compared to controls. These results suggest that the increasingly-edited sites are particularly dependent on ADAR 1/2 mechanisms.

It is interesting to see these results from the perspective of a previously-proposed model of RNA editing. Previous studies (Daniel et al. 2012; Daniel et al. 2014) proposed that the long hairpin structure of RNA induces A-to-I editing by promoting the recruitment of ADAR enzymes to the transcript. This model, however, does not fully explain how different RNA editing rates are regulated. For example, A-to-I editing sites in Gabra3 and NEIL1 that were considered in a previous study (Daniel et al. 2014) as examples for the proposed model show two different editing patterns (increasing and stable-high, respectively) in brain development. The defined quantitative representation of double-stranded structure of RNAs may improve the model by adding another feature of RNA hairpin structure that has not been noted before.

## 4.4. Methods

**RNA-seq processing**

For RNA-seq data obtained from previous studies, STAR (Dobin et al. 2013) is used for alignments. To estimate ADAR expression levels, RPKM (Reads Per Kilobase per Million) were calculated using HTSeq (Anders et al. 2014).

**Targeted DNA sequencing**

This was done with the Fluidigm Access Array and the Illumina MiSeq through which PCR amplicons around target sites were generated on microfluidic chips and then sequenced. Only uniquely-mapped reads determined by BWA were used to determine genotypes. If a proportion of reference allele-supporting reads are greater than 90% of total sequencing reads at a site, its genotype is declared as homozygous with the reference allele.

**RNA secondary structure analyses**

The degree of double-stranded structure around and its distance from a given A-to-I editing site was calculated from computationally-predicted RNA-structure. First, RNA structure was predicted by RNAfold in Vienna RNA package 2.0 (Lorenz et al. 2011) using a pre-mRNA sequence spanning 800 bp upstream and downstream from a given editing site. Second, the number of nucleotides in a double-stranded configuration was counted within a flanking 100 bp region for every position on a pre-mRNA sequence. Finally, after the site associated with the maximum value within 500 bp upstream and downstream from an editing site was identified, the associated value was declared as the degree of double-stranded structure and the distance from the site to an editing site was called as the distance between a double-stranded structure and an editing site. Figure 4.8 summarizes the method.
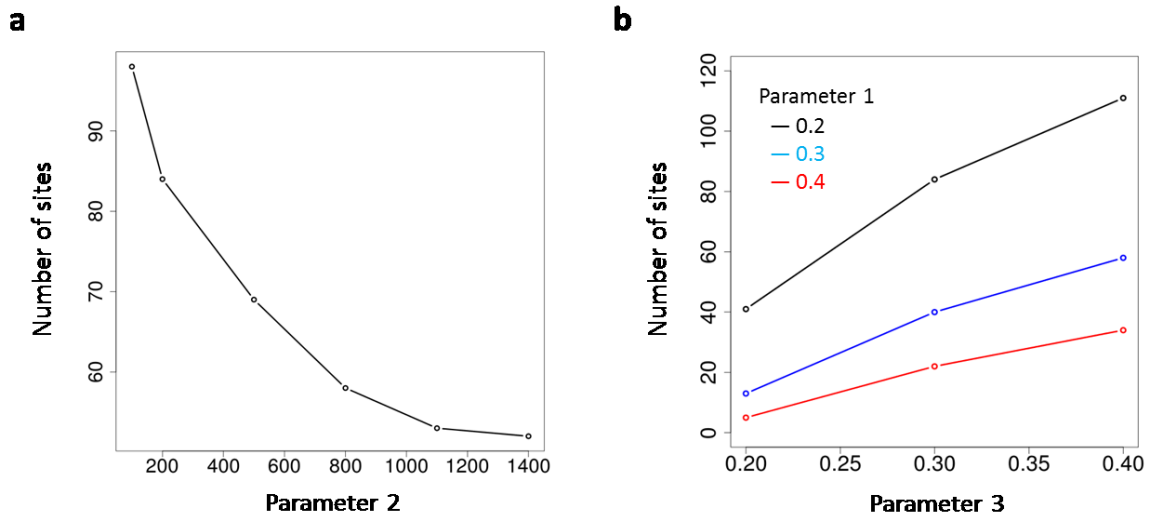
**Figure 4.1 Criteria for selecting the representative sites for the increasing A-to-I editing patterns.** (a) As parameter 2 increases, the number of selected sites is decreasing for given parameters 1 and 3. Here, parameter 1 and 3 are 0.3 and 0.4, respectively. (b) For parameter 1 and 3, several values were tried with parameter 2 fixed at 800 bp.
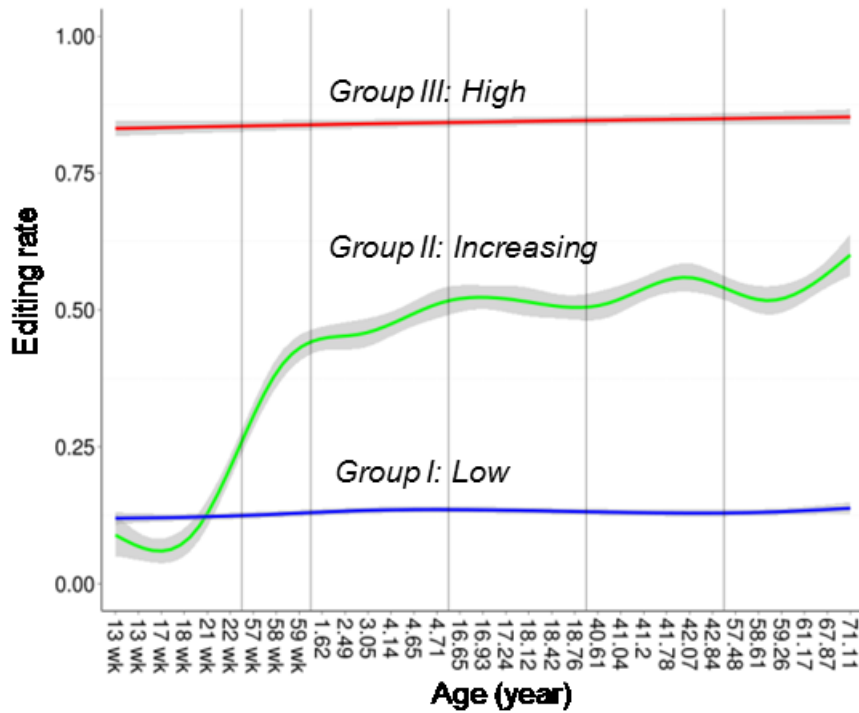
**Figure 4.2 Representative sites for developmental A-to-I editing patterns.** Regression lines generated with the selected sites explicitly represent each pattern. Specifically, lines are generated by locally weighted scatterplot smoothing (LOESS) regression with shades indicating 95% confidence interval.
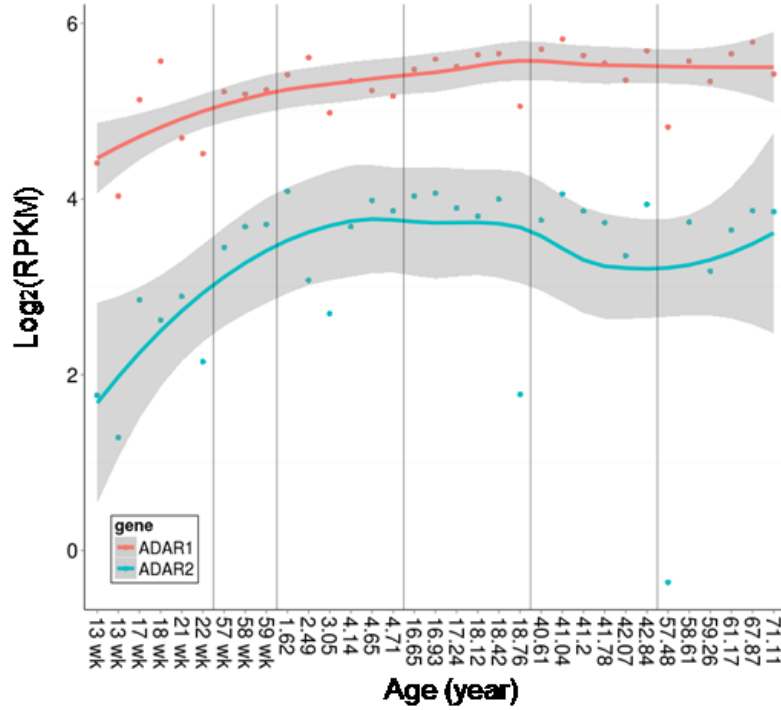
**Figure 4.3 ADAR expression across human brain development.** mRNA expression levels of ADAR enzymes, ADAR1 (red) and ADAR2 (blue) across development. Each dot represents an individual brain.
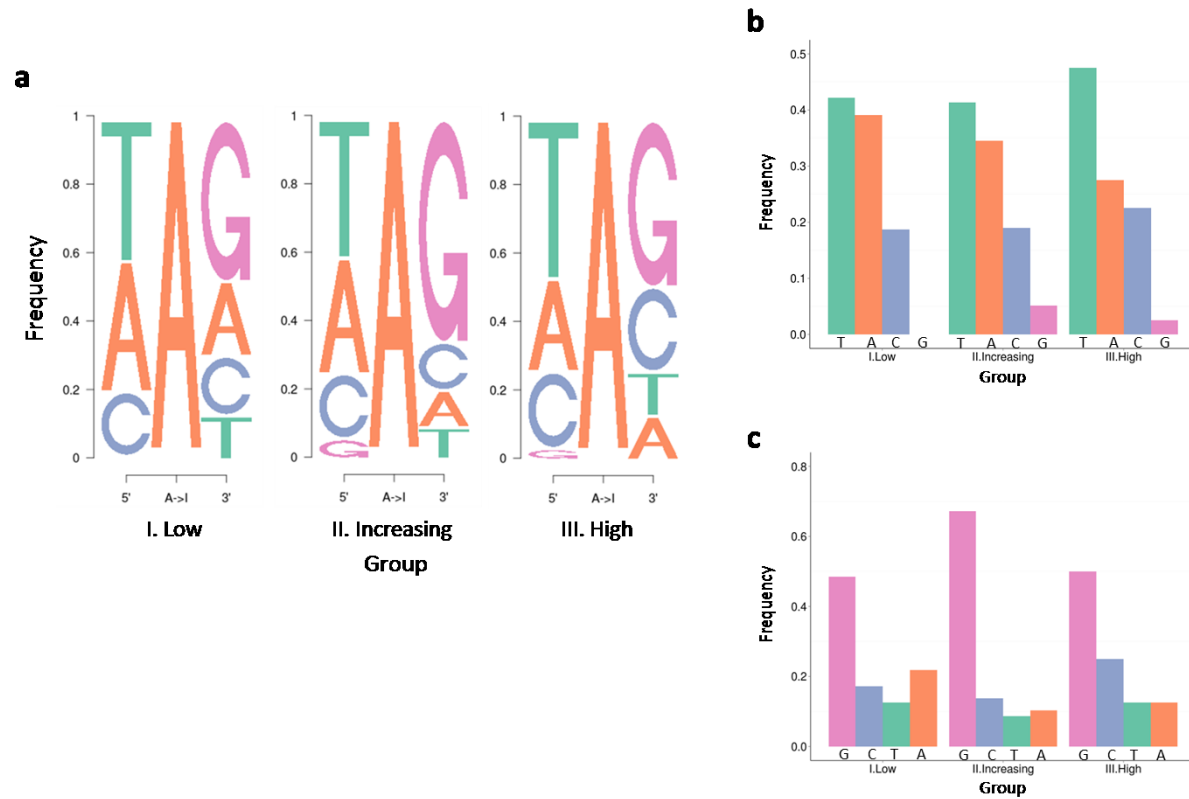
**Figure 4.4 Sequence motif around A-to-I editing sites.** (a) 5' and 3' sequence features flanking the sites in the three groups of selected sites, (b and c) Details of sequence frequency at 5' and 3' immediate neighbor respectively. Chi-Square test was performed to determine the significance of motif differences among three groups: p-values are 0.59 and 0.29 for 5' and 3' immediate neighbor, respectively.
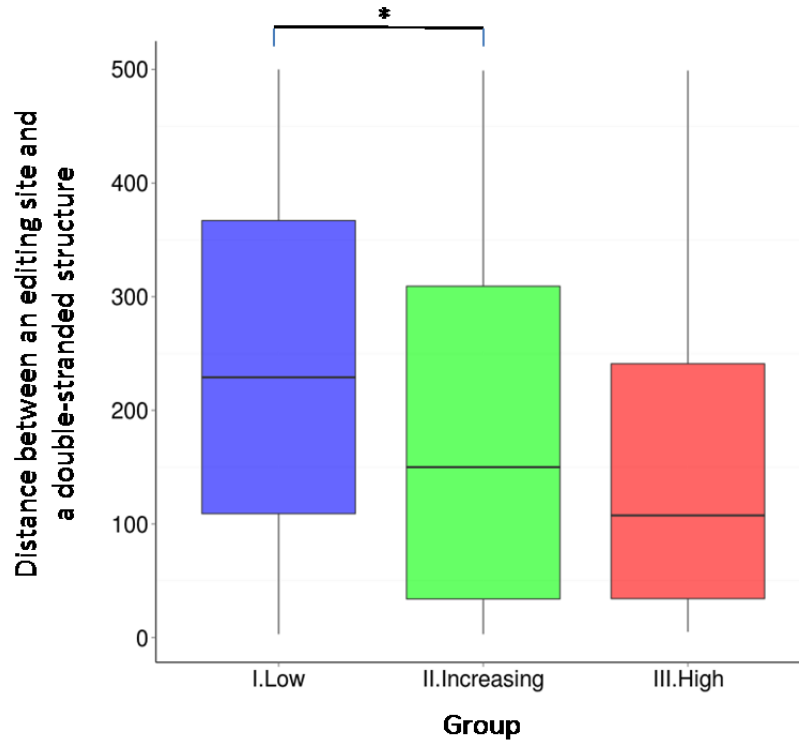
**Figure 4.5 Distance between an editing site and a double-stranded structure.** The groups were compared in terms of the distance between an editing site and a double-stranded structure, which is defined by the distance from an editing site to the site associated with the maximum number of local base-pairings ('the double-stranded degree'). * indicates *p-value* $\leq 0.05$ in two-sample t-tests.
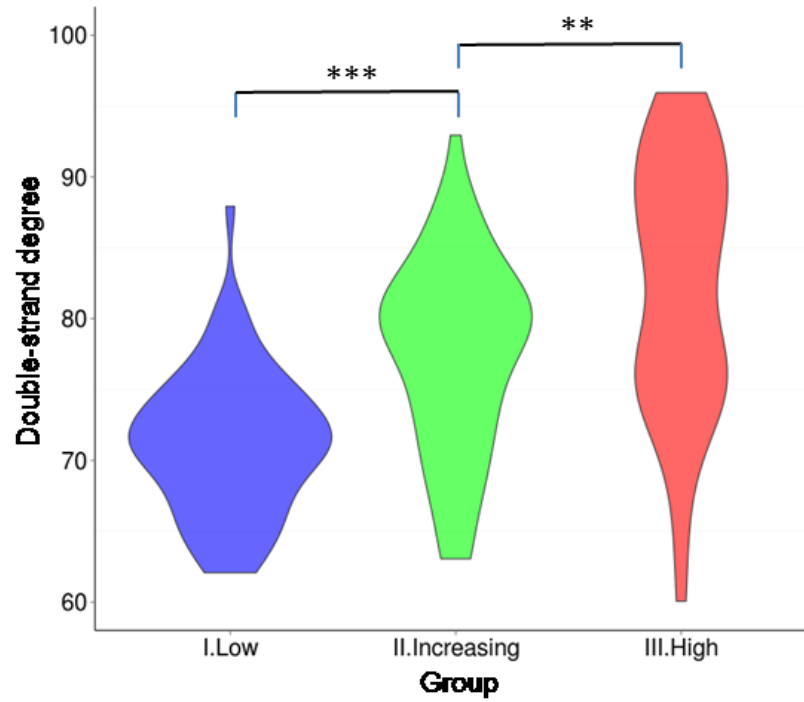
**Figure 4.6 Distribution of the degree of double-stranded structures (double-stranded degree) among the three groups of selected sites.** The double-stranded degree represents the maximum number of local base-pairings around editing sites. *** and ** indicate *p-value* ≤0.001 and *p-value* ≤0.01 in two-sample t-tests, respectively.
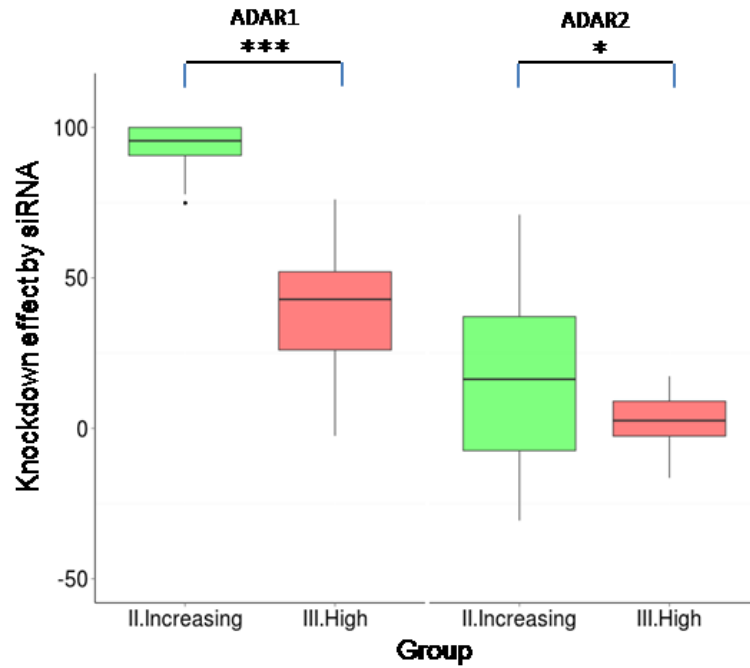
**Figure 4.7 Effect of ADAR enzymes knockdown by siRNA in B-cells.** The knockdown effect is defined by the percent decrease of editing rate in ADAR siRNA-treatments compared to control experiments. *** and * indicate *p-value* ≤0.001 and *p-value* ≤0.05 in two-sample t-tests, respectively.
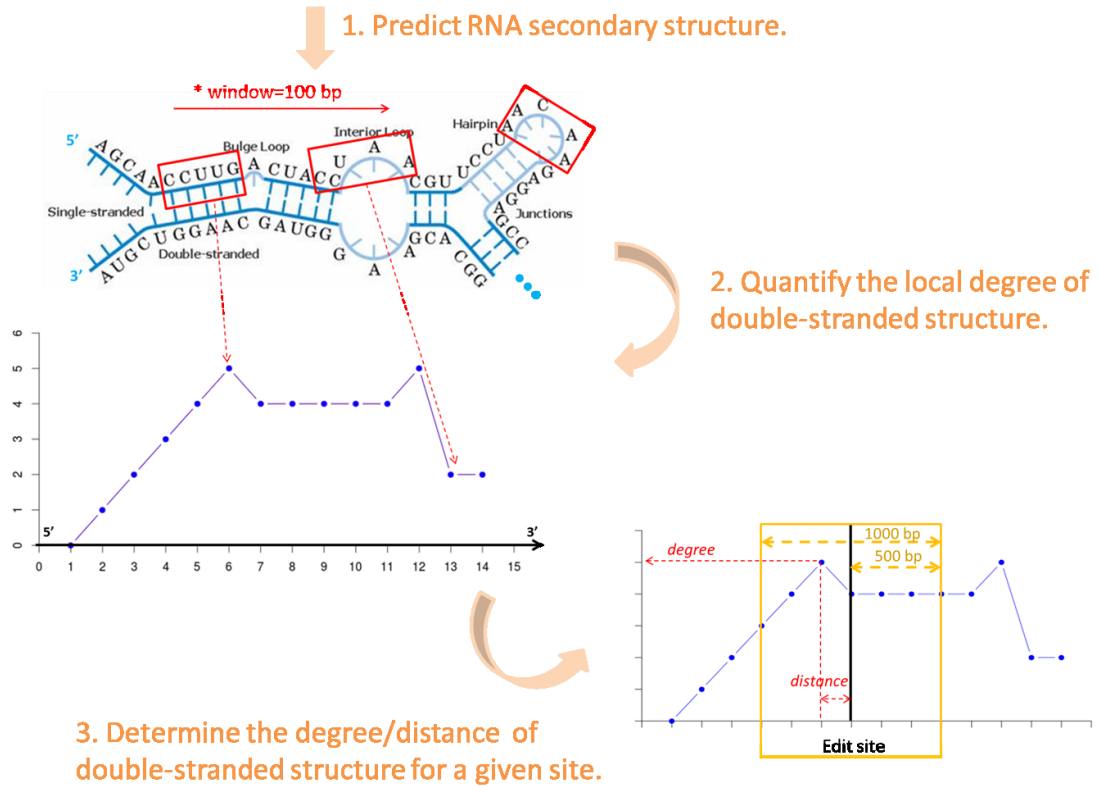
**Figure 4.8 Quantification of degree and distance of double-stranded structure around RNA-editing sites.** The degree of double-stranded structure for a given A-to-I editing site was calculated from computationally-predicted RNA-structure. First, RNA structure was predicted by *RNAfold* (rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) using a pre-mRNA sequence spanning 800 bp upstream and downstream from a given editing site. Second, the number of nucleotides in a double-stranded configuration was counted within a flanking 100 bp region for every position on a pre-mRNA sequence (moving-window search). In the figure, this is depicted by the rolling window of 5 nucleotides for simplicity. Finally, after the site associated with the maximum value within 500 bp upstream and downstream from an editing site was identified, the associated value was taken as the degree of double-stranded structure while the distance from the site to an editing site was called as the distance between a double-stranded structure and an editing site.

| Site | Ref. | D0199 (paired to R3404 in fetal) | | | | D1390 (paired to R3594 in infant) | | | | D1408 (paired to R5824 in teen) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Total read count | Ref. read count | Ref. read rate | Genotype | Total read count | Ref. read count | Ref. read rate | Genotype | Total read count | Ref. read count | Ref. read rate | Genotype |
| chr1:40147856 | T | 7707 | 7685 | 0.997 | TT | 4756 | 4746 | 0.998 | TT | 3546 | 3539 | 0.998 | TT |
| chr1:53289852 | A | 7894 | 7850 | 0.994 | AA | 7884 | 7854 | 0.996 | AA | 7871 | 7838 | 0.996 | AA |
| chr1:53291420 | A | 7994 | 7963 | 0.996 | AA | 7478 | 7452 | 0.997 | AA | 7995 | 7958 | 0.995 | AA |
| chr1:67874689 | T | 7994 | 7976 | 0.998 | TT | 7994 | 7980 | 0.998 | TT | 7987 | 7956 | 0.996 | TT |
| chr1:67874696 | T | 7992 | 7965 | 0.997 | TT | 7996 | 7967 | 0.996 | TT | 7985 | 7959 | 0.997 | TT |
| chr1:109748657 | A | 7787 | 7745 | 0.995 | AA | 7785 | 7739 | 0.994 | AA | 7804 | 7768 | 0.995 | AA |
| chr10:15120275 | T | 7975 | 7969 | 0.999 | TT | 6898 | 6895 | 1.000 | TT | 4984 | 4979 | 0.999 | TT |
| chr10:15120306 | T | 7980 | 7957 | 0.997 | TT | 6952 | 6945 | 0.999 | TT | 5013 | 5008 | 0.999 | TT |
| chr10:102777342 | T | 7994 | 7976 | 0.998 | TT | 7988 | 7971 | 0.998 | TT | 7884 | 7865 | 0.998 | TT |
| chr12:107280405 | A | 0 | 0 | NA | NA | 0 | 0 | NA | NA | 0 | 0 | NA | NA |
| chr14:26917530 | T | 6618 | 6605 | 0.998 | TT | 3114 | 3109 | 0.998 | TT | 2393 | 2389 | 0.998 | TT |
| chr14:31916464 | T | 3877 | 3863 | 0.996 | TT | 2126 | 2126 | 1.000 | TT | 1308 | 1305 | 0.998 | TT |
| chr16:23476581 | T | 7964 | 7957 | 0.999 | TT | 7968 | 7951 | 0.998 | TT | 7960 | 7944 | 0.998 | TT |
| chr16:23477179 | T | 0 | 0 | NA | NA | 0 | 0 | NA | NA | 0 | 0 | NA | NA |
| chr16:67715890 | T | 7961 | 7929 | 0.996 | TT | 7940 | 7914 | 0.997 | TT | 7922 | 7892 | 0.996 | TT |
| chr16:89630026 | A | 8004 | 7986 | 0.998 | AA | 7997 | 7981 | 0.998 | AA | 7018 | 7003 | 0.998 | AA |
| chr17:3763779 | T | 8005 | 7990 | 0.998 | TT | 4209 | 4203 | 0.999 | TT | 3294 | 3291 | 0.999 | TT |
| chr17:29862208 | A | 7985 | 7957 | 0.996 | AA | 7975 | 7946 | 0.996 | AA | 7987 | 7960 | 0.997 | AA |
| chr17:79780692 | T | 7962 | 7944 | 0.998 | TT | 7950 | 7931 | 0.998 | TT | 7950 | 7933 | 0.998 | TT |
| chr19:30191863 | T | 2579 | 2574 | 0.998 | TT | 1124 | 1121 | 0.997 | TT | 767 | 766 | 0.999 | TT |
| chr19:38887763 | A | 3919 | 3911 | 0.998 | AA | 3346 | 3332 | 0.996 | AA | 2070 | 2066 | 0.998 | AA |
| chr19:38888055 | A | 2251 | 2244 | 0.997 | AA | 787 | 787 | 1.000 | AA | 548 | 546 | 0.996 | AA |
| chr19:38889352 | A | 5076 | 5067 | 0.998 | AA | 3574 | 3563 | 0.997 | AA | 1980 | 1923 | 0.971 | AA |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| chr2:172605884 | A | 8002 | 7989 | 0.998 | AA | 7676 | 7655 | 0.997 | AA | 5823 | 5817 | 0.999 | AA |
| chr2:176791181 | T | 7991 | 7957 | 0.996 | TT | 7998 | 7883 | 0.986 | TT | 7992 | 7962 | 0.996 | TT |
| chr2:176791182 | T | 7991 | 7971 | 0.997 | TT | 7993 | 7968 | 0.997 | TT | 7993 | 7971 | 0.997 | TT |
| chr2:176791183 | T | 7985 | 7956 | 0.996 | TT | 7996 | 7973 | 0.997 | TT | 7989 | 7967 | 0.997 | TT |
| chr2:202486541 | T | 7999 | 7980 | 0.998 | TT | 6933 | 6919 | 0.998 | TT | 5338 | 5329 | 0.998 | TT |
| chr20:5175539 | A | 7989 | 7985 | 0.999 | AA | 5163 | 5159 | 0.999 | AA | 3160 | 3158 | 0.999 | AA |
| chr21:30953750 | T | 8000 | 7972 | 0.997 | TT | 7022 | 6996 | 0.996 | TT | 5256 | 5236 | 0.996 | TT |
| chr21:44452594 | A | 0 | 0 | NA | NA | 0 | 0 | NA | NA | 0 | 0 | NA | NA |
| chr3:10195096 | A | 2625 | 2617 | 0.997 | AA | 1392 | 1388 | 0.997 | AA | 1230 | 1227 | 0.998 | AA |
| chr4:100801747 | T | 7881 | 7832 | 0.994 | TT | 7866 | 7821 | 0.994 | TT | 7844 | 7805 | 0.995 | TT |
| chr4:158281294 | A | 7988 | 7945 | 0.995 | AA | 6723 | 6693 | 0.996 | AA | 4947 | 4931 | 0.997 | AA |
| chr5:68576744 | T | 5825 | 5819 | 0.999 | TT | 4113 | 4110 | 0.999 | TT | 3052 | 3048 | 0.999 | TT |
| chr5:156904831 | T | 7982 | 7931 | 0.994 | TT | 7978 | 7950 | 0.996 | TT | 7983 | 7956 | 0.997 | TT |
| chr5:156904853 | T | 7972 | 7959 | 0.998 | TT | 7977 | 7959 | 0.998 | TT | 7981 | 7959 | 0.997 | TT |
| chr5:156904922 | T | 7896 | 7862 | 0.996 | TT | 7897 | 7848 | 0.994 | TT | 7887 | 7850 | 0.995 | TT |
| chr5:156904950 | T | 7924 | 7903 | 0.997 | TT | 7915 | 7879 | 0.995 | TT | 7908 | 7872 | 0.995 | TT |
| chr5:156905396 | T | 7951 | 7881 | 0.991 | TT | 7944 | 7910 | 0.996 | TT | 7951 | 7897 | 0.993 | TT |
| chr5:156905398 | T | 7932 | 7909 | 0.997 | TT | 7934 | 7906 | 0.996 | TT | 7927 | 7904 | 0.997 | TT |
| chr5:156905405 | T | 7973 | 7949 | 0.997 | TT | 7966 | 7920 | 0.994 | TT | 7971 | 7918 | 0.993 | TT |
| chr5:156905560 | T | 7999 | 7964 | 0.996 | TT | 7999 | 7972 | 0.997 | TT | 7999 | 7958 | 0.995 | TT |
| chr6:18129214 | T | 7999 | 7969 | 0.996 | TT | 8000 | 7964 | 0.996 | TT | 8004 | 7947 | 0.993 | TT |
| chr6:52965317 | A | 0 | 0 | NA | NA | 0 | 0 | NA | NA | 0 | 0 | NA | NA |
| chr6:90344222 | T | 7975 | 7926 | 0.994 | TT | 7976 | 7943 | 0.996 | TT | 7962 | 7935 | 0.997 | TT |
| chr6:90344571 | T | 386 | 382 | 0.990 | TT | 330 | 327 | 0.991 | TT | 325 | 320 | 0.985 | TT |
| chr6:90344605 | T | 385 | 385 | 1.000 | TT | 329 | 328 | 0.997 | TT | 324 | 322 | 0.994 | TT |
| chr6:90344706 | T | 385 | 385 | 1.000 | TT | 331 | 330 | 0.997 | TT | 323 | 323 | 1.000 | TT |
| chr6:102337702 | A | 7993 | 7978 | 0.998 | AA | 6698 | 6693 | 0.999 | AA | 6185 | 6181 | 0.999 | AA |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| chr6:109784286 | T | 7965 | 7932 | 0.996 | TT | 7946 | 7914 | 0.996 | TT | 7374 | 7357 | 0.998 | TT |
| chr6:109784327 | T | 4489 | 4432 | 0.987 | TT | 3773 | 3747 | 0.993 | TT | 3069 | 3048 | 0.993 | TT |
| chr7:5662192 | T | 0 | 0 | NA | NA | 0 | 0 | NA | NA | 0 | 0 | NA | NA |
| chr8:38828267 | A | 121 | 118 | 0.975 | AA | 158 | 154 | 0.975 | AA | 136 | 134 | 0.985 | AA |
| chr8:48889633 | A | 7983 | 7941 | 0.995 | AA | 7976 | 7952 | 0.997 | AA | 7977 | 7943 | 0.996 | AA |
| chr8:48890109 | A | 7996 | 7986 | 0.999 | AA | 8001 | 7990 | 0.999 | AA | 7989 | 7965 | 0.997 | AA |
| chrX:19931744 | T | 7992 | 7969 | 0.997 | TT | 7459 | 7445 | 0.998 | TT | 6676 | 6667 | 0.999 | TT |
| chrX:151358319 | T | 6097 | 6083 | 0.998 | TT | 3194 | 3193 | 1.000 | TT | 2918 | 2915 | 0.999 | TT |

**Table 4.1 Genotype confirmation of sites in *Group II. Increasing.*** Among 12 samples genotyped, three samples are shown as examples. Site and Reference (Ref.): location and reference sequence at A-to-I editing sites in UCSC human genome hg19; Total read count, Reference (Ref.) read count, Reference (Ref.) read rate, Genotype in each sample: uniquely-mapped reads are counted. The numbers and the proportions of reads supporting the reference allele are described. If a proportion of reference allele-supporting reads are greater than 90% at a site, its genotype is declared as homozygous with the reference allele.

| Group | Chromosome | | Gene | Gene region | Edit rate | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Mean | Prenatal mean | Post Infant mean | Mean diff. |
| Group III: High | chr1 | 1595586 | SLC35E2B | 3' UTR | 0.822 | 0.889 | 0.800 | -0.089 |
| Group III: High | chr1 | 3730400 | CEP104 | 3' UTR | 0.770 | 0.712 | 0.788 | 0.076 |
| Group III: High | chr1 | 36067817 | PSMB2 | 3' UTR | 0.903 | 0.831 | 0.927 | 0.095 |
| Group III: High | chr1 | 36067886 | PSMB2 | 3' UTR | 0.735 | 0.676 | 0.757 | 0.081 |
| Group III: High | chr1 | 226791920 | C1orf95 | 3' UTR | 0.710 | 0.755 | 0.691 | -0.064 |
| Group III: High | chr1 | 226793282 | C1orf95 | 3' UTR | 0.731 | 0.665 | 0.759 | 0.094 |
| Group III: High | chr2 | 201843471 | FAM126B | 3' UTR | 0.805 | 0.804 | 0.809 | 0.006 |
| Group III: High | chr3 | 179115627 | GNB4 | 3' UTR | 0.859 | 0.792 | 0.867 | 0.075 |
| Group III: High | chr4 | 17803019 | DCAF16 | 3' UTR | 0.725 | 0.729 | 0.737 | 0.008 |
| Group III: High | chr4 | 17803537 | DCAF16 | 3' UTR | 0.969 | 0.961 | 0.972 | 0.011 |
| Group III: High | chr4 | 158257875 | GRIA2 | CDS | 0.927 | 0.928 | 0.921 | -0.007 |
| Group III: High | chr5 | 131287375 | ACSL6 | 3' UTR | 0.937 | 0.842 | 0.959 | 0.117 |
| Group III: High | chr5 | 131289291 | ACSL6 | 3' UTR | 0.973 | 0.981 | 0.968 | -0.012 |
| Group III: High | chr6 | 158619523 | GTF2H5 | 3' UTR | 0.822 | 0.730 | 0.846 | 0.116 |
| Group III: High | chr10 | 15118748 | ACBD7 | 3' UTR | 0.851 | 0.829 | 0.855 | 0.026 |
| Group III: High | chr10 | 15118774 | ACBD7 | 3' UTR | 0.761 | 0.713 | 0.767 | 0.054 |
| Group III: High | chr11 | 16778023 | C11orf58 | 3' UTR | 0.886 | 0.849 | 0.889 | 0.040 |
| Group III: High | chr12 | 98942688 | TMPO | 3' UTR | 0.885 | 0.811 | 0.905 | 0.094 |
| Group III: High | chr12 | 98943033 | TMPO | 3' UTR | 0.954 | 0.940 | 0.953 | 0.013 |
| Group III: High | chr12 | 117014187 | MAP1LC3B2 | 3' UTR | 0.938 | 0.956 | 0.926 | -0.030 |
| Group III: High | chr12 | 120899011 | GATC | 3' UTR | 0.876 | 0.823 | 0.896 | 0.073 |
| Group III: High | chr13 | 50487444 | SPRYD7 | 3' UTR | 0.776 | 0.854 | 0.767 | -0.086 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Group III: High | chr15 | 75646086 | NEIL1 | CDS | 0.931 | 0.928 | 0.931 | 0.004 |
| Group III: High | chr15 | 90375494 | AP3S2 | 3' UTR | 0.937 | 0.896 | 0.949 | 0.053 |
| Group III: High | chr15 | 90375568 | AP3S2 | 3' UTR | 0.759 | 0.781 | 0.760 | -0.020 |
| Group III: High | chr15 | 90375859 | AP3S2 | 3' UTR | 0.840 | 0.818 | 0.850 | 0.032 |
| Group III: High | chr17 | 2320651 | METTL16 | 3' UTR | 0.837 | 0.795 | 0.846 | 0.051 |
| Group III: High | chr17 | 49042252 | SPAG9 | 3' UTR | 0.916 | 0.826 | 0.938 | 0.112 |
| Group III: High | chr19 | 4654380 | TNFAIP8L1 | 3' UTR | 0.703 | 0.650 | 0.717 | 0.067 |
| Group III: High | chr19 | 10742170 | SLC44A2 | CDS | 0.703 | 0.819 | 0.704 | -0.115 |
| Group III: High | chr19 | 13883381 | MRI1 | 3' UTR | 0.857 | 0.843 | 0.859 | 0.015 |
| Group III: High | chr19 | 39981298 | TIMM50 | 3' UTR | 0.966 | 0.938 | 0.972 | 0.033 |
| Group III: High | chr19 | 40537196 | ZNF780B | 3' UTR | 0.716 | 0.712 | 0.721 | 0.009 |
| Group III: High | chr19 | 54488967 | CACNG8 | 3' UTR | 0.723 | 0.712 | 0.726 | 0.014 |
| Group III: High | chr20 | 3851209 | MAVS | 3' UTR | 0.919 | 0.910 | 0.916 | 0.006 |
| Group III: High | chr20 | 43706947 | STK4 | 3' UTR | 0.829 | 0.839 | 0.820 | -0.019 |
| Group III: High | chr21 | 34636361 | IFNAR2 | 3' UTR | 0.917 | 0.834 | 0.935 | 0.102 |
| Group III: High | chr21 | 34636384 | IFNAR2 | 3' UTR | 0.950 | 0.974 | 0.944 | -0.031 |
| Group III: High | chr22 | 18572675 | PEX26 | 3' UTR | 0.894 | 0.831 | 0.910 | 0.080 |
| Group III: High | chrX | 118672671 | CXorf56 | 3' UTR | 0.710 | 0.677 | 0.719 | 0.042 |
| Group II: Increasing | chr1 | 40147856 | HPCAL4 | 3' UTR | 0.345 | 0.067 | 0.412 | 0.345 |
| Group II: Increasing | chr1 | 53289852 | ZYG11B | 3' UTR | 0.395 | 0.056 | 0.488 | 0.432 |
| Group II: Increasing | chr1 | 53291420 | ZYG11B | 3' UTR | 0.547 | 0.130 | 0.653 | 0.523 |
| Group II: Increasing | chr1 | 67874689 | SERBP1 | 3' UTR | 0.679 | 0.128 | 0.818 | 0.690 |
| Group II: Increasing | chr1 | 67874696 | SERBP1 | 3' UTR | 0.308 | 0.034 | 0.379 | 0.345 |
| Group II: Increasing | chr1 | 109748657 | KIAA1324 | 3' UTR | 0.600 | 0.240 | 0.689 | 0.449 |
| Group II: Increasing | chr2 | 172605884 | DYNC1I2 | 3' UTR | 0.345 | 0.063 | 0.418 | 0.354 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Group II: Increasing | chr2 | 176791181 | KIAA1715 | 3' UTR | 0.350 | 0.028 | 0.449 | 0.421 |
| Group II: Increasing | chr2 | 176791182 | KIAA1715 | 3' UTR | 0.505 | 0.109 | 0.613 | 0.503 |
| Group II: Increasing | chr2 | 176791183 | KIAA1715 | 3' UTR | 0.265 | 0.018 | 0.345 | 0.327 |
| Group II: Increasing | chr2 | 202486541 | TMEM237 | 3' UTR | 0.282 | 0.025 | 0.352 | 0.327 |
| Group II: Increasing | chr3 | 10195096 | VHL | 3' UTR | 0.391 | 0.134 | 0.454 | 0.320 |
| Group II: Increasing | chr4 | 100801747 | LAMTOR3 | 3' UTR | 0.704 | 0.215 | 0.823 | 0.608 |
| Group II: Increasing | chr4 | 158281294 | GRIA2 | CDS | 0.477 | 0.166 | 0.554 | 0.388 |
| Group II: Increasing | chr5 | 68576744 | CCDC125 | 3' UTR | 0.304 | 0.034 | 0.379 | 0.345 |
| Group II: Increasing | chr5 | 156904831 | ADAM19 | 3' UTR | 0.538 | 0.040 | 0.675 | 0.635 |
| Group II: Increasing | chr5 | 156904853 | ADAM19 | 3' UTR | 0.279 | 0.016 | 0.360 | 0.344 |
| Group II: Increasing | chr5 | 156904922 | ADAM19 | 3' UTR | 0.368 | 0.112 | 0.443 | 0.331 |
| Group II: Increasing | chr5 | 156904950 | ADAM19 | 3' UTR | 0.304 | 0.039 | 0.381 | 0.342 |
| Group II: Increasing | chr5 | 156905396 | ADAM19 | 3' UTR | 0.422 | 0.084 | 0.512 | 0.427 |
| Group II: Increasing | chr5 | 156905398 | ADAM19 | 3' UTR | 0.370 | 0.015 | 0.450 | 0.435 |
| Group II: Increasing | chr5 | 156905405 | ADAM19 | 3' UTR | 0.308 | 0.033 | 0.377 | 0.344 |
| Group II: Increasing | chr5 | 156905560 | ADAM19 | 3' UTR | 0.493 | 0.111 | 0.609 | 0.497 |
| Group II: Increasing | chr6 | 18129214 | TPMT | 3' UTR | 0.401 | 0.008 | 0.515 | 0.507 |
| Group II: Increasing | chr6 | 52965317 | FBXO9 | 3' UTR | 0.349 | 0.025 | 0.443 | 0.419 |
| Group II: Increasing | chr6 | 90344222 | LYRM2 | 3' UTR | 0.586 | 0.122 | 0.716 | 0.594 |
| Group II: Increasing | chr6 | 90344571 | LYRM2 | 3' UTR | 0.348 | 0.037 | 0.425 | 0.388 |
| Group II: Increasing | chr6 | 90344605 | LYRM2 | 3' UTR | 0.393 | 0.104 | 0.473 | 0.369 |
| Group II: Increasing | chr6 | 90344706 | LYRM2 | 3' UTR | 0.266 | 0.045 | 0.324 | 0.279 |
| Group II: Increasing | chr6 | 102337702 | GRIK2 | CDS | 0.645 | 0.216 | 0.758 | 0.542 |
| Group II: Increasing | chr6 | 109784286 | ZBTB24 | 3' UTR | 0.443 | 0.159 | 0.508 | 0.349 |
| Group II: Increasing | chr6 | 109784327 | ZBTB24 | 3' UTR | 0.445 | 0.180 | 0.509 | 0.330 |

81

| Group | chr | Position | Gene | Region | | | | |
|---|---|---|---|---|---|---|---|---|
| Group II: Increasing | chr7 | 5662192 | RNF216 | 3' UTR | 0.340 | 0.059 | 0.426 | 0.367 |
| Group II: Increasing | chr8 | 38828267 | PLEKHA2 | 3' UTR | 0.327 | 0.040 | 0.387 | 0.346 |
| Group II: Increasing | chr8 | 48889633 | MCM4 | 3' UTR | 0.480 | 0.100 | 0.586 | 0.485 |
| Group II: Increasing | chr8 | 48890109 | MCM4 | 3' UTR | 0.659 | 0.182 | 0.786 | 0.605 |
| Group II: Increasing | chr10 | 15120275 | ACBD7 | 3' UTR | 0.458 | 0.139 | 0.546 | 0.406 |
| Group II: Increasing | chr10 | 15120306 | ACBD7 | 3' UTR | 0.324 | 0.118 | 0.376 | 0.258 |
| Group II: Increasing | chr10 | 102777342 | PDZD7 | CDS | 0.613 | 0.144 | 0.734 | 0.590 |
| Group II: Increasing | chr12 | 107280405 | RIC8B | 3' UTR | 0.391 | 0.067 | 0.488 | 0.421 |
| Group II: Increasing | chr14 | 26917530 | NOVA1 | CDS | 0.245 | 0.015 | 0.309 | 0.294 |
| Group II: Increasing | chr14 | 31916464 | DTD2 | 3' UTR | 0.496 | 0.182 | 0.577 | 0.394 |
| Group II: Increasing | chr16 | 23476581 | GGA2 | 3' UTR | 0.683 | 0.213 | 0.803 | 0.590 |
| Group II: Increasing | chr16 | 23477179 | GGA2 | 3' UTR | 0.310 | 0.077 | 0.380 | 0.304 |
| Group II: Increasing | chr16 | 67715890 | GFOD2 | 3' UTR | 0.452 | 0.111 | 0.538 | 0.427 |
| Group II: Increasing | chr16 | 89630026 | RPL13 | 3' UTR | 0.515 | 0.141 | 0.626 | 0.484 |
| Group II: Increasing | chr17 | 3763779 | CAMKK1 | 3' UTR | 0.483 | 0.069 | 0.608 | 0.539 |
| Group II: Increasing | chr17 | 29862208 | RAB11FIP4 | 3' UTR | 0.535 | 0.162 | 0.629 | 0.466 |
| Group II: Increasing | chr17 | 79780692 | FAM195B | 3' UTR | 0.526 | 0.069 | 0.655 | 0.586 |
| Group II: Increasing | chr19 | 30191863 | C19orf12 | 3' UTR | 0.357 | 0.100 | 0.419 | 0.319 |
| Group II: Increasing | chr19 | 38887763 | SPRED3 | 3' UTR | 0.249 | 0.047 | 0.306 | 0.259 |
| Group II: Increasing | chr19 | 38888055 | SPRED3 | 3' UTR | 0.360 | 0.050 | 0.452 | 0.402 |
| Group II: Increasing | chr19 | 38889352 | SPRED3 | 3' UTR | 0.340 | 0.032 | 0.422 | 0.390 |
| Group II: Increasing | chr20 | 5175539 | CDS2 | 3' UTR | 0.327 | 0.040 | 0.406 | 0.366 |
| Group II: Increasing | chr21 | 30953750 | GRIK1 | CDS | 0.517 | 0.130 | 0.599 | 0.469 |
| Group II: Increasing | chr21 | 44452594 | PKNOX1 | 3' UTR | 0.283 | 0.040 | 0.346 | 0.305 |
| Group II: Increasing | chrX | 19931744 | CXorf23 | 3' UTR | 0.529 | 0.153 | 0.633 | 0.480 |

82

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Group II: Increasing | chrX | 151358319 | GABRA3 | CDS | 0.694 | 0.201 | 0.804 | 0.603 |
| Group I: Low | chr1 | 20978457 | DDOST | 3' UTR | 0.131 | 0.230 | 0.108 | -0.122 |
| Group I: Low | chr1 | 109474780 | CLCC1 | 3' UTR | 0.184 | 0.148 | 0.191 | 0.043 |
| Group I: Low | chr1 | 160112527 | ATP1A2 | 3' UTR | 0.143 | 0.171 | 0.142 | -0.029 |
| Group I: Low | chr1 | 179070646 | ABL2 | 3' UTR | 0.179 | 0.188 | 0.174 | -0.014 |
| Group I: Low | chr2 | 102508477 | MAP4K4 | 3' UTR | 0.110 | 0.109 | 0.108 | -0.001 |
| Group I: Low | chr2 | 166730667 | TTC21B | 3' UTR | 0.131 | 0.062 | 0.157 | 0.095 |
| Group I: Low | chr3 | 49452787 | TCTA | 3' UTR | 0.108 | 0.143 | 0.105 | -0.038 |
| Group I: Low | chr3 | 101545798 | NXPE3 | 3' UTR | 0.134 | 0.120 | 0.138 | 0.018 |
| Group I: Low | chr3 | 155480786 | C3orf33 | 3' UTR | 0.171 | 0.172 | 0.180 | 0.008 |
| Group I: Low | chr4 | 7059305 | TADA2B | 3' UTR | 0.109 | 0.085 | 0.118 | 0.033 |
| Group I: Low | chr4 | 166000148 | TMEM192 | 3' UTR | 0.136 | 0.068 | 0.153 | 0.085 |
| Group I: Low | chr6 | 42175084 | MRPS10 | 3' UTR | 0.107 | 0.126 | 0.104 | -0.022 |
| Group I: Low | chr6 | 42175097 | MRPS10 | 3' UTR | 0.103 | 0.098 | 0.106 | 0.008 |
| Group I: Low | chr7 | 65618305 | CRCP | 3' UTR | 0.141 | 0.148 | 0.139 | -0.009 |
| Group I: Low | chr7 | 65618306 | CRCP | 3' UTR | 0.136 | 0.134 | 0.138 | 0.004 |
| Group I: Low | chr7 | 65618332 | CRCP | 3' UTR | 0.101 | 0.089 | 0.103 | 0.013 |
| Group I: Low | chr7 | 73150707 | ABHD11 | 3' UTR | 0.121 | 0.163 | 0.110 | -0.054 |
| Group I: Low | chr7 | 73646106 | RFC2 | 3' UTR | 0.137 | 0.122 | 0.142 | 0.020 |
| Group I: Low | chr7 | 73646121 | RFC2 | 3' UTR | 0.112 | 0.100 | 0.116 | 0.016 |
| Group I: Low | chr7 | 73646147 | RFC2 | 3' UTR | 0.145 | 0.144 | 0.148 | 0.004 |
| Group I: Low | chr7 | 73646229 | RFC2 | 3' UTR | 0.130 | 0.089 | 0.141 | 0.052 |
| Group I: Low | chr7 | 92166602 | RBM48 | 3' UTR | 0.115 | 0.135 | 0.110 | -0.024 |
| Group I: Low | chr7 | 102089037 | ORAI2 | 3' UTR | 0.123 | 0.114 | 0.125 | 0.011 |
| Group I: Low | chr7 | 102090451 | ORAI2 | 3' UTR | 0.131 | 0.135 | 0.132 | -0.003 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Group I: Low | chr8 | 42884354 | HOOK3 | 3' UTR | 0.152 | 0.081 | 0.172 | 0.091 |
| Group I: Low | chr8 | 42884422 | HOOK3 | 3' UTR | 0.112 | 0.116 | 0.114 | -0.003 |
| Group I: Low | chr8 | 42884423 | HOOK3 | 3' UTR | 0.162 | 0.119 | 0.174 | 0.054 |
| Group I: Low | chr8 | 104411941 | SLC25A32 | 3' UTR | 0.112 | 0.107 | 0.115 | 0.008 |
| Group I: Low | chr9 | 132590069 | C9orf78 | 3' UTR | 0.105 | 0.085 | 0.104 | 0.019 |
| Group I: Low | chr10 | 82192318 | FAM213A | 3' UTR | 0.172 | 0.204 | 0.166 | -0.038 |
| Group I: Low | chr10 | 82282216 | TSPAN14 | 3' UTR | 0.102 | 0.064 | 0.110 | 0.046 |
| Group I: Low | chr10 | 126451032 | METTL10 | CDS | 0.143 | 0.085 | 0.154 | 0.069 |
| Group I: Low | chr11 | 8707840 | RPL27A | 3' UTR | 0.153 | 0.153 | 0.156 | 0.003 |
| Group I: Low | chr11 | 61567700 | FADS1 | 3' UTR | 0.103 | 0.099 | 0.105 | 0.006 |
| Group I: Low | chr11 | 63724420 | NAA40 | 3' UTR | 0.136 | 0.048 | 0.159 | 0.111 |
| Group I: Low | chr11 | 111653930 | ALG9 | 3' UTR | 0.125 | 0.110 | 0.135 | 0.025 |
| Group I: Low | chr12 | 54629209 | CBX5 | 3' UTR | 0.115 | 0.098 | 0.116 | 0.019 |
| Group I: Low | chr13 | 46090371 | COG3 | CDS | 0.106 | 0.031 | 0.120 | 0.090 |
| Group I: Low | chr14 | 23303976 | MRPL52 | 3' UTR | 0.228 | 0.150 | 0.244 | 0.094 |
| Group I: Low | chr16 | 28976933 | NFATC2IP | 3' UTR | 0.108 | 0.127 | 0.098 | -0.029 |
| Group I: Low | chr16 | 69390514 | TERF2 | 3' UTR | 0.125 | 0.104 | 0.125 | 0.021 |
| Group I: Low | chr16 | 70407018 | DDX19A | 3' UTR | 0.133 | 0.121 | 0.134 | 0.013 |
| Group I: Low | chr16 | 70413821 | ST3GAL2 | 3' UTR | 0.118 | 0.130 | 0.110 | -0.020 |
| Group I: Low | chr17 | 1368284 | MYO1C | 3' UTR | 0.106 | 0.119 | 0.096 | -0.023 |
| Group I: Low | chr17 | 1368288 | MYO1C | 3' UTR | 0.105 | 0.118 | 0.097 | -0.020 |
| Group I: Low | chr17 | 20217878 | SPECC1 | 3' UTR | 0.134 | 0.057 | 0.151 | 0.094 |
| Group I: Low | chr17 | 20217935 | SPECC1 | 3' UTR | 0.132 | 0.073 | 0.154 | 0.081 |
| Group I: Low | chr17 | 25640011 | WSB1 | 3' UTR | 0.106 | 0.107 | 0.107 | 0.000 |
| Group I: Low | chr17 | 28513019 | NSRP1 | 3' UTR | 0.109 | 0.080 | 0.113 | 0.034 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Group I: Low | chr18 | 11882606 | GNAL | 3' UTR | 0.149 | 0.101 | 0.166 | 0.065 |
| Group I: Low | chr19 | 5206255 | PTPRS | 3' UTR | 0.111 | 0.167 | 0.108 | -0.058 |
| Group I: Low | chr19 | 40022726 | EID2B | 3' UTR | 0.107 | 0.085 | 0.115 | 0.030 |
| Group I: Low | chr19 | 41828971 | CCDC97 | 3' UTR | 0.119 | 0.033 | 0.135 | 0.102 |
| Group I: Low | chr19 | 58774627 | ZNF544 | 3' UTR | 0.134 | 0.093 | 0.150 | 0.057 |
| Group I: Low | chr20 | 3805458 | AP5S1 | 3' UTR | 0.109 | 0.214 | 0.083 | -0.132 |
| Group I: Low | chr20 | 3805459 | AP5S1 | 3' UTR | 0.200 | 0.286 | 0.171 | -0.115 |
| Group I: Low | chr20 | 3805511 | AP5S1 | 3' UTR | 0.146 | 0.129 | 0.153 | 0.024 |
| Group I: Low | chr20 | 3853956 | MAVS | 3' UTR | 0.115 | 0.088 | 0.129 | 0.041 |
| Group I: Low | chr20 | 3853967 | MAVS | 3' UTR | 0.165 | 0.124 | 0.184 | 0.060 |
| Group I: Low | chr20 | 3854106 | MAVS | 3' UTR | 0.145 | 0.108 | 0.159 | 0.051 |
| Group I: Low | chr21 | 34728113 | IFNAR1 | 3' UTR | 0.124 | 0.079 | 0.124 | 0.045 |
| Group I: Low | chr21 | 34728136 | IFNAR1 | 3' UTR | 0.133 | 0.190 | 0.109 | -0.081 |
| Group I: Low | chr21 | 37666181 | DOPEY2 | 3' UTR | 0.122 | 0.106 | 0.125 | 0.019 |
| Group I: Low | chr22 | 21244926 | SNAP29 | 3' UTR | 0.104 | 0.133 | 0.099 | -0.034 |
| Group I: Low | chrX | 123044597 | XIAP | 3' UTR | 0.112 | 0.058 | 0.129 | 0.070 |

**Table 4.2 The representative sites for developmental A-to-I editing patterns.** Group: a pattern that a site represents, corresponding to figure 4.2; Chromosome and Coordinate: location of A-to-I editing sites in UCSC human genome hg19; Gene and Gene region: gene annotation based on RefSeq; Mean, Prenatal mean, Post-infant mean and Mean difference (diff.) in Editing rate: summary of editing rates in 33 samples.

| Site | Group | Depth | Edit rate | | | Knock down effect | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | Control | ADAR1 KD | ADAR2 KD | ADAR1 KD | ADAR2 KD |
| chr1:1595586 | Group III: High | 138 | 0.78 | 0.56 | 0.74 | 0.29 | 0.06 |
| chr1:3730400 | Group III: High | 54 | 0.93 | 0.22 | 0.84 | 0.76 | 0.10 |
| chr1:36067817 | Group III: High | 449 | 0.82 | 0.40 | 0.80 | 0.51 | 0.02 |
| chr1:36067886 | Group III: High | 320 | 0.75 | 0.51 | 0.81 | 0.32 | -0.09 |
| chr2:201843471 | Group III: High | 43 | 0.88 | 0.67 | 0.76 | 0.25 | 0.14 |
| chr4:17803019 | Group III: High | 55 | 0.80 | 0.59 | 0.72 | 0.27 | 0.10 |
| chr4:17803537 | Group III: High | 44 | 0.84 | 0.72 | 0.98 | 0.15 | -0.16 |
| chr12:98942688 | Group III: High | 243 | 0.93 | 0.50 | 0.92 | 0.46 | 0.01 |
| chr12:98943033 | Group III: High | 262 | 0.93 | 0.72 | 0.97 | 0.22 | -0.05 |
| chr12:117014187 | Group III: High | 41 | 0.98 | 1.00 | 1.00 | -0.03 | -0.03 |
| chr12:120899011 | Group III: High | 277 | 0.89 | 0.51 | 0.87 | 0.42 | 0.02 |
| chr13:50487444 | Group III: High | 34 | 0.76 | 0.40 | 0.76 | 0.48 | 0.00 |
| chr15:90375494 | Group III: High | 44 | 0.96 | 0.74 | 0.91 | 0.22 | 0.05 |
| chr15:90375568 | Group III: High | 46 | 0.78 | 0.63 | 0.83 | 0.19 | -0.05 |
| chr15:90375859 | Group III: High | 45 | 0.86 | 0.38 | 0.78 | 0.56 | 0.09 |
| chr17:2320651 | Group III: High | 134 | 0.75 | 0.37 | 0.72 | 0.50 | 0.04 |
| chr17:49042252 | Group III: High | 70 | 0.94 | 0.49 | 0.78 | 0.48 | 0.16 |
| chr19:4654380 | Group III: High | 120 | 0.74 | 0.26 | 0.72 | 0.65 | 0.04 |
| chr19:10742170 | Group III: High | 399 | 0.52 | 0.49 | 0.47 | 0.06 | 0.09 |
| chr19:13883381 | Group III: High | 99 | 0.98 | 0.58 | 0.84 | 0.41 | 0.15 |
| chr19:39981298 | Group III: High | 42 | 1.00 | 0.32 | 0.98 | 0.68 | 0.02 |
| chr19:40537196 | Group III: High | 32 | 0.66 | 0.37 | 0.54 | 0.43 | 0.17 |
| chr20:3851209 | Group III: High | 50 | 0.98 | 0.47 | 0.94 | 0.52 | 0.04 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| chr20:43706947 | Group III: High | 728 | 0.61 | 0.40 | 0.62 | 0.35 | 0.00 |
| chr21:34636361 | Group III: High | 371 | 0.65 | 0.21 | 0.69 | 0.68 | -0.06 |
| chr21:34636384 | Group III: High | 357 | 0.86 | 0.51 | 0.88 | 0.41 | -0.03 |
| chr22:18572675 | Group III: High | 352 | 0.68 | 0.32 | 0.77 | 0.52 | -0.14 |
| chrX:118672671 | Group III: High | 119 | 0.68 | 0.26 | 0.66 | 0.61 | 0.03 |
| chr1:53289852 | Group II: Inc. | 32 | 0.31 | 0.00 | 0.14 | 1.00 | 0.55 |
| chr1:53291420 | Group II: Inc. | 75 | 0.31 | 0.08 | 0.40 | 0.75 | -0.30 |
| chr1:67874689 | Group II: Inc. | 225 | 0.68 | 0.03 | 0.72 | 0.95 | -0.06 |
| chr1:67874696 | Group II: Inc. | 232 | 0.26 | 0.00 | 0.28 | 0.99 | -0.08 |
| chr2:176791181 | Group II: Inc. | 30 | 0.30 | 0.00 | 0.09 | 1.00 | 0.71 |
| chr2:176791182 | Group II: Inc. | 30 | 0.33 | 0.00 | 0.18 | 1.00 | 0.47 |
| chr3:10195096 | Group II: Inc. | 93 | 0.62 | 0.04 | 0.45 | 0.93 | 0.28 |
| chr4:100801747 | Group II: Inc. | 41 | 0.39 | 0.00 | 0.49 | 1.00 | -0.26 |
| chr5:68576744 | Group II: Inc. | 59 | 0.24 | 0.00 | 0.17 | 1.00 | 0.26 |
| chr5:156904831 | Group II: Inc. | 182 | 0.22 | 0.03 | 0.14 | 0.86 | 0.36 |
| chr5:156904922 | Group II: Inc. | 129 | 0.29 | 0.03 | 0.14 | 0.91 | 0.53 |
| chr5:156904950 | Group II: Inc. | 147 | 0.12 | 0.00 | 0.08 | 1.00 | 0.39 |
| chr5:156905396 | Group II: Inc. | 119 | 0.20 | 0.03 | 0.26 | 0.85 | -0.31 |
| chr5:156905560 | Group II: Inc. | 145 | 0.35 | 0.03 | 0.40 | 0.92 | -0.16 |
| chr6:90344222 | Group II: Inc. | 67 | 0.51 | 0.11 | 0.42 | 0.78 | 0.17 |
| chr6:90344605 | Group II: Inc. | 77 | 0.31 | 0.00 | 0.20 | 1.00 | 0.37 |
| chr6:90344706 | Group II: Inc. | 67 | 0.22 | 0.00 | 0.21 | 1.00 | 0.06 |
| chr6:109784286 | Group II: Inc. | 99 | 0.27 | 0.01 | 0.22 | 0.96 | 0.16 |
| chr6:109784327 | Group II: Inc. | 104 | 0.41 | 0.00 | 0.50 | 1.00 | -0.22 |
| chr8:38828267 | Group II: Inc. | 178 | 0.33 | 0.03 | 0.19 | 0.90 | 0.43 |

| Site | Group | Depth | | | | | |
|------|-------|-------|---|---|---|---|---|
| chr8:48889633 | Group II: Inc. | 481 | 0.13 | 0.01 | 0.09 | 0.89 | 0.34 |
| chr8:48890109 | Group II: Inc. | 479 | 0.27 | 0.02 | 0.23 | 0.92 | 0.16 |
| chr14:31916464 | Group II: Inc. | 79 | 0.59 | 0.05 | 0.63 | 0.92 | -0.06 |
| chr16:23476581 | Group II: Inc. | 250 | 0.35 | 0.02 | 0.38 | 0.94 | -0.07 |
| chr16:67715890 | Group II: Inc. | 23 | 0.31 | 0.00 | 0.22 | 1.00 | 0.29 |
| chr16:89630026 | Group II: Inc. | 154 | 0.31 | 0.01 | 0.34 | 0.97 | -0.09 |
| chr19:30191863 | Group II: Inc. | 26 | 0.35 | 0.00 | 0.18 | 1.00 | 0.49 |
| chr20:5175539 | Group II: Inc. | 69 | 0.25 | 0.06 | 0.23 | 0.78 | 0.07 |

**Table 4.3 Effect of ADAR knock down (KD) by siRNA on developmental A-to-I editing patterns.** Site: location of A-to-I editing sites in UCSC human genome hg19; Group: a pattern that a site represents, corresponding to figure 4.2, 'Group II: Inc.' stands for 'Group II. Increasing'; Depth: minimum sequencing depth among control, ADAR1 KD, ADAR2 KD RNA-seq experiments; Editing rate and Knock down effect: The knock down effect for a given site is the percent decrease of editing rate in ADAR1/2 KD experiments compared to controls.

# Chapter 5

# Functional implications of increasing A-to-I editing pattern

In this chapter, I explore the functional implications of the developmental A-to-I editing patterns. First, I report that the increasing editing pattern is much less apparent in other organ tissues and is conserved in mouse brain development, highlighting the importance of the increasing pattern in the nervous system. Second, the increasing editing pattern is analyzed at the cellular level with the differentiating human and mouse embryonic stem cells, which reveals its temporal association with the growth of cortical layers and neuronal maturation. Gene Ontology (GO) analyses implicate genes with the increasing A-to-I editing pattern in vesicle/organelle membrane and glutamate signaling pathways. Third, possible roles of the increasing editing pattern are presented at a molecular level. The computational analyses show that editing rates are correlated with gene expression levels and have potential to affect miRNA-binding. Finally, I demonstrate that the increasing editing pattern involves genes associated with neurodevelopmental disorders and is perturbed in spinal cord injury and glioblastoma, noting potential clinical relevance.

## 5.1. Tissue variation of the increasing editing pattern

In order to explore the functional implications of the developmental A-to-I editing patterns, I first investigated how the three patterns manifest in other human tissues, using available RNA-seq datasets (see section 5.6 Methods). Editing rates at the sites showing the increasing pattern in brain were compared across brain, heart and liver, each consisting of one fetal and one adult sample. It should be noted that here independent brain samples were used from the initial 33 discovery brain tissues, which allows further validation of the patterns. Interestingly, the brain shows the clearest separation of the three editing patterns, with the increasing editing pattern found principally in brain (Figure. 5.1a), and with very little of the increasing pattern in heart (Figure. 5.1b) and less in liver (Figure. 5.1c). In contrast, high-edited sites and low-edited sites show similar editing rates in fetus and adult regardless of tissue types. This analysis was extended to a total of 6 different organ tissues with multiple fetal and adult samples in every tissue (Figure. 5.2). Brain tissue has again the most significant distinction of the increasing A-to-I editing pattern (ANOVA and post-hoc two-sample t-test). It turned out that only brain tissue has higher expression levels of both ADAR1 and ADAR2 in adult samples compared to fetal samples (Figure. 5.3). In short, the increasing pattern of RNA editing, though found to a small degree in some other tissues, is strongest in brain tissue, at least at these specific A-to-I editing sites we found changing across brain development.  But it is also confirmed that some other tissues have their own set of sites showing a developmentally increasing editing pattern (see 5.6 Methods, Figure 5.4). Specifically, lung and liver also show relative enrichment of increasingly-edited sites, compared to heart and muscle. But the sites showing an increasing editing pattern in these tissues are mostly different in each tissue and for the most part not the same as those in brain (Figure 5.5). Therefore, it should be noted that the discovery of the A-to-I editing sites showing the increasing pattern in brain development does not mean that increasing

90

editing patterns per se are unique to brain. Increasing patterns might be seen in the development of different organs at different editing sites.

## 5.2. The increasing editing pattern in mouse brain development

Available mouse RNA-seq data (Sauvageau et al. 2013; Fertuzinhos et al. 2014) were also surveyed to see if the increasing A-to-I editing pattern in human brain development is conserved in mouse brain development. Among 742 A-to-I editing sites showing the increasing pattern in human brain, 95 sites are found to be conserved in the mouse genome (Table 5.1). The 64 sites with adequate sequencing depth (median depth$\geq$20) were investigated and a similar increasing pattern was mainly found at the sites in CDS regions (Figures 5.6 and 5.7 with Table 5.2). As expected, this increasing is correlated with mouse ADAR expression levels, especially for ADAR2 (Figure 5.8). One exceptional site in a CDS, showing marked loss of the increasing pattern, is the editing site in NEIL1, which may be understood in terms of Alu repeats in the human genome. The Alu repeats neighboring this site form long double-stranded structures (Figure 5.9) and are believed to induce RNA editing only in human tissues (Daniel et al. 2014). These results show that although the number of editing sites showing the increasing pattern is relatively limited in mouse, the pattern itself is conserved in both human and mouse brain development.

But it should be also noted that sites with increasing pattern that is found only in human genome, not in model organisims such as mouse, are likely human or primate-specific editing sites with potential functionalities. One possible example is the novel RNA editing site in PDZD7 (PDZ-domain containing protein 7). This site has a developmentally-increasing editing rate and

causes the loss of a stop codon (UAG to UGG) in the gene isoform specific to humans (Figure 5.10). In fact, PDZD7, a scaffolding protein implicated in Usher syndrome is known to be involved in the cilia compartment which is important in the developing brain (Sarkisian & Guadiana 2015; Valente et al. 2013).

## 5.3. Cellular understanding of the increasing editing pattern

The relative brain-selectivity of the evolutionary-conserved developmentally-increasing A-to-I editing pattern implies a potentially unique functional role in the nervous system. To explore this possibility at a cellular level, I investigated *in-vitro* differentiation of human and mouse embryonic stem cells (hESCs and mESCs) into cortical neurons from available RNA-seq datasets (van de Leemput et al. 2014; Hubbard et al. 2013). Although the amount of increasing is modest compared to the post-mortem brain samples (Figure 5.11), I confirmed similar increasing patterns of A-to-I editing such that the relative differences within the increasing pattern found in brain tissues are replicated in the *in-vitro* corticogenesis with hESCs, as marked by the matching numbers between figure 3.4 (also in figure 3.6) and figure 5.12. The in-vitro differentiation of mouse ESCs shows the increasing pattern more clearly (Figure 5.13). This *in-vitro* recapitulation of the increasing pattern enables us to annotate a developmental stage manifesting this pattern. Specifically, the increasing pattern appears to occur coincidentally with deep cortical layer (DL) formation after cortical speciation (CS) according to previously-defined cellular phenotypes as shown in a top bar in figure. 5.12 (van de Leemput et al. 2014). Indeed, it correlates with expression of a deep cortical layer marker, TBR1 and an axon marker, MAPT (Figure 5.14), showing that cortical layer development with neuronal maturation may be a specific period involving the emergence of the increasing RNA editing pattern. The relevance for neuronal

maturation is further supported by the change in editing rate found in primary culture of mouse neurons (Figure 5.15). A reanalysis of the previous data (Rybak-Wolf et al. 2015) reveals that the editing rates are increasing as early neurons mature with the increasing expression of synaptic markers such as SYN1 and SYN2 (Figure 5.16). These results are also consistent with the observation from *in-vivo* mouse brain tissues (Figure 5.6) where the increasing pattern emerges in embryonic day 15 to 18 and continues into early post-natal days, a period characterized by cortical layer expansion and neuronal maturation.

The GO terms associated with genes showing the increasing RNA editing pattern (Figure 5.17 and Table 5.3), such as glutamate signaling and organelle or vesicle-related transport, are cellular processes known to be active during cortical circuit development that we highlight. For example, glutamate signaling participates in neuronal migration in cortex (Luhmann et al. 2015) and also stimulates the growth of functional spines in developing cortex (Kwon & Sabatini 2011). Organelles and vesicles are critical in regulating neuronal morphogenesis including neurite growth and synaptic plasticity (Sekine et al. 2009). Interestingly, the genes in the increasing pattern have different functional implications depending on the location of the editing site in the gene. Editing sites in 3' UTR's are associated with localization-related GO terms, while CDS-residing A-to-I editing sites are involved primarily with synapse or glutamate-related GO terms (Tables 5.4 and 5.5).

## 5.4. Molecular understanding of the increasing pattern

Functional implications of the increasing pattern were also interrogated at a molecular level. First, the impact of A-to-I editing in CDS on protein function was computationally

predicted. Seventy-seven percent (24 out of 31 in CDS) of these sites cause non synonymous shifts including one stop codon deletion and six damaging amino acid changes (Table 5.6). Second, a potential relationship between editing rates and mRNA abundances was investigated. The increasingly-edited sites have some correlation with host mRNA abundances, either positive or negative (Figure 5.18a). Interestingly, the distributions of correlation coefficients of editing rates with their respect mRNAs are significantly different between CDS-residing editing sites and 3'UTR-associated editing sites (Figure 5.18b, p-value $\leq 0.05$ by Kolmogorov-Smirnov test), implying that RNA editing in 3' UTRs may have relatively richer regulatory potential to affect RNA abundances through either RNA stabilization or degradation depending on genes, than ones in CDS regions, as generally expected. Finally, the potential of increasingly-edited sites to perturb miRNA-binding was explored. After computationally predicting the binding energy between miRNA and mRNA regions flanking editing sites (see 5.6 methods), it was found that edited mRNA regions generally have lower binding energy with miRNA, compared to mRNA regions without RNA editing (Figure 5.19, p-value$\leq 0.001$ by two-sample t-test). These results implicate possible regulatory roles of RNA editing in the context of brain development.

## 5.5. The increasing pattern in brain disorders

RNA editing is known to be involved in several human diseases (Slotkin & Nishikura 2013). I hypothesized that the increasing pattern of A-to-I editing may be especially relevant to clinical conditions that involve brain development or damage. First, I looked into two neuropathological conditions in which RNA editing has already been implicated, glioblastoma and spinal cord injury (SCI). The increasing pattern is selectively and significantly disrupted in

both conditions. Specifically, in the previous mouse model of spinal cord injury (Chen et al. 2013), the editing sites in CDS regions, which show the increasing pattern in mouse brain development (Figure 5.6), tend to decrease as the disease progresses from acute and subacute phases (Figure 5.20). Also, the editing rates of sites in *Group II. Increasing*, found in the developing brain, are decreased in the glioblastoma samples compared to the flanking non-tumor tissues of the same individual, while the editing sites in the *Group I. Low* and the *Group III. High* keep their editing rates stable regardless of conditions (Fig. 5.21). Given the increase in glia after injury and in this tumor, it might be questioned whether this change in cellular composition accounts for these disease-associated findings. However, as shown in the chapter 3 (specifically, 3.5), simple compositional changes do not seem to be the principle explanation for the developmental pattern and its selective alteration in glioblastoma and spinal cord injury. Rather, the observation that these pathological states involve reversals of a specific developmental editing pattern in brain, i.e. the increasing editing pattern, may provide insights into how the imbalance of RNA editing influences these pathological conditions. For example, the aberration of the increasing RNA editing might be invovled in the dedifferentiation of cell types in glioblastoma. Also, the increasingly-edited sites can have direct clinical significance as a recent study showed that two A-to-I editing sites, R/G site in GRIA2 and I/V site in COG3, showing the increasing pattern in our dataset affects drug sensitivity of cancer cell lines (Han et al. 2015).

Second, the enrichment test of genes showing the increasing pattern was performed in terms of previously-defined gene sets associated with various neurodevelopmental disorders including autism, schizophrenia, bipolar disorder, intellectual disability (ID) as well as disorders thought not to be neurodevelopmental, e.g. type 2 diabetes (T2D), Alzheimer disease (AD) and Parkinson disease (PD) (Birnbaum et al. 2014; Ripke et al. 2014). Several genes with increasing A-to-I editing patterns were involved in neurodevelopmental disorders (Tables 5.7 and 5.8), while they are generally not associated with T2D, AD and PD (with one exception, AP3S2 for T2D).

This tendency is not limited to the genes with the increasing pattern but in general with genes having A-to-I editing sites. While these results do not suggest that editing is a pathogenic factor in these disorders, it implies that editing plays a role in the regulation of at least some genes that are involved in these conditions. Regarding this, a recent study showed that the balance of RNA editing is perturbed by FMRP, a key protein involved in intellectual disability (Bhogal et al. 2011). Perhaps related to this observation, some genes with the increasing editing pattern are implicated in intellectual disabilities as well as potentially other neurodevelopmental disorders.

## 5.6. Methods

**Datasets from public database**

RNA-seq raw datasets for various samples were downloaded from the following references: Illumina Body Map 2.0 project, GSE69360 (Choy et al. 2015), NIH Roadmap Epigenomics (http://www.roadmapepigenomics.org/) and ENCODE project (Bernstein et al. 2012) for human organ tissues, *in-vitro* differentiation of human embryonic stem cells (hESCs) (van de Leemput et al. 2014) and mouse embryonic stem cells (mESCs) into cortical neurons (Hubbard et al. 2013), *in-vivo* mouse brain tissues (Sauvageau et al. 2013; Fertuzinhos et al. 2014), primary culture of mouse cortical neurons (Rybak-Wolf et al. 2015), cell type specific RNA-seq with mouse cortical tissues (Zhang et al. 2014), single cell RNA-seq with human brain tissues (Darmanis et al. 2015), mouse model of spinal cord injury (Chen et al. 2013). The details are found in Table 5.9. DNA methylation data used to estimate cellular composition in brain tissues can be found in the reference (Jaffe et al. 2016).

**RNA-seq data preprocessing**

RNA-seq data from previous studies are aligned by STAR (Dobin et al. 2013). RPKM (Reads Per Kilobase per Million) were calculated to estimate gene expression level using HTSeq (Anders et al. 2014).

**Genome-wide editing rate differences between fetal and adult samples in multiple tissues.**

The computational tool described in Chapter 2 was modified to identify RNA editing sites for a simple but thorough comparison of RNA editing rates between fetal and adult samples as follows: First, an initial call of RNA editing sites was made, which include the sites with at least five sequencing reads with at least two variant-supporting reads. Second, the possible genomic variants were removed by excluding SNP sites (except for SNPs of molecular type 'cDNA'), sites only shown in a single sample, and sites with multiple variants (by removing those sites whose numbers of sequencing reads supporting the major and the minor allele are less than 95% of total sequencing reads). Finally, A-to-I editing rates were compared between fetal and adult samples if the median depth of a site is greater than 20 and a site is in mRNA regions (5'UTR, CDS and 3'UTR).

**miRNA binding prediction**

Whether a miRNA binds to the mRNA regions around RNA editing sites was computationally predicted using the software *miRanda* (Enright et al. 2003). Specifically, mature miRNA sequences are obtained from the well-known database, miRBase (Griffiths-Jones et al. 2008). For mRNA target sequences, two types of sequences are prepared with flanking regions (50 bp upstream and downstream) of editing sites in all mRNA transcripts in refSeq: the reference

sequence ('reference') and the sequence affected by A-to-I editing ('edited'). The binding energies were calculated between a miRNA and both a 'reference' and 'edited' mRNA. The comparison was performed with all the predicted binding pairs of miRNAs and mRNA targets.

**Bioinformatics for functional analyses**

The coordinates in the human genome (UCSC hg19) was converted to the mouse genome (UCSC mm10) using UCSC coordinate converter or *liftOver* (https://genome.ucsc.edu/cgi-bin/hgLiftOver). The effect of CDS-associated RNA editing on protein functions was predicted by PolyPhen-2 (Maathuis et al. 2000). Enrichment of genes in pre-defined gene sets was evaluated by the Fisher exact test.
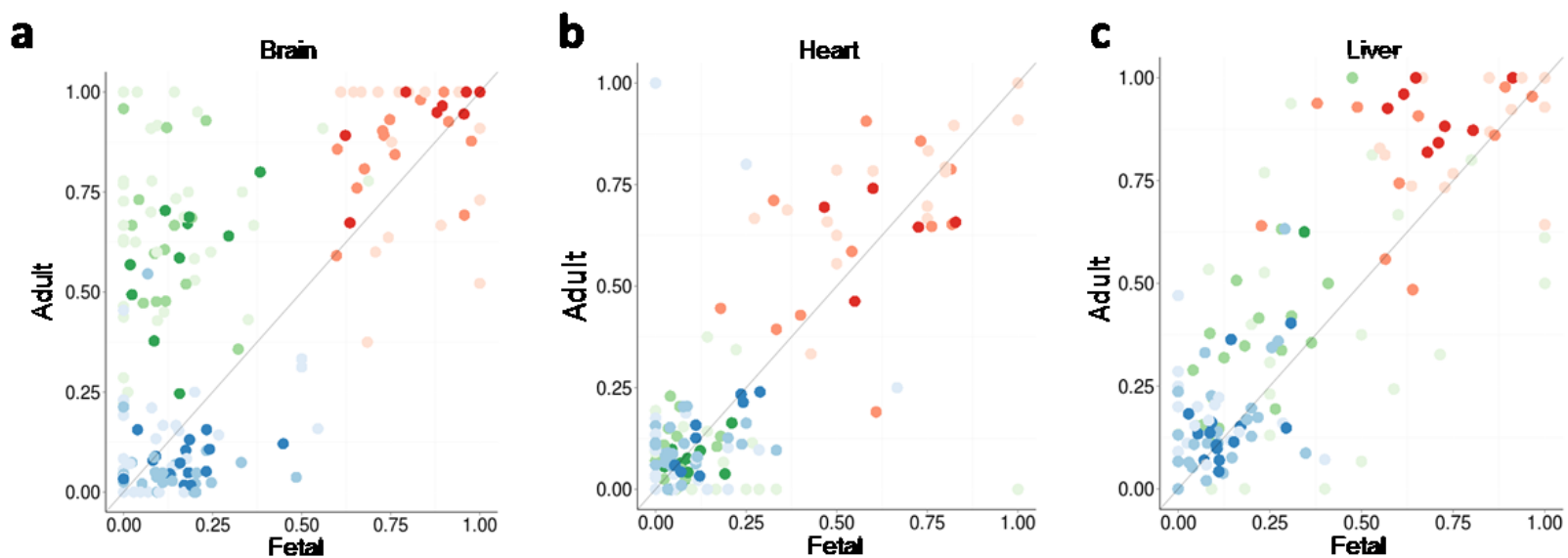
**Figure 5.1 Fetal and adult editing rates at sites in 'Group II. increasing'.** The editing rates at the sites in the groups in Figure 4.2 were compared between fetal and adult tissues for brain (a), heart (b) and liver (c). Only sites with sequencing depth greater than 20 in both fetal and adult tissues were used. Note that these brain samples are independent from the 33 discovery samples in Figure 3.4. The same colors from Figure 4.2 are used to indicate the three developmental pattern groups: 'Group I. Low' (blue), 'Group II. Increasing' (green), 'Group III. High' (red). Color shade in a,b,c and point size in e are proportional to three categories of sequencing depth: low (less than 20), medium (20 to 50), high (greater than 50).
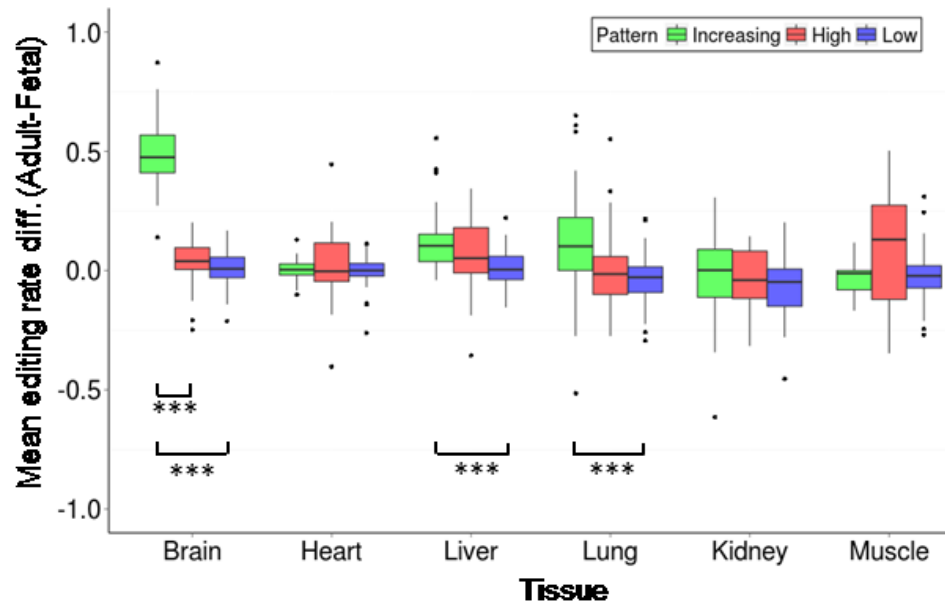
**Figure 5.2 Differences of mean editing rates between fetal and adult samples across multiple tissues.** At least three samples, except for a liver fetal tissue, are used to calculate mean editing rates for fetal and adult samples respectively. Note that 3 fetal and 3 adult samples were selected from our 33 discovery brain tissues. ANOVA reveals that brain, liver and lung show clear separation among patterns (*p-value* ≤ 0.01), though the increasing pattern is much more marked in brain as evaluated by post-hoc two-sample t-tests. *** indicates *p-value* ≤ 0.001.
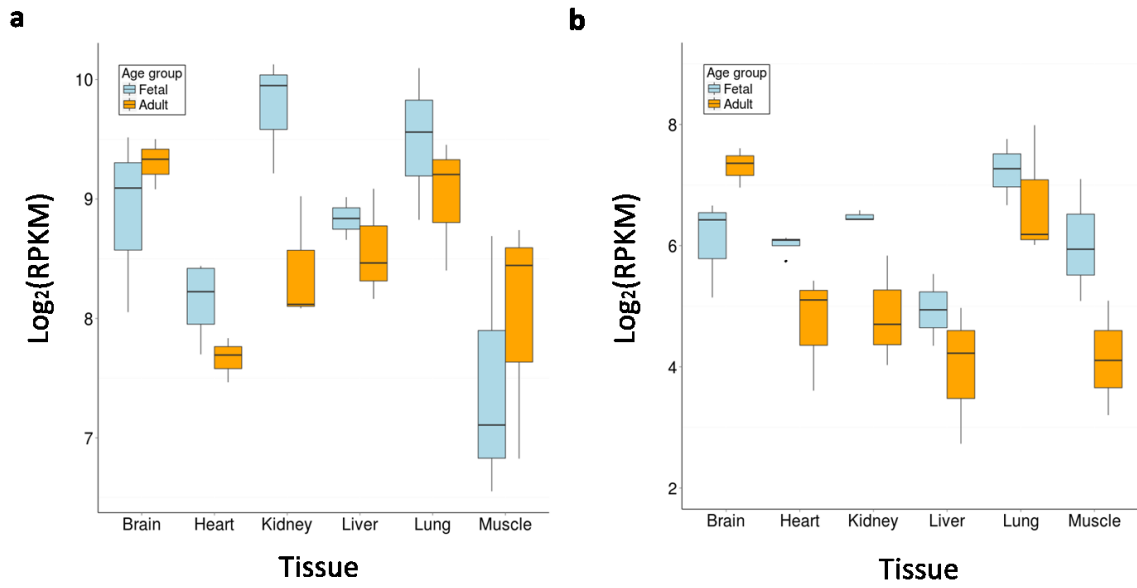
**Figure 5.3 mRNA expression levels of ADAR enzymes.** In each tissue, fetal and adult samples are compared: (a) ADAR1 (b) ADAR2. Note that boxplots are generated for visualization, even with 3 fetal and 3 adult samples in every tissue, except for adult heart (4 samples) and fetal liver (2 samples) tissues.
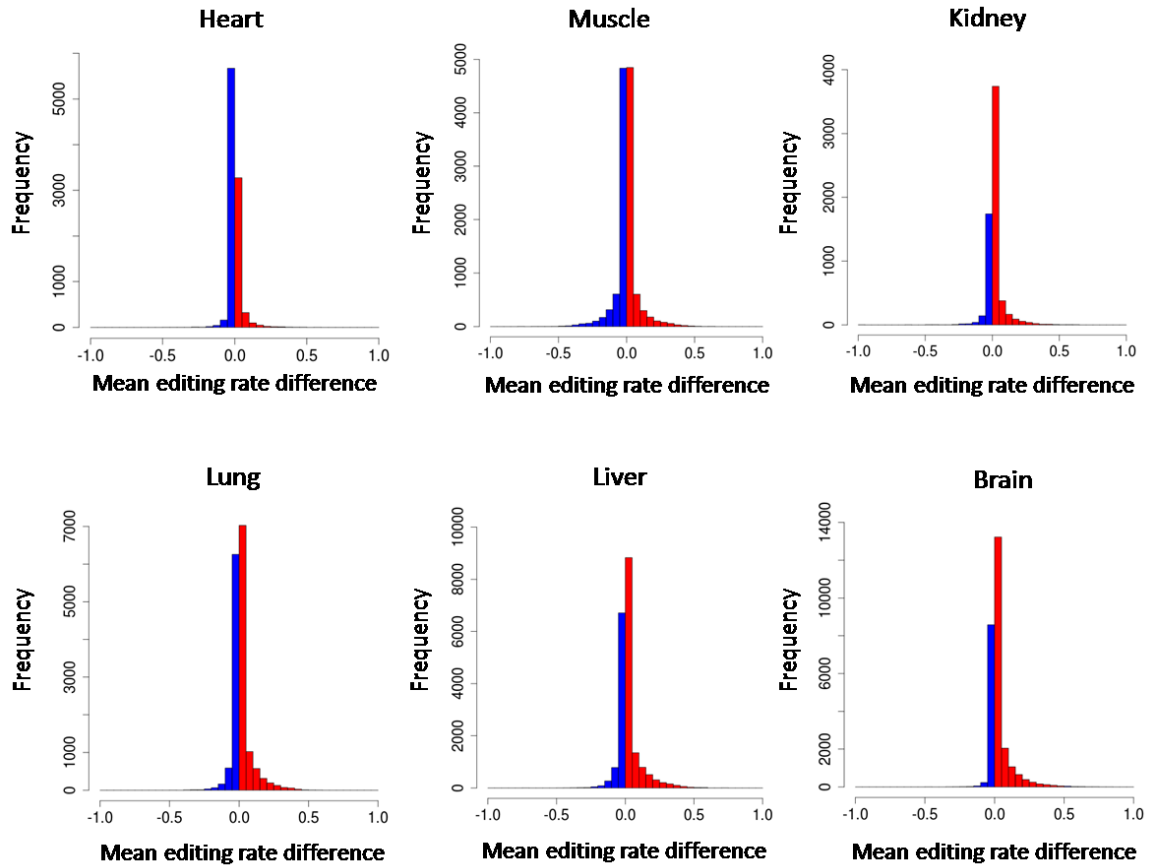
**Figure 5.4 Genome-wide editing rate differences between fetal and adult samples in multiple tissues.** (a) Histogram of mean editing rate difference between fetal and adult samples. Mean editing rate difference is defined by the mean editing rate of adult samples minus the mean editing rate of fetal samples. The blue is used to describe a site whose mean editing rate is greater in fetal compared to adult while the red indicates the opposite.
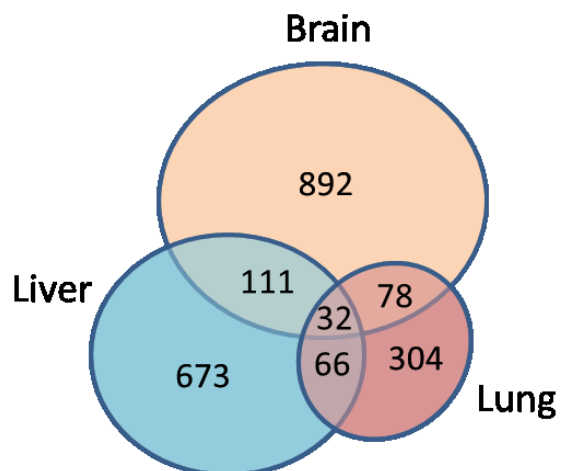
**Figure 5.5 Venn diagram showing the overlap of sites across brain, liver and lung.** Sites whose editing rate differences are greater than 0.2 are analyzed. Numbers indicate the number of editing sites.
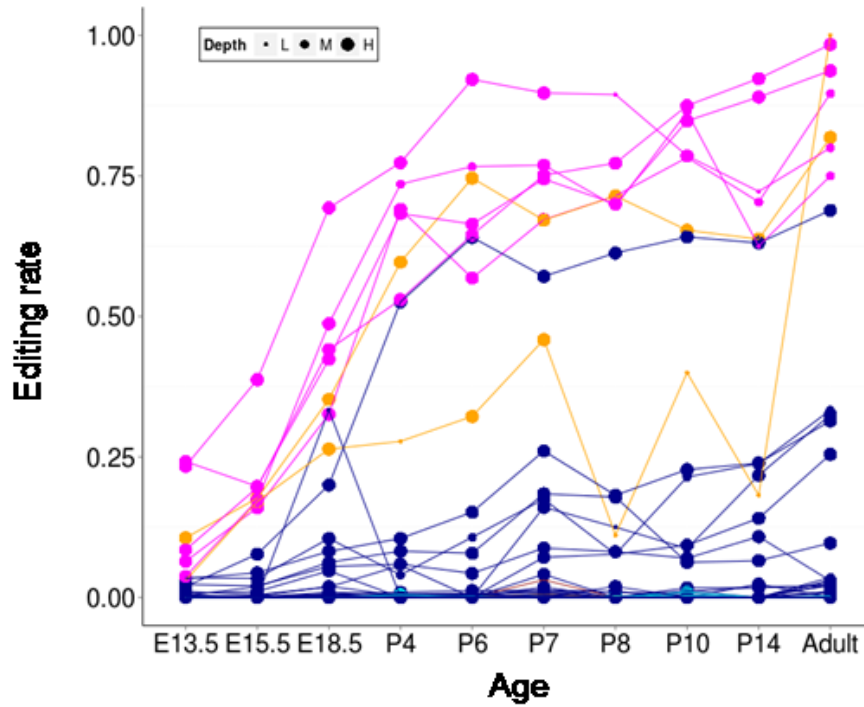
**Figure 5.6 Increasing pattern in mouse brain development.** The increasing pattern is conserved in mouse brain development (E: embryonic day, P: postnatal day in x-axis). Each line represents a conserved A-to-I editing site showing the increasing pattern in human brain development. The same colors as in Figure 3.6 are used to show conserved relative editing rates within the increasing pattern.
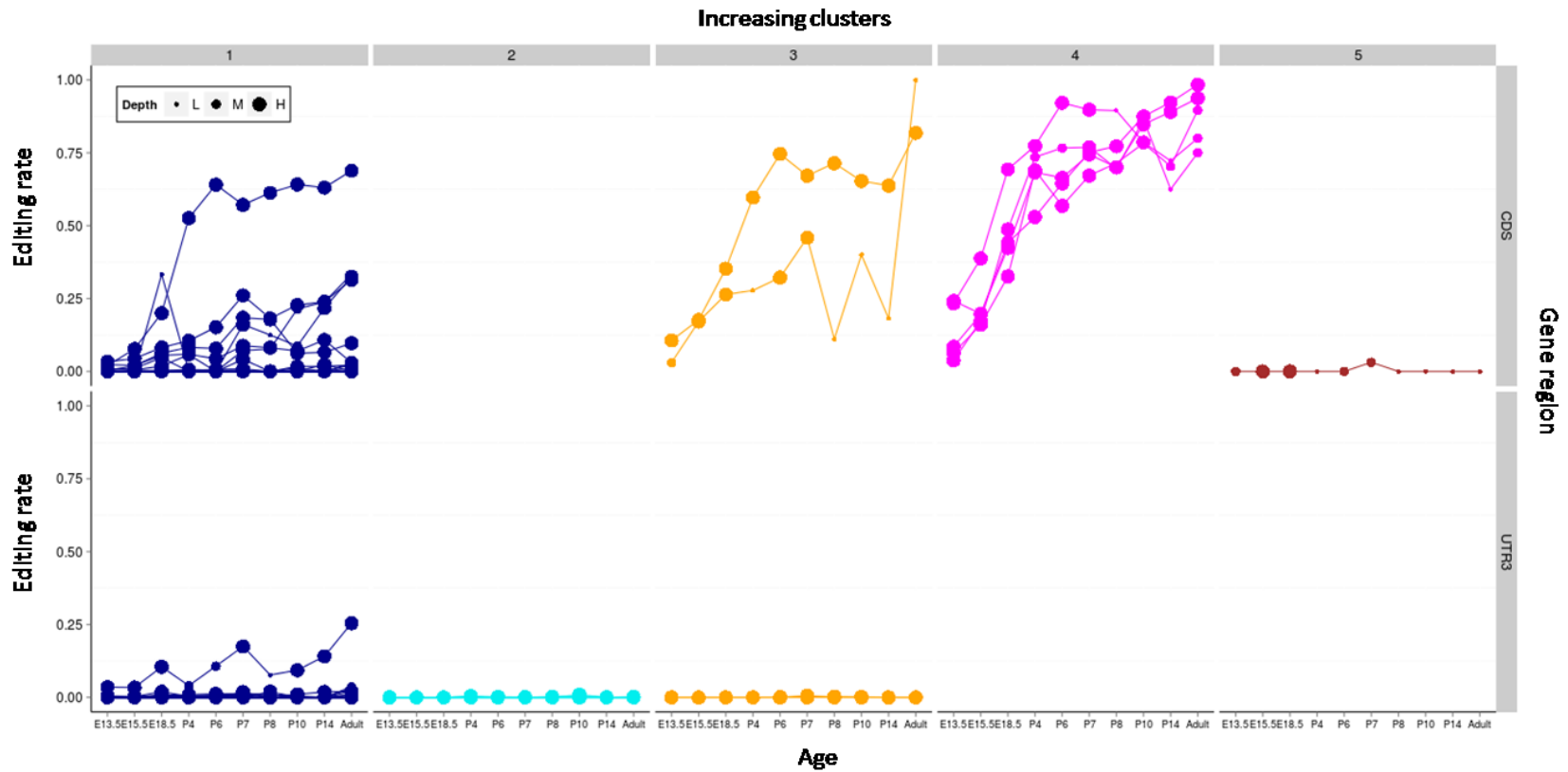
**Figure 5.7 Increasing pattern in mouse brain development according to gene regions.** The conserved A-to-I editing sites are classified according to clusters of increasing pattern (Figure 3.6) and gene regions. Most of the A-to-I editing sites in CDS regions show an increasing pattern (except for the site in NEIL1 in the cluster 5, denoted with a brown line). Dot size is proportional to three categories of sequencing depth: low (less than 20), medium (20 to 50), high (greater than 50).
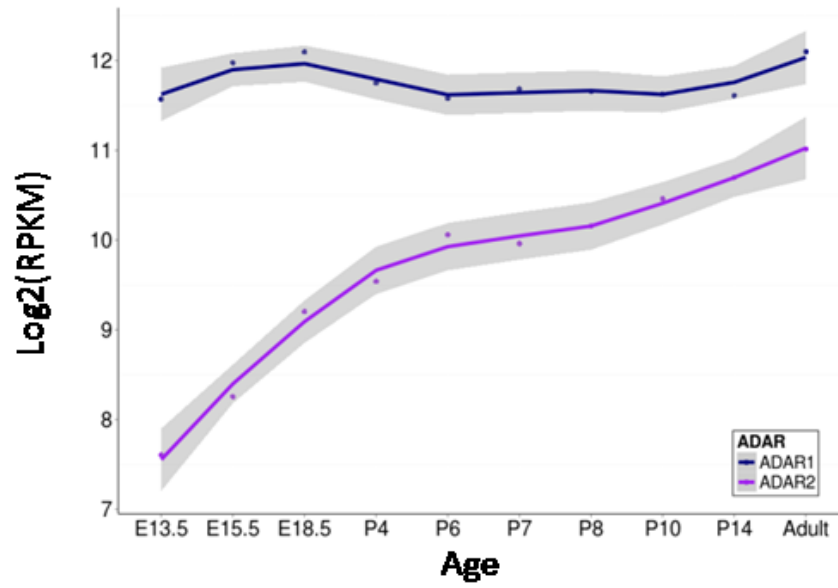
**Figure 5.8 mRNA expression levels of ADAR1 (blue) and ADAR2 (purple) in mouse brain development.** Lines are generated by locally weighted scatterplot smoothing (LOESS) regression with shades indicating 95% confidence interval.
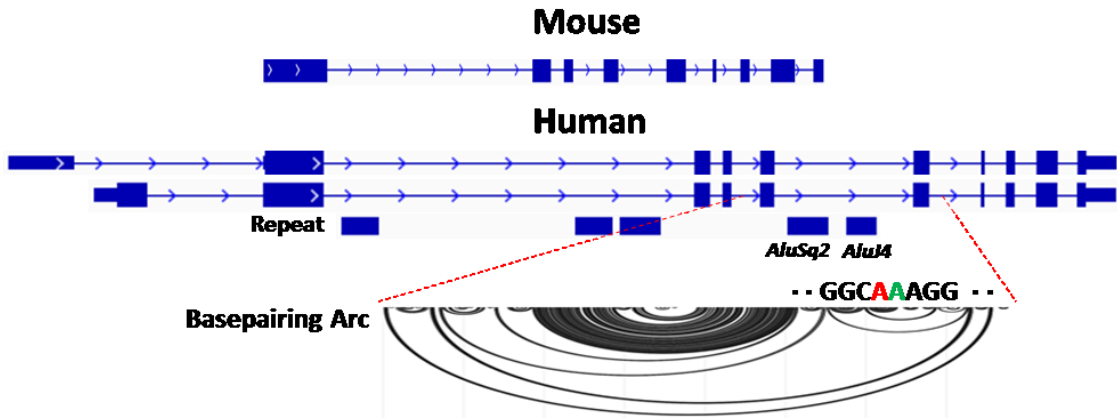
**Figure 5.9 A-to-I editing sites in NEIL1 in human brain development.** NEIL1 in the human genome has Alu repeats forming double-stranded structures near A-to-I editing sites. A double-stranded structure is depicted by arcs, which connect two base-paired nucleotides ('baseparing arcs') using an R library, R-CHIE (www.e-rna.org/r-chie). A-to-I editing sites with two different patterns are found in NEIL1: a green (increasing) and a red (stable high).
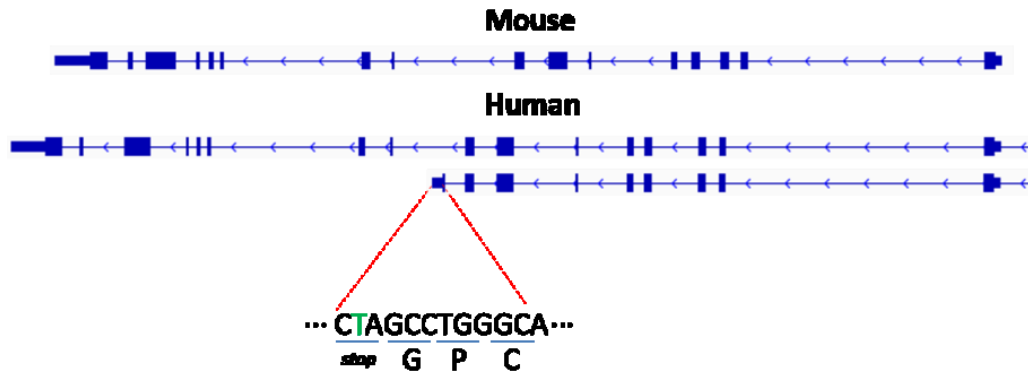
**Figure 5.10 A-to-I editing sites in PDZD7 in human brain development.** The A-to-I editing site in the PDZD7 is only found in a transcript specific to the human genome. This A-to-I editing can destroy a stop codon.
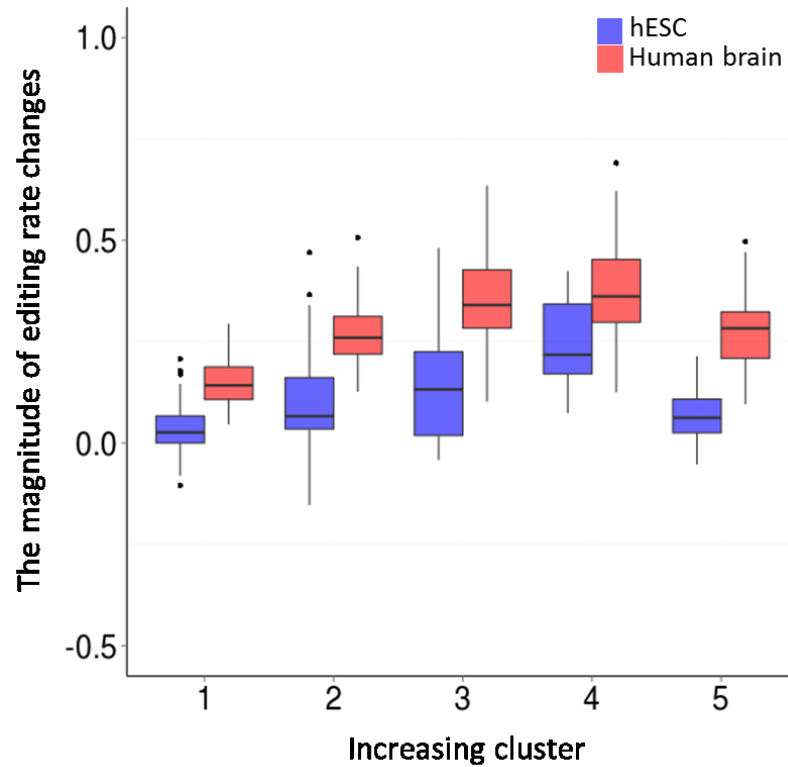
**Figure 5.11 The magnitude of editing rate changes found at increasingly-edited sites.** The magnitude of editing rate changes is defined as the mean editing rate difference between fetal and post-infant samples for human brain development while it is defined as the mean editing rate differences between samples of in-vitro day 0 and 7 and samples of in-vitro day 33 and 49 for hESC samples. The comparisons between samples in human brain development and samples in in-vitro differentiation of hESC into cortical neurons are performed separately according to the five increasing clusters defined in figures 3.4 and 3.6. The x-axis numbers indicate the clusters, which is the same as the numbers shown in the figures 3.4, 3.6 and the following figure 5.12.
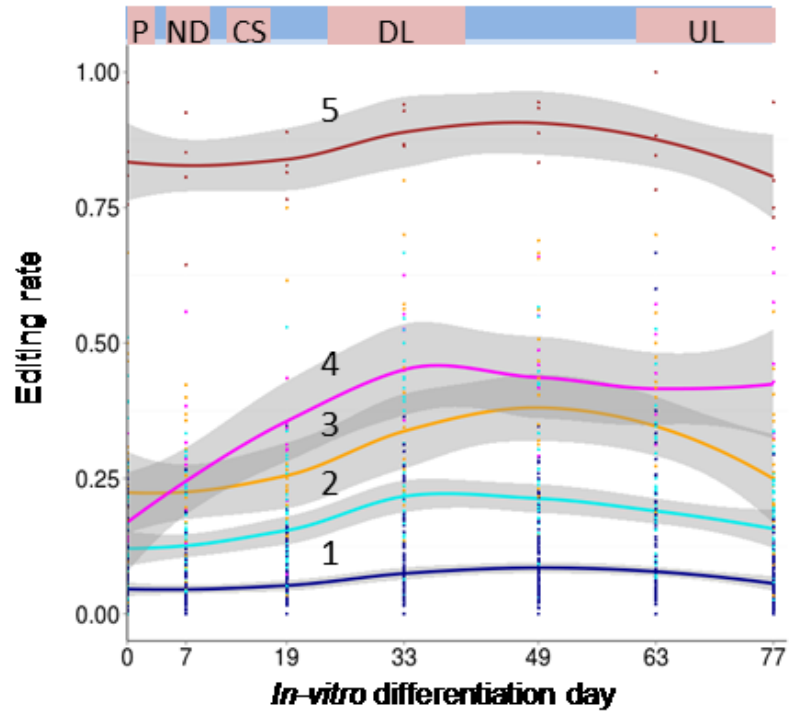
**Figure 5.12 The increasing editing pattern in the differentiation of human embryonic stem cells into cortical neurons.** Numbers are matched with Figures 3.4 and 3.6. The letters in the upper bar stands for the following (van de Leemput et al. 2014): P, pluripotency; ND, neural differentiation; CS, cortical specification; DL, deep layer; UL, upper layer. Note that colors indicate the corresponding clusters in Figure 3.6.
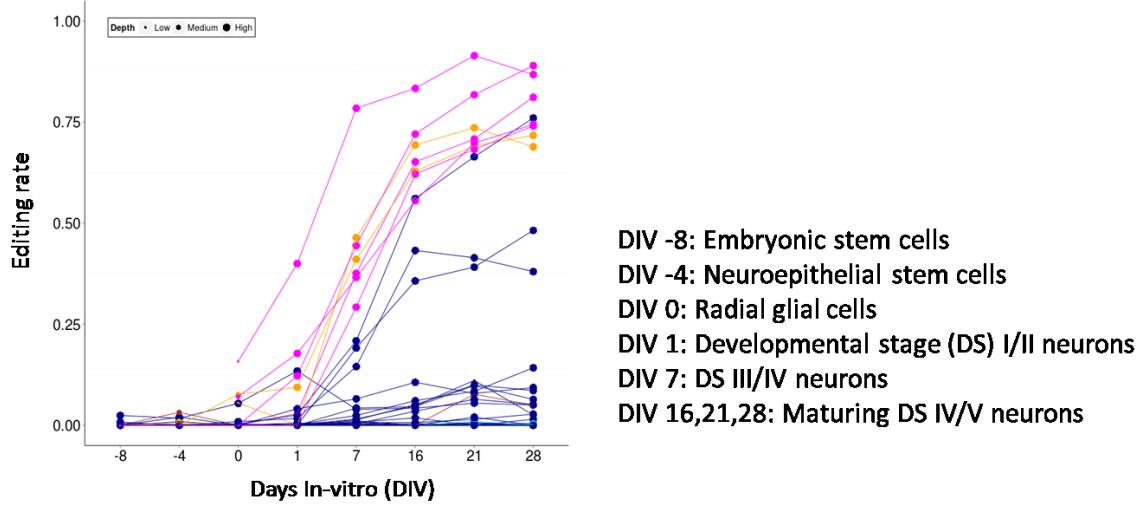
**Figure 5.13 Editing rate changes in *in-vitro* differentiation of mESCs to cortical neurons.** Different colors indicate different clusters in Figures 3.4 and 3.6 to show conserved relative editing rates within the increasing pattern. *Days In-vitro* (DIV) descriptions were adapted from (Hubbard et al. 2013). Dot size is proportional to three categories of sequencing depth: low (less than 20), medium (20 to 50), high (greater than 50).
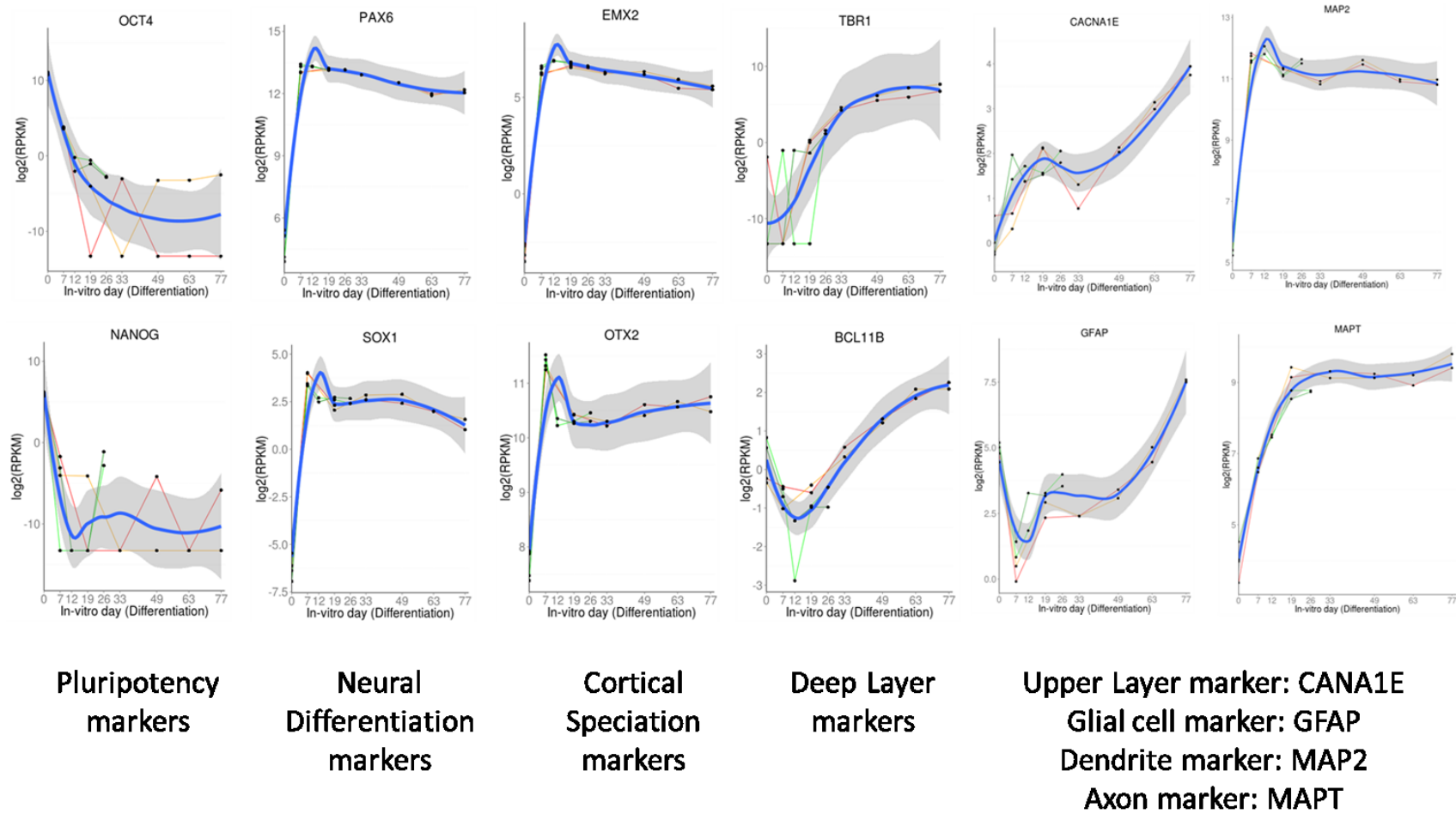
**Figure 5.14 Cellular markers in in-vitro differentiation of hESCs to cortical neurons.** Expression levels of markers are measured by RNA-seq with an RPKM unit: OCT4 and Nanog for pluripotency, PAX6 and Sox1 for neural differentiation, EMX2 and OTX2 for cortical specification,

TBR1 and BCL11B (CTIP2) for deep layer neurons, CACNE1 for upper layer neurons, GFAP for glial cells, MAP2 for dendrites, MAPT for axons. Lines are generated by LOESS regression with shades indicating 95% confidence interval.
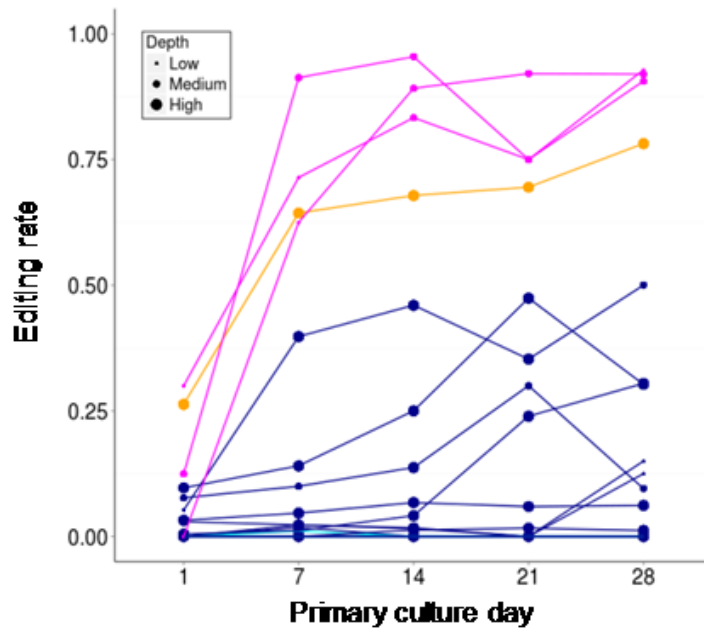
**Figure 5.15 Editing rate change during the culture of primary mouse neuron.** Only the conserved sites whose sequencing depth is sufficient to estimate editing rate reliably are depicted. Note that colors indicate the corresponding clusters in Figure 3.6. Point sizes are proportional to three categories of sequencing depth: low (less than 20), medium (20 to 50), high (greater than 50).
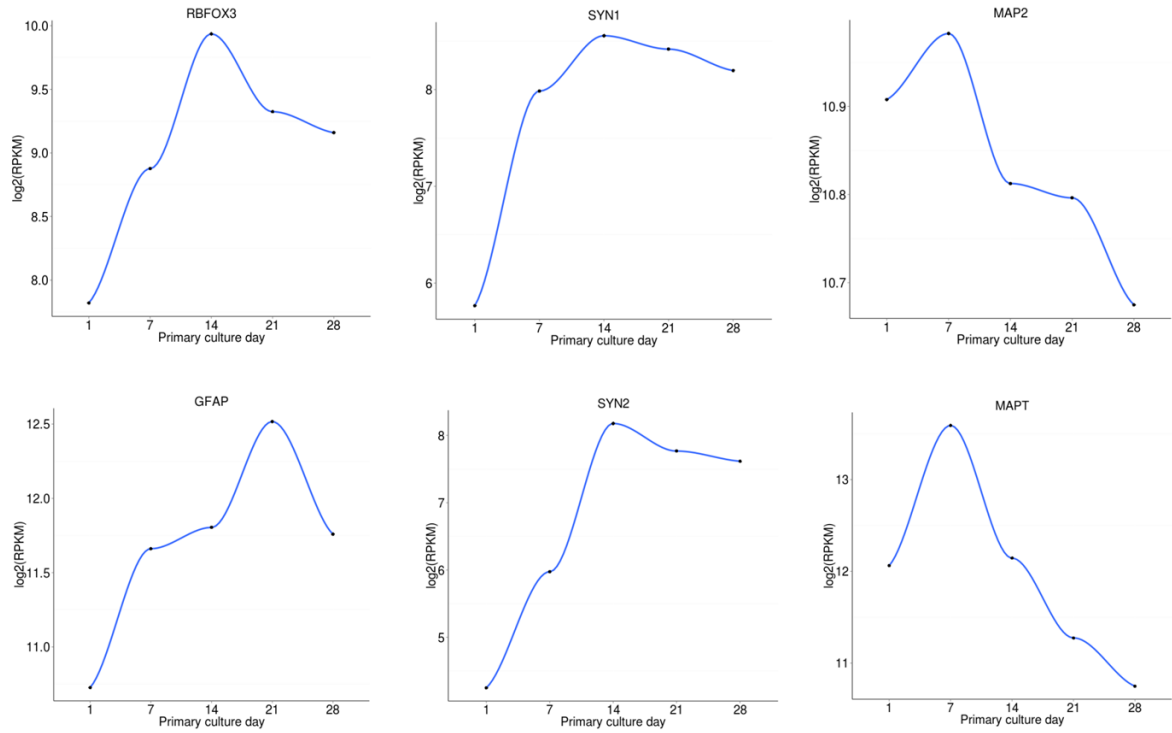
**Figure 5.16 Cellular markers in primary culture of mouse cortical neurons.** Expression levels of markers are measured by RNA-seq with an RPKM unit: Rbfox3 (NeuN) for neurons, Gfap for glial cells, Syn1 and Syn2 for synaptic formation, MAP2 for dendrites, MAPT for axons.
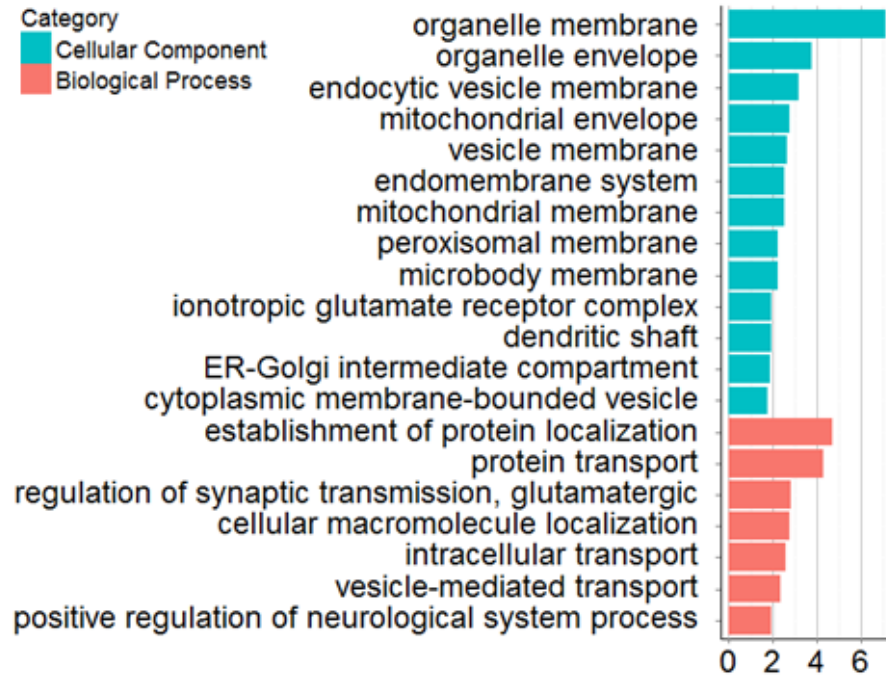
**Figure 5.17 The enriched GO terms for genes with the increasing pattern.** Two GO categories, cellular component and biological process denoted by different colors, are described.
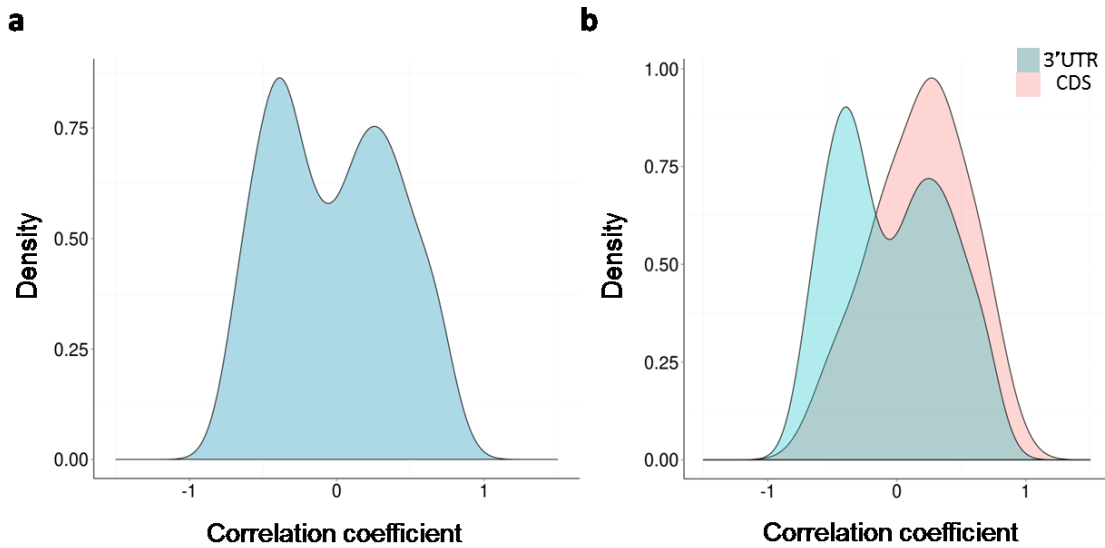
**Figure 5.18 Correlation of editing rates with mRNA expression levels.** Spearman correlation coefficient between editing rates and mRNA expression levels are calculated at the sites showing the increasing pattern. When there are multiple editing sites in a gene, the highest correlation value is assigned to a gene. (a) Distribution of correlation coefficient at the total sites with the increasing pattern (b) Distributions of correlation coefficients are separately described according to the two gene regions associated with the sites: CDS and 3' UTR.
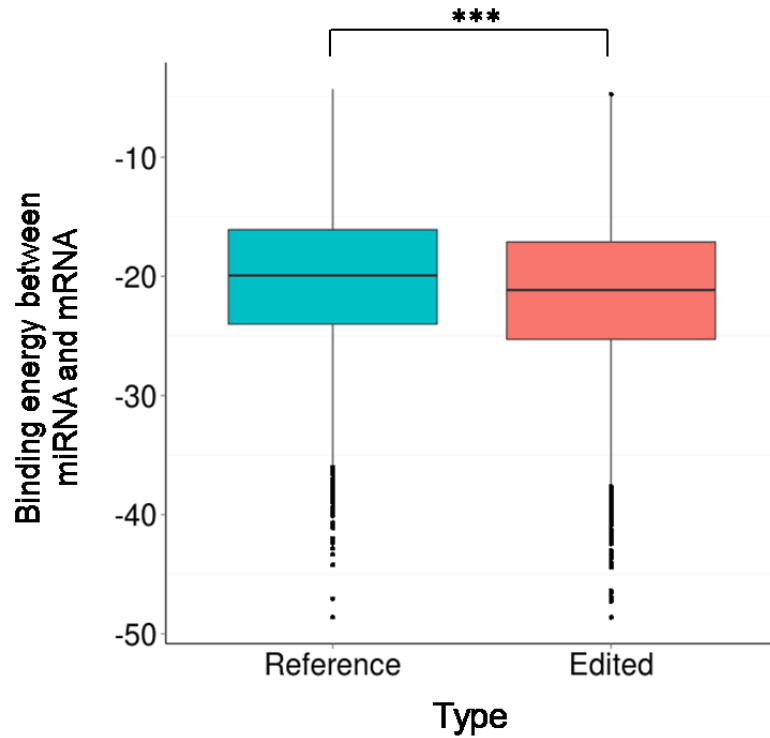
**Figure 5.19 Binding energy between miRNA and mRNA target.** Binding of miRNA to an mRNA region spanning increasingly-edited sites are computationally predicted and the associated binding energy is compared between canonical mRNA sequence ('Reference') and mRNA sequence affected by RNA editing ('Edited'). *** indicates *p-value* ≤0.001 in two-sample t-tests.
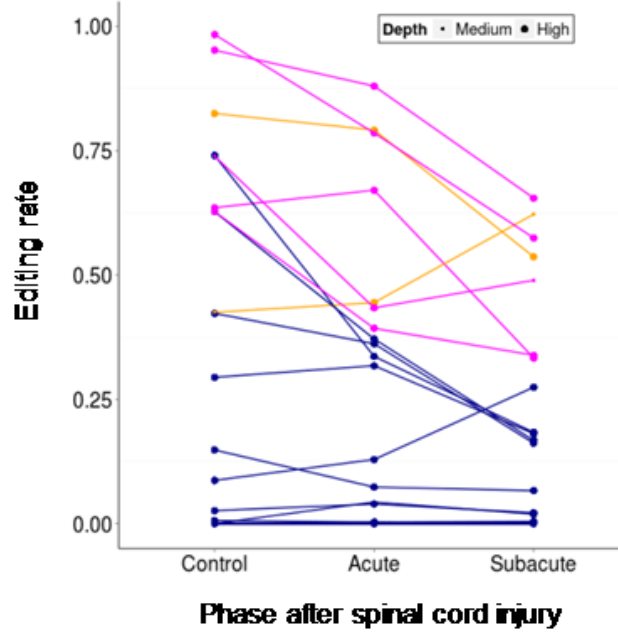
**Figure 5.20 Editing rate changes in the mouse model of spinal cord injury at the CDS-residing conserved editing sites in the increasing pattern.** Sites with enough sequencing depth are only described. Colors indicate the corresponding clusters in Figure 3.6. Point sizes are proportional to three categories of sequencing depth: low (less than 20), medium (20 to 50), high (greater than 50).
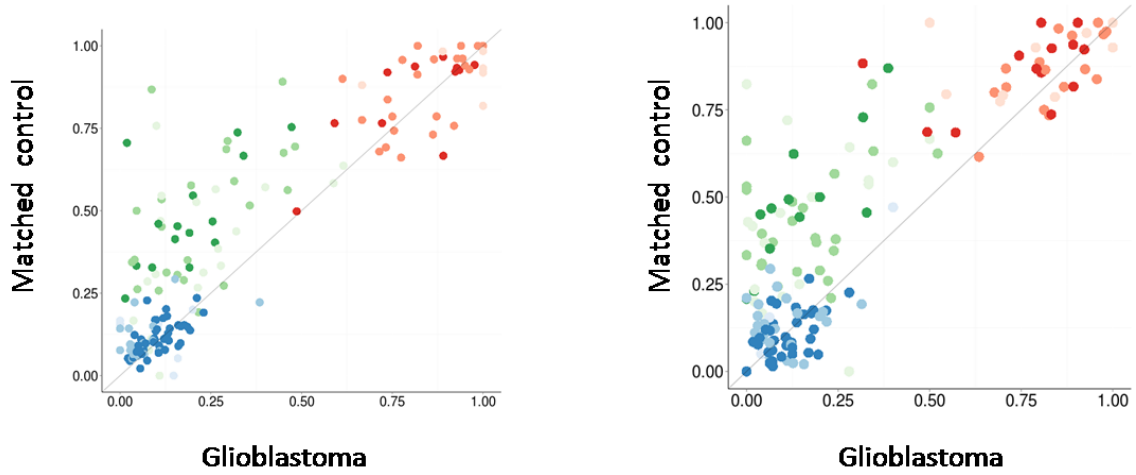
**Figure 5.21 The comparison of editing rates in the groups of selected sites between a glioblastoma and neighboring non-tumor tissue.** Each figure is generated with an independent patient. Color denote group same as Figure 4.2. Color shade is proportional to three categories of sequencing depth: low (less than 20), medium (20 to 50), high (greater than 50).

| Chromosome | Coordinate | Gene | Gene reion | Mouse genome | |
|---|---|---|---|---|---|
| | | | | Location | Reference |
| chr1 | 2436080 | PLCH2 | CDS | chr4:154984234 | G |
| chr1 | 40147856 | HPCAL4 | 3' UTR | chr4:123191224 | A |
| chr1 | 41089410 | RIMS3 | 3' UTR | chr4:120893807 | A |
| chr1 | 160185788 | DCAF8 | 3' UTR | chr1:172196141 | G |
| chr1 | 160302244 | COPA | CDS | chr1:172092348 | A |
| chr1 | 160319987 | NCSTN | CDS | chr1:172074326 | T |
| chr2 | 25381529 | EFR3B | 3' UTR | chr12:3962972 | C |
| chr2 | 73171432 | SFXN5 | 3' UTR | chr6:85215002 | G |
| chr2 | 172605884 | DYNC1I2 | 3' UTR | chr2:71264165 | A |
| chr2 | 202486541 | TMEM237 | 3' UTR | chr18:48047793 | A |
| chr3 | 42590546 | SEC22C | 3' UTR | chr9:121681599 | A |
| chr3 | 119545199 | GSK3B | 3' UTR | chr16:38241020 | A |
| chr3 | 179093028 | MFN1 | CDS | chr3:32561485 | A |
| chr4 | 77979680 | CCNI | CDS | chr5:93189584 | T |
| chr4 | 158257879 | GRIA2 | CDS | chr3:80706908 | T |
| chr4 | 158281294 | GRIA2 | CDS | chr3:80692286 | T |
| chr5 | 156904831 | ADAM19 | 3' UTR | chr11:46146839 | A |
| chr5 | 156904833 | ADAM19 | 3' UTR | chr11:46146837 | A |
| chr5 | 156904853 | ADAM19 | 3' UTR | chr11:46146821 | A |
| chr5 | 156904922 | ADAM19 | 3' UTR | chr11:46146777 | A |
| chr5 | 156904929 | ADAM19 | 3' UTR | chr11:46146770 | T |
| chr5 | 156904947 | ADAM19 | 3' UTR | chr11:46146752 | A |
| chr5 | 156904950 | ADAM19 | 3' UTR | chr11:46146749 | A |
| chr5 | 156904965 | ADAM19 | 3' UTR | chr11:46146734 | A |
| chr5 | 156905396 | ADAM19 | 3' UTR | chr11:46146102 | A |
| chr5 | 156905397 | ADAM19 | 3' UTR | chr11:46146101 | A |
| chr5 | 156905398 | ADAM19 | 3' UTR | chr11:46146100 | A |
| chr5 | 156905411 | ADAM19 | 3' UTR | chr11:46146089 | A |
| chr5 | 156905556 | ADAM19 | 3' UTR | chr11:46145955 | T |
| chr5 | 156905560 | ADAM19 | 3' UTR | chr11:46145951 | G |
| chr6 | 44120349 | TMEM63B | CDS | chr17:45662949 | T |
| chr6 | 102337702 | GRIK2 | CDS | chr10:49272776 | T |
| chr6 | 102372572 | GRIK2 | CDS | chr10:49244347 | T |
| chr6 | 102372589 | GRIK2 | CDS | chr10:49244330 | T |
| chr7 | 25160004 | CYCS | 3' UTR | chr6:50562746 | T |
| chr8 | 9639522 | TNKS | 3' UTR | chr8:34826795 | T |
| chr8 | 12886333 | KIAA1456 | 3' UTR | chr8:36518849 | A |
| chr9 | 127712691 | SCAI | 3' UTR | chr2:39073337 | T |
| chr9 | 136229572 | SURF4 | 3' UTR | chr2:26921169 | T |

| chr9 | 139335555 | SEC16A | 3' UTR | chr2:26410275 | T |
|---|---|---|---|---|---|
| chr9 | 139335597 | SEC16A | 3' UTR | chr2:26410306 | C |
| chr9 | 139335623 | SEC16A | 3' UTR | chr2:26410329 | C |
| chr9 | 139335882 | SEC16A | 3' UTR | chr2:26410574 | C |
| chr10 | 102121601 | SCD | 3' UTR | chr19:44304213 | G |
| chr11 | 68523904 | CPT1A | 3' UTR | chr19:3384947 | T |
| chr11 | 68524809 | CPT1A | 3' UTR | chr19:3384135 | G |
| chr11 | 75316685 | MAP6 | 3' UTR | chr7:99317823 | G |
| chr11 | 75316752 | MAP6 | 3' UTR | chr7:99317759 | A |
| chr11 | 75316759 | MAP6 | 3' UTR | chr7:99317752 | A |
| chr11 | 82868441 | PCF11 | 5' UTR | chr7:92669661 | T |
| chr11 | 105804694 | GRIA4 | CDS | chr9:4456006 | T |
| chr11 | 119532688 | PVRL1 | 3' UTR | chr9:43806689 | A |
| chr12 | 5021742 | KCNA1 | CDS | chr6:126642158 | T |
| chr12 | 48736610 | ZNF641 | 3' UTR | chr15:98288379 | T |
| chr12 | 56633548 | ANKRD52 | 3' UTR | chr10:128394072 | A |
| chr12 | 120531332 | CCDC64 | 3' UTR | chr5:115649059 | A |
| chr12 | 132407137 | ULK1 | 3' UTR | chr5:110785005 | T |
| chr14 | 20920211 | OSGEP | CDS | chr14:50919694 | T |
| chr14 | 26917530 | NOVA1 | CDS | chr12:46700334 | T |
| chr15 | 40710778 | IVD | 3' UTR | chr2:118880849 | G |
| chr15 | 65425334 | PDCD7 | CDS | chr9:65346924 | G |
| chr15 | 75646087 | NEIL1 | CDS | chr9:57144307 | T |
| chr15 | 89739330 | ABHD2 | 3' UTR | chr7:79360823 | A |
| chr17 | 3763779 | CAMKK1 | 3' UTR | chr11:73041884 | G |
| chr17 | 29862338 | RAB11FIP4 | 3' UTR | chr11:79695576 | C |
| chr17 | 37826529 | PNMT | CDS | chr11:98387978 | G |
| chr17 | 42154831 | HDAC5 | 3' UTR | chr11:102195028 | T |
| chr17 | 42154907 | HDAC5 | 3' UTR | chr11:102195105 | C |
| chr17 | 42475813 | GPATCH8 | CDS | chr11:102479070 | T |
| chr17 | 56049824 | VEZF1 | 3' UTR | chr11:88083835 | A |
| chr17 | 56049829 | VEZF1 | 3' UTR | chr11:88083830 | G |
| chr17 | 79780692 | FAM195B | 3' UTR | chr11:120543288 | A |
| chr18 | 51061986 | DCC | 3' UTR | chr18:71253914 | T |
| chr18 | 51061990 | DCC | 3' UTR | chr18:71253910 | C |
| chr19 | 14593605 | GIPC1 | CDS | chr8:83661176 | A |
| chr19 | 14593693 | GIPC1 | CDS | chr8:83661088 | T |
| chr19 | 18892789 | CRTC1 | 3' UTR | chr8:70382669 | T |
| chr19 | 38887763 | SPRED3 | 3' UTR | chr7:29160661 | T |
| chr19 | 38888055 | SPRED3 | 3' UTR | chr7:29160385 | C |
| chr19 | 38888069 | SPRED3 | 3' UTR | chr7:29160371 | T |

| chr19 | 38888095 | SPRED3 | 3' UTR | chr7:29160345 | T |
|-------|----------|--------|--------|---------------|---|
| chr19 | 38889352 | SPRED3 | 3' UTR | chr7:29159681 | T |
| chr19 | 54485579 | CACNG8 | CDS | chr7:3415084 | A |
| chr20 | 2128469 | STK35 | 3' UTR | chr2:129831568 | G |
| chr20 | 5175539 | CDS2 | 3' UTR | chr2:132309649 | A |
| chr20 | 5175805 | CDS2 | 3' UTR | chr2:132309873 | A |
| chr20 | 44802643 | CDH22 | 3' UTR | chr2:165111769 | A |
| chr20 | 44802728 | CDH22 | 3' UTR | chr2:165111853 | T |
| chr21 | 30953750 | GRIK1 | CDS | chr16:87940543 | T |
| chr22 | 37765609 | ELFN2 | 3' UTR | chr15:78668715 | G |
| chrX | 54955567 | TRO | CDS | chrX:150648527 | T |
| chrX | 84346885 | APOOL | 3' UTR | chrX:112384042 | A |
| chrX | 84346892 | APOOL | 3' UTR | chrX:112384049 | A |
| chrX | 122598962 | GRIA3 | CDS | chrX:41654252 | A |
| chrX | 151358319 | GABRA3 | CDS | chrX:72445292 | T |

**Table 5.1 The conserved sites between human and mouse among the sites with the increasing pattern.** Among 742 A-to-I editing sites showing the increasing pattern in human brain, 95 sites are found to be conserved in the mouse genome (UCSC mm10).

**Table 5.2 Increasing pattern in mouse brain development.** See appendix 2. Site: corresponding location of A-to-I editing sites found in human brain tissues in UCSC mouse genome mm10; Increasing cluster (Clu.): an index of cluster described in figures 3.4 and 3.6; Gene and gene region (Reg.): gene annotation of A-to-I editing sites based on RefSeq; Editing rate (ER) and Depth: editing rate with sequencing depth at a given site are shown according to embryonic (E) day or postnatal (P) day.

| Category | ID | Term | Count | Percentage | p-value | FDR | Genes |
|---|---|---|---|---|---|---|---|
| GOTERM_BP_FAT | GO:0031646 | establishment of protein localization | 25 | 11.31 | 2.02E-05 | 0.033 | COPA, GRIK2, SNX1, VPS53, GIPC1, TIMM50, ZFYVE20, HOOK3, SEC16A, SEC22C, PEX13, TNKS, RAB2B, VPS41, AP4S1, RAB11FIP4, PHAX, GRIA2, RAB36, GSK3B, PEX26, GGA2, NUP43, SRP9, SSR3 |
| GOTERM_BP_FAT | GO:0051971 | protein localization | 27 | 12.22 | 2.36E-05 | 0.038 | COPA, GRIK2, SNX1, VPS53, GIPC1, TIMM50, ZFYVE20, HOOK3, SEC16A, SEC22C, PEX13, TNKS, COX18, RAB2B, VPS41, AP4S1, RAB11FIP4, PHAX, GRIA2, ULK1, RAB36, GSK3B, PEX26, GGA2, NUP43, SRP9, SSR3 |
| GOTERM_BP_FAT | GO:0050806 | protein transport | 24 | 10.86 | 5.13E-05 | 0.083 | RAB2B, COPA, GRIK2, VPS53, SNX1, GIPC1, VPS41, TIMM50, ZFYVE20, AP4S1, HOOK3, RAB11FIP4, PHAX, RAB36, GSK3B, SEC16A, PEX26, SEC22C, PEX13, TNKS, GGA2, NUP43, SRP9, SSR3 |
| GOTERM_BP_FAT | GO:0006886 | regulation of synaptic transmission, glutamatergic | 4 | 1.81 | 0.001495514 | 2.384 | GRIA2, GRIK1, GRIK2, GRIA4 |
| GOTERM_BP_FAT | GO:0016192 | cellular protein localization | 14 | 6.33 | 0.0015849 5 | 2.525 | COPA, GRIK2, SNX1, GIPC1, VPS41, AP4S1, GSK3B, PEX26, PEX13, TNKS, COX18, GGA2, SRP9, SSR3 |
| GOTERM_BP_FAT | GO:0007215 | cellular macromolecule localization | 14 | 6.33 | 0.0016905 91 | 2.691 | COPA, GRIK2, SNX1, GIPC1, VPS41, AP4S1, GSK3B, PEX26, PEX13, TNKS, COX18, GGA2, SRP9, SSR3 |
| GOTERM_BP_FAT | GO:0046907 | intracellular transport | 18 | 8.14 | 0.0027618 31 | 4.361 | COPA, GRIK2, SNX1, GIPC1, VPS41, ZFYVE20, AP4S1, TAPBP, HOOK3, PHAX, GSK3B, PEX26, PEX13, SEC22C, GGA2, GOLGA3, SRP9, SSR3 |

125

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| GOTERM_BP_FAT | GO:0070727 | glutamate signaling pathway | 4 | 1.81 | 0.00336452 6 | 5.289 | GRIK1, GRIK2, GRIA3, GRIA4 |
| GOTERM_BP_FAT | GO:0034613 | vesicle-mediated transport | 16 | 7.24 | 0.00459710 4 | 7.160 | RAB2B, COPA, CRCP, SNX1, VPS41, ZFYVE20, AP4S1, TAPBP, HOOK3, RIMS3, GRIA2, ULK1, SEC16A, SEC22C, GGA2, GOLGA3 |
| GOTERM_BP_FAT | GO:0051966 | intracellular protein transport | 12 | 5.43 | 0.00617209 9 | 9.500 | COPA, GRIK2, GSK3B, PEX26, SNX1, VPS41, PEX13, GIPC1, AP4S1, GGA2, SRP9, SSR3 |
| GOTERM_BP_FAT | GO:0015031 | positive regulation of synaptic transmission | 4 | 1.81 | 0.00808043 3 | 12.261 | GRIA2, GRIK1, GRIK2, GRIA4 |
| GOTERM_BP_FAT | GO:0008104 | positive regulation of transmission of nerve impulse | 4 | 1.81 | 0.01021694 7 | 15.259 | GRIA2, GRIK1, GRIK2, GRIA4 |
| GOTERM_BP_FAT | GO:0045184 | positive regulation of neurological system process | 4 | 1.81 | 0.01180633 6 | 17.427 | GRIA2, GRIK1, GRIK2, GRIA4 |
| GOTERM_CC_FAT | GO:0031090 | organelle membrane | 36 | 16.29 | 7.75E-08 | 0.000 | MAVS, ACOX1, COPA, APOOL, MTDH, TIMM50, GIPC1, SFXN5, HOOK3, TAPBP, CDS2, ATP5S, PEX13, TNKS, COX18, GOLGA3, ACSL6, RAB2B, CACNG8, SCD, GRIA3, VPS41, GRIA4, SYNJ2BP, NDUFA10, CPT1A, NCSTN, MFN1, GRIA2, ULK1, L2HGDH, PEX26, MDM2, TMPO, SRP9, SSR3 |
| GOTERM_CC_FAT | GO:0031967 | organelle envelope | 20 | 9.05 | 1.81E-04 | 0.239 | MAVS, APOOL, MTDH, SCD, CYCS, TIMM50, SYNJ2BP, NDUFA10, CPT1A, SFXN5, CBX5, CDS2, MFN1, L2HGDH, ATP5S, TNKS, TMPO, COX18, NUP43, ACSL6 |
| GOTERM_ | GO:0031975 | envelope | 20 | 9.05 | 1.88E-04 | 0.249 | MAVS, APOOL, MTDH, SCD, CYCS, |

| CC_FAT | | | | | | | TIMM50, SYNJ2BP, NDUFA10, CPT1A, SFXN5, CBX5, CDS2, MFN1, L2HGDH, ATP5S, TNKS, TMPO, COX18, NUP43, ACSL6 |
|---|---|---|---|---|---|---|---|
| GOTERM_CC_FAT | GO:0030666 | endocytic vesicle membrane | 5 | 2.26 | 6.47E-04 | 0.855 | GRIA2, CACNG8, MDM2, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0030139 | endocytic vesicle | 6 | 2.71 | 7.26E-04 | 0.959 | GRIA2, CACNG8, MDM2, GIPC1, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0005829 | cytosol | 30 | 13.57 | 0.00124027 | 1.632 | COPA, XIAP, PNMT, GNE, RPL13, GIPC1, GSR, PGPEP1, SORBS1, PSMB2, EEF2K, GUCY1A3, RPS20, INPP5B, SMAD9, VHL, CYCS, EIF2S3, VPS41, RIC8B, TPMT, PHAX, PKNOX1, ULK1, PPIA, GSK3B, MTR, MDM2, APAF1, DYNC1I2 |
| GOTERM_CC_FAT | GO:0005740 | mitochondrial envelope | 14 | 6.33 | 0.00176424 1 | 2.314 | MAVS, APOOL, CYCS, TIMM50, SYNJ2BP, NDUFA10, CPT1A, SFXN5, CDS2, MFN1, L2HGDH, ATP5S, COX18, ACSL6 |
| GOTERM_CC_FAT | GO:0044429 | mitochondrial part | 17 | 7.69 | 0.00233576 2 | 3.053 | MAVS, APOOL, PDP2, CYCS, TIMM50, SYNJ2BP, NDUFA10, CPT1A, SFXN5, CDS2, MFN1, L2HGDH, IVD, PPM1K, ATP5S, COX18, ACSL6 |
| GOTERM_CC_FAT | GO:0012506 | vesicle membrane | 8 | 3.62 | 0.00241113 1 | 3.150 | COPA, GRIA2, CACNG8, ULK1, MDM2, GIPC1, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0012505 | endomembrane system | 20 | 9.05 | 0.00291114 6 | 3.792 | RAB2B, COPA, MTDH, CACNG8, SCD, GIPC1, GRIA3, GRIA4, AP4S1, CBX5, TAPBP, GRIA2, ULK1, MDM2, TNKS, TMPO, NUP43, GOLGA3, SRP9, SSR3 |
| GOTERM_CC_FAT | GO:0031966 | mitochondrial membrane | 13 | 5.88 | 0.00306975 8 | 3.994 | MAVS, APOOL, TIMM50, SYNJ2BP, NDUFA10, CPT1A, SFXN5, CDS2, |

| | | | | | | MFN1, L2HGDH, ATP5S, COX18, ACSL6 |
|---|---|---|---|---|---|---|
| GOTERM_CC_FAT | GO:0005739 | mitochondrion | 24 | 10.86 | 0.0059924 92 | 7.660 | MAVS, ACOX1, APOOL, PDP2, CXORF23, VHL, CYCS, AASS, TIMM50, SYNJ2BP, NDUFA10, MRPL30, CPT1A, SFXN5, CDS2, GSR, MFN1, L2HGDH, IVD, PPM1K, ATP5S, CTSB, COX18, ACSL6 |
| GOTERM_CC_FAT | GO:0031903 | microbody membrane | 4 | 1.81 | 0.0061294 88 | 7.828 | ACOX1, PEX26, PEX13, ACSL6 |
| GOTERM_CC_FAT | GO:0005778 | peroxisomal membrane | 4 | 1.81 | 0.0061294 88 | 7.828 | ACOX1, PEX26, PEX13, ACSL6 |
| GOTERM_CC_FAT | GO:0030659 | cytoplasmic vesicle membrane | 7 | 3.17 | 0.0069121 95 | 8.786 | COPA, GRIA2, CACNG8, ULK1, MDM2, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0008328 | ionotropic glutamate receptor complex | 3 | 1.36 | 0.0120833 89 | 14.886 | GRIK1, GRIK2, GRIA4 |
| GOTERM_CC_FAT | GO:0043198 | dendritic shaft | 3 | 1.36 | 0.0120833 89 | 14.886 | GSK3B, GIPC1, GRIA3 |
| GOTERM_CC_FAT | GO:0005793 | ER-Golgi intermediate compartment | 4 | 1.81 | 0.0133120 79 | 16.279 | GNPNAT1, SURF4, UGGT1, GOLGA3 |
| GOTERM_CC_FAT | GO:0044439 | peroxisomal part | 4 | 1.81 | 0.0151434 66 | 18.316 | ACOX1, PEX26, PEX13, ACSL6 |
| GOTERM_CC_FAT | GO:0044438 | microbody part | 4 | 1.81 | 0.0151434 66 | 18.316 | ACOX1, PEX26, PEX13, ACSL6 |
| GOTERM_CC_FAT | GO:0016023 | cytoplasmic membrane-bounded vesicle | 14 | 6.33 | 0.0163378 05 | 19.620 | TGOLN2, COPA, CACNG8, GIPC1, VPS41, GRIA3, GRIA4, NCSTN, SPAG9, GRIA2, ULK1, SLC30A4, MDM2, CTSB |

**Table 5.3 Gene ontology (GO) terms associated with genes with increasing editing pattern.** Category: GO category, GOTERM_BP_FAT for biological process and GOTERM_CC_FAT for cellular component; ID: GO ID; Term: GO term; Count: the number of genes associated with a given term; Percentage: The proportion of genes with increasing pattern in a given term; p-value and FDR (False Discovery Rate): based on Fisher-exact test with DAVID tool (https://david.ncifcrf.gov/summary.jsp); Genes: list of genes with increasing pattern associated with a given term.

| Category | ID | Term | Count | Percentage | p-value | FDR | Genes |
|---|---|---|---|---|---|---|---|
| GOTERM_BP_FAT | GO:0008104 | protein localization | 23 | 11.62 | 1.56E-04 | 0.247 | RAB2B, AP3S2, VPS53, SNX1, VPS41, TIMM50, ZFYVE20, AP4S1, HOOK3, RAB11FIP4, PHAX, RAB36, ULK1, GSK3B, SEC16A, PEX26, SEC22C, PEX13, TNKS, COX18, GGA2, NUP43, SSR3 |
| GOTERM_BP_FAT | GO:0015031 | protein transport | 21 | 10.61 | 1.64E-04 | 0.259 | RAB2B, AP3S2, VPS53, SNX1, VPS41, TIMM50, ZFYVE20, AP4S1, HOOK3, RAB11FIP4, PHAX, RAB36, GSK3B, SEC16A, PEX26, SEC22C, PEX13, TNKS, GGA2, NUP43, SSR3 |
| GOTERM_BP_FAT | GO:0045184 | establishment of protein localization | 21 | 10.61 | 1.85E-04 | 0.292 | RAB2B, AP3S2, VPS53, SNX1, VPS41, TIMM50, ZFYVE20, AP4S1, HOOK3, RAB11FIP4, PHAX, RAB36, GSK3B, SEC16A, PEX26, SEC22C, PEX13, TNKS, GGA2, NUP43, SSR3 |
| GOTERM_BP_FAT | GO:0016192 | vesicle-mediated transport | 15 | 7.58 | 0.003338 | 5.154 | RAB2B, AP3S2, CRCP, SNX1, VPS41, ZFYVE20, AP4S1, TAPBP, HOOK3, RIMS3, ULK1, SEC16A, SEC22C, GGA2, GOLGA3 |
| GOTERM_BP_FAT | GO:0046907 | intracellular transport | 15 | 7.58 | 0.01031 | 15.126 | AP3S2, SNX1, VPS41, ZFYVE20, AP4S1, TAPBP, HOOK3, PHAX, GSK3B, PEX26, PEX13, SEC22C, GGA2, GOLGA3, SSR3 |
| GOTERM_BP_FAT | GO:0034613 | cellular protein localization | 11 | 5.56 | 0.012491 | 18.039 | GSK3B, PEX26, AP3S2, SNX1, VPS41, TNKS, PEX13, AP4S1, COX18, GGA2, SSR3 |
| GOTERM_BP_FAT | GO:0070727 | cellular macromolecule localization | 11 | 5.56 | 0.013099 | 18.822 | GSK3B, PEX26, AP3S2, SNX1, VPS41, TNKS, PEX13, AP4S1, COX18, GGA2, SSR3 |
| GOTERM_CC_FAT | GO:0031090 | organelle membrane | 28 | 14.14 | 2.89E-05 | 0.037 | MAVS, APOOL, ACOX1, MTDH, TIMM50, SFXN5, HOOK3, TAPBP, CDS2, ATP5S, PEX13, TNKS, COX18, GOLGA3, ACSL6, RAB2B, CACNG8, SCD, VPS41, SYNJ2BP, NDUFA10, CPT1A, ULK1, L2HGDH, PEX26, MDM2, TMPO, SSR3 |
| GOTERM_CC_F | GO:00319 | organelle | 19 | 9.60 | 9.93E- | 0.128 | MAVS, APOOL, MTDH, SCD, CYCS, TIMM50, |

| | | | | | | |
|---|---|---|---|---|---|---|
| AT | 67 | envelope | | | 05 | | SYNJ2BP, NDUFA10, CPT1A, SFXN5, CBX5, CDS2, L2HGDH, ATP5S, TNKS, TMPO, COX18, NUP43, ACSL6 |
| GOTERM_CC_FAT | GO:0031975 | envelope | 19 | 9.60 | 1.03E-04 | 0.133 | MAVS, APOOL, MTDH, SCD, CYCS, TIMM50, SYNJ2BP, NDUFA10, CPT1A, SFXN5, CBX5, CDS2, L2HGDH, ATP5S, TNKS, TMPO, COX18, NUP43, ACSL6 |
| GOTERM_CC_FAT | GO:0044429 | mitochondrial part | 16 | 8.08 | 0.001647 | 2.106 | MAVS, APOOL, PDP2, CYCS, TIMM50, SYNJ2BP, NDUFA10, CPT1A, SFXN5, CDS2, L2HGDH, IVD, PPM1K, ATP5S, COX18, ACSL6 |
| GOTERM_CC_FAT | GO:0005740 | mitochondrial envelope | 13 | 6.57 | 0.001706 | 2.181 | MAVS, CDS2, APOOL, L2HGDH, ATP5S, CYCS, TIMM50, SYNJ2BP, NDUFA10, COX18, ACSL6, SFXN5, CPT1A |
| GOTERM_CC_FAT | GO:0005739 | mitochondrion | 23 | 11.62 | 0.002456 | 3.127 | MAVS, ACOX1, APOOL, PDP2, CXORF23, VHL, CYCS, AASS, TIMM50, SYNJ2BP, NDUFA10, MRPL30, CPT1A, SFXN5, CDS2, GSR, L2HGDH, IVD, PPM1K, ATP5S, CTSB, COX18, ACSL6 |
| GOTERM_CC_FAT | GO:0031966 | mitochondrial membrane | 12 | 6.06 | 0.003223 | 4.084 | MAVS, CDS2, APOOL, L2HGDH, ATP5S, TIMM50, SYNJ2BP, NDUFA10, COX18, SFXN5, CPT1A, ACSL6 |
| GOTERM_CC_FAT | GO:0005829 | cytosol | 26 | 13.13 | 0.003322 | 4.207 | XIAP, GNE, RPL13, GSR, PGPEP1, PSMB2, EEF2K, GUCY1A3, RPS20, INPP5B, SMAD9, VHL, CYCS, EIF2S3, VPS41, RIC8B, TPMT, PHAX, PKNOX1, ULK1, PPIA, GSK3B, MTR, MDM2, APAF1, DYNC1I2 |
| GOTERM_CC_FAT | GO:0031903 | microbody membrane | 4 | 2.02 | 0.004259 | 5.364 | ACOX1, PEX26, PEX13, ACSL6 |
| GOTERM_CC_FAT | GO:0005778 | peroxisomal membrane | 4 | 2.02 | 0.004259 | 5.364 | ACOX1, PEX26, PEX13, ACSL6 |
| GOTERM_CC_FAT | GO:0005793 | ER-Golgi intermediate compartment | 4 | 2.02 | 0.009351 | 11.427 | GNPNAT1, SURF4, UGGT1, GOLGA3 |
| GOTERM_CC_FAT | GO:0044438 | microbody part | 4 | 2.02 | 0.01066 | 12.927 | ACOX1, PEX26, PEX13, ACSL6 |

131

| GOTERM_CC_FAT | GO:0044439 | peroxisomal part | 4 | 2.02 | 0.01066 | 12.927 | ACOX1, PEX26, PEX13, ACSL6 |
|---|---|---|---|---|---|---|---|
| GOTERM_CC_FAT | GO:0005794 | Golgi apparatus | 18 | 9.09 | 0.011081 | 13.404 | TGOLN2, RAB2B, MGAT4A, GNPNAT1, AP3S2, VPS53, SNX1, VPS41, AP4S1, TAPBP, HOOK3, NMT2, RAB36, SEC16A, TNKS, MAP6, GGA2, GOLGA3 |
| GOTERM_CC_FAT | GO:0005743 | mitochondrial inner membrane | 9 | 4.55 | 0.016547 | 19.387 | CDS2, APOOL, L2HGDH, ATP5S, TIMM50, NDUFA10, COX18, SFXN5, CPT1A |

**Table 5.4 Gene ontology (GO) terms associated with genes with increasing editing pattern in 3' UTR.** The description of columns is same as table 5.3.

| Category | ID | Term | Count | Percentage | p-value | FDR | Genes |
|---|---|---|---|---|---|---|---|
| GOTERM_BP_FAT | GO:0007268 | synaptic transmission | 8 | 2.27 | 3.01E-07 | 0.000 | GRIA2, GRIK1, GRIK2, GABRA3, KCNA1, GIPC1, GRIA4, NOVA1 |
| GOTERM_BP_FAT | GO:0019226 | transmission of nerve impulse | 8 | 2.27 | 8.92E-07 | 0.001 | GRIA2, GRIK1, GRIK2, GABRA3, KCNA1, GIPC1, GRIA4, NOVA1 |
| GOTERM_BP_FAT | GO:0051966 | regulation of synaptic transmission, glutamatergic | 4 | 1.13 | 3.56E-06 | 0.005 | GRIA2, GRIK1, GRIK2, GRIA4 |
| GOTERM_BP_FAT | GO:0007215 | glutamate signaling pathway | 4 | 1.13 | 8.39E-06 | 0.012 | GRIK1, GRIK2, GRIA3, GRIA4 |
| GOTERM_BP_FAT | GO:0050806 | positive regulation of synaptic transmission | 4 | 1.13 | 2.16E-05 | 0.030 | GRIA2, GRIK1, GRIK2, GRIA4 |
| GOTERM_BP_FAT | GO:0051971 | positive regulation of transmission of nerve impulse | 4 | 1.13 | 2.80E-05 | 0.038 | GRIA2, GRIK1, GRIK2, GRIA4 |
| GOTERM_BP_FAT | GO:0007267 | cell-cell signaling | 8 | 2.27 | 3.10E-05 | 0.043 | GRIA2, GRIK1, GRIK2, GABRA3, KCNA1, GIPC1, GRIA4, NOVA1 |
| GOTERM_BP_FAT | GO:0031646 | positive regulation of neurological system process | 4 | 1.13 | 3.28E-05 | 0.045 | GRIA2, GRIK1, GRIK2, GRIA4 |
| GOTERM_BP_FAT | GO:0050804 | regulation of synaptic transmission | 5 | 1.42 | 6.21E-05 | 0.085 | GRIA2, GRIK1, GRIK2, GIPC1, GRIA4 |
| GOTERM_BP_FAT | GO:0051969 | regulation of transmission of nerve impulse | 5 | 1.42 | 8.41E-05 | 0.115 | GRIA2, GRIK1, GRIK2, GIPC1, GRIA4 |
| GOTERM_BP_FAT | GO:0031644 | regulation of | 5 | 1.42 | 9.82E-05 | 0.135 | GRIA2, GRIK1, GRIK2, GIPC1, |

| | | neurological system process | | | | | GRIA4 |
|---|---|---|---|---|---|---|---|
| GOTERM_CC_FAT | GO:0044456 | synapse part | 7 | 1.98 | 1.48E-06 | 0.002 | GRIA2, GRIK1, GRIK2, GABRA3, GIPC1, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0045211 | postsynaptic membrane | 6 | 1.70 | 1.67E-06 | 0.002 | GRIA2, GRIK1, GRIK2, GABRA3, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0030139 | endocytic vesicle | 5 | 1.42 | 1.88E-06 | 0.002 | GRIA2, CACNG8, GIPC1, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0012506 | vesicle membrane | 6 | 1.70 | 2.89E-06 | 0.003 | COPA, GRIA2, CACNG8, GIPC1, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0045202 | synapse | 7 | 1.98 | 1.23E-05 | 0.015 | GRIA2, GRIK1, GRIK2, GABRA3, GIPC1, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0030666 | endocytic vesicle membrane | 4 | 1.13 | 1.73E-05 | 0.021 | GRIA2, CACNG8, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0043235 | receptor complex | 5 | 1.42 | 2.79E-05 | 0.033 | SORBS1, GRIK1, GRIK2, GRIA4, SRP9 |
| GOTERM_CC_FAT | GO:0030659 | cytoplasmic vesicle membrane | 5 | 1.42 | 5.67E-05 | 0.067 | COPA, GRIA2, CACNG8, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0030054 | cell junction | 7 | 1.98 | 1.03E-04 | 0.121 | GRIA2, SORBS1, GRIK1, GRIK2, GABRA3, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0030425 | dendrite | 5 | 1.42 | 1.05E-04 | 0.125 | GRIK1, GRIK2, GIPC1, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0031090 | organelle membrane | 9 | 2.55 | 1.41E-04 | 0.166 | NCSTN, MFN1, COPA, GRIA2, CACNG8, GIPC1, GRIA3, GRIA4, SRP9 |
| GOTERM_CC_FAT | GO:0044459 | plasma membrane part | 12 | 3.40 | 1.42E-04 | 0.168 | NCSTN, COPA, GRIA2, SORBS1, GRIK1, CACNG8, GRIK2, GABRA3, TRO, KCNA1, GRIA3, GRIA4 |
| GOTERM_CC_FAT | GO:0016023 | cytoplasmic membrane-bounded vesicle | 7 | 1.98 | 1.43E-04 | 0.169 | NCSTN, COPA, GRIA2, CACNG8, GIPC1, GRIA3, GRIA4 |

**Table 5.5 Gene ontology (GO) terms associated with genes with increasing editing pattern in CDS.** The description of columns is same as table 5.3.

| Chromosome | Coordinate | Gene | CDS | | |
| | | | Effect | Description | Nonsynonymous Prediction |
|---|---|---|---|---|---|
| chr1 | 2436080 | PLCH2 | nonsynonymous | R->G | Neutral |
| chr1 | 160302244 | COPA | nonsynonymous | I->V | Neutral |
| chr1 | 160319987 | NCSTN | nonsynonymous | S->G | Neutral |
| chr1 | 225974581 | SRP9 | synonymous | L->L | NA |
| chr3 | 179093028 | MFN1 | synonymous | A->A | NA |
| chr4 | 77979680 | CCNI | nonsynonymous | R->G | Deleterious |
| chr4 | 158257879 | GRIA2 | synonymous | Q->Q | NA |
| chr4 | 158281294 | GRIA2 | nonsynonymous | R->G | Neutral |
| chr6 | 44120349 | TMEM63B | nonsynonymous | Q->R | Deleterious |
| chr6 | 102337702 | GRIK2 | nonsynonymous | Y->C | Neutral |
| chr6 | 102372572 | GRIK2 | synonymous | G->G | NA |
| chr6 | 102372589 | GRIK2 | nonsynonymous | Q->R | Neutral |
| chr10 | 102777342 | PDZD7 | stoploss | X->W | NA |
| chr11 | 105804694 | GRIA4 | nonsynonymous | R->G | Neutral |
| chr12 | 5021742 | KCNA1 | nonsynonymous | I->V | Deleterious |
| chr14 | 20920211 | OSGEP | nonsynonymous | I->M | Deleterious |
| chr14 | 26917530 | NOVA1 | nonsynonymous | S->G | Neutral |
| chr15 | 65425334 | PDCD7 | synonymous | A->A | NA |
| chr15 | 75646087 | NEIL1 | synonymous | K->K | NA |
| chr17 | 37826529 | PNMT | nonsynonymous | S->G | Neutral |
| chr17 | 42475813 | GPATCH8 | nonsynonymous | K->R | Deleterious |
| chr19 | 14593605 | GIPC1 | nonsynonymous | T->A | Neutral |
| chr19 | 14593693 | GIPC1 | synonymous | P->P | NA |
| chr19 | 54485579 | CACNG8 | nonsynonymous | S->G | Neutral |
| chr21 | 30953750 | GRIK1 | nonsynonymous | Q->R | Neutral |
| chrX | 54955567 | TRO | nonsynonymous | S->G | Deleterious |
| chrX | 122598962 | GRIA3 | nonsynonymous | R->G | Neutral |
| chrX | 151358319 | GABRA3 | nonsynonymous | I->M | Neutral |

**Table 5.6 Effect of A-to-I editing sites in CDS region.** Chromosome and Coordinate: location of A-to-I editing sites in UCSC human genome hg19; Gene: gene annotation based on RefSeq; Effect: synonymous or nonsynonymous; Description and Nonsynonymous prediction: amino acid change by A-to-I editing and its predicted effect.

| Disease | Gene set size | Genomic background (N=23368) | | Brain-expressed (N=17244) | | Overlapped genes |
|---|---|---|---|---|---|---|
| | | Number of overlap | p-value | Number of overlap | p-value | |
| ASD | 408 | 7 | 0.134 | 7 | 0.368 | MTR, CCDC64, DCTN5, CDH22, GSK3B, GRIK2, MCPH1 |
| SCZ | 722 | 4 | 0.336 | 4 | 0.088 | SLC4A8, CBX5, GFOD2, PLCH2 |
| BPAD | 123 | 3 | 0.127 | 3 | 0.214 | MGAT4A, MRPL30, PTPRT |
| ID | 88 | 3 | 0.059 | 3 | 0.119 | AP4S1, GRIA3, GRIK2 |
| T2D | 66 | 1 | NA | 1 | NA | AP3S2 |
| AD | 42 | 0 | NA | 0 | NA | |
| PD | 41 | 0 | NA | 0 | NA | |

**Table 5.7 Disease association of genes with increasing editing pattern.** Enrichment of genes with the increasing pattern in previously-defined gene sets associated with various neurodevelopmental disorders including autism (ASD), schizophrenia (SCZ), biplolar (BPAD), intellectual disability (ID) and disorders thought not to be neurodevelopmental: type 2 diabetes (T2D), Alzheimer disease (AD) and Parkinson disease (PD). Brain-expressed genes indicate genes whose RPKM are greater than 1 at least one samples in the 33 samples. p-value is obtained with Fisher-exact test.

| Disease-associated genes | | | | RNA editing site showing increasing pattern | |
|---|---|---|---|---|---|
| Gene | Description | Disease | Genetic evidence | | |
| MTR | 5-methyltetrahydrofolate-homocysteine methyltransferase | ASD DATABASE | Functional | chr1 | 237066314 |
| CCDC64 | coiled-coil domain containing 64 | ASD DATABASE | Genetic Association | chr12 | 120531332 |
| DCTN5 | dynactin 5 (p25) | ASD DATABASE | Rare Single Gene variant | chr16 | 23684186 |
| | | | | chr16 | 23684193 |
| | | | | chr16 | 23684265 |
| | | | | chr16 | 23684625 |
| | | | | chr16 | 23684636 |
| CDH22 | cadherin-like 22 | ASD DATABASE | Genetic Association | chr20 | 44802643 |
| | | | | chr20 | 44802728 |
| GSK3B | glycogen synthase kinase 3 beta | ASD DATABASE | Functional | chr3 | 119545199 |
| GRIK2 | glutamate receptor, ionotropic, kainate 2 | ASD DATABASE, ID | Genetic Association | chr6 | 102337702 |
| | | ASD DATABASE, ID | Genetic Association | chr6 | 102372572 |
| | | ASD DATABASE, ID | Genetic Association | chr6 | 102372589 |
| MCPH1 | microcephalin 1 | ASD DATABASE | Rare Single Gene variant | chr8 | 6500879 |
| SLC4A8 | solute carrier family 4, sodium bicarbonate cotransporter, member 8 | SCZ SNV | Exome sequencing | chr12 | 51903940 |
| | | | | chr12 | 51905425 |
| | | | | chr12 | 51907308 |
| CBX5 | chromobox homolog 5 (HP1 alpha homolog, Drosophila) | SCZ SNV | Exome sequencing | chr12 | 54634284 |
| GFOD2 | glucose-Fructose Oxidoreductase Domain Containing 2 | SCZ PGC2 | GWAS | chr16 | 67715404 |
| | | | | chr16 | 67715585 |
| | | | | chr16 | 67715890 |

| Gene | Description | Disease | Genetic evidence | Chr | Position |
|---|---|---|---|---|---|
| PLCH2 | phospholipase C, Eta 2 | SCZ PGC2 | GWAS | chr1 | 2436080 |
| MGAT4A | mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isozyme A | BPAD GWAS | GWAS | chr2 | 99239382 |
| MRPL30 | mitochondrial ribosomal protein L30 | BPAD GWAS | GWAS | chr2 | 99812336 |
| | | | | chr2 | 99812429 |
| | | | | chr2 | 99812815 |
| | | | | chr2 | 99812897 |
| | | | | chr2 | 99812978 |
| | | | | chr2 | 99812988 |
| | | | | chr2 | 99813003 |
| | | | | chr2 | 99813015 |
| | | | | chr2 | 99813020 |
| PTPRT | protein tyrosine phosphatase, receptor type, T | BPAD GWAS | GWAS | chr20 | 40705281 |
| AP4S1 | adaptor-related protein complex 4, sigma 1 subunit | ID | NA | chr14 | 31563916 |
| | | | | chr14 | 31564294 |
| | | | | chr14 | 31564650 |
| | | | | chr14 | 31564700 |
| | | | | chr14 | 31564752 |
| | | | | chr14 | 31565032 |
| GRIA3 | glutamate receptor, ionotrophic, AMPA 3 | ID | NA | chrX | 122598962 |
| AP3S2 | adaptor-Related Protein Complex 3 Subunit Sigma-2 | Type 2 Diabete | NA | chr15 | 90375419 |
| | | | | chr15 | 90375513 |

**Table 5.8 Genes involved in neurodevelopmental disorders with increasing A-to-I editing patterns.** Detail description of genes associated with diseases. Disease and Genetic evidence is obtained from (Birnbaum et al. 2014).

| Section | Description | Reference | Accession |
|---|---|---|---|
| 5.1 Tissue variation | Brain: adult (77 year old) | Illumina BodyMap2 | ERR030882, ERR030890 |
| | Kidney: adult (60 year old) | | ERR030885, ERR030893 |
| | Lung: adult (65 year old) | | ERR030879, ERR030896 |
| | Skeletal muscle: adult (77 year old) | | ERR030876, ERR030899 |
| | Brain (PFC): fetal (22 week) | ENCODE | ENCLB181ZZZ |
| | Heart: adult (34 year) | | ENCLB172ZZZ |
| | Heart: fetal (28 weak) | | ENCLB183ZZZ |
| | Kidney (metanephros): fetal (24 week) | | ENCLB186ZZZ |
| | Liver: adult (32 year) | | ENCLB174ZZZ |
| | Liver: fetal (20 week) | | ENCLB187ZZZ |
| | Lung: fetal (20 week) | | ENCLB189ZZZ |
| | Skeletal muscle: fetal (22 week) | | ENCLB196ZZZ |
| 5.2 Mouse brain development | In-vivo mouse E13.5 (pooled) | GSE49581 | GSM1202240, GSM1202241 |
| | In-vivo mouse E15.5 (pooled) | | GSM1202244, GSM1202245 |
| | In-vivo mouse E18.5 (pooled) | | GSM1202232, GSM1202233 |
| | In-vivo mouse P7 (pooled) | | GSM1202249, GSM 1202250 |
| | In-vivo mouse P4 (pooled) | SRP031888 | SRR1016207,SRR1016210,SRR1016215,SRR1016218,SRR1016221,SRR1016224 |
| | In-vivo mouse P6 (pooled) | | SRR1016227,SRR1016230,SRR1016232,SRR1016234,SRR1016236,SRR1016238 |
| | In-vivo mouse P8 (pooled) | | SRR1016245,SRR1016247,SRR1016258,SRR1016263,SRR1016268,SRR1016271 |
| | In-vivo mouse P10 (pooled) | | SRR1016273,SRR1016341,SRR1016343,SRR1016345,SRR1016347,SRR1016355 |
| | In-vivo mouse P14 (poolded) | | SRR1016362,SRR1016364,SRR1016366,SRR1016368,SRR1016370,SRR1016372 |
| | In-vivo mouse Adult (pooled) | | SRR1016375,SRR1016380,SRR1016382,SRR1016384,SRR1016386,SRR1016388 |
| 5.3 Cellular understanding | In-vitro differentiation of hESC to cortical neuron | GSE56796 | All |
| | In-vitro differentiation of mESC to cortical neuron | PRJNA185305 | All |
| | Mouse primary cortical neuron culture day1 | GSE65926 | SRR1993667 |
| | Mouse primary cortical neuron culture day7 | | SRR1993668 |
| | Mouse primary cortical neuron | | SRR1993669 |

| | culture day14 | | |
|---|---|---|---|
| | Mouse primary cortical neuron culture day21 | | SRR1993670 |
| | Mouse primary cortical neuron culture day28 | | SRR1993671 |
| 5.5 Disease | Mouse model of acute spinal cord injury, CTRL (pooled) | GSE45376 | SRR789190,SRR789191 |
| | Mouse model of acute spinal cord injury, 2day (pooled) | | SRR789193,SRR789194,SRR789195 |
| | Mouse model of acute spinal cord injury, 7day (pooled) | | SRR789196,SRR789198,SRR922121 |

**Table 5.9 Previous RNA-seq datasets.** Section: section in the chapter 5; Description: summary for previous samples; Reference and Accession: database identifier for accessibility.

# Chapter 6

# Conclusions

This study reveals systematic and dynamic aspects of RNA editing in human brain across development and in two disease states. Further studies of the molecular mechanisms underlying dynamic RNA editing will give new insights into the regulation of sequence diversity in the nervous system that is not genomically encoded.

The recent expansion of RNA sequencing datasets has led to the identification of a huge number of RNA editing sites (Ramaswami & Li 2014). Among the emergent questions that have arisen from this identification are which of these sites are functional and how are they regulated. Constructing a spatiotemporal atlas of RNA editing is instrumental to answering such questions (Li & Church 2013). This study explored the genome-wide landscape of RNA editing in human brain development to advance understanding of the function and regulation of A-to-I editing. The genome-wide approach discovered that there are different levels of RNA editing rates both at a given site across development and at a given state of cell differentiation across multiple genes, which shows the 'dynamic' nature of RNA editing.

The dynamic aspects of A-to-I editing were summarized by three distinct editing patterns: stable high, stable low and increasing across cortical development. Identification of these patterns helps us to appreciate how A-to-I editing is regulated and functionally-implicated in the development of the nervous system. Specifically, the temporal expression of ADAR enzymes and the secondary structures of RNA species, as potential *trans-* and *cis-* regulatory mechanisms respectively, likely account at least in part for the developmental A-to-I editing patterns. Interestingly, the increasing A-to-I editing pattern across brain development is associated temporally with the growth of cortical layers and neuronal maturation.

However, the biochemical meaning of the proposed mechanisms for dynamic A-to-I editing patterns remains to be explored. Specifically, although the suggested *cis-* regulatory mechanism - RNA secondary structure – is computationally quantified in this study, a molecular exploration has yet to be completed. Also, the mechanism behind stable high-editing pattern throughout brain development is still unclear as many sites with low degrees of double-stranded structure have high editing rates during fetal life despite relatively-low expression of ADAR enzymes. There might be cofactors to enhance A-to-I editing (Garncarz et al. 2013) or unknown facilitatory mechanisms associated with high-edited sites. For example, the distance between

143

editing sites and the hairpin structure may affect editing rates synergistically with the degree of double-stranded structure.

In addition, although this study showed the correlation of A-to-I editing changes with neuronal maturation in neurogenesis or pathological conditions in brain disorders, mechanistic understanding on how the A-to-I editing change mediates those cellular phenotypes is still very limited. Some possiblilities were investigated including amino acid changes in functional domains or perturbation of miRNA-binding potential, but further studies of molecular mechanisms are necessary to fully appreciate their functional importance.

Finally, it should be noted that an emphasis on human samples in this study identifies many developmentally-regulated editing sites in primate-specific regions of the genome, especially in Alu repeats. So far, many functional studies on RNA editing have focused on evolutionarily-conserved sites, which are relatively few (Pinto et al. 2014). However, considering that the number of RNA editing sites has expanded in primates, especially in brain tissues, it is important to identify which sites are functionally relevant. The increasing editing pattern might provide a clue to this question. For example, RNA editing sites showing developmentally-increasing editing patterns only in human-specific transcripts, such as one in found in PDZD7 (see section 5.2) can be candidates for molecular experiments. In addition, a comparison of Alu-containing transcripts with and without the increasing A-to-I editing pattern will promote the understanding of the A-to-I editing in Alu-containing transcripts, and more generally, the role of RNA editing in human-specific aspects of brain development.

# References

Anders, S., Pyl, P.T. & Huber, W., 2014. HTSeq A Python framework to work with high-throughput sequencing data. Bioinformatics, 31(2), pp.166–169.

Athanasiadis, A., Rich, A. & Maas, S., 2004. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS biology, 2(12), e391.

Auwera, G.A. Van Der et al., 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline A. Bateman et al., eds. Current Protocols in Bioinformatics, 43:11.10, pp.1–33.

Backus, J.W. & Smith, H.C., 1992. Three distinct RNA sequence elements are required for efficient apolipoprotein B (apoB) RNA editing in vitro. Nucleic acids research, 20(22), pp.6007–14.

Barbon, A. & Barlati, S., 2011. Glutamate receptor RNA editing in health and disease. Biochemistry. Biokhimii͡a, 76(8), pp.882–9.

Bass, B.L. & Weintraub, H., 1987. A developmentally regulated activity that unwinds RNA duplexes. Cell, 48(4), pp.607–613.

Bass, B.L. & Weintraub, H., 1988. An unwinding activity that covalently modifies its double-stranded RNA substrate. Cell, 55(6), pp.1089–1098.

Ben-Ari, Y., 2014. The GABA excitatory/inhibitory developmental sequence: A personal journey. Neuroscience, 279, pp.187–219.

Benne, R. et al., 1986. Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell, 46(6), pp.819–826.

Berger, M., Gray, J.A. & Roth, B.L., 2009. The expanded biology of serotonin. Annual Review of Medicine, 60, pp.355–366.

Bernstein, B.E. et al., 2012. An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414), pp.57–74.

Bhalla, T. et al., 2004. Control of human potassium channel inactivation by editing of a small mRNA hairpin. Nature structural & molecular biology, 11(10), pp.950–6.

Bhogal, B. et al., 2011. Modulation of dADAR-dependent RNA editing by the Drosophila fragile X mental retardation protein. Nature Neuroscience, 14(12), pp.1517–1524.

Birnbaum, R. et al., 2014. Prenatal expression patterns of genes associated with neuropsychiatric disorders. American Journal of Psychiatry, 171(7), pp.758–767.

Blanc, V. & Davidson, N.O., 2010. APOBEC-1-mediated RNA editing. Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 2(5), pp.594–602.

Blow, M. et al., 2004. A survey of RNA editing in human brain. Genome research, 14(12), pp.2379–87.

Brusa, R. et al., 1985. Early-Onset Epilepsy and Postnatal Lethality Associated with an Editing-Deficient GluR-B Allele in Mice. Science, 270(5242), pp.1677–1680.

Burns, C.M. et al., 1997. Regulation of serotonin-2C receptor G-protein coupling by RNA editing. Nature, 387, pp.303–308.

Chen, K. et al., 2013. RNA-seq characterization of spinal cord injury transcriptome in acute/subacute phases: a resource for understanding the pathology at the systems level. PloS one, 8(8), p.e72567.

Choy, J.Y.H. et al., 2015. A resource of ribosomal RNA-depleted RNA-Seq data from different normal adult and fetal human tissues. Scientific data, 2, p.150063.

Daniel, C. et al., 2012. A distant cis acting intronic element induces site-selective RNA editing. Nucleic acids research, 40(19), pp.9876–86.

Daniel, C. et al., 2011. Adenosine-to-inosine RNA editing affects trafficking of the gamma-aminobutyric acid type A (GABA(A)) receptor. The Journal of biological chemistry, 286(3), pp.2031–40.

Daniel, C. et al., 2014. Alu elements shape the primate transcriptome by cis-regulation of RNA editing. Genome biology, 15(2), p.R28.

Darmanis, S. et al., 2015. A survey of human brain transcriptome diversity at the single cell level. Proceedings of the National Academy of Sciences, 112(23), pp.7285–7290.

Davidson, N.O. & Shelness, G.S., 2000. APOLIPOPROTEIN B: mRNA editing, lipoprotein assembly, and presecretory degradation. Annu Rev Nutr, 20, pp.169–193.

DeCerbo, J. & Carmichael, G.G., 2005. Retention and repression: Fates of hyperedited RNAs in the nucleus. Current Opinion in Cell Biology, 17(3), pp.302–308.

DePristo, M. a et al., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics, 43(5), pp.491–8.

Dillman, A.A. et al., 2013. mRNA expression, splicing and editing in the embryonic and adult mouse cerebral cortex. Nature neuroscience, 16(4), pp.499–506.

Dobin, A. et al., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), pp.15–21.

Eggington, J.M., Greene, T. & Bass, B.L., 2011. Predicting sites of ADAR editing in double-stranded RNA. Nature communications, 2(May), p.319.

Enright, A.J. et al., 2003. MicroRNA targets in Drosophila. Genome Biology, 5(R1), pp.1–14.

Eran, A. et al., 2013. Comparative RNA editing in autistic and neurotypical cerebella. Molecular Psychiatry, 18(9), pp.1041–8.

Fertuzinhos, S. et al., 2014. Laminar and temporal expression dynamics of coding and noncoding RNAs in the mouse neocortex. Cell reports, 6(5), pp.938–50.

Gallo, A. & Locatelli, F., 2012. ADARs: allies or enemies? The importance of A-to-I RNA editing in human disease: from cancer to HIV-1. Biological reviews of the Cambridge Philosophical Society, 87(1), pp.95–110.

Garncarz, W. et al., 2013. A high-throughput screen to identify enhancers of ADAR-mediated RNA-editing. RNA biology, 10(2), pp.192–204.

Garrett, S. & Rosenthal, J.J.C., 2012. RNA Editing Underlies Temperature Adaptation in K+ Channels from Polar Octopuses. Science, 335(6070), pp.848–851.

Gommans, W.M., Mullen, S.P. & Maas, S., 2009. RNA editing: a driving force for adaptive evolution? BioEssays, 31(10), pp.1137–45.

Gray, M.W., 2012. Evolutionary origin of RNA editing. Biochemistry, 51(26), pp.5235–42.

Griffiths-Jones, S. et al., 2008. miRBase: Tools for microRNA genomics. Nucleic Acids Research, 36(SUPPL. 1), pp.154–158.

Han, L. et al., 2015. The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. Cancer Cell, 28(4), pp.515–528.

Hartner, J.C. et al., 2004. Liver Disintegration in the Mouse Embryo Caused by Deficiency in the RNA-editing Enzyme ADAR1. Journal of Biological Chemistry, 279(6), pp.4894–4902.

Higuchi, M. et al., 2000. Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. Nature, 406, pp.1998–2001.

Horsch, M. et al., 2011. Requirement of the RNA-editing enzyme ADAR2 for normal physiology in mice. The Journal of biological chemistry, 286(21), pp.18614–22.

Hubbard, K.S. et al., 2013. Longitudinal RNA sequencing of the deep transcriptome during neurogenesis of cortical glutamatergic neurons from murine ESCs. F1000Research, 2, p.35.

Hundley, H. a. & Bass, B.L., 2010. ADAR editing in double-stranded UTRs and other noncoding RNA sequences. Trends in Biochemical Sciences, 35(7), pp.377–383.

Jaffe, A.E. et al., 2015. Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. Nature Neuroscience, 18(1), pp.154–161.

Jaffe, A.E. et al., 2016. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. Nature Neuroscience, 19(19), pp.40–47.

Jan, L.Y. & Jan, Y.N., 2012. Voltage-gated potassium channels and the diversity of electrical signalling. The Journal of physiology, 590(Pt 11), pp.2591–9.

Kappler, M. et al., 2002. Control of kinetic properties of GluR2 flop AMPA-type channels : impact of R / G nuclear editing. European Journal of Neuroscience, 15, pp.51–62.

Kawahara, Y. & Kwak, S., 2004. RNA editing and death of motor neurons. Nature, 427, pp.801.

Kim, D.D.Y. et al., 2004. Widespread RNA editing of embedded alu elements in the human transcriptome. Genome research, 14(9), pp.1719–25.

Kleinman, C.L. & Majewski, J., 2012. Comment on "Widespread RNA and DNA sequence differences in the human transcriptome". Science, 335(6074), p.1302.

Knisbacher, B. a. & Levanon, E.Y., 2015. DNA and RNA editing of retrotransposons accelerate mammalian genome evolution. Annals of the New York Academy of Sciences, 1341(1), pp.115–125.

Kwon, H.-B. & Sabatini, B.L., 2011. Glutamate induces de novo growth of functional spines in developing cortex. Nature, 474(7349), pp.100–4.

van de Leemput, J. et al., 2014. CORTECON: A Temporal Transcriptome Analysis of In Vitro Human Cerebral Cortex Development from Human Embryonic Stem Cells. Neuron, 83(1), pp.51–68

Lev-Maor, G. et al., 2007. RNA-editing-mediated exon evolution. Genome biology, 8(2), p.R29.

Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14), pp.1754–60.

Li, J.B. et al., 2009. Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science, 324(5931), pp.1210–1213.

Li, J.B. & Church, G.M., 2013. Deciphering the functions and regulation of brain-enriched A-to-I RNA editing. Nature neuroscience, 16(11), pp.1518–22.

Li, M. et al., 2011. Widespread RNA and DNA sequence differences in the human transcriptome. Science, 333(6038), pp.53–8.

Li, Z. et al., 2013. Evolutionary and ontogenetic changes in RNA editing in human, chimpanzee, and macaque brains. RNA, 19(12), pp.1693–702.

Liang, H.A.N. & Landweber, L.F., 2007. Hypothesis : RNA editing of microRNA target sites in humans ? RNA, pp.463–467.

Lorenz, R. et al., 2011. ViennaRNA Package 2.0. Algorithms for molecular biology : AMB, 6, p.26.

Luhmann, H.J., Fukuda, A. & Kilb, W., 2015. Control of cortical neuronal migration by glutamate and GABA. Frontiers in Cellular Neuroscience, 9, pp.1–15.

Lyddon, R., Navarrett, S. & Dracheva, S., 2012. Ionotropic glutamate receptor mRNA editing in the prefrontal cortex: no alterations in schizophrenia or bipolar disorder. Journal of psychiatry & neuroscience : JPN, 37(4), pp.267–72.

Maas, S. et al., 2001. Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. Proceedings of the National Academy of Sciences of the United States of America, 98(25), pp.14687–92.

Maathuis, M. et al., 2000. A method and server for predicting damaging missense mutations. Ann. Stat. Cell Statist. Soc. Ser. B J. Roy. Statist. Soc. Ser. B Biol, 37(16), pp.3133–3164.

Nishikura, K., 2010. Functions and regulation of RNA editing by ADAR deaminases. Annual review of biochemistry, 79, pp.321–49.

Niswender, C.M. et al., 2001. RNA editing of the human serotonin 5-HT2C receptor: Alterations in suicide and implications for serotonergic pharmacotherapy. Neuropsychopharmacology, 24(5), pp.478–491.

O'Neil, R.T. & Emeson, R.B., 2012. Quantitative analysis of 5HT 2C receptor RNA editing patterns in psychiatric disorders. Neurobiology of Disease, 45(1), pp.8–13.

Pachernegg, S. et al., 2015. GluA2 is rapidly edited at the Q/R site during neural differentiation in vitro. Frontiers in cellular neuroscience, 9(March), p.69.

Palladino, M.J. et al., 2000. A-to-I Pre-mRNA Editing in Drosophila Is Primarily Involved in Adult Nervous System Function and Integrity. Cell, 102(4), pp.437–449.

Pandey, R. & Mukerji, M., 2011. From "JUNK" to just unexplored noncoding knowledge: The case of transcribed Alus. Briefings in Functional Genomics, 10(5), pp.294–311.

Park, E. et al., 2012. RNA editing in the human ENCODE RNA-seq data. Genome research, pp.1626–1633.

Paul, M.S. & Bass, B.L., 1998. Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. The EMBO journal, 17(4), pp.1120–1127.

Pinto, Y., Cohen, H.Y. & Levanon, E.Y., 2014. Mammalian conserved ADAR targets comprise only a small fragment of the human editosome. Genome biology, 15(1), p.R5.

Powell, L.M. et al., 1987. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. Cell, 50(6), pp.831–840.

Prasanth, K. V et al., 2005. Regulating gene expression through RNA nuclear retention. Cell, 123(2), pp.249–63.

Pullirsch, D. & Jantsch, M.F., 2010. Proteome diversification by adenosine to inosine RNA editing. RNA biology, 7(2), pp.205–212.

Ramaswami, G. et al., 2013. Identifying RNA editing sites using RNA sequencing data alone. Nature methods, 10(2), pp.128–32.

Ramaswami, G. & Li, J.B., 2014. RADAR: a rigorously annotated database of A-to-I RNA editing. Nucleic acids research, 42(Database issue), pp.D109–13.

Rebagliati, M.R. & Melton, D. a., 1987. Antisense RNA injections in fertilized frog eggs reveal an RNA duplex unwinding activity. Cell, 48(4), pp.599–605.

Ripke, S. et al., 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature, 511, pp.421–427.

Rosenberg, B.R. et al., 2011. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs. Nature Structural & Molecular Biology, 18(2), pp.230‑236.

Rosenthal, J.J.C. & Seeburg, P.H., 2012. A-to-I RNA editing: effects on proteins key to neural excitability. Neuron, 74(3), pp.432–9.

Roth, B.L. ed., 2006. The Serotonin Receptors: From Molecular Pharmacology to Human Therapeutics, Humana Press.

Rueter, S.M., Dawson, T.R. & Emeson, R.B., 1999. Regulation of alternative splicing by RNA editing. Nature, 399(May), pp.75–80.

Rula, E.Y. et al., 2008. Developmental modulation of GABA(A) receptor function by RNA editing. The Journal of neuroscience the official journal of the Society for Neuroscience, 28(24), pp.6196–201.

Rybak-Wolf, A. et al., 2015. Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. Molecular Cell, 58, pp.870–885.

Sakurai, M. et al., 2014. A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. Genome research, 24(3), pp.522–34.

Sansam, C.L. & Emerson, R.B., 2005. mRNA editing. Encyclopedia of Life Sciences, pp.1–9.

Sarkisian, M.R. & Guadiana, S.M., 2015. Influences of primary cilia on cortical morphogenesis and neuronal subtype maturation. The Neuroscientist, 21(2), pp.136–51.

Sauvageau, M. et al., 2013. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. eLife, 2, p.e01749.

Savva, Y.A., Rieder, L.E. & Reenan, R.A., 2012. The ADAR protein family. Genome Biology, 13(12), p.252.

Scadden, A. D.J., 2007. Inosine-Containing dsRNA Binds a Stress-Granule-like Complex and Downregulates Gene Expression In trans. Molecular Cell, 28(3), pp.491–500.

Scadden, A.D.J., 2005. The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. Nature Structural & Molecular Biology, 12(6), pp.489–496.

Schoft, V.K., Schopoff, S. & Jantsch, M.F., 2007. Regulation of glutamate receptor B pre-mRNA splicing by RNA editing. Nucleic acids research, 35(11), pp.3723–32.

Sekine, S., Miura, M. & Chihara, T., 2009. Organelles in developing neurons: essential regulators of neuronal morphogenesis and function. The International journal of developmental biology, 53(1), pp.19–27.

Sherry, S.T. et al., 2001. dbSNP : the NCBI database of genetic variation. , 29(1), pp.308–311.

Slotkin, W. & Nishikura, K., 2013. Adenosine-to-inosine RNA editing and human disease. Genome medicine, 5(11), p.105.

Sodhi, M.S. et al., 2001. RNA editing of the 5-HT(2C) receptor is reduced in schizophrenia. Molecular psychiatry, 6(4), pp.373–379.

Sommer, B. et al., 1991. RNA Editing in Brain Controls of Ion Flow in Glutamate-Gated a Determinant Channels. Cell, 67, pp.11–19.

Tonkin, L. a. et al., 2002. RNA editing by ADARs is important for normal behavior in Caenorhabditis elegans. EMBO Journal, 21(22), pp.6025–6035.

Trapnell, C., Pachter, L. & Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics, 25(9), pp.1105–11.

Valente, E.M. et al., 2013. Primary cilia in neurodevelopmental disorders. Nature Reviews Neurology, 10(1), pp.27–36.

Vissel, B. et al., 2001. The Role of RNA Editing of Kainate Receptors in Synaptic Plasticity and Seizures. Neuron, 29(1), pp.217–227.

Vitali, P. & Scadden, A. D.J., 2010. Double-stranded RNAs containing multiple IU pairs are sufficient to suppress interferon induction and apoptosis. Nature structural & molecular biology, 17(9), pp.1043–1050.

Wagnert, R.W. et al., 1989. A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells. Proceedings of the National Academy of Sciences, 86, pp.2647–2651.

Wahlstedt, H. et al., 2009. Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. Genome research, pp.978–986.

Wahlstedt, H. & Öhman, M., 2011. Site-selective versus promiscuous A-to-I editing. Wiley Interdisciplinary Reviews: RNA, 2(6), pp.761–771.

Wang, I.X. et al., 2013. ADAR regulates RNA editing, transcript stability, and gene expression. Cell reports, 5(3), pp.849–60.

Wang, Q. et al., 2013. ADAR1 regulates ARHGAP26 gene expression through RNA editing by disrupting miR-30b-3p and miR-573 binding. RNA, 19(11), pp.1525–36.

Wang, Q. et al., 2004. Stress-induced Apoptosis Associated with Null Mutation of ADAR1 RNA Editing Deaminase Gene. Journal of Biological Chemistry, 279(6), pp.4952–4961.

Whitney, N.P. et al., 2008. Calcium-permeable AMPA receptors containing Q/R-unedited GluR2 direct human neural progenitor cell differentiation to neurons. FASEB, 22(8), pp.2888–900.

Yamaguchi, K. et al., 1999. The reversible change of GluR2 RNA editing in gerbil hippocampus in course of ischemic tolerance. Journal of cerebral blood flow and metabolism, 19(4), pp.370–5.

Zhang, Y. et al., 2014. An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. The Journal of neuroscience, 34(36), pp.11929–47.

Zhang, Z. & Carmichael, G.G., 2001. The fate of dsRNA in the Nucleus: A p54nrb-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. Cell, 106(4), pp.465–475.

# CURRICULUM VIATE FOR Ph. D. CANDIDATES

The Johns Hopkins University School of Medicine

**Taeyoung Hwang**                                                                 **3/22/2015**

taeyoungh@gmail.com

## Educational History

Ph. D. expected          2016    Department of Biomedical Engineering, Johns Hopkins University School of Medicine (Mentor: Daniel R. Weinberger, M.D.)

M.S.                              2009    Interdisciplinary program in Bioinformatics, Seoul National University, Korea

B.S.                              2006    Department of Electrical Engineering, Seoul National University, Korea

## Other Professional Experience

Research rotation          2011    Dr. Takanari Inoue's laboratory, Department of Cell Biology, Johns Hopkins University

Research rotation          2010    Dr. Andre Levchenko's laboratory, Department of Biomedical Engineering, Johns Hopkins University

Researcher                    2010    Dr. Kwang-Hyun Cho's laboratory, KAIST, Korea

Researcher                    2009    Research Institute of Basic Science, Seoul National University, Korea

## Scholarships, fellowships or other external funding

SOM-APL Precision medicine award (PI: Daniel Weinberger and Christopher Bradurne), 11/20/2015, $100,000 for 12 months: research support (participating as a lead researcher).

## Academic and other honors at Hopkins and elsewhere

10/22/2015      First prize in poster session      9th Genomics and Bioinformatics Symposium hosted by Center for Computational Biology, Johns Hopkins University

10/6/2015    Reviewer's choice in poster presentation    Annual meeting in American Society of Human Genetics (ASHG 2015).

## Publications, peer reviewed

**T. Hwang**, C.-K. Park, A. K. L. Leung, Y. Gao, T. M Hyde, J. E Kleinman, A. Rajpurohit, R. Tao, J. H. Shin, D. R Weinberger, Dynamic regulation of RNA editing in human brain development and disease, *under review*.

D. Mathios**\*, T. Hwang\***, J. Phallen, A. P. See, P. Burger, K. L. Mcdonald, G. L. Gallia, M. Kai, H. Brem, D. Pardoll, P. Ha, V. E. Velculescu, C. Bettegowda, C.-K. Park, M. Lim, Genome-wide investigation of intragenic methylation identifies ZMIZ1 gene as a novel prognostic marker in cancer, *Submitted*. \* equally contributed

**T. Hwang**, T. Park (2009) Identification of differentially expressed sub-networks based on multivariate ANOVA scoring method. BMC Bioinformatics 10:128.

## Publications, chapters and other non-peer reviewed

**T. Hwang**, T. Park (2010) Multivariate analysis of microarray data: Application of Manova. In: Medical Biostatistics for Complex Diseases (Frank Emmert-Streib and Matthias Dehmer, eds), Wiley Online Library

## Posters and presentations

**T. Hwang**, J. Shin, D. R. Weinberger (2016) Dynamic regulation of RNA-editing in human brain development and disease, *poster presentation*, Cold Spring Harbor conference on Systems Biology: Global Regulation of Gene Expression, Mar. 2016.

**T. Hwang**, J. Shin, D. R. Weinberger (2015) Dynamic regulation of RNA-editing in human brain development and disease, *oral presentation*, 2015 IEEE International Conference on Bioinformatics and Biomedicine (IEEE BIBM), Nov. 2015.

**T. Hwang**, J. Shin, D. R. Weinberger (2015) Dynamic regulation of RNA-editing in human brain development and disease, *poster presentation*, Annual meeting in American Society of Human Genetics (ASHG 2015), Oct. 2015.

**T. Hwang**, J. Shin, D. R. Weinberger (2014) A dynamic regulation of RNA-editing in human brain development, *poster presentation*, Gordon Research Conference, Jul. 2014.

A. P. See\*, **T. Hwang\***, D. Shin, J. Phallen, Y. Gao, M. Lim (2011) Genome-wide analysis of clinically significant CpG methylation sites in GBM, *poster presentation*, *Society for NeuroOncology Annual Meeting*, Nov. 2011. \* equally contributed

## Service and leadership

2014-15        Co-director of Professional Development, BME EDGE (BioMedical Engineering Extramural Development in Graduate Education)

BME EDGE was established to provide additional training opportunities for the biomedical engineering (BME) PhD students in Johns Hopkins University (JHU). As a part of BME EDGE executive team in 2014-15, I coordinated events called Speakers' series where speakers from inside/outside  JHU gave a talk and interact with PhD students.

2013 Intersession        Instructor, EN 580.102.12 Intro to Next Generation Sequencing

I organized the freshmen or sophomore-level course in Johns Hopkins University to introduce next generation sequencing (NGS) technologies and its application. This class presented various examples showing how NGS technologies are utilized to answer biomedical problems. Also, computational tools with real datasets were introduced to help understand the analyses of data generated by NGS technologies.