DYNAMIC ALTERNATIVE SPLICING EVENTS IN THE DORSOLATERAL PREFRONTAL CORTEX DURING ADOLESCENCE-YOUNG ADULTHOOD PERIOD AND IMPLICATIONS FOR SCHIZOPHRENIA

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

NEUROSCIENCE

By

Kübra Çelikbaş

November 2020

DYNAMIC ALTERNATIVE SPLICING EVENTS IN THE DORSO-LATERAL PREFRONTAL CORTEX DURING ADOLESCENCE-YOUNG ADULTHOOD PERIOD AND IMPLICATIONS FOR SCHIZOPHRENIA

By Kübra Çelikbaş

November 2020

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Timothea Toulopoulou(Advisor)

Ali Osmay Güre

Kerem Mert Şenses

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

DYNAMIC ALTERNATIVE SPLICING EVENTS IN THE DORSOLATERAL PREFRONTAL CORTEX DURING ADOLESCENCE-YOUNG ADULTHOOD PERIOD AND IMPLICATIONS FOR SCHIZOPHRENIA

Kübra Çelikbaş
M.S. in Neuroscience
Advisor: Timothea Toulopoulou
November 2020

Alternative splicing (AS) or differential exon usage (DEU) is a regular process after gene expression and it contributes to the diversity of the genome by generating multiple protein isoforms. According to recent studies, the majority (92-94%) of all human multi-exon genes undergo AS and the brain, especially the neocortex, has the highest number of AS events compared to other tissues. While contributing to the complexity of the brain, AS may lead to neuropsychiatric disorders such as schizophrenia or autism if dysregulated. Adolescence and young adulthood (AYA) period which nearly covers age range between 15 to 24 years old, is known to be a critical time to develop several neuropsychiatric disorders including schizophrenia and depression. Therefore, it is important to know developmental changes in AS events that occur in healthy brains in order to understand what is disrupted in a diseased brain. Although there are many studies investigating the possible roles of AS in the function of specific neuron types and during neurogenesis, there are only a few studies investigating AS changes in the human brain during different developmental periods. Therefore, in this study we first compared DEU that occur in the dorsolateral prefrontal cortex (DLPFC) of psychologically healthy individuals during AYA period to other developmental periods: infancy, early childhood, middle and late childhood, young adulthood, middle adulthood, and late adulthood. Additionally we compared DEU that occur in the DLPFC of schizophrenia patients to psychologically healthy individuals. Then we found exons that show both developmental and schizophrenia related DEU changes. Our results revealed 4 exons that belong to 3 different genes: AKAP7, BAIAP3 and SEMA3B. If further investigated, these exons can help us better understand the pathophysiology of schizophrenia and be possible early markers of the disease.

Keywords: schizophrenia, alternative splicing, differential exon usage, adolescence, young adulthood, dorsolateral prefrontal cortex.

ÖZET

ERGENLİK-ERKEN YETİŞKİNLİK DÖNEMİNDE DORSOLATERAL PREFRONTAL KORTEKSTE GERÇEKLEŞEN DİNAMİK ALTERNATİF GEN KIRPILMA OLAYLARI VE ŞİZOFRENİ İLE İLGİLİ ÇIKARIMLAR

Kübra Çelikbaş
Nörobilim, Yüksek Lisans
Tez Danışmanı: Timothea Toulopoulou
Ekim 2020

Alternatif kırpılma (AK) gen ifadesinden sonra gerçekleşen ve birçok protein isoformu oluşturarak genomun çeşitliliğine katkı sağlayan normal bir aşamadır. Yakın zamanda yapılan araştırmalara göre çok ekzonlu genlerin büyük bir kısmı (% 92-94) alternatif kırpılmaya uğramaktadır ve beyin, özellikle neokorteks, diğer dokularla kıyaslandığında alternatif kırpılma olaylarının en çok görüldüğü dokudur. Beynin karmaşık yapısına katkı sağlamasının yanısıra, alternatif kırpılma bozulması halinde otizm ve şizofreni gibi nörogelişimsel hastalıklara yol açabilmektedir. Yaklaşık olarak 15 ile 24 yaş arasını kapsayan ergenlik ve erken yetişkinlik (EEY) dönemi şizofreni ve depresyon gibi nöropsikolojik hastalıklara yakalanmak için kritik bir zaman dilimidir. Bu yüzden sağlıklı bireylerin beyinlerinde AK olaylarında görülen gelişimsel değişmeleri bilmek hasta beyinlerdeki bozuklukları anlamamız için önemlidir. Nöron oluşumu sırasında ve değişik nöron tiplerindeki AK olayları ile ilgili pek çok araştırma olmasına rağmen, insan beyninde AK olaylarında farklı gelişimsel dönemlerde meydana gelen değişiklikleri inceleyen çok az çalışma bulunmaktadır. Bu yüzden, bu çalışmada ilk olarak psikolojik açıdan sağlıklı bireylerin dorsolateral prefrontal korteksinde (DLPFK) EEY döneminde görülen AK olaylarını diğer gelişimsel dönemlerle karşılaştırdık: bebeklik , erken çocukluk, orta ve geç çocukluk, erken yetişkinlik, orta yetişkinlik ve ileri yetişkinlik. Ek olarak şizofreni hastalarında DLPFK bölgesinde görülen AK olayları ile psikolojik açıdan sağlıklı bireylerinkiler ile karşılaştırdık. Sonrasında hem gelişimsel hem de şizofreni ile ilintili AK değişikliği gösteren ekzonları tespit ettik. Sonuçlarımız bu özelliği gösteren 3 gene farklı gene ait 4 ekzon olduğunu gösterdi: AKAP7, BAIAP3 ve SEMA3B.

Eğer daha fazla araştırılırsa, bu ekzonlar şizofreni hastalığının patofizyolojisini daha iyi anlamamıza yardımcı ve hastalığın erken biyoişaretçilerinden olabilirler.

Anahtar sözcükler: şizofreni, alternatif gen kırpılması, ergenlik, erken yetişkinlik, dorsolateral prefrontal korteks.

Acknowledgement

Firstly, I would like to thank my advisor Prof. Dr. Timothea Toulopoulou for giving me the opportunity to study at Bilkent University and guiding me through my master. Besides my advisor, I want to thank Assoc. Prof. Ali Osmay Güre and Asst. Prof. Kerem Mert Şenses for their invaluable contributions to my thesis.

I want to especially thank Kerem Mert Şenses for all the things that he taught me and for guiding me throughout my master period. Also, I want to thank Murat İşbilen for his willingness to help whenever we needed.

I want to thank all of my friends Kübra Fırtına, Ceren Bilge Çelebi, Ilgım Özerk, Gülin Sayal, Simge Kelekçi, Büşra Korkmaz, Rabia Şen, Damla Güneş, Hazal Beril Çatalak, Melike Demir, Sena Atıcı, Gizem Sunar, Vildan Güler, Melike Uzuntaş, Özlem Aybüke Işık and Muntader Jihad for their priceless support whenever I needed. Thank you all for making this journey enjoyable and memorable for me.

Last but not the least, I want to thank my whole family: my mother Nermin Çelikbaş, my father Yüksel Çelikbaş, my sisters Emine Hatun and Büşra Çelikbaş, my little brother Ahmet Efe Çelikbaş and my cat Badem for being in my life and for their precious support. My greatest thanks to my husband Özgür Yılmaz for his endless emotional support and for his help during my analyses and throughout my master period.

Contents

1 Introduction **1**

1.1 Alternative Splicing 1

1.2 Types of Alternative Splicing 4

1.3 Regulation of Alternative Splicing 5

1.4 Developmental Changes of Alternative Splicing in Brain 7

1.5 Alternative Splicing in Brains of Individuals with Schizophrenia . 10

1.6 Research Question and Rationale 19

2 Methods **21**

2.1 Exon Microarray Data 21

2.2 Affymetrix Human Exon 1.0 ST Arrays 24

2.3 Alternative Splicing Analysis 24

2.4 RNA Sequencing Data 27

2.5 RNA-Sequencing Data Retrieval 28

2.6 RNA-Seq Data Analysis 29

2.6.1 Quality Control 30

2.6.2 Trimming of Low Quality Reads 30

2.6.3 Alignment of Reads to the Reference Genome 31

2.6.4 Annotation and Counting of Reads 32

2.6.5 Alternative Splicing Analysis 33

2.7 Investigation of the Effect of Alternatively Spliced Exons 34

3 Results 35

3.1 Results of Exon Microarray Analysis 35

3.1.1 Developmental Alternative Splicing of PALM Gene 37

3.1.2 Developmental Alternative Splicing of MAPT Gene 38

3.1.3 Developmental Alternative Splicing of NRXN1 Gene . . . 40

3.2 Results of RNA-Seq Analysis 41

3.2.1 Alternative Splicing of GRIN1 Gene between schizophrenia and Healthy Groups 42

3.3 Genes That Show both Developmental and Schizophrenia- Associated Alternative Splicing 45

3.3.1 Developmental and Schizophrenia-Associated Alternative Splicing of AKAP7 Gene 46

3.3.2 Developmental and Schizophrenia-Associated Alternative Splicing of BAIAP3 Gene 52

3.3.3 Developmental and Schizophrenia-Associated Alternative Splicing of SEMA3B Gene 57

4 Discussion 63

5 Supplementary Material 68

5.1 Supplementary Figures 68

5.2 Supplementary Tables 75

List of Figures

1.1 **Sequences Involved in Splicing.** Important sequences on mRNA that are known to be involved in splicing process in metazoans. 2

1.2 **Lariat Formation and Cleavage.** Demonstration of the intron cleavage through lariat formation. 3

2.1 **The Schema of AltAnalyze Pipeline.** This figure explains the alternative splicing analysis pipeline of AltAnalyze software. RMA: Robust Multichip Analysis, CEL: extension of microarray files, DABG: Detection Above Background 25

3.1 **Summary of the Exon Array Analysis.** These bar graphs summarizes results of group comparisons at exon level for both prefrontal cortex (PFC) and dorsolateral prefrontal cortex (DLPFC). (a) The number of significantly differentially expressed probesets (corrected p ¡0.05) as a result of comparison between adolescence and young adulthood (AYA) period and other periods listed in the x-axis (Infancy (I), early childhood (EC), middle and late childhood (MLC), young adulthood (YA), middle adulthood (MA) and late adulthood (LA) including array data coming from all regions listed in Table 2.1. (b) The number of significantly differentially expressed genes (corrected $p < 0.05$) as a result of comparison between AYA period and other periods listed in the x-axis including array data coming only from DFC. Numbers on the top of each bar indicate exact exon probeset numbers as a result of each comparison. 36

3.2 **The Expression of Exons of PALM Gene.** Log2 expressions of exons of PALM gene during infancy (red), early childhood (orange) and adolescence and young adulthood (dark green) periods. 37

3.3 **The Expression of Exons of MAPT Gene.** Log2 expressions of exons of MAPT gene during infancy (red), early childhood (orange), middle and late childhood (light green) and adolescence and young adulthood (dark green) periods. 38

3.4 **The Expression of Exons of NRXN1 Gene.** Log2 expressions of exons of NRXN1 gene during infancy (red), early childhood (orange), and adolescence and young adulthood (dark green) periods. 40

3.5 **The Expression of Exons of GRIN1 Gene.** Expression graph on the top shows fitted expression of exons of GRIN: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all exons of GRIN1 transcripts and genomic locations of them: pink bars indicate differentially expressed exons (FDR corrected p value ¡0.1) between schizophrenia and healthy groups. 42

3.6 **The Expression of GRIN1 Gene.** Expression graph on the top shows fitted expression of exons of GRIN with the effect of overall gene expression: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all exons of GRIN1 transcripts and genomic locations of them: pink bars indicate differentially expressed exons (FDR corrected p value ¡0.1) between schizophrenia and healthy groups. 44

3.7 **The Expression of Exons of AKAP7 Gene.** Log2 expressions of exons of AKAP7 gene during infancy (red), and adolescence and young adulthood (dark green) periods. 46

3.8 **The Expression of Exons of AKAP7 Gene.** Expression graph on the top shows fitted expression of exons of AKAP7: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all exons of AKAP7 transcripts and genomic locations of them: pink bars indicate differentially expressed exons (FDR corrected p value < 0.1) between schizophrenia and healthy groups. 47

3.9 **The Expression of Exons of AKAP7 Gene and Its Transcripts.** Expression graph on the top shows fitted expression of exons of AKAP7 with the effect of overall gene expression: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, different transcripts of AKAP7 gene can be seen and pink bars indicate differentially expressed exons (FDR corrected p value < 0.1) between schizophrenia and healthy groups. 49

3.10 **Ensembl AKAP7 Transcripts Expressed in the Frontal Cortex.** Ensembl IDs and exon combinations of AKAP7 transcripts that are expressed in the frontal cortex. 50

3.11 **The Expression of Exons of BAIAP3 Gene.** Log2 expressions of exons of BAIAP3 gene during early childhood (orange), and adolescence and young adulthood (dark green) periods. 52

3.12 **The Expression of Exons of BAIAP3 Gene.** Expression graph on the top shows fitted expression of exons of BAIAP3: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all exons of BAIAP3 transcripts and genomic locations of them: pink bars indicate differentially expressed exons (FDR corrected p value < 0.1) between schizophrenia and healthy groups. 53

3.13 **The Expression of Exons of BAIAP3 Gene and Its Transcripts.** Expression graph on the top shows fitted expression of exons of BAIAP3 gene with the effect of overall gene expression: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, different transcripts of BAIAP3 gene can be seen and pink bars indicate differentially expressed exons (FDR corrected p value < 0.1) between schizophrenia and healthy groups. 55

3.14 **Ensembl BAIAP3 Transcripts Expressed in the Frontal Cortex.** Ensembl IDs and exon combinations of BAIAP3 transcripts that are expressed in the frontal cortex. 56

3.15 **The Expression of Exons of SEMA3B Gene.** Log2 expressions of exons of SEMA3B gene during infancy (red), and adolescence and young adulthood (dark green) periods. 57

3.16 **The Expression of Exons of SEMA3B Gene.** Expression graph on the top shows fitted expression of exons of SEMA3B and MIR6872: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all exons of SEMA3B transcripts and genomic locations of them: pink bars indicate differentially expressed exons (FDR corrected p value < 0.1) between schizophrenia and healthy groups. 59

3.17 **The Expression of Exons of SEMA3B Gene and Its Transcripts.** Expression graph on the top shows fitted expression of exons of SEMA3B gene with the effect of overall gene expression: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, different transcripts of SEMA3B gene can be seen and pink bars indicate differentially expressed exons (FDR corrected p value < 0.1) between schizophrenia and healthy groups. 61

3.18 **Ensembl SEMA3B Transcripts Expressed in the Frontal Cortex.** Ensembl IDs and exon combinations of SEMA3B transcripts that are expressed in the frontal cortex. 62

5.1 **The Normalized DLPFC Counts of AKAP7 Gene.** Expression graph on the top shows normalized expression of DLPFCs of AKAP7 gene for all samples: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all DLPFCs of AKAP7 transcripts and genomic locations of them: pink bars indicate differentially expressed DLPFCs (FDR corrected p value < 0.1) between schizophrenia and healthy groups. 69

5.2 **The Normalized DLPFC Counts of BAIAP3 Gene.** Expression graph on the top shows normalized expression of DLPFCs of BAIAP3 gene for all samples: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all DLPFCs of BAIAP3 transcripts and genomic locations of them: pink bars indicate differentially expressed DLPFCs (FDR corrected p value < 0.1) between schizophrenia and healthy groups. 70

5.3 **The Normalized DLPFC Counts of SEMA3B Gene.** Expression graph on the top shows normalized expression of DLPFCs of SEMA3B and MIR6872 gene for all samples: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all DLPFCs of SEMA3B transcripts and genomic locations of them: pink bars indicate differentially expressed DLPFCs (FDR corrected p value < 0.1) between schizophrenia and healthy groups. 71

5.4 **The Expression of DLPFC Counts of AKAP7 Gene for All Developmental Periods.** Log2 expressions of DLPFCs of AKAP7 gene during different developmental periods. 72

5.5 **The Expression of DLPFC Counts of BAIAP3 Gene for All Developmental Periods.** Log2 expressions of DLPFCs of BAIAP3 gene during different developmental periods. 73

5.6 **The Expression of DLPFC Counts of SEMA3B Gene for All Developmental Periods.** Log2 expressions of DLPFCs of SEMA3B gene during different developmental periods. 74

List of Tables

1.1 **List of alternatively spliced genes in schizophrenia patiens** 17

2.1 **Brain regions included in the study and array numbers for each region across the groups including right and left hemispheres** 22

2.2 **Sample information of RNA-Seq data included in the study** 27

3.1 **Length and Motifs of AKAP7 Transcripts Expressed in the Frontal Cortex** 51

5.1 **GSM accession codes, developmental group, sex, age, hemisphere and region information for each array included in the study** 75

5.2 **Exon IDs, Probeset numbers and Genomic Locations of AKAP7 Provided by AltAnalyze** 83

5.3 **Exon IDs, Probeset numbers and Genomic Locations of BAIAP3 Provided by AltAnalyze** 84

5.4 **Exon IDs, Probeset numbers and Genomic Locations of SEMA3B Provided by AltAnalyze** 86

5.5 **Exon IDs and Genomic Locations of AKAP7 Provided by DEXseq** 87

5.6 **Exon IDs and Genomic Locations of BAIAP3 Provided by DEXseq** 88

5.7 **Exon IDs and Genomic Locations of SEMA3B Provided by DEXseq** 91

5.8 **Exon IDs and Genomic Locations of SEMA3B Provided by DEXseq** 92

Chapter 1

Introduction

1.1 Alternative Splicing

Transcription of mRNAs from DNA produces nascent mRNA (pre-mRNA) molecules that should be further processed in order to be functional. These post-transcriptional modifications include 5' capping, 3' polyadenylation and RNA splicing. While 5' capping and 3' polyadenylation serve to protect ends of pre-mRNAs from attacks of ribonucleases, RNA splicing is required in order to obtain final mRNA code that will be translated into a protein. This step is required because unlike other prokaryotic mRNAs, eukaryotic mRNAs are discontinuous with exon regions that will be translated and with noncoding intron regions that need to be removed [1].

RNA splicing is carried out by large RNA-protein complexes (RNPs) called spliceosomes. The conventional U2-dependent spliceosome is composed of five small nuclear RNPs (snRNPs) U1, U2, U4, U5 and U6 and many accessory proteins (Staley and Guthrie, 1998; Jurica and Moore, 2003). Each snRNP molecule contains an RNA component (snRNA), a set of seven Sm proteins (B/B', D3, D2, D1, E, F, and G) and varying number of other accessory proteins [2].

Splicing is carried out by a series of snRNP and mRNA interactions. Certain snRNPs can bind to conserved sequences on mRNAs through base pairing with their RNA components. These sequences, demonstrated in Fig. 1.1, are; 5' and 3' splice sites, branch point sequence (BPS) and polypyrimidine tract (PPT). 5' and 3' splice sites are found at the beginning and at the end of introns, defining the exon-intron boundaries. They can change according to the type of intron but most commonly occurring U2-type introns have GU at their 5' site and AG at their 3' end. Branch point sequence can be found anywhere from 18 to 40 nucleotides upstream of the 3' end of an intron. It is not very well conserved and has a typical sequence "YNYYRAY", where Y indicates a pyrimidine, N is any nucleotide, R indicates any purine, and A stands for adenine. Polypyrimidine tract lies between BPS and 3' splice site of an intron, located 5 to 40 nucleotides upstream of the 3' end of an intron. It is rich in terms of pyrimidine nucleotides, especially uracil, and usually 15–20 base pairs long [3].

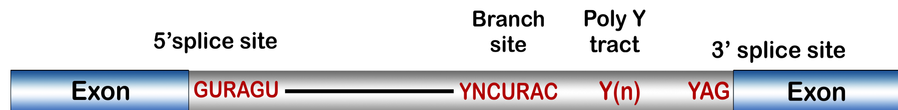


Figure 1.1: **Sequences Involved in Splicing.** Important sequences on mRNA that are known to be involved in splicing process in metazoans.

If an intron is short (<200–250 nts), spliceosome machinery forms across this intron [4]. At the early stages of splicing, U1 snRNP binds to 5' splice site (GU) and other accessory proteins SF1/mBBP and U2AF bind to the BPS and PPT, respectively. At this step spliceosome is called complex E. Then, U2 snRNP binds to adenine base of the branch point sequence (BPS), forming complex A (prespliceosome). A trimer containing U4/U5, U6 interacts with both U1 and U2, forming complex B (precatalytic spliceosome). Then the release of U1 and U4 allows other components of the spliceosome, especially U6 snRNP, to come into close position to 5' splice site. The complex which is now called complex B2 is activated and can carry out splicing reaction. Then the first catalytic step occurs: 2'OH group of Adenine at the BPS which is brought to closer proximity

to the 5’ splice site attacks phosphodiester bond of Guanine at the 5’ splice site. As a result, intron is cleaved from the 5’ splice position, releasing the first exon that it is bound to while at the 3’ splice site it is still bound to the second exon as a loop structure called lariat. Then the second catalytic step occurs: Free 3’ OH group of the released first exon attacks the phosphodiester bond at the 3’ splice site (AG), releasing the lariat which is rapidly degraded. As a result of this process, two exons are ligated to each other [5]. At the end, spliceosome is disassembled until next splicing reaction.

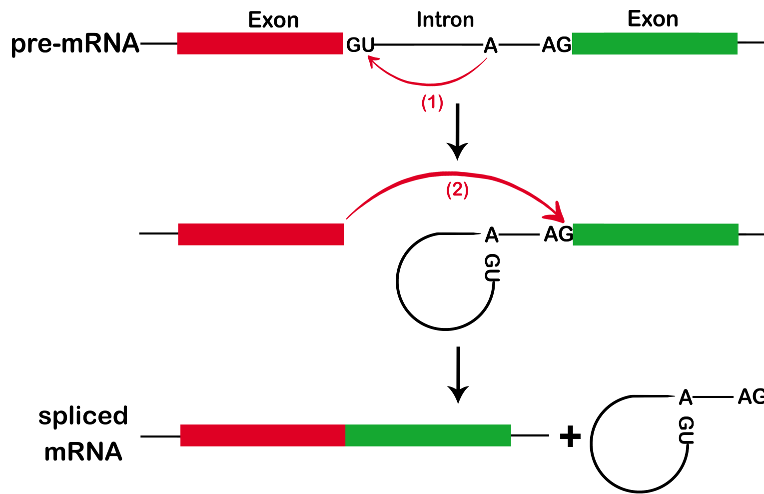


Figure 1.2: **Lariat Formation and Cleavage.** Demonstration of the intron cleavage through lariat formation.

However, if an intron is longer (> 200-250 nts) like the most of eukaryotic introns, the spliceosome machinery is first formed on an exon through a process called exon definition. In this case, U1 binds to the 5’ splice site while U2AF interacts with the PTT sequence of the upstream intron, defining the beginning and the end of exon. Then, U2 is recruited to the BPS of upstream intron. Finally, with the recruitment of other accessory proteins, exon-defined spliceosome is stabilized on exon but in order to cleave an intron, 5’ splice site of spliceosome machinery should interact with the downstream 3’ splice site of the same intron. This transition from exon-defined to intron-defined spliceosome complex is currently not well understood [5].

There is an alternative spliceosome which uses different snRNPs than the conventional one and cleaves the minor class of introns. Introns with splice sites 5'GU- 3'AG comprise the majority of introns but there are introns with splice sites 5'AU-3'AC and 5'GU-3'AG. These minor classes of introns are spliced by an alternative spliceosome containing U11 and U12 snRNPs, and therefore called U12-dependent spliceosomes [5]. Splicing takes place during the transcription process in order to ensure ordered removal of introns as they are released from the transcription complex but it does not always produce a same mature mRNA from the same transcript. There are two general types of RNA splicing. One is called constitutive splicing through which all introns are removed and all exons are ligated together to form a mature mRNA. The other is called alternative splicing, and as it name implies in this type of splicing, some exons and introns can be included and/or excluded in different combinations, creating diverse splice variants from one transcript. This alternative splicing process is thought to evolve in order to increase protein diversity in complex organisms. Although gene numbers differ little across different species, the protein diversity varies much more due to posttrancriptional modifications including splicing, probably due to increased intron length and number as species become complex [6].

1.2 Types of Alternative Splicing

There are several types of alternative splicing; cassette exon, intron retention, mutually exclusive exon, alternate 3' and 5' splice sites, mutually exclusive 3' and 5' untranslated regions (UTRs). Cassette exon events occur when one or more exons are skipped while mutually exclusive alternative splicing events result when pre-mRNA cannot contain both at the same time but can contain each separately. If different 5' and 3' competing splice sites are available, one of them can be alternatively selected over others. Mutually exclusive 5' UTRs occurs when alternative promoters alter the transcription start site, therefore the first exon of pre-mRNA. Similarly, alternative polyadenylation sites can alter the transcription end site (the last exon), and therefore 3' UTR. Intron retention, as it name

implies, occurs when one or more introns are not removed but kept in the pre-mRNA [7].

1.3 Regulation of Alternative Splicing

Alternative splicing is regulated by several factors including the strength of splice sites, cis-regulatory sequences on pre-mRNAs and trans-acting factors. If an intron contains conserved splice site sequences it can be easily detected by spliceosome machinery and it is cleaved almost every time, resulting in constitutive splicing. However, introns containing weak, i.e. non-conserved, splice sites need other factors like cis-regulatory sequences and trans-acting factors in order to be stably recognized by spliceosome.

Cis-regulatory sequences can be found on either exons or introns, and can act as either silencers or activators. Exonic splicing enhancers (ESEs) or exonic splicing silencers (ESSs) serve to facilitate or inhibit the retention of exons in which they reside, respectively. Similarly, intronic splicing enhancers (ISEs) and intronic splicing silencers (ISSs) serve to facilitate or inhibit the retention of exons from intronic regions. They carry out these functions by recruiting trans-acting factors (RNA binding proteins (RBPs)) which can either activate or suppress the activity of spliceosome.

ESEs are found on nearly all exons and can contain varying range of sequences [8]. Generally, ESEs function by recruiting SR family of proteins (trans-acting factors). These proteins bind to ESEs with their one terminal while facilitating the binding of accessory proteins that can promote spliceosome assembly with their other terminal called RS domain. On the other hand, ESSs bind to hnRNP family of proteins and they may contain several different sequences that can bind to RNA. They can inhibit splicing in various ways including preventing U1 and U2 interaction and displacing snRNP on exons.

There are several sequences and functions for ISEs: clusters of guanine (either

3 or more) can facilitate the recognition of nearby splice sites [9]; CA repeats can facilitate splicing of upstream exons [10]; UGCAUG hexanucleotides or sometimes variations of it are found in the introns downstream of neuron-specific exons [11]. Similar to ESEs ISSs function by recruiting hnRNP family of proteins. Both intronic elements (ISEs and ISSs) contain sequences that can bind to tissue-specific splicing factors [12].

Intron retention is also regulated by the splicing regulatory elements (SREs) mentioned above. For example, 5' splice site like sequences on an exon can promote the retention of downstream intron [13]. Also G clusters on ISEs can regulate intron retention by facilitating splicing of retained introns on some genes [14].

SREs regulate alternative splicing events in a context dependent manner. Their function can change according to their location on mRNAs. For example, G clusters when they are on introns facilitate splicing; however, when they are located on exons they inhibit splicing [15]. Also SR family of proteins can change their function according to their distance to nearest splice site. Considering the complexity and size of spliceosome, this is reasonable since the activity of these trans-acting factors will be affected according to their distance to the spliceosome complex. For example, G cluster binding hnRNP family of proteins facilitates splicing when they bind to G clusters located at the downstream of 5' splice sites, but prevents splicing when they bind to G clusters on exons [16]. Also, considering the abundance of the SRE elements across the genome, not all SRE elements can be recognized by trans-acting factors and deciding which SRE elements are recognized and what factors contribute to it is still unknown. However, one major suspect that can affect SRE function is the secondary structures of mRNAs that can make both SREs and splice sites more readily accessible. For example, the loop structure of exon 10 of tau gene affects its splicing by revealing or hiding the 5' splice site adjacent to this exon [17]. Finally, the ultimate activity of SREs are affected by the availability of trans-acting factors which are most probably the result of tissue and cell type specific splicing patterns [18].

The regulation of splicing can be also affected by transcription machinery since

many splicing events are thought to occur cotranscriptionally. It is found that alternative splicing of some genes is affected by the mutated RNA polymerase II that has a slow elongation rate [19]. Therefore, it seems the splicing regulation mechanisms are complex with many different factors involved in it.

These regulation processes are both temporarily and spatially regulated. For example, among all genes expressed in the brain 0.1% of them exhibit alternative splicing differences only across different regions of the brain, 19.5% of these genes exhibit alternative splicing differences only across different developmental periods and 70.6% exhibits both region and time specific expression [20]. These time and region specific alternative splicing is achieved by the expression of splicing regulatory molecules to be expressed in a timely and tissue specific manner. What affects the expression of these splicing regulatory molecules is probably both environmental and genetic factors, which are still not well known.

1.4 Developmental Changes of Alternative Splicing in Brain

According to recent studies, the majority (92-94%) of all human multi-exon genes undergoes AS [21] and brain, especially the neocortex, is among the tissues showing higher number of AS events compared to other tissues [22]. Similar to developmental gene expression changes, most of the alternative splicing events (83%) take place during prenatal development [20]. However, there are still ongoing changes during postnatal development and these changes might be important in order to understand neurodevelopmental disorders.

Alternative splicing studies can be carried out by using several methods: exon microarrays, RNA-Sequencing and quantitative PCR. Exon microarray method depends on hybridization between probes on chips that are complementary to exons and fluorescently label cDNA molecules. When this hybridization occurs, a detector recognizes where the signal is coming from and how strong it is in order to

determine the identity and amount of exon in a sample. Although there are many different RNA-Sequencing methods, nearly all depend on replicating cDNAs by using fluorescently labeled nucleotides. Each nucleotide contains a different fluorescent color and when they bind to the replication chain they emit color, and this way whole sequence of cDNAs can be detected. Quantitative PCR method depends on replicating cDNA seuences over many cycles and expression is determined by using fluorescent dyes that bind to double-stranded DNA molecules. If the cycle number of a cDNA in which fluorescent dye is detected (CT value) is small, its expression is higher compared to another cDNA molecule with a larger CT value. While exon arrays and RNA-Sequencing methods allow wide screening of expressed mRNAs, qPCR analysis can be used to assess a limited number of mRNAs. By using these 3 methods, alternative splicing analysis can be carried on different tissue types including postmortem tissues, blood and cell lines that are generated from patient derived pluripotent stem cells. Literature review is restricted in a way that this section (1.4 Developmental Changes of Alternative Splicing in Brain) and the next section (1.5 Alternative Splicing in Brains of Individuals with Schizophrenia) contain only the AS studies using postmortem tissues, mostly brain tissue.

There are several studies investigating temporal changes in alternative splicing events during postnatal normal development of the human brain by using postmortem brain samples.

In 2012, Tao et al. investigated the developmental changes in the expression of KCC2 gene in postmortem DLPFC tissue of individuals with the age range from gestational weeks 14 to 20 (fetal periods) and from birth up to 78 years of age (after birth) by using PCR. They identified the expression of 11 novel transcripts (which are first reported in this study) and they check the expression of 4 of them since they share a common 5' UTR region: a transcript called AK098371 which lacks the exon7 of full-length KCC2, another transcript called transcriptExon2B which is different than AK098371 only with its new second exon, transcriptΔExon6 which is different than AK098371 only by lacking exon6 and lastly transcriptExon6B which is different than AK098371 only with

its new sixth exon. The results showed that expressions of AK098371, transcriptEXON6B, and transcriptEXON2B splice variants were low at fetal periods and increased during birth, remaining high afterwards. However, expression of transcriptΔExon6 was at its highest before birth, and decreased after birth [23].

In one study, researchers investigated the changes in alternative splicing patterns by using RNA-Sequencing data of prefrontal cortex and cerebellum of 30 postmortem tissues of healthy individuals. They divided the data into different age groups: pool1 (2 days to 35 days), pool2 (182 days to 274 days), pool 3 (16 years to 20 years), pool 4 (25 years to 28 years), pool 5 (70 years to 80 years) and pool 6 (88 years to 98 years). They found that one third of the genes (1456 genes) expressed in the brain shows alternative splicing across these periods, and 15% of them show AS between two brain regions. Furthermore, they validated alternative splicing of 24 genes that they found in silico by using PCR [24].

Neuregulin 1 gene (NRG1) is among the highly studied genes since its association with schizophrenia and bipolar disease is long known. In 2014, Paterson et al. studied the expression changes of isoforms of NRG1; NRG1 I-IV and NRG1-IVNV in DLPFC region throughout postnatal (birth-83 years old) periods by using postmortem brain tissues and PCR method. They found that while the expression of NRG1-I was almost stable throughout aging, NRG1-II and NRG1-III were highest at birth and decreased by aging. The expression of NRG1-IVNV was similarly higher at birth but decreased during early infancy and disappeared after 3 years of age [25].

In a 2017 study, Tao et al. investigated the developmental expression changes of 10 splice variants of GAD1 gene in the postmortem DLPFC and hippocampus tissues of psychologically healthy people by using RNA-Sequencing data. They observed that transcripts with novel exons (which are first reported in this study) are mainly expressed in fetal brains (14-20 gestational week) and transcripts which are generated by exon skipping are mainly expressed in postnatal brains (birth-85 years old). The expression of full-length transcript of GAD1 (encoding GAD67 protein) was lower during prenatal period but it gradually increased towards 20 years of ages and remained high throughout adulthood [26].

A recent large study investigated age related AS events by analyzing nearly 8500 RNA-seq samples across 48 tissues obtained from postmortem tissues of 544 individuals. The study revealed that blood and skin tissue have the most age related splicing changes compared to other tissues. And, compared to Mazin et al. (2013) they identified a lesser number of alternative splicing events in the similar brain regions, which might be due to differences in their methodologies, and sample sizes. They also carried out gene expression analysis across tissues and found that splicing changes are better predictor of biological age than gene expression changes, showing the importance of alternative splicing [27].

In another recent study, Mazin et al. investigated developmental splicing changes that take place in the prefrontal cortex of humans, chimpanzees and rhesus macaques. Their results revealed that %7.3 of splicing variation can be explained by age while %38 of splicing variation is due to differences that occur among species. Among the genes that are alternatively spliced between species, snoRNA-host gene SNHG11 was remarkable since it showed two human specific and one chimpanzee specific splicing events and these events were confirmed with semi quantitative PCR. Results of age related splicing analysis showed that age-dependent intron retention events were twice as common in humans as in chimpanzees and macaques. Therefore, it seems like intron retention regulates gene-expression in PFC region in a human-specific manner and distortions might be important for understanding neurodevelopmental diseases [28].

1.5 Alternative Splicing in Brains of Individuals with Schizophrenia

Alternative splicing is an important phenomenon which affects many aspects of neurodevelopment including synaptogenesis [29] and axon guidance [30]. Therefore, it is important to study disruptions in AS events that might lead to neurodevelopmental diseases such as schizophrenia and autism. Indeed, there are several

studies investigating abnormal AS events that occur in patients with schizophrenia, and many genes with abnormal splice patterns have been detected.

Dopamine receptor DRD3 which is a target for antipsychotics was long known to exhibit different AS patterns. One study by using PCR method found that the AS variant lacking 98 nucleotides of exon 7 is more abundant among schizophrenia patients while the untruncated transcript was found to be lost in the parietal cortex [31].

Exon10 of GABRB2 gene encoding one of the three β subunits of GABA receptor A was shown to be decreased in the dorsolateral prefrontal cortex of individuals with schizophrenia [32]. Other splice variants one with intron 9 retention and exon 3 skipping and the other with exon 10 and 11 skipping for this gene were found to be increased and decreased in the same brain region in schizophrenia group compared to normal subjects, respectively [33]. Not only β subunits but also γ subunits also show different splicing patterns in schizophrenia patients. Exon 9 of the GABA receptor $\gamma 2$ subunit (GABRG2) was shown to be reduced to half in DLPFC of schizophrenia patients [34].

GRIN1 gene which encodes one of the components of ionotropic glutamate receptors (NMDA-type) was shown to exhibit alternative splicing patterns in the thalamus of people with schizophrenia. Expression of a variant containing exon22 and lacking exon5 and exon21 was found to be reduced in the thalamus of schizophrenia patients compared to normal group [35]. Also, another variant lacking both exon21 and exon22 was found to be expressed at higher levels in the superior temporal gyrus of patients with schizophrenia [36].

A variant of RGS4 gene which is a negative regulator of G-proteins lacking some part of exon2 (RGS4-3) was found to be expressed in lower levels in dorsolateral prefrontal cortex of schizophrenia patients [37].

One of the splice variants of alphaN-catenin (CTNNA2) gene, lacking exon 17 and retaining intron 6, is significantly low in the hippocampi of schizophrenic non-smokers compared to schizophrenic smokers and control smokers [38].

One of the three major isoforms of NCAM1 (neural cell adhesion molecule 1) gene, NCAM-180, was found to be decreased in the Brodmann Area 46 of schizophrenia patients with a short illness duration, less than 7 years, compared to healthy controls [39].

Full- length isoform of KCNH2 gene which is a potassium channel was found to be expressed lower in DLPFC and hippocampi of schizophrenia patients. Another isoform of this gene with a 5' extension of exon 3 along with all downstream exons (KCNH2-3.1) was found to be expressed in higher levels in the hippocampi of schizophrenia patients [40].

Type I NRG1 (neuregulin 1) isoform was found to be expressed in higher levels in the hippocampus [41] and dorsolateral prefrontal cortex [42] of individuals with schizophrenia. Also, type IV variant of NRG1 gene is associated with schizophrenia in the brain through its schizophrenia-associated SNP containing promoter [43]. Not only NRG1, but also NRG3 isoforms I and IV were found to be expressed in higher levels in DLPFC of schizophrenia patients compared to controls [44].

Two variants of ERBB4, NRG1 receptor, were found to be increased in the dorsolateral prefrontal cortex of individuals with schizophrenia compared to controls. One of these variants contains exon16 while the other containing exon26 [45].

Although yielded controversial results in different cohorts, expression of metabotropic glutamate receptor 3 (GRM3) variant that lacks exon 4 is found to be affected by the presence of schizophrenia-associated SNP located on the 3rd exon of the gene and found to be higher in the dorsolateral prefrontal cortex of individuals with schizophrenia [46].

Three splice variants of disrupted-In-Schizophrenia-1 (DISC1) gene were found to be high during fetal development compared to adult human brains and their expression were higher in the hippocampi of individuals with schizophrenia. These variants include one with lacking exon3, other lacking exon7 and exon8 and the

last one which is called extra short variant terminates with a unique exon3a [47].

Quaking homolog KH domain RNA binding (QKI) protein was also shown to exhibit different AS patterns in the frontal cortex of schizophrenia patients. Two splice variants (one containing a unique exon 7 and the other containing a unique exon6) were found to be decreased in the frontal cortex of schizophrenia patients compared to healthy controls. Moreover, the expression of these two splice variants were significantly lower among patients using atypical neuroleptics and untreated patients compared to patients using typical epileptics [48].

Short isoform of myelin associated glycoprotein (MAG) which contains only first 4 exons was found to be significantly decreased in the frontal cortex and BA8 of schizophrenia patients compared to healthy controls [49].

One of the splice variants of presenilin2 gene (PS2) lacking exon5 was expressed in higher levels in some cases in the frontal cortex of individuals with schizophrenia [50].

One of the four splice variants of KCC2 gene which shows developmental expression changes (mentioned in the below section 3.2) also shows significant expression changes between schizophrenia and control group. Exon6B containing a unique exon 7 was expressed at significantly lower levels in schizophrenia patients compared to control while the expression of other variants were not affected [23]. KCC2 gene is also involved in GABA function. It is necessary for the developmental change of GABA from excitatory to inhibitory [51].

Four genes MCPH1, ZC3H13, BICD2, and DLG3 were found to exhibit alternative splicing differences in DLPFC of schizophrenia patients compared to controls. Exon8 of DLG3, exon6 of MCPH1, exon 13 of ZC3H13 and exon1 of BICD2 were found to be expressed lower in both tissue types of patients with schizophrenia [52].

An isoform of schizophrenia-associated gene ZNF804A that lacks exon1 and exon2 of the full-length transcript and containing new sequence from intron 2 was found to be expressed in lower levels in DLPFC of schizophrenia patients [53].

In 2012, Cohen et al. investigated AS differences between schizophrenia and control samples in a large scale by using Affymetrix Human Gene 1.0 ST arrays. They identified 43 genes and 31 genes that exhibit AS patterns between two conditions in Brodmann Area 10 (BA10) and caudate, respectively. Validation by qPCR analysis showed that 3' UTR region of CPNE3 gene was shorter among schizophrenia group in both brain regions. Similar qPCR analysis verified a decrease in exon11 of ENAH and increase in exon10 of KLHL5 gene among schizophrenia group only in BA10. Their results only confirmed dysregulation of the expression of ERBB4 exons from previous studies but not GRM3, ESR1, NRG1, DISC1 and CTNNA2 [54].

Another study investigated alternative splicing differences between superior temporal gyrus of 9 schizophrenia patients and 9 controls by using RNA-sequencing. They identified 1032 genes showing significantly different alternative splicing patterns between two conditions. PLP1 gene which is involved in myelination formation and DCLK1 gene which takes roles in neural migration and synaptic plasticity were among the genes showing differences in alternative splicing pattern and with neurological relevance [55].

An isoform of schizophrenia-associated DLG1 gene containing a new sequence from intron 3 was lower in DLPFC of early-onset schizophrenia patients with a disease onset age lesser than 18 years old compared to control. There was no significant difference between control and non-early-onset schizophrenia patients [56].

In a 2015 study carried on Chinese population found that short isoform of DRD2 gene containing exon 6 was found to be expressed higher in DLPFC region of schizophrenia patients due to schizophrenia-associated SNP located in intron 6 of the gene [57]. However, this result could not be repeated with non-Chinese populations, possibly indicating a population-specific role.

AS3MT gene which is located in a schizophrenia-associated locus also found to exhibit different alternative splicing patterns in schizophrenia patients compared to controls. A splice variant of this gene lacking exon2 and 3 was found to be

increased in DLPFC of schizophrenia patients together [58].

In 2018, Gandal et al. investigated aberrant splicing events between prefrontal cortex of 95 schizophrenia patients and 259 non-psychiatric controls. They identified 472 genes with DEU and these genes were enriched in pathways related to neuron development, cytoskeleton regulation and guanosine triphosphatase receptor. They specifically mentioned exon 4 of the GRIN1 gene, which was found to be expressed lower in schizophrenia patients, because of its known relevance to schizophrenia [59].

Schizophrenia-associated risk locus 11q25 and a protein-coding gene SNX19 which is the closest to the risk SNP in that locus were studied in another study. It was found that rare transcripts of SNX19 gene were expressed higher in DLPFC of schizophrenia patients. Among these transcripts, there are the ones that skip exon9 and that contain additional exons between exon8 and exon9. The study also revealed that DNA methylation regions at the transcription start site and inside exon2 were affected by the presence of risk alleles and control the expression of transcript isoforms [60].

Another group of researchers investigated the unfolded protein response (UPR) proteins and they found that spliced isoform of XBP1 gene was found to be higher in DLPFC of schizophrenia patients, which suggests an aberrant UPR activity in schizophrenia patients [61].

In a large recent study including 1497 post-mortem RNA-Seq data spanning 13 different brain regions found differentially spliced 4 genes: CYP2D6, SNX19, ARL6IP4 and APOPT1. The CYP2D6 gene was especially highlighted since it has known that it affects psychotic symptoms and cognitive performance of schizophrenia patients. The study revealed that variant of this gene lacking exon3 was found to be expressed higher in SCZ patients and this situation was found to be correlated with the presence of a particular SNP [62].

Although all of the above studies found gene/genes to be differentially expressed between schizophrenia and control groups, there is one study which could

not find a positive result but succeeded to be published. Moody et al. wanted to investigate adenosine kinase (ADK) gene since it is a long suspected hypothesis that adenosine by modulating dopamine and glutamate signaling may contribute to schizophrenia pathology. They couldn't find any difference in ADK splicing variants between in DLPFC of schizophrenia and controls [63].

All of the alternatively spliced genes are listed in Table 1.1.

Table 1.1: **List of alternatively spliced genes in schizophrenia patiens**

Gene	Study	Variant Property	R	Brain Region	Method
DRD3	Schmauss,1996	lacking 98 nts from exon 7	U	parietal cortex	PCR
GABRB2	Zhao et al.,2007	containing exon 10	D	DLPFC	qPCR
GABRB2	Zhao et al.,2009	containing intron 9,lacking exon 3	U	DLPFC	qPCR
GABRB2	Zhao et al.,2009	lacking exon 10 and exon 11	D	DLPFC	qPCR
GABRG2	Huntsman et al.,1998	containing exon 9	D	DLPFC	PCR
GRIN1	Clinton et al.,2003	containing exon 22,lacking exon 5 and 21	D	thalamus	in-situ h.
GRIN1	Le Corre et al.,2000	lacking exon 21 and exon 22	D	STG	in-situ h.
RGS4	Ding et al.,2009	lacking exon 2	D	DLPFC	qPCR
CTNNA2	Mexal et al.,2008	lacking exon 17,containing intron 6	D	hippocampus	qPCR
NCAM1	Gibbons et al.,2009	containing exon 18	D	BA46	qPCR
KCNH2	Huffaker et al.,2009	containing 5' extension of exon 3	U	hippocampus	qPCR
NRG1	Law et al.,2006	containing exon 2 (Type I)	U	hippocampus	qPCR
NRG1	Hashimoto et al.,2004	containing exon 2	U	DLPFC	qPCR
NRG3	Kao et al.,2010	containing exon 1 (Type I)	U	DLPFC	qPCR
NRG3	Kao et al.,2010	containing exon 3 (type IV)	U	DLPFC	qPCR
ERBB4	Law et al.,2007	containing exon 16 and exon 26	U	DLPFC	qPCR
GRM3	Sartorius et al.,2008	lacking exon 4	U	DLPFC	qPCR
DISC1	Nakata et al.,2009	lacking exon 3	U	hippocampus	qPCR
DISC1	Nakata et al.,2009	lacking exon 7 and exon 8	U	hippocampus	qPCR
DISC1	Nakata et al.,2009	containing unique exon 3	U	hippocampus	qPCR
QKI	Aberg et al.,2006	containing unique exon 7	D	frontal cortex	qPCR

Gene	Study	Variant Property	R	Brain Region	Method
QKI	Aberg et al.,2006	containing unique exon 6	D	frontal cortex	qPCR
MAG	Aberg et al.,2006	short isoform(containing first 4 exons)	D	frontal cortex	qPCR
PS2	Smith et al.,2004	lacking exon 5	U	frontal cortex	qPCR
KCC2	Tao et al.,2012	containing unique exon 7	U	DLPFC	qPCR
MCPH1	Oldmeadow et al.,2014	containing exon 6	D	DLPFC	exon array
DLG3	Oldmeadow et al.,2014	containing exon 8	D	DLPFC	exon array
ZC3H13	Oldmeadow et al.,2014	containing exon 13	D	DLPFC	exon array
BICD2	Oldmeadow et al.,2014	containing exon 1	D	DLPFC	exon array
ZNF804A	Tao et al.,2014	containing intron 2,lacking exon 1 and exon 2	D	DLPFC	RNA-Seq
CPNE3	Cohen et al.,2012	short 3'UTR region	U	BA10	exon array
ENAH	Cohen et al.,2012	containing exon 11	D	BA10	exon array
KLHL5	Cohen et al.,2012	containing exon 10	U	BA10	exon array
DLG1	Uezato et al.,2015	containing unique sequence from intron 3	D	DLPFC	qPCR
DRD2	Cohen et al.,2015	short isoform containing exon 6	U	DLPFC	qPCR
AS3MT	Li et al.,2016	lacking exon2 and exon 3	U	DLPFC	qPCR
SNX19	Ma et al.,2019	rare transcripts	U	DLPFC	RNA-Seq
XBP1	Kim et al.,2019	lacking 26 nts unconventional intron	D	DLPFC	qPCR
CYP2D6	Ma et al.,2020	lacking exon 3	U	DLPFC	RNA-Seq

R:Regulation Status, U:Upregulation, D:Downregulation, DLPFC: Dorsolateral Prefrontal Cortex, BA10: Broadmann Area 10, BA46: Broadmann Area 46, STG: Superior Temporal Gyrus, UTR: Untranslated Region, nts: nucleotides, in-situ h: in-situ hybridization

1.6 Research Question and Rationale

Schizophrenia is a complex disease which occurs through interaction of genetics and environmental stimuli according to one hypothesis of the disease [64]. Although there are many loci (108) associated with schizophrenia through GWAS studies, only a small fraction of them are located in exonic parts of genes and most of them are either intergenic or intronic [65]. Therefore, it is thought that most of the associated loci exert their effect through gene expression regulation, e.g. by affecting alternative splicing, by interacting with cis-acting elements etc. Among them AS is an appealing candidate to be studied because of several reasons: (1) splicing occurs more frequently in the brain than other tissues [66, 67], showing its importance for the brain functioning; (2) studies show that environmental stimuli can affect alternative splicing mechanisms, e.g. mice that are exposed to environmental stress exhibited AS changes in neuroxin gene in the hippocampi [68]; (3) Although most of the splicing events occurs during prenatal development (%83), there is still evidence showing postnatal AS changes [20]; (4) disruptions in AS events are found in many neuropsychiatric disorders including schizophrenia, bipolar disorder and autism [59]. As a result of reasons that are explained above we wanted to investigate aberrations in alternative splicing that may contribute to schizophrenia pathophysiology. Literature research revealed that there are different ways to study AS in schizophrenia:(1) Investigating the effects of schizophrenia-associated SNPs on AS of genes that are related to these SNPs; (2) Investigating AS events in large scale (by using microarrays, RNA-Seq data) and use PCR to prove the most significant ones; (3) Investigating AS of schizophrenia-associated genes in psychiatrically healthy brains and then look at them in schizophrenia brains; (4) Large-scale investigation of AS events in psychiatrically healthy brains and then to at schizophrenia brain. Our approach is more similar to the last (4) study type since we think that in order to understand what is disrupted in a diseased brain, we should first understand what can be considered as a "normal/healthy" in a healthy brain. In the aim of this, we determined our research question as "What are alternative splicing changes that occur in brains of healthy individuals during postnatal developments, especially during the AYA period, and if these changes help us better understand schizophrenia

pathophysiology?". To answer this question we first investigated AS changes that take place in the brains of healthy individuals by comparing AS events in adolescence and young adulthood period (AYA) that comprises the age range between 15 and 23 to other developmental time periods: Infancy, early childhood, middle and late childhood, young adulthood, middle adulthood and lastly late adulthood. The reason why we compare the AYA period to all other developmental periods is that it represents the critical time period for schizophrenia since many people develop the disease during this period. In order to address the second part of our research question we investigated the genes that show developmental AS changes in healthy brains in a schizophrenia dataset by comparing RNA-Seq data of schizophrenia patients to healthy controls.

Chapter 2

Methods

2.1 Exon Microarray Data

Table 2.1: **Brain regions included in the study and array numbers for each region across the groups including right and left hemispheres**

Groups	DLPFC	OFC	MFC	VFC	M1C
I	7(2R,5L,n=5)	7(2R,5L,n=5)	7(2R,5L,n=5)	7(2R,5L,n=5)	6(2R,4L,n=4)
EC	6(3R,3L,n=4)	5(2R,3L,n=3)	6(3R,3L,n=4)	7 (3R,4L,n=5)	7(3R,4L,n=5)
MLC	5(1R,4L,n=4)	5(1R,4L,n=4)	5(1R,4L,n=4)	5(1R,4L,n=4)	4(1R,3L,n=3)
AYA	9(3R,6L,n=6)	9(3R,6L,n=6)	9(3R,6L,n=6)	9(3R,6L,n=6)	9(3R,6L,n=6)
YA	9(4R,5L,n=5)	11(5R,6L,n=6)	11(5R,6L,n=6)	8(4R,4L,n=4)	8(4R,4L,n=4)
MA	6(2R,4L,n=4)	6(2R,4L,n=4)	6(2R,4L,n=4)	6(2R,4L,n=4)	5(2R,3L,n=3)
LA	6(3R,3L,n=3)	5(3R,2L,n=3)	6(3R,3L,n=3)	5(3R,2L,n=3)	6(3R,3L,n=3)

OFC: Orbital Prefrontal Cortex, MFC: Medial Prefrontal Cortex, DLPFC: Dorsolateral Prefrontal Cortex, VFC: Ventrolateral Prefrontal Cortex, M1C: Primary Motor Cortex; the numbers outside parentheses indicate array numbers; R: right hemisphere, L: left hemisphere, n: number of individuals providing brain regions

For the analysis of differential exon usage (DEU) changes in the frontal cortex of healthy individuals throughout their lifetime, publically available Affymetrix Human Exon 1.0 ST array data with a study accession code GSE25219 provided by Gene Expression Omnibus (GEO) website (https://www.ncbi.nlm.nih.gov/geo/) was used. This dataset contains 1,340 samples coming from 57 healthy postmortem human brains with age range from 8 PCW (post-conceptual week) to 82 years old. It comprises 16 different brain regions including 11 regions of neocortex, hippocampus, amygdala, thalamus, striatum, and the cerebellar cortex. For the specific purposes of this research, data was filtered in a way that it will include arrays that are coming from frontal cortex regions (OFC: Orbital Prefrontal Cortex, DLPFC: Dorsolateral Prefrontal Cortex, VFC: Ventrolateral Prefrontal Cortex, MFC: Medial Prefrontal Cortex, M1C: Primary Motor (M1) Cortex), and covering age range from birth to late adulthood (0 to 82 yr.). Then, filtered data was separated into developmental age groups: Infancy (I) $0 < \text{Age} \leq 10\text{M}$, Early Childhood (EC) $1\text{Y} \leq \text{Age} \leq 4\text{Y}$, Middle and Late Childhood $8\text{Y} \leq \text{Age} \leq 13\text{Y}$, Adolescence and Young Adulthood $15\text{Y} \leq \text{Age} \leq 23\text{Y}$, Young Adulthood $27\text{Y} \leq \text{Age} \leq 37\text{Y}$, Middle Adulthood $40\text{Y} \leq \text{Age} \leq 55\text{Y}$, Late Adulthood $60\text{Y} \leq \text{Age} \leq 82\text{Y}$. These developmental groups were created based on developmental groups used in Kang et al. (2011) article [20]. However, we created novel developmental groups which are appropriate for our research question. One of the periods which normally do not exist in the article but we created is adolescence and young adulthood period covering age range between 15 to 24 years old. This novel period was created since it is known to be a critical time to develop schizophrenia. We compared the alternative splicing changes that occur during AYA period to all other developmental periods listed above. Also in the article young adulthood covers the age range between $20 \leq \text{Age} < 40$ but the AYA period contains individuals younger than 24 years old; and therefore the remaining individuals older than 24 years old were included in a separate period called young adulthood. The characteristics and detailed information about accession number, age, sex of individual arrays for each group can be found in the supplementary table 5.1.

2.2 Affymetrix Human Exon 1.0 ST Arrays

Exon microarray method depends on hybridization between probes on chips that are complementary to exons and fluorescently label cDNA molecules. When this hybridization occurs, a detector recognizes where the signal is coming from and how strong it is in order to determine the identity and amount of exon in a sample. Unlike conventional microarray chips which depends on the 3' polyA tail of mRNAs for hybridization, exon arrays are designed in a way that they can bind to any cDNA generated by using random hexamers. These arrays have 5.4 million probes which are grouped into 1.4 million probesets, spanning over 1 million exon clusters. Probes are selected from regions called as Probe Selection Regions (PSR) ranging in size from 123 bp to 25 bp. Most of the time PSRs reflect an exon but sometimes, because of overlapping exon structures, they may reflect the subset of an exon. Nearly 90% of PSRs is reflected by 4 probes called as a probe set. These probes are perfect match (PM) to the PSRs and in order to detect background noise a set of background probes having sequences that are not present in the human genome and having the same GC content with each PM probe are used.

2.3 Alternative Splicing Analysis

Alternative splicing analysis is carried out by comparing alternative exon usage that occurs during adolescence and young adulthood (AYA) period to other developmental groups listed in Table 2.1 by using AltAnalyze software [69]. It is a user-friendly tool to analyze alternative splicing events for splicing sensitive platforms such as RNA-seq and exon microarrays.

Analysis at the exon level requires the correct gene level and exon level intensities. This is achieved through AltAnalyze via a pipeline that is schematized below in Fig. 2.1:

Gene level intensities were calculated by RMA (Robust Multichip Analysis)

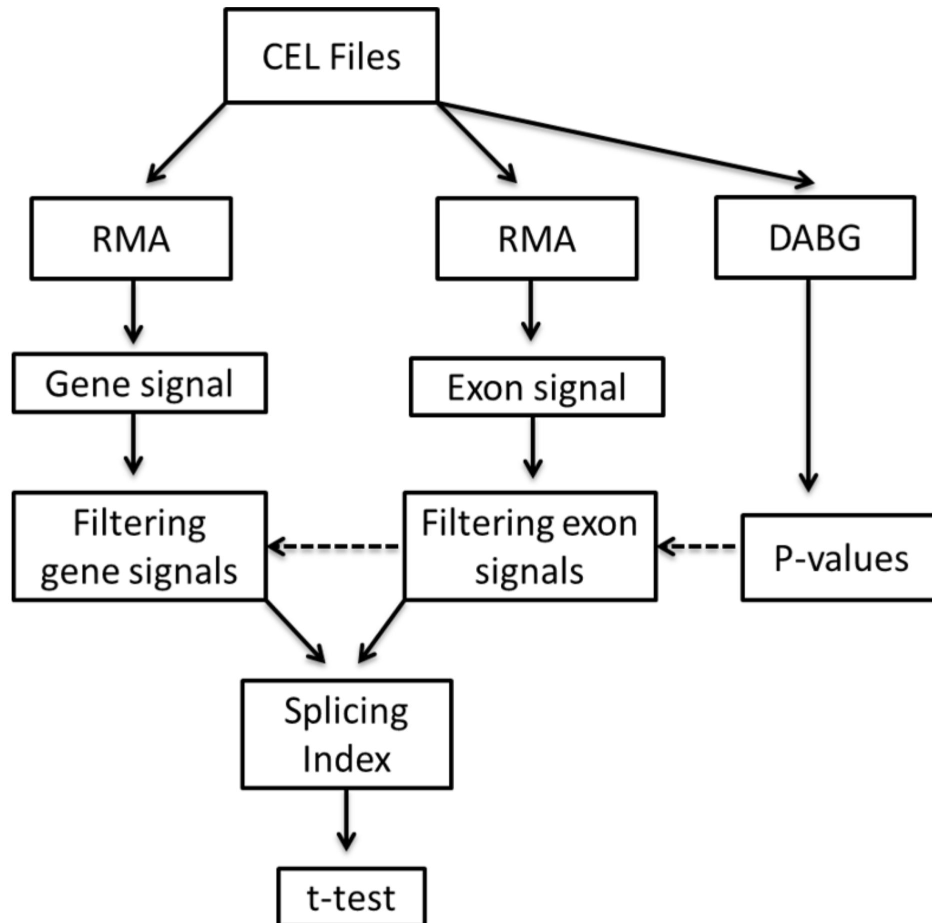


Figure 2.1: **The Schema of AltAnalyze Pipeline.** This figure explains the alternative splicing analysis pipeline of AltAnalyze software. RMA: Robust Multichip Analysis, CEL: extension of microarray files, DABG: Detection Above Background

method using constitutive probesets which are defined as exons that are common to all known transcript of a gene. Meanwhile, DABG (Detection Above Background) p-values are generated for each probeset using background probes. Any probeset that has a mean DABG p-value > 0.05 in one of the biological groups are excluded.

Exon level intensities were calculated by RMA method considering all probesets including core, extended and full probesets. These names reflect the evidence levels of exons.

Core probesets reflect the well-validated RefSeq transcripts and full length mRNAs while extended probesets refers to cDNA-based transcript annotations and full probesets refers to probesets which are predicted computationally. Any non-constitutive probesets that has a mean DABG p-value > 0.05 for both compared biological groups were excluded.

After, exon and gene level intensities are normalized and filtered according to user-defined parameters alternative splicing analysis is carried out by calculating splicing index (SI) values for each probeset. In order to calculate SI values, firstly gene level normalized (NI) intensities of probesets for each biological group are calculated by dividing the probeset intensity to the gene intensity that it belongs to:

$$GeneLevelNormalizedIntensity(NI) = \frac{Probesetintensity}{Geneintensity} \tag{2.1}$$

Then SI values are calculated by taking the differences of log2 of NI values for each biological sample:

$$SplicingIndex(SI) = log2\frac{NIofSample1}{NIofSample2} \tag{2.2}$$

To identify probesets that are significantly differentially expressed between two

groups, t-test is applied to NIs of each sample and corrected p-values (Benjamini-Hochberg) are reported together with SI values. At this step, several parameters are needed to be defined: one of them is minimum alternative exon score which is set to 2, indicating that SI ≥ 2 will be reported in the results; other is maximum absolute gene expression change which is set to 3 (non-log), indicating that if gene expression change is more than 3 fold between two samples, it will not be reported as alternatively spliced. This is important since if a gene is differentially expressed between two samples, it is possible that many of its constitutive exons will be reported as alternatively spliced.

As indicated above alternative splicing analysis was first carried out by including all regions of prefrontal cortex (OFC, MFC, DLPFC, VFC and M1C), and then the same analysis was repeated by including only dorsolateral prefrontal cortex (DLPFC) region in order to make results more comparable to RNA-Seq results since RNA-Seq data only contains DLPFC region.

2.4 RNA Sequencing Data

One of the aims of this research is to find if developmental DEU changes that occur in healthy individuals are different for individuals with schizophrenia. For this purpose, available splicing sensitive datasets that belong to frontal cortex of individuals with schizophrenia were searched. Unfortunately, there were no exon microarray dataset. Among the RNA-Seq datasets available for download and downstream in silico analyses, they were all from dorsolateral prefrontal cortex (DLPFC) region. Therefore, RNA-Seq dataset with a project code PRJNA319583 was chosen since it contains more samples compared to others.

Table 2.2: **Sample information of RNA-Seq data included in the study**

	Number of Individuals	Brain Region	Age	Gender
Schizophrenia	24	DLPFC	42.67±9.89	3F 21M
Control	24	DLPFC	50.25±12.24	3F 21M

DLPFC: Dorsolateral Prefrontal Cortex,F: female, M: male

PRJNA319583 public dataset contains 352 samples coming from postmortem brains of 24 individuals with schizophrenia, 24 individuals with bipolar disorder, 24 individuals with major depression disorder, and 24 healthy individuals. It comprises 3 different brain regions including the nucleus accumbens, dorsolateral prefrontal cortex and anterior cingulate cortex. For the specific purposes of this research, only DLPFC samples that belong to schizophrenia and healthy controls are included in the study. Information related to schizophrenia and control healthy groups are summarized in table 2.2.

2.5 RNA-Sequencing Data Retrieval

It is important to know how RNA-seq data is obtained since this affects downstream analyses. Total RNA from postmortem brain tissues was isolated and purified. RNA-seq libraries were prepared by using poly(A) selection and transposase-based non-directional library construction. With poly(A) selection method only mRNAs that contain poly(A) tails on their 3' terminal by using poly(T) primers were amplified and they were converted into cDNAs. Double-stranded (ds) cDNAs were incubated with transposon complexes (transposase enzyme and transposon that contain a sequence including Illumina adapter sequences). With the help of hyperactive transposases, ds cDNA molecules were fragmented and adapter sequences were ligated to the ends of these fragments. This type of library construction is inherently not strand specific, i.e. the information about the strand origin of transcripts is lost [70]. The prepared library were sequenced with Illumina Hiseq 2000 sequencing machine, producing paired-end and 50 bp long reads. How these reads are generated is explained below:

After cDNAs are ready, they are hybridized onto Illumina chips (flowcells) containing oligos complementary to one of the adapters found on the cDNA ends. Once they are hybridized, a polymerase synthesizes the complementary strands of these cDNAs, making all double stranded. Then by washing out original templates, newly synthesized sequences are now tethered to the ground of chips. Now cDNAs are amplified through a process called bridge amplification. This

time adapter region at the free ends of cDNAs hybridizes to the second type of oligo on the chips, making bridge-like appearance. Polymerases synthesize the complementary strands, making a double-stranded bridge. Then these bridges are denatured and bridge amplification process is repeated many times until many copies of cDNAs are obtained. After this clonal amplification, reverse strands are cleaved and washed away, leaving only forward strands on the chips. Free 3' ends are blocked to prevent hybridization with second type of oligos. Then by binding of sequencing primers, polymerases begin to add fluorescently-tagged nucleotides to the growing chain, and each nucleotide releases a characteristic light that can be recognized by a detector, allowing the sequence of nucleotides in the chain. After synthesis, read products are washed away and blocking of 3' ends are removed. Free ends now hybridizes with second type of oligos on the chips. By bridge amplification reverse strands are synthesized and forward strands are washed away, leaving only reverse strands on the template this time. Reverse templates are sequenced in the same manner with forward strands. Since both reverse and forward strands are sequenced, this data contains pairs and called paired-end.

2.6 RNA-Seq Data Analysis

1.Quality Control

2.Trimming of Low Quality Reads

3.Alignment of Reads to the Reference Genome

4.Counting of Reads

5.Alternative Splicing Analysis

2.6.1 Quality Control

For quality control of fastq files FastQC tool was used. If we run the below code this tool will provide us two web-based result in a specified output directory.

```
"path/to/file/fastQC" -o "path/to/output/directory" "path/to/fastq1.gz"
"path/to/fastq2.gz"
```

It checks for per base sequence quality, per base GC content, duplicated or overrepresented sequences, adapter content etc. In our analysis, when fastq files were first checked with fastQC, there were no files containing adapter sequences at a problematic level. Therefore, we did not use adapter trimming tool.

2.6.2 Trimming of Low Quality Reads

Although adapter content was not a problem, some sequences especially at the ends of the reads were problematic since their quality scores were low. We used Trimmomatic tool in order to remove and filter low quality reads according to user defined parameters.

```
java -jar "path/to/file/trimmomatic" PE
"path/to/fastq1.gz" "path/to/fastq2.gz"
"path/to/output/directory/fileP1.fastq.gz"
"path/to/output/directory/fileU1.fastq.gz"
"path/to/output/directory/fileP2.fastq.gz"
"path/to/output/directory/fileU2.fastq.gz" LEADING:30 TRAILING:30
SLIDINGWINDOW:4:20 MINLEN:30
```

The above code takes 2 fastq files containing paired-end reads (indicated as PE) and trims low quality reads. First, bases at both ends of a read are trimmed if their phred scores are below 30 (specified by leading and trailing in the code). Trimmomatic also scans reads with a window size 4 and if the average phred score of these 4 bases drops under 20, it removes all 4 bases in that window and

remaining reads till the 3' ends of reads (specified by sliding window:4:20 in the code). If minimum length of a read drops under 30 bp, after this trimming process, they are also eliminated (specified by minlen:30 in the code). When trimming and filtering is completed, four files are generated as output: FileP1 contains paired reads in forward reads while FileU1 contains unpaired forward reads. Similarly, FileP2 contains paired reverse reads and FileU2 contains unpaired reverse reads. The reason of the presence of unpaired reads may be the elimination of a pair due to low quality or its pair may not be present at all.

In our analysis when fastq files were trimmed, the percentage of paired reads were between 60%-90%. Only one pair showed 0% pairing ratio most likely due to they were wrong pairs. This pair was eliminated from further analysis.

2.6.3 Alignment of Reads to the Reference Genome

After removing low quality reads, the next step is to align reads to a reference genome. We have chosen HISAT2 software for the alignment process since it is a splice-aware aligner tool which was suitable for studying differential exon usage.

```
"path/to/file/hisat2" -x "path/to/file/grch38_tran/genome_tran" -1
"path/to/output/directory/fileP1.fastq.gz" -2
"path/to/output/directory/fileP2.fastq.gz" -S
"path/to/output/directory/aligned.sam"
```

As a reference genome, we used release 84 version of GRCh38 Homo sapiens (human) reference genome from Ensembl. We continued the analysis with paired-end reads (fileP1 and fileP2) since for splicing analysis, it is recommended to use paired end and long reads [71].

The above code takes 2 paired and trimmed fastq files and align them to the reference genome and give output as one file with an extension .sam. For all of the fastq files, overall alignment rate were greater than 90%. For only one sam file was problematic with 0% overall alignment. This file was eliminated from

further analysis.

2.6.4 Annotation and Counting of Reads

Once reads are aligned to the correct positions on the genome, they need to be annotated and counted. We used HTSeq python package for this purpose since it is recommended by DEXseq R package that we will use after this step. DEXseq provides two HTSeq pyton scripts that we can use for annotation and for counting.

In order to prepare annotation file an appropriate GTF file is required. GTF file contains start and end positions of every feature (exons, genes, and transcripts) and their attributes (gene id, transcript id, exon number etc.). To be used as an annotation file this GTF file should be converted to a GFF file which contains collapsed exon counting bins. Counting bins refers to parts of exons since one exon can be a part of different transcripts with different boundaries, and with this process information about varying boundaries is stored in counting bins. If a read overlaps with more than one counting bin they are all counted in order to obtain count number for this read. We obtained GTF file from Ensembl since we also used Ensembl reference genome for alignment.

```
python
"path/to/library/DEXseq/python_scripts/DEXseq_prepare_annotation.py"
"path//to/file/Homo_sapiens.GRCh38.98.gtf"
"path/to/file/Homo_sapiens.GRCh38.98.gff"
```

The above code takes GTF file that we provide and converts it to GFF file.

After GFF file is ready, the next step is counting.

```
python "path/to/library/DEXseq/python_scripts/DEXseq_count.py" -p yes -s
no "path/to/file/Homo_sapiens.GRCh38.98.gff"
"path/to/file/aligned.sam" "path/to/output/directory/count.txt"
```

The above code takes sam file and prepared annotation file (GFF) and outputs a txt file which contains two columns. The first column includes counting bin IDs which are composed by gene ID and exon bin number, e.g. ENSG00000000003:001. The second column contains count numbers corresponding to each counting bin. Since our data is paired end (-p yes) and strand information is not protected during library preparation (-s no), these information should be provided to code.

2.6.5 Alternative Splicing Analysis

Until now, all the analyses were carried out in a linux environment; but, from now on we will continue with R software in windows. As indicated above, differential exon usage analysis was carried out by DEXseq package of R which is specifically developed for this kind of analysis.

DEXseq first reads count.txt files generated for all sam files into R environment. Also we need to read GFF file generated and a sample table containing information related to conditions (i,e. which files belong to control group and which to schizophrenia) into R environment. DEXseq creates an object called "dxd" that contains data arranged in a matrix. Then data was normalized. As a normalization method, DEXseq does'nt use classical RNA-seq data normalization methods such as RPKM/FPKM and TPM since a total number of reads is not always a good option especially when few highly and differentially expressed exons may have a big effect on total read count. Instead, DEXseq calculates a size factor (S) for each sample by taking medians of all counts (of counting bins) divided by geometric mean of each in all samples which can be explained by the below equation:

$$S_{sample} = Median \frac{count for exon 1}{geometric mean of exon 1 in all samples} \frac{count for exon 2}{geometric mean of exon 2 in all samples} \ldots \tag{2.3}$$

The counts are then normalized by dividing them by the size factors of their

samples. After normalization, DEXseq carries out dispersion estimation analysis. This is necessary to be able to distinguish real expression changes from noise that occur among biological replicates. Therefore, it calculates dispersion (variation) within biological replicates to catch the real effects of conditions (in our case controls vs schizophrenia) on DEU. With the information of size factors and dispersion estimates, DEU analysis is carried out by fitting generalized linear models for each exon bin. Finally differentially spliced exons between schizophrenia patients and psychologically healthy control group are found by chi-square likelihood ratio test.

Finally, results were visualized by plotting functions of DEXseq.

All information regarding RNA-seq analysis is written by the help of DEXseq online manual [72].

2.7 Investigation of the Effect of Alternatively Spliced Exons

In order to find the effects of alternatively spliced exons on protein function, we first tried to find which transcript forms are expressed in the frontal cortex region from GTEx website. GTEx website is a publically available online tool which includes expression of exons/transcripts/genes for 54 different healthy tissues that come from nearly 1000 individuals (https://gtexportal.org/home/) [73]. After finding the transcripts that are expressed in the frontal cortex, cDNA sequences were found from the Ensembl webpage. Expasy translate online tool was used to translate cDNA sequences into protein sequences. Motif scan online tool was used to find motifs that are found on the transcripts by using protein sequences. However for long protein sequences this website was not functional: therefore, another online tool called ScanProsite was used for this purpose. Then the effects of amino acids that are encoded by the alternatively used exons that we identified were investigated.

Chapter 3

Results

3.1 Results of Exon Microarray Analysis

We have compared differential exon usage between adolescence and young adulthood (AYA) period to all other developmental periods listed in table 2.1 (Infancy (I), Early Childhood (EC), Middle and Late Childhood (MLC), Young Adulthood (YA), Middle Adulthood (MA), and Late Adulthood (LA)). We first carried out the analysis by including all regions of prefrontal cortex (OFC, MFC, DLPFC, VFC and M1C), and then the same analysis was repeated by including only dorsolateral prefrontal cortex (DLPFC) region in order to make results more comparable to RNA-Seq results since RNA-Seq data only contains DLPFC region.

Both PFC and DLPFC regions displayed a similar pattern of developmental exon expression. As we move away from our reference period (AYA) the number of differentially expressed exon level probesets increases for both regions. According to the literature, most of the exon expression changes occurs early in the development [20, 24] and this can also be observed in Fig. 3.1, the greatest exon difference being observed between I and AYA groups. As it is expected since only one region included, DLPFC region comparisons reveal lesser number (1286 vs

865) of exon probesets compared to PFC region comparisons.

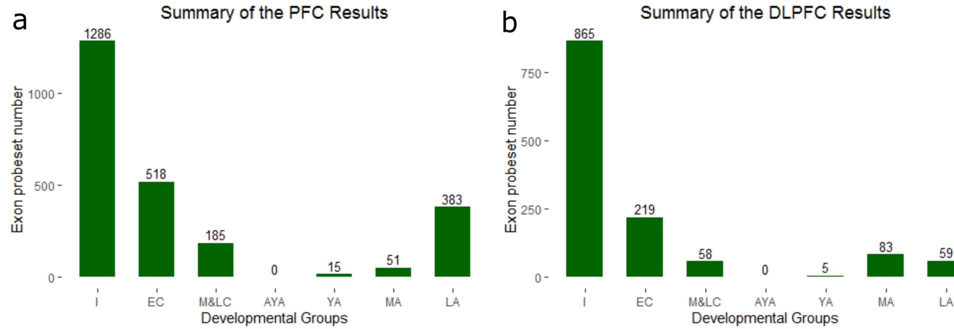


Figure 3.1: **Summary of the Exon Array Analysis.** These bar graphs summarizes results of group comparisons at exon level for both prefrontal cortex (PFC) and dorsolateral prefrontal cortex (DLPFC). (a) The number of significantly differentially expressed probesets (corrected p ¡0.05) as a result of comparison between adolescence and young adulthood (AYA) period and other periods listed in the x-axis (Infancy (I), early childhood (EC), middle and late childhood (MLC), young adulthood (YA), middle adulthood (MA) and late adulthood (LA) including array data coming from all regions listed in Table 2.1. (b) The number of significantly differentially expressed genes (corrected $p < 0.05$) as a result of comparison between AYA period and other periods listed in the x-axis including array data coming only from DFC. Numbers on the top of each bar indicate exact exon probeset numbers as a result of each comparison.

3.1.1 Developmental Alternative Splicing of PALM Gene

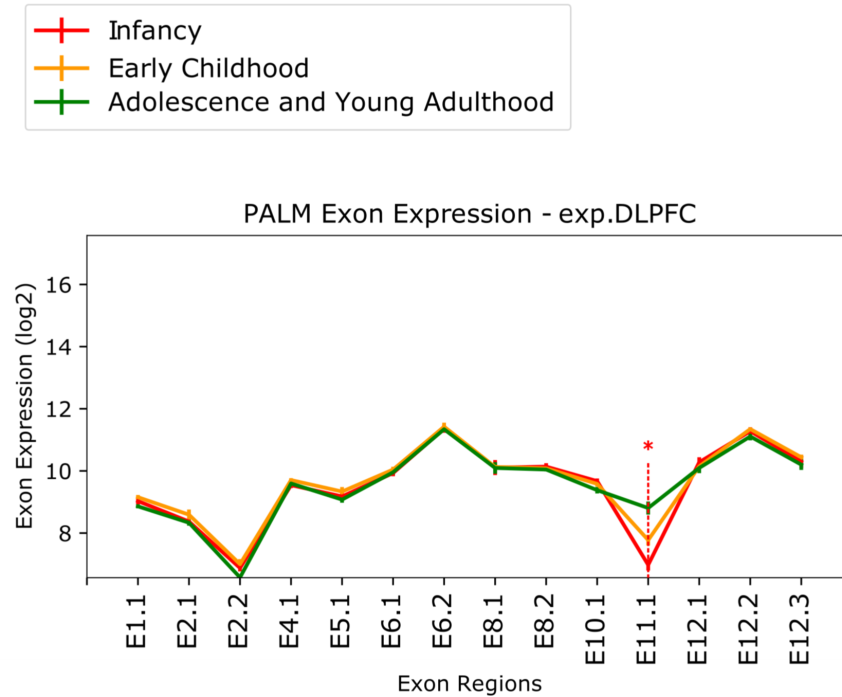


Figure 3.2: **The Expression of Exons of PALM Gene.** Log2 expressions of exons of PALM gene during infancy (red), early childhood (orange) and adolescence and young adulthood (dark green) periods.

Fig. 3.2 shows the expression of exon probests of PALM gene in different developmental periods in DLPFC. It can be seen that 8th exon of PALM gene (location: chr19: 740,352-740,483, Hg38), which is shown as E11.1 in the graph, is expressed significantly lower during infancy and early childhood periods compared to adolescence and young adulthood period. According to this graph, it seems that the expression of this exon increases gradually as the brain grows from infancy through AYA period in the dorsolateral prefrontal cortex while the expressions of other exons does not significantly change through development. This result is compatible with early findings of Mazin et al. (2012). They found that the expression of the same exon is low in newborns (2 days to 35 days) but increases

in young and old samples [24] according to both RNA-Seq and PCR results.

3.1.2 Developmental Alternative Splicing of MAPT Gene

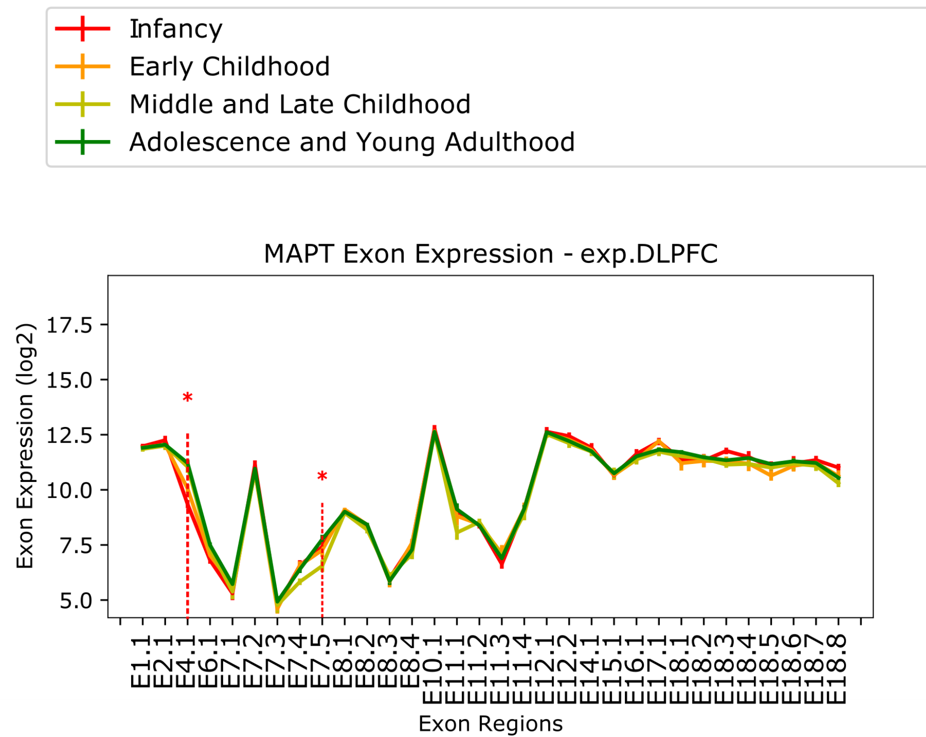


Figure 3.3: **The Expression of Exons of MAPT Gene.** Log2 expressions of exons of MAPT gene during infancy (red), early childhood (orange), middle and late childhood (light green) and adolescence and young adulthood (dark green) periods.

Fig. 3.3 shows the expression of exon probests of MAPT gene in different developmental periods in DLPFC. It can be seen that 3rd exon of MAPT gene (location: chr17: 45,971,876- 45,971,945 Hg38), which is shown as E4.1 in the graph, is expressed significantly lower during infancy and early childhood periods compared to adolescence and young adulthood period. Another exon which is not defined among ensemble transcripts for this gene but located between 4th exon and 5th exon of MAPT gene (location: chr17: 45,979,996-45,980,105 Hg38),

which is shown as E7.5 in the graph, is expressed significantly lower during middle and late childhood period compared to AYA period.

Although Mazin et al. found MAPT gene in their study as developmentally spliced, they have found different regions than our study: One region (location: chr17: 45,972,027-45,972,086) which is located between 3rd and 4th exons is low in newborns but increases in young (25 years to 28 years) samples according to both RNA-Seq and PCR results; another region (location: chr17: 46,010,478-46,010,570) which is located between 7th and 8th exons is low in newborns but immediately increases in 182 days to 274 days old samples according to both RNA-Seq and PCR results [24].

3.1.3 Developmental Alternative Splicing of NRXN1 Gene

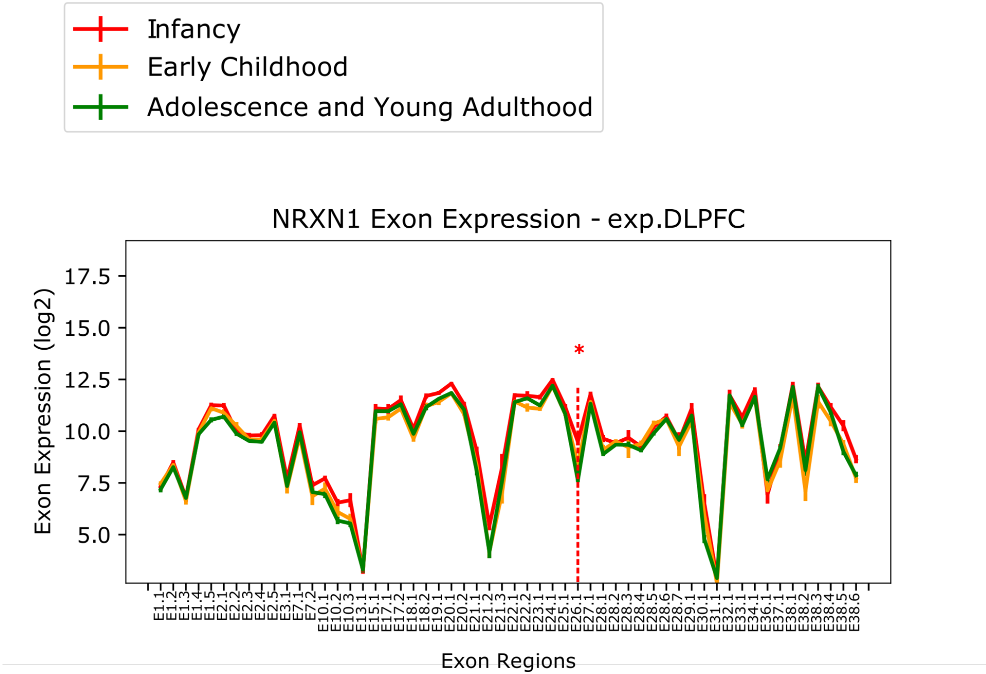


Figure 3.4: **The Expression of Exons of NRXN1 Gene.** Log2 expressions of exons of NRXN1 gene during infancy (red), early childhood (orange), and adolescence and young adulthood (dark green) periods.

Fig. 3.4 shows the expression of exon probests of NRXN1 gene in different developmental periods in DLPFC. An exon sequence (location: chr2: 50,466,455-50,466,493) located between the 16th exon and 17th exon of NRXN1 gene and depicted as E26.1 in the graph is significantly expressed higher during infancy and early childhood periods compared to adolescence and young adulthood period. Although there are many studies related to NRXN1 gene splice variants, there is no study finding developmental expression change for this specific exon.

Although only the exon expression graphs of MAPT, PALM and NRXN1 genes are shown here, results indicate that many neuropsychiatric disorder-related

genes show dynamic alternative splicing changes during different developmental stages. These include: SHANK2 associated with several neuropsychiatric disorders including autism spectrum disorder, schizophrenia and bipolar disorder [74]; SEMA3B associated with schizophrenia through GWAS study [75]; CLASP2 regulates neuronal polarity and synaptic function [76]. These indicate that developmental splicing changes may be important in the pathophysiology of many neuropsychiatric disorders.

3.2 Results of RNA-Seq Analysis

Differential exon usage analysis between schizophrenia and psychologically-healthy groups revealed 3340 exons (1648 genes) which have FDR corrected p values smaller than 0.05. This number corresponds to nearly 0.5% of exons that are being tested and nearly 3% of all genes tested are found to be alternatively spliced.

3.2.1 Alternative Splicing of GRIN1 Gene between schizophrenia and Healthy Groups

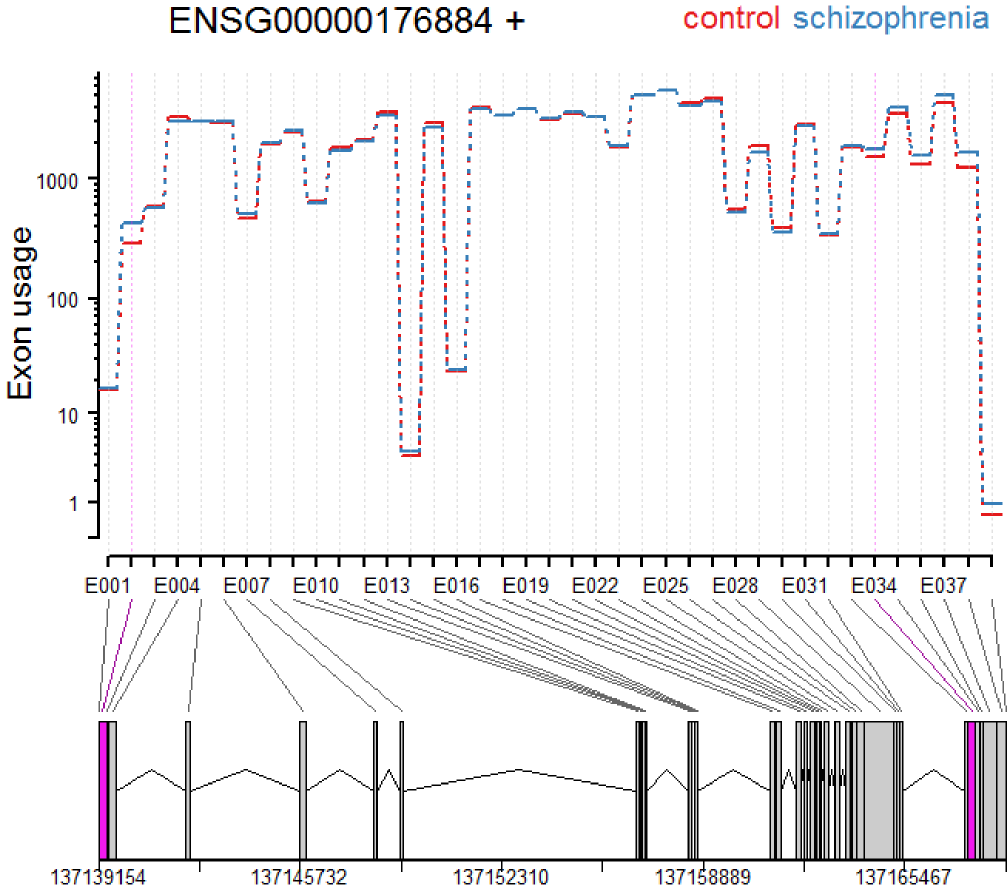


Figure 3.5: **The Expression of Exons of GRIN1 Gene.** Expression graph on the top shows fitted expression of exons of GRIN: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all exons of GRIN1 transcripts and genomic locations of them: pink bars indicate differentially expressed exons (FDR corrected p value ¡0.1) between schizophrenia and healthy groups.

Two exons of GRIN1 gene which is depicted as E002 (location: chr9:137,139,175-137,139,409) and E034 (location: chr9:137,167,535-137,167,773) in Fig. 3.5

are differentially used between schizophrenia and healthy groups. DEXseq takes FDR corrected p values smaller than 0.1 as significant, therefore it takes both E002 (FDR corrected p value=0.03110320) and E034 (FDR corrected p value=0.09083872) as significant. This gene was previously implicated in schizophrenia to be alternatively spliced. By in situ hybridization with probes that target exon 21 (location: chr9:137,167,530-137,167,574) and exon 22 ((location: chr9:137,167,636-137,167,680), Le Corre et al. found that expression of these exons were higher in the superior temporal gyrus of schizophrenia patients [36]. Although our data contains dorsolateral prefrontal cortex, the same conclusion can be driven since E034, which covers the same region indicated by the exon21 and exon22 in the study of Le Corre et al., is expressed higher among schizophrenia sample. Different than literature, another exon shown as E002 in the graph is also found to be expressed higher among schizophrenia sample.

The expressions of the exons which are depicted as E001, E014, E016 and E039 are very low compared to other exons, which is probably due to low expression of transcripts containing these exons in the brain.

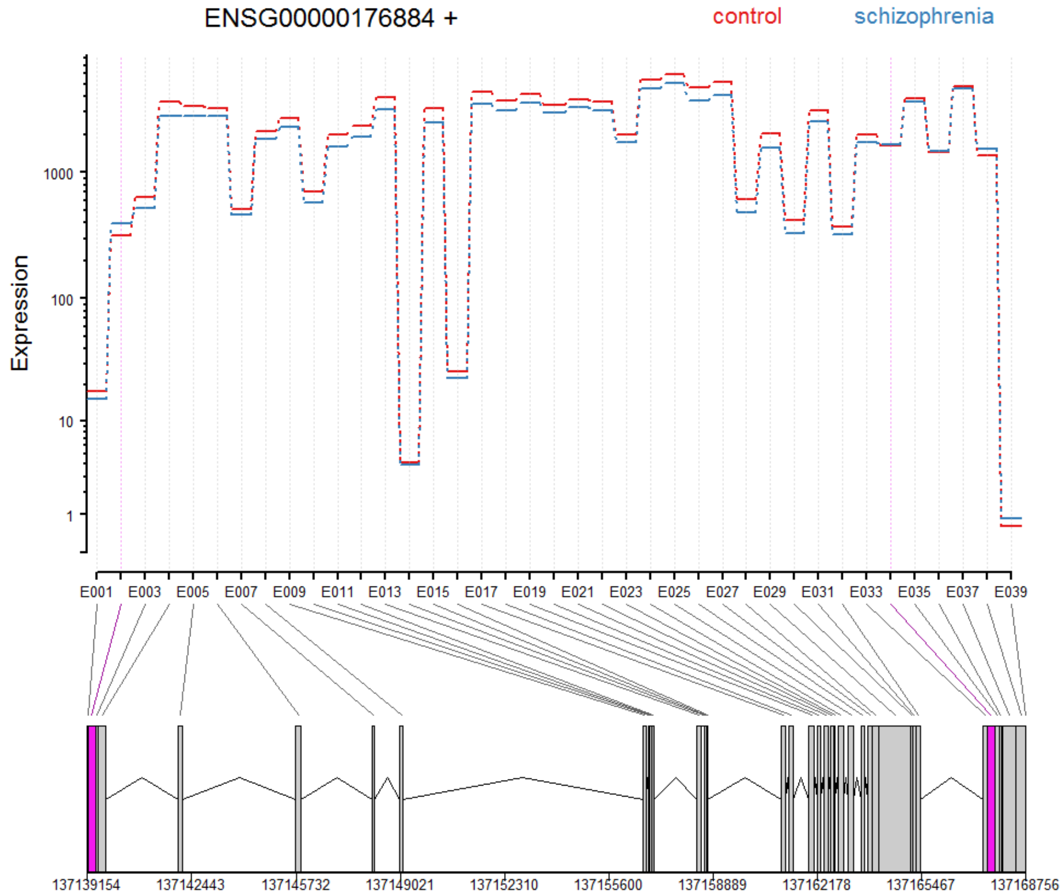


Figure 3.6: **The Expression of GRIN1 Gene.** Expression graph on the top shows fitted expression of exons of GRIN with the effect of overall gene expression: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all exons of GRIN1 transcripts and genomic locations of them: pink bars indicate differentially expressed exons (FDR corrected p value ¡0.1) between schizophrenia and healthy groups.

Although in Fig. 3.5 overall gene expression effect cannot be seen in order to see the differential exon usage better, DEXseq also provides exon expression graph including overall gene expression effect of GRIN1 gene which can be seen in Fig. 3.6. It can be seen that overall gene expression is lower for GRIN1 gene among schizophrenia samples but E002 and E034 behaves differently than

other exons of the GRIN1 gene: the fitted expression of E002 is higher among schizophrenia patients while the fitted expression of E034 is the same for both groups, indicating their differential usage.

Other than GRIN1 gene, many differentially spliced genes in the context of schizophrenia are detected in our study including RGS4, NCAM1, KCNH2, NRG1, QK1, MAG, DLG3, PLP1 and SNX19 as differentially spliced between schizophrenia and healthy groups. We only investigated splicing of the GRIN1 gene as an example here since our research question is different.

3.3 Genes That Show both Developmental and Schizophrenia- Associated Alternative Splicing

In the pursuit of our research question "What are alternative splicing changes that occur in brains of healthy individuals during postnatal development, especially during the AYA period, and if these changes help us better understand schizophrenia pathophysiology", we wanted to find which exons that are differentially expressed between schizophrenia and healthy groups also show developmental expression change. Therefore we found the common exons between these two studies. Only 4 exons that belong to 3 different genes were common: AKAP7, BAIAP3, and SEMA3B. While investigating the common exons between two studies, only exons with an FDR corrected p values smaller than 0.05 were included.

3.3.1 Developmental and Schizophrenia-Associated Alternative Splicing of AKAP7 Gene

3.3.1.1 Developmental Alternative Splicing of AKAP7 Gene

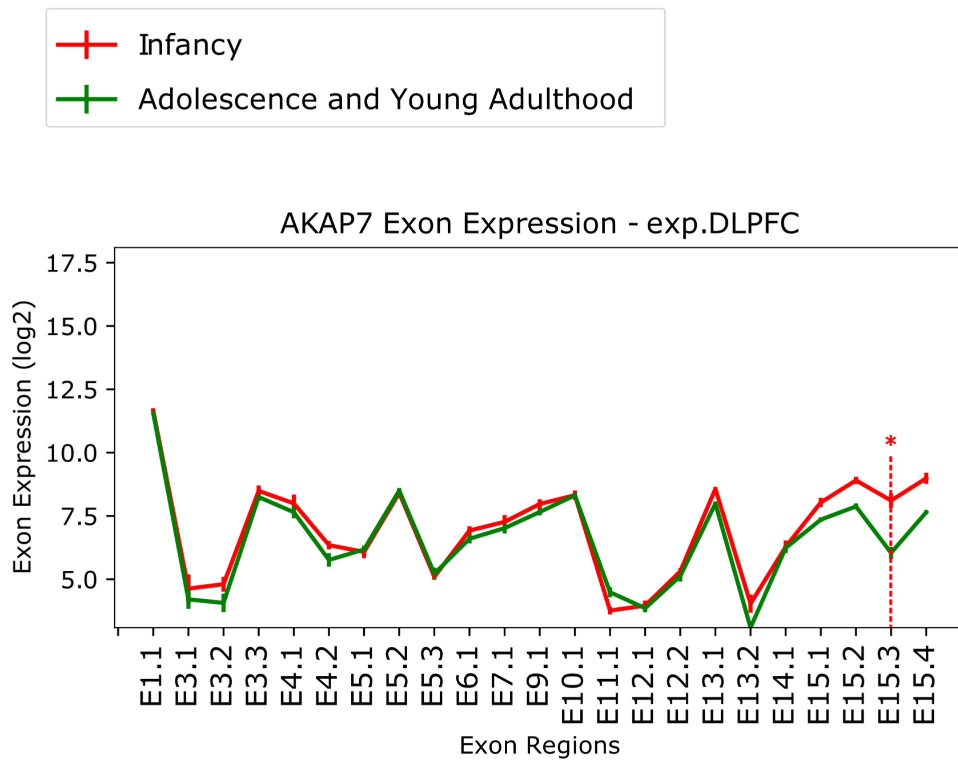


Figure 3.7: **The Expression of Exons of AKAP7 Gene.** Log2 expressions of exons of AKAP7 gene during infancy (red), and adolescence and young adulthood (dark green) periods.

Fig. 3.7 shows the expression of exon probests of AKAP7 gene in different developmental periods in DLPFC. It can be seen that 8th and the last exon of AKAP7 gene (location: chr6: 131,282,152- 131,282,358), which is shown as E15.3 in the graph, is expressed significantly higher during infancy and period compared to adolescence and young adulthood period. Expression graph which shows the

expression of AKAP7 exons for all developmental periods can be found in supplementary Fig. 5.4. Genomic location for all of the AKAP7 exon probesets can be found in supplementary table 5.3.

3.3.1.2 Schizophrenia-Associated Alternative Splicing of AKAP7 Gene

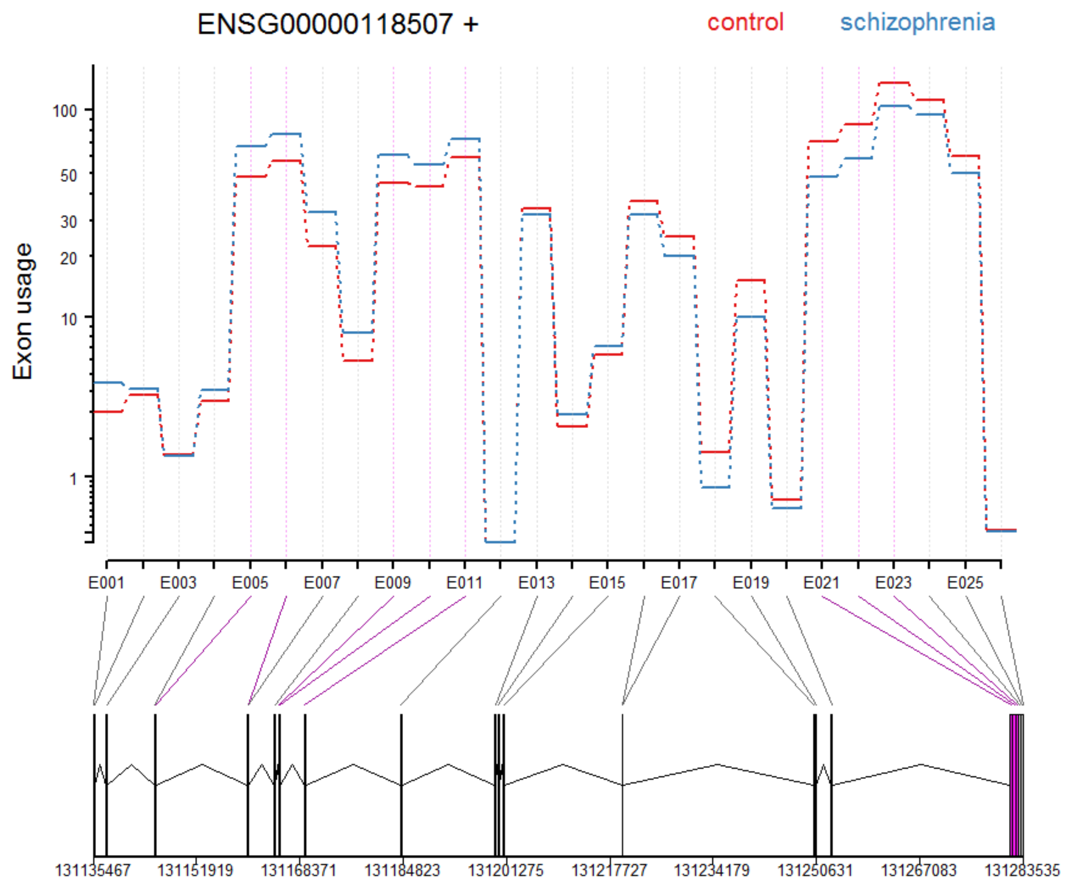


Figure 3.8: **The Expression of Exons of AKAP7 Gene.** Expression graph on the top shows fitted expression of exons of AKAP7: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all exons of AKAP7 transcripts and genomic locations of them: pink bars indicate differentially expressed exons (FDR corrected p value < 0.1) between schizophrenia and healthy groups.

According to Fig. 3.8, many exons of AKAP7 including E005, E006, E009, E010, E011, E021, E022 and E023 are differentially used between schizophrenia and healthy groups but only E022 (location: chr6: 131,281,944-131,282,398) coincides with the E15.3 from figure 8. Graph of normalized counts of AKAP7 exons can be found in supplementary Fig. 5.1. Genomic location for all of the AKAP7 exons can be found in supplementary table 5.6.

Therefore, it can be said that an exon which is possibly found near the 3' end of AKAP7 gene is expressed higher in early stages of development but its expression decreases by age (supplementary Fig. 5.4) in the DLPFC regions of psychologically healthy people. The same exon is expressed lower among schizophrenia patients compared to healthy controls.

3.3.1.3 Investigation of the Effect of Alternatively Spliced Exon on AKAP7 Function

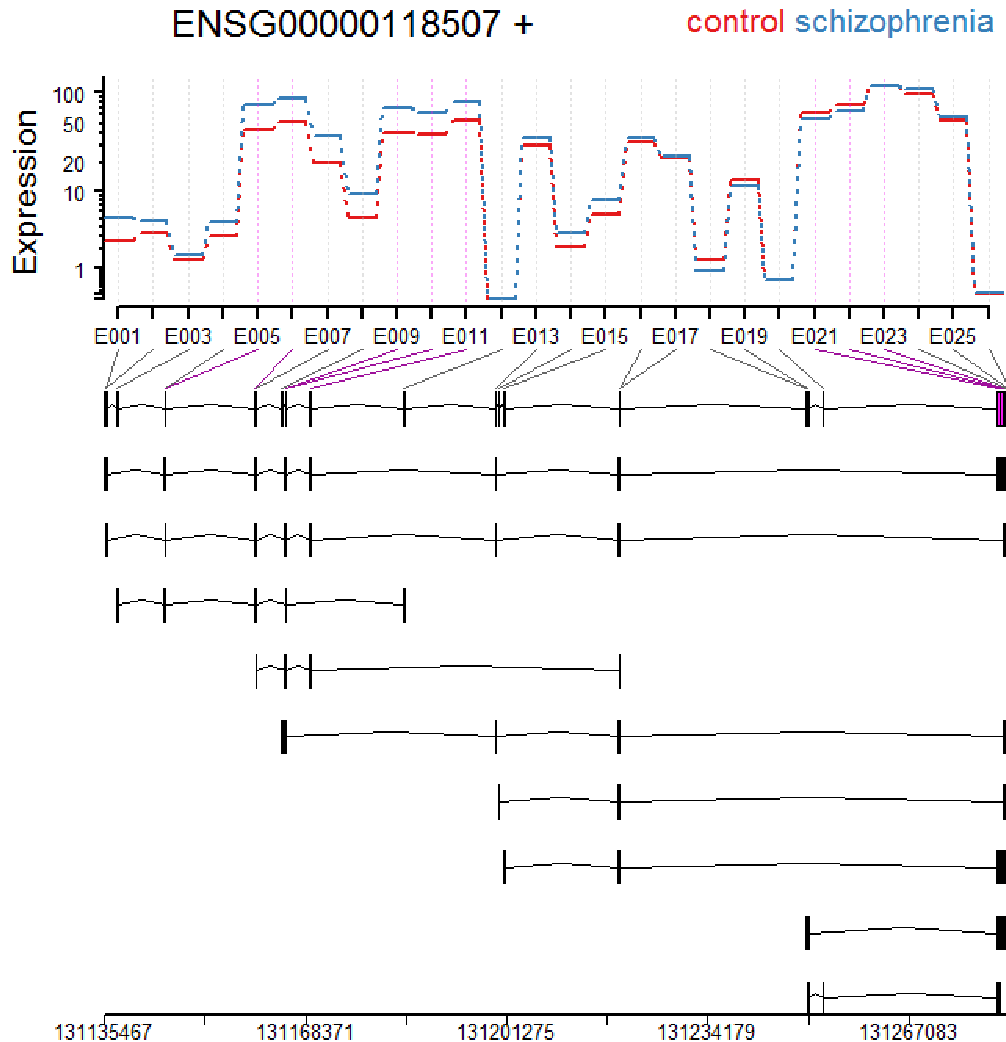


Figure 3.9: **The Expression of Exons of AKAP7 Gene and Its Transcripts.** Expression graph on the top shows fitted expression of exons of AKAP7 with the effect of overall gene expression: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, different transcripts of AKAP7 gene can be seen and pink bars indicate differentially expressed exons (FDR corrected p value < 0.1) between schizophrenia and healthy groups.

When we look at Fig. 3.9, we can see all defined transcripts of AKAP7 gene together with expression graph. It can be seen that overall expression of exons before E012 is generally higher among SCZ patients. This might be due to overexpression of short transcripts that contain these exons among SCZ patients. However, since there are many transcripts with different combinations, it is highly difficult to find which one of the transcript/s are affected.

Furthermore, when we look at the defined transcripts of AKAP7 gene at Fig. 3.9, we can see that the region that we have found (E022,location: chr6: 131,281,944-131,282,398) is not found as a separate exon in any of the defined transcripts but it is found as a whole last exon containing from E021 to E025 (location: chr6: 131,281,530-131,283,532) in several transcripts. Also, according to The Genotype-Tissue Expression (GTEx) project website which includes tissue specific gene/transcript/exon expression, in the frontal cortex 4 different transcripts of AKAP7 are expressed. These are ENST00000541650.5, ENST00000342266.4, ENST00000263050.3 and ENST00000431975.6 transcripts which are shown in the below Fig. 3.10 obtained from GTEx website [73]. As it can be noticed the last exon can be found in two forms the long form (also contains our exonic region) and the short form.

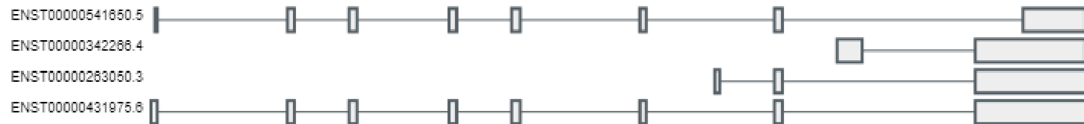


Figure 3.10: **Ensembl AKAP7 Transcripts Expressed in the Frontal Cortex.** Ensembl IDs and exon combinations of AKAP7 transcripts that are expressed in the frontal cortex.

The below graph 3.1 summarizes the properties of these transcripts including motifs they include and amino acid length of each. Although there are many other sequences detected, especially in long transcripts such as glycosylation and phosphorylation sites, the below graph includes only nuclear localization signal and PKA binding sites since only these two regions are included in the region covered by the last exon.

Table 3.1: **Length and Motifs of AKAP7 Transcripts Expressed in the Frontal Cortex**

Transcript ID	Length	Nuclear Localization Signal	PKA binding site
ENST00000541650.5	351 aa	+ (37-282)	-
ENST00000342266.4	82 aa	-	+ (21-81)
ENST00000263050.3	85 aa	+ (1-19)	+ (24-84)
ENST00000431975.6	349 aa	+ (38-283)	+ (288-348)

aa: amino acid, + sign indicates that the motif is found in the transcript and - sign indicates that the motif is not found in the transcript, numbers inside parentheses shows the location of the motif along the amino acid sequence of each transcript.

As it can be noticed, transcripts which include the bigger last exon that also cover the exonic region that we have found (ENST00000342266.4, ENST00000263050.3 and ENST00000431975.6) include region involved in PKA binding in their N-terminal encoded by the last exon. However, the transcript which contains the short last exon (ENST00000541650.5) does not have this PKA binding region. Therefore it can be concluded that the transcripts which contain long last exon (ENST00000342266.4, ENST00000263050.3 and ENST00000431975.6) and highly expressed in the frontal cortex may be expressed lower during AYA period and among schizophrenia patients compared to healthy controls. Since this long last exon is involved in interaction with PKA protein, it is possible that this interaction is developmentally regulated and disrupted in schizophrenia patients.

3.3.2 Developmental and Schizophrenia-Associated Alternative Splicing of BAIAP3 Gene

3.3.2.1 Developmental Alternative Splicing of BAIAP3 Gene

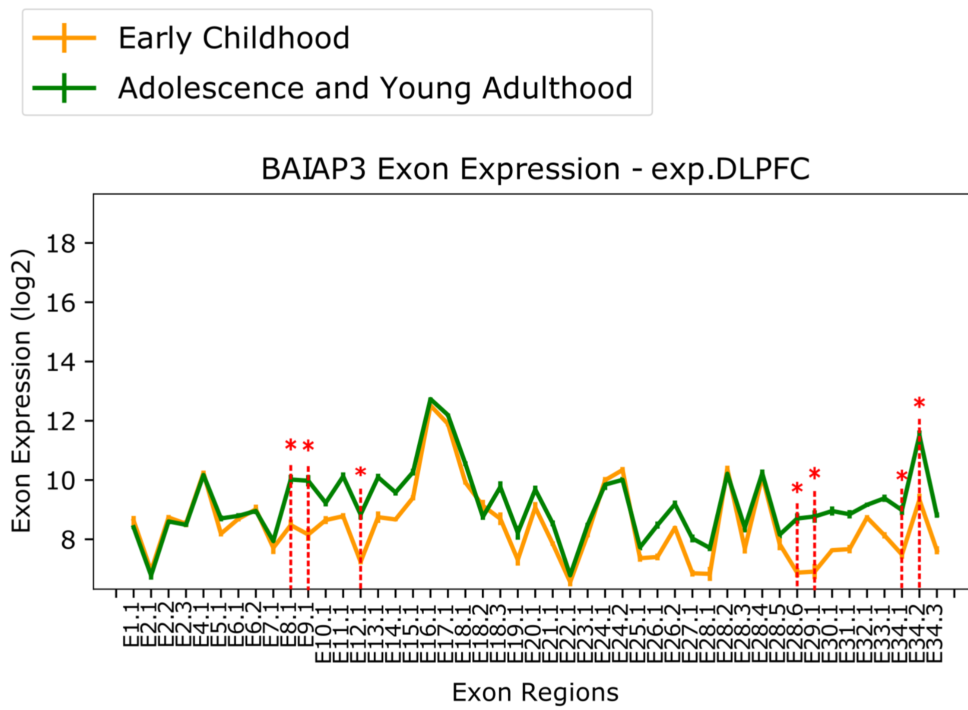


Figure 3.11: **The Expression of Exons of BAIAP3 Gene.** Log2 expressions of exons of BAIAP3 gene during early childhood (orange), and adolescence and young adulthood (dark green) periods.

3.3.2.2 Schizophrenia-Associated Alternative Splicing of BAIAP3 Gene

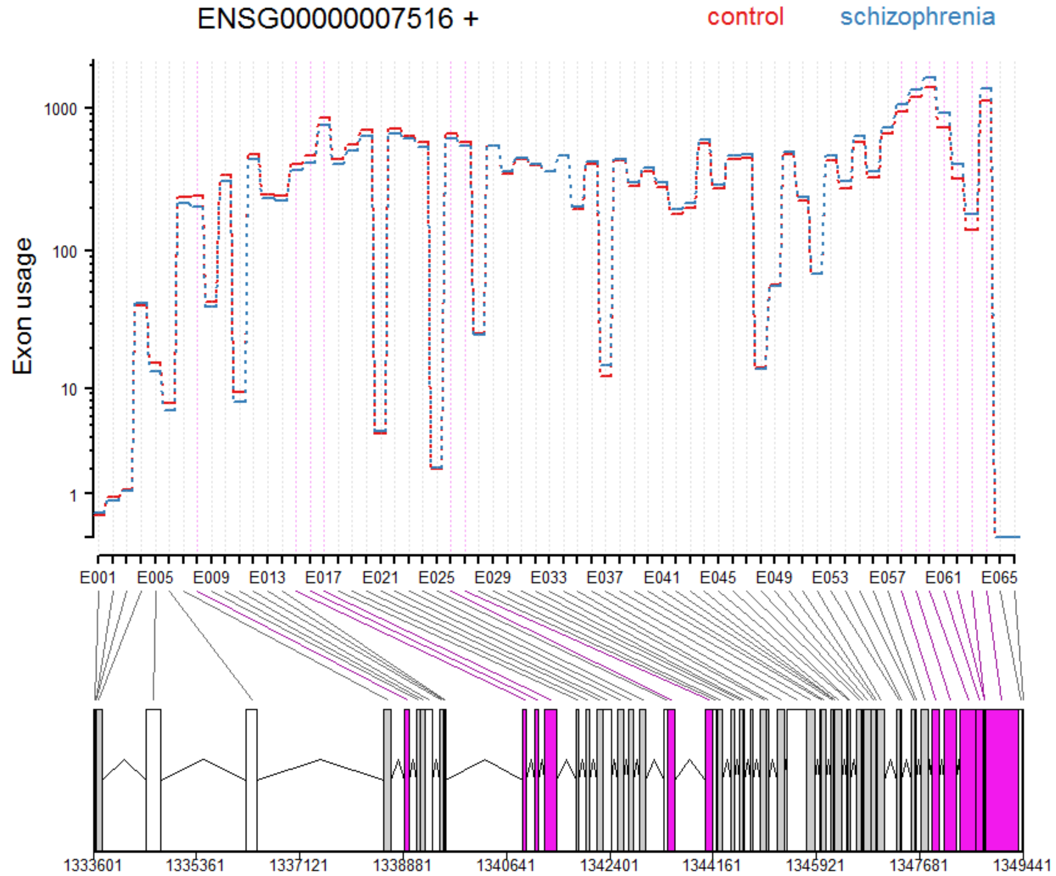


Figure 3.12: **The Expression of Exons of BAIAP3 Gene.** Expression graph on the top shows fitted expression of exons of BAIAP3: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all exons of BAIAP3 transcripts and genomic locations of them: pink bars indicate differentially expressed exons (FDR corrected p value < 0.1) between schizophrenia and healthy groups.

According to Fig. 3.12, many exons of BAIAP3 gene including E026, E058, E060, E061, E062, E063, and E064 are differentially used between schizophrenia and healthy groups but only E060 (location: chr16:1,348,379-1,348,640) coincides

with the E34-1 (chr16:1,348,421-1,348,480) from Fig. 3.11. Graph of normalized counts of BAIAP3 exons can be found in supplementary Fig. 5.2. Genomic location for all of the BAIAP3 exons can be found in supplementary table 5.3.

Although other exons do not exactly overlap with each other in both graphs 3' end of the BAIAP3 gene seems to be differentially expressed between groups. Therefore, it can be said that exon/s which are possibly found near the 3' end of BAIAP3 gene are expressed lower in early stages of development but its expression increases by age (Fig. 5.5) in the DLPFC regions of psychologically healthy people. The same exon/s are expressed higher among schizophrenia patients compared to healthy controls.

3.3.2.3 Investigation of the Effect of Alternatively Spliced Exon on BAIAP3 Function

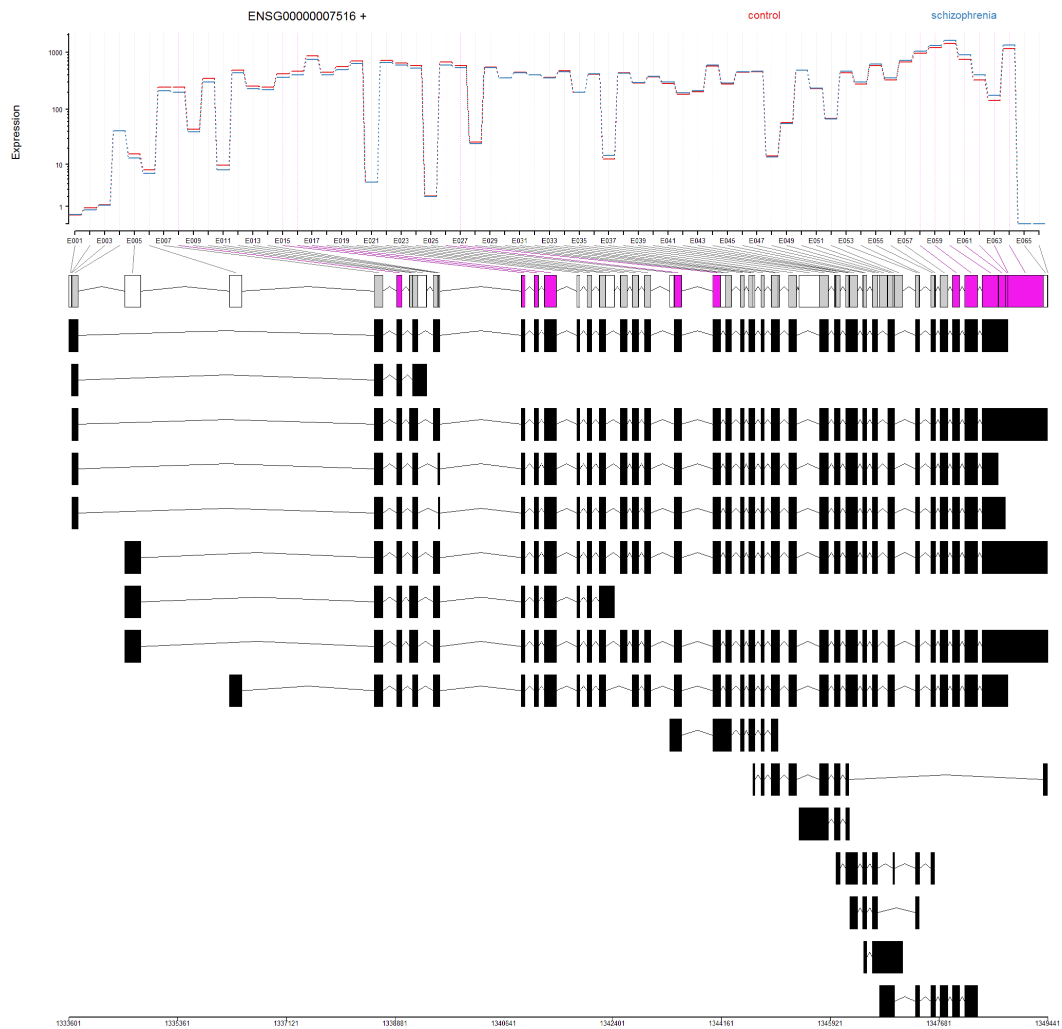


Figure 3.13: **The Expression of Exons of BAIAP3 Gene and Its Transcripts.** Expression graph on the top shows fitted expression of exons of BAIAP3 gene with the effect of overall gene expression: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, different transcripts of BAIAP3 gene can be seen and pink bars indicate differentially expressed exons (FDR corrected p value < 0.1) between schizophrenia and healthy groups.

When we look at Fig. 3.13, there is no overall gene expression change between schizophrenia and healthy samples since the expression of exons come closer when the effect of overall gene expression is included into GLM equations. When we look at the transcript isoforms of BAIAP3 gene, we can see that there are 15 different transcripts. It is not possible to decide which one/s are affected and differentially used between two groups.

According to The Genotype-Tissue Expression (GTEx) project website which includes tissue specific gene/transcript/exon expression, in the frontal cortex 7 different transcripts of BAIAP3 are expressed. These are ENST00000397488.6, ENST00000324385.9, ENST00000565665.5, ENST00000568198.1, ENST00000566-389.1, ENST00000567203.1 and ENST00000561602.1 transcripts which are shown in the below Fig. 3.14 obtained from GTEx website [73].

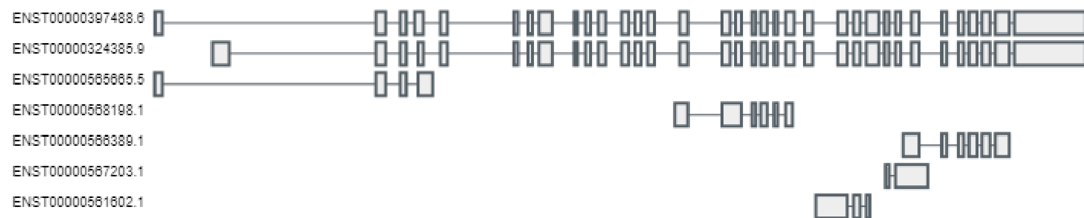


Figure 3.14: **Ensembl BAIAP3 Transcripts Expressed in the Frontal Cortex.** Ensembl IDs and exon combinations of BAIAP3 transcripts that are expressed in the frontal cortex.

The exonic region that we have found (E060, location: chr16:1,348,379-1,348,640) normally found as a unique exon in one of the BAIAP3 transcripts (ENST00000562208.5). However, the expression of it is 0 in the frontal cortex. Out of 7 BAIAP3 transcripts expressed in the frontal cortex, 2 of them contain long last exon including also E060 while 4 of them does not.

The last exon which encodes the last 34 amino acids at the N-terminal does not correspond with any motif found on the transcripts ENST00000397488.6 and ENST00000324385.9. Both of these two transcripts contain two C2 domains and two Munc Homology domains (MHD). Since it does not correspond with a novel motif, it is difficult to comment on its function on the protein. However shorter

transcripts, e.g., ENST00000566389.1 transcript contains one C2 domain and one MHD, which may affect its functional efficiency.

3.3.3 Developmental and Schizophrenia-Associated Alternative Splicing of SEMA3B Gene

3.3.3.1 Developmental Alternative Splicing of SEMA3B Gene

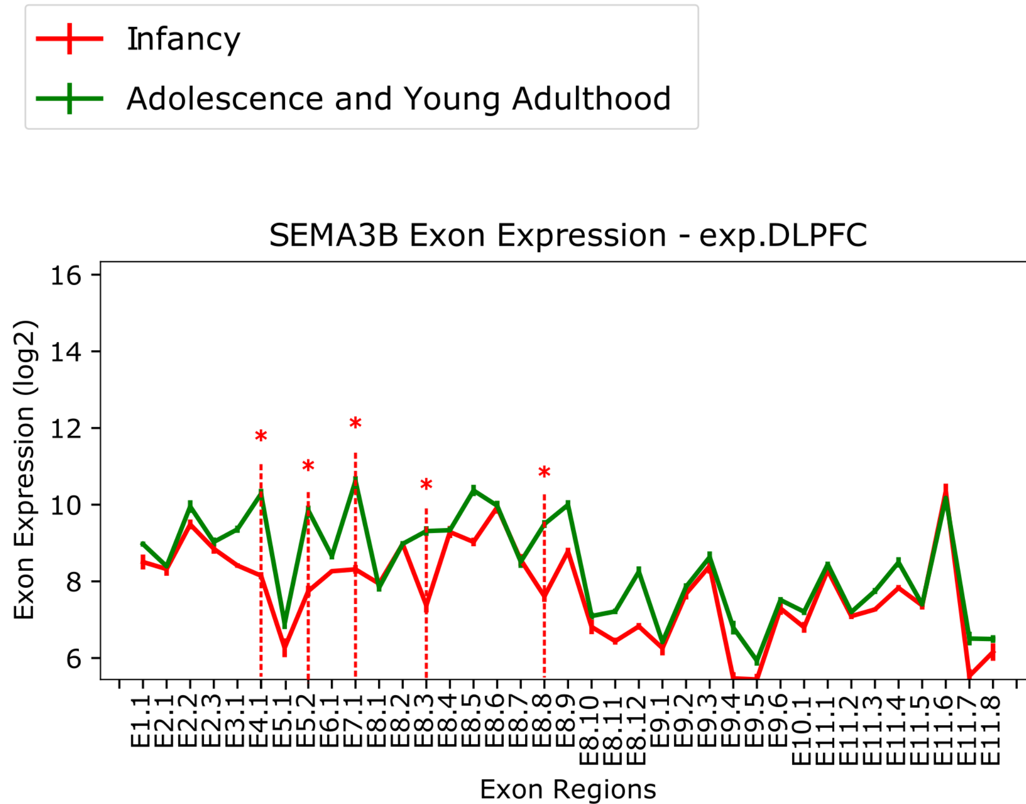


Figure 3.15: **The Expression of Exons of SEMA3B Gene.** Log2 expressions of exons of SEMA3B gene during infancy (red), and adolescence and young adulthood (dark green) periods.

Fig. 3.15 shows the expression of exon probests of SEMA3B gene in different developmental periods in DLPFC. It can be seen that many of the exon probesets are differentially expressed between I and AYA: E4-1, E5-2, E7-1, E8-3, and E8-8. Genomic locations of these probesets can be found in supplementary table 5.5. All of these exon probesets are expressed lower during infancy period compared to AYA period. Expression graph which shows the expression of SEMA3B exons for all developmental periods can be found in supplementary Fig. 5.6.

3.3.3.2 Schizophrenia-Associated Alternative Splicing of SEMA3B Gene

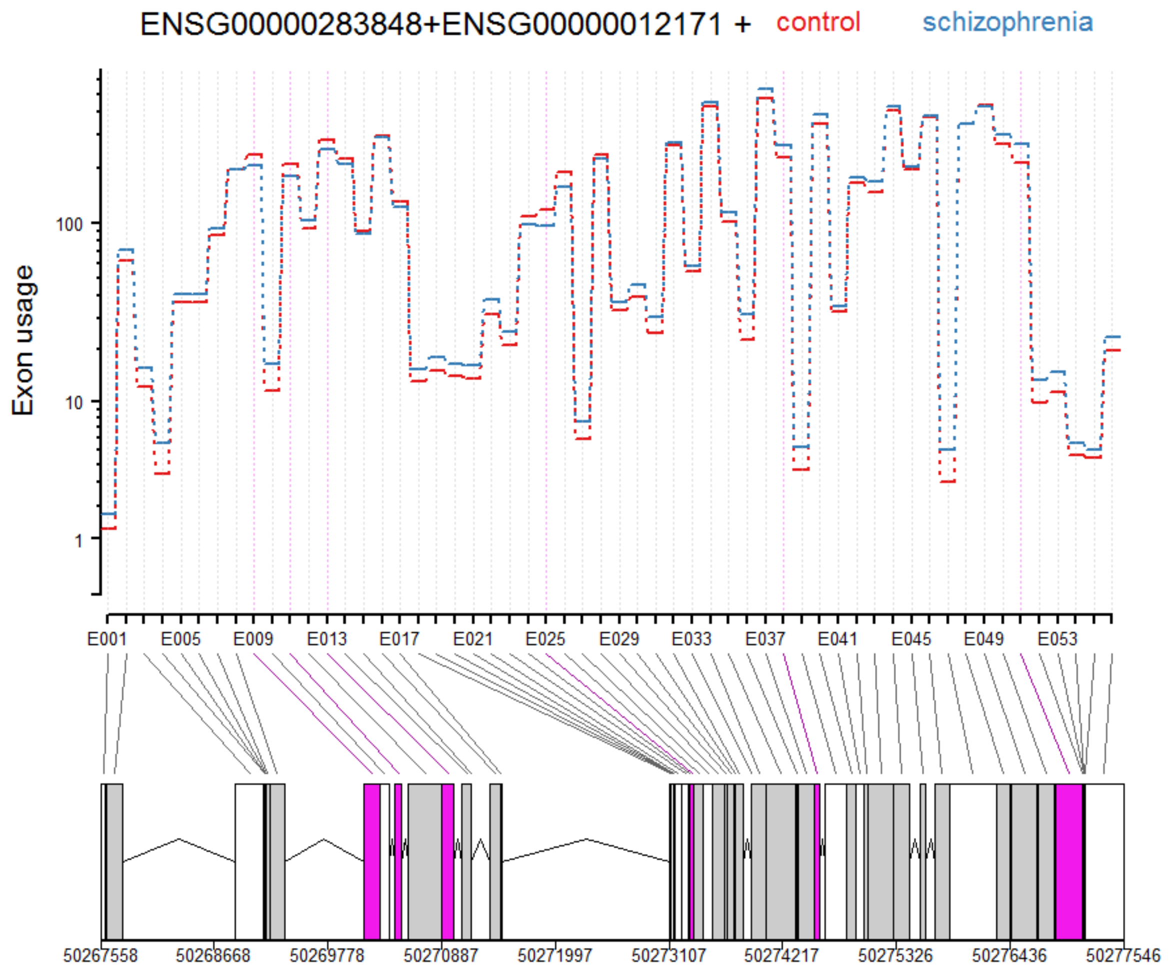


Figure 3.16: **The Expression of Exons of SEMA3B Gene.** Expression graph on the top shows fitted expression of exons of SEMA3B and MIR6872: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all exons of SEMA3B transcripts and genomic locations of them: pink bars indicate differentially expressed exons (FDR corrected p value < 0.1) between schizophrenia and healthy groups.

According to Fig. 3.16, many exons of SEMA3B gene including E009, E011, E013, E025, E038, and E051 are differentially used between schizophrenia and healthy groups but only E011 and E013 exons have FDR corrected p values that are smaller than 0.05. When we compare these exons in terms of developmental expression we see: E011 (chr3:50,270,433-50,270,495) coincides with the E4-1 (chr3:50,270,448-50,270,478); E013 (chr3:50,270,890-50,271,009) coincides with E5-2 (chr3: 50,270,900-50,270,970); E025 (chr3:50,273,315-50,273,345) coincides with E8-3 (chr3:50,273,317-50,273,441). The graph of normalized counts of SEMA3B exons can be found in supplementary Fig. 5.3. Genomic location for all exons of SEMA3B can be found in supplementary table 5.8.

Therefore, it can be said that E011, E013 and E025 exons of SEMA3B gene are expressed lower in early stages of development but their expression increases by age (Fig. 5.6) in the DLPFC regions of psychologically healthy people. It seems that the expression of the same exons remains to be lower among schizophrenia patients compared to healthy controls.

As it can be noticed, the title of Fig. 3.16 contains 2 different ensemble gene ids. One of them belongs to SEMA3B gene (ENSG00000012171) while the other belongs to a microRNA MIR6872 (ENSG00000283848) which is synthesized from the exon depicted as E023 (chr3:50,273,236-50,273,297). Since its expression does not differ between two conditions, it is not needed to be considered in this study.

3.3.3.3 Investigation of the Effect of Alternatively Spliced Exon on SEMA3B Function

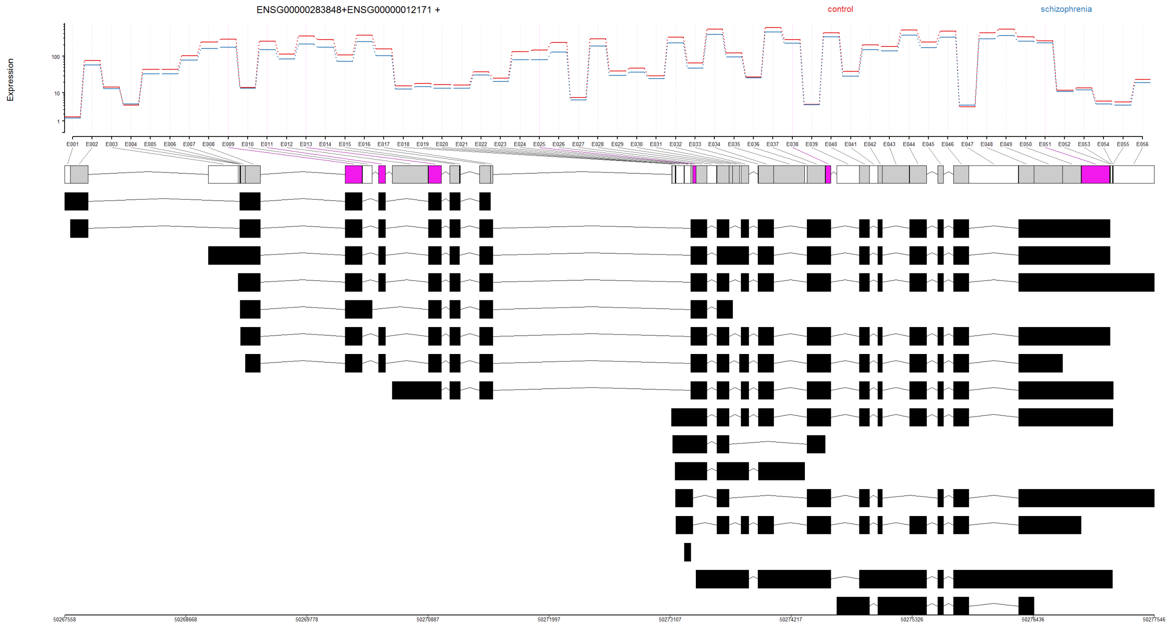


Figure 3.17: **The Expression of Exons of SEMA3B Gene and Its Transcripts.** Expression graph on the top shows fitted expression of exons of SEMA3B gene with the effect of overall gene expression: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, different transcripts of SEMA3B gene can be seen and pink bars indicate differentially expressed exons (FDR corrected p value < 0.1) between schizophrenia and healthy groups.

When we look at Fig. 3.17, it can be seen that nearly all of the exons of SEMA3B gene are expressed lower among schizophrenia patients compared to healthy group. However, since DEXseq also calculates this possibility, it tries to normalize the overall gene expression effect and give the results accordingly. The last four transcripts may be higher expressed among schizophrenia patients since they do not include E011, E013 and E025.

According to The Genotype-Tissue Expression (GTEx) project website which includes tissue specific gene/transcript/exon expression, in the frontal cortex 12

different transcripts of SEMA3B are expressed. These are ENST00000618865.4, ENST00000456560.6, ENST00000621029.4, ENST00000441915.5, ENST0000043-3753.4, ENST00000616701.4, ENST00000612509.4, ENST00000456210.5, ENST-00000619119.4, ENST00000434030.1, ENST00000419007.5 and ENST000004162-95.1 transcripts which are shown in the below Fig. 3.14 obtained from GTEx website [73].

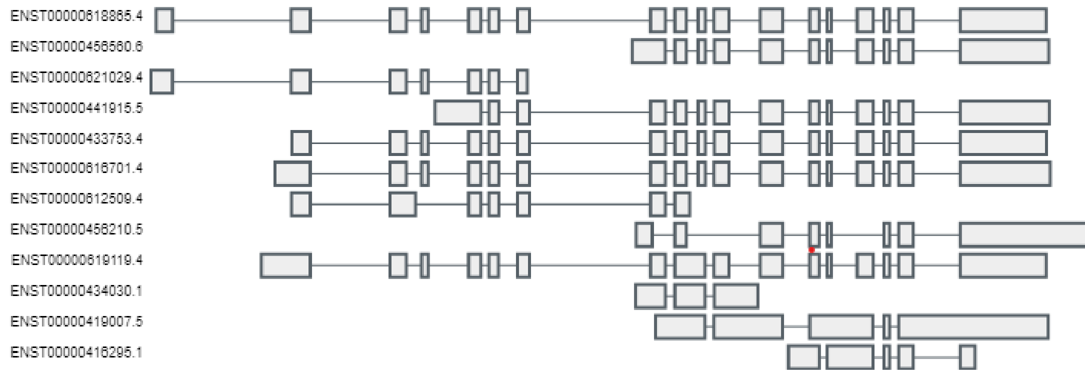


Figure 3.18: **Ensembl SEMA3B Transcripts Expressed in the Frontal Cortex.** Ensembl IDs and exon combinations of SEMA3B transcripts that are expressed in the frontal cortex.

When we look at the motifs of ENST00000618865.4 transcript which is 780 amino acids long, it contains two domains: one is Sema domain and the other Ig-like domain. While Sema domain covers amino acids from the 30th position till the 153th position which also include the two exonic regions that we have found, Ig-like domain covers amino acids from the 573th position till the 659th position which does not include the exonic regions we found. E011 encodes amino acids between 87-107 and E013 encodes 108-148. As it may be noticed Sema domain covers a bigger region than these two exonic regions but this domain have 4 disulfide bridges between cysteine residues and 2 of them involves cysteine residues from E011 and E013. One disulfide bridge form between cysteine residues at 102nd position (on E011) and at 133rd position (on E013), and other disulfide bridge forms between cysteine residues at 131st position (on E013) and at 140th position (on E013). Since two of the disulfide bonds are found on two exonic regions that we have found, the absence of them may alter protein structure and therefore protein function.

Chapter 4

Discussion

In this study, we take a different perspective to investigate DLPFC-specific splicing aberrations that may contribute to the pathophysiology of schizophrenia. Instead of simply comparing schizophrenia and psychologically healthy people, we tried to understand what are the normal splicing changes that occur throughout development and then we investigated those splicing changes among schizophrenia patients by using publically available RNA-seq and exon-microarray data. Although many of the splicing changes occur prenatally, there are still postnatal changes in splicing [20]. Our perspective may help to understand postnatal splicing changes in the brains of healthy people and the effects of disruptions to these changes to neurodevelopmental disorders like schizophrenia and autism. Results revealed 3 different genes that show splicing changes in the early development and schizophrenia.

One of them is A-kinase anchoring protein 7 (AKAP7) gene which belongs to a group of proteins that bind to the regulatory subunit of cAMP-dependent protein kinase A (PKA) and carry the enzyme to the specific parts of cells. PKA enzyme is known to regulate synaptic plasticity through interaction with calcium channels. AKAPs also regulate this process by localizing PKAs and by recruiting other proteins that PKAs interact with [77]. AKAPs contain a PKA-binding site together with motifs that target PKAs to different cellular localizations. According

to domains listed in Uniprot, PKA-binding site amino acids cover the last 294-348 amino acids in the carboxyl end of the protein (transcript=ENST00000431975.7). However, localization signals are not as obviously defined as PKA-binding site but it has been shown that different isoforms of AKAP7 with the same PKA-binding site target different cellular parts [78]. Our results revealed that an exon (location: chr6: 131,281,944-131,282,398) of AKAP7 gene covers the region where PKA-binding site is located is developmentally regulated and differentially used in schizophrenia. Its expression is higher during early stages of development but decreases by age and lower among schizophrenia patients compared to healthy controls. It is possible that the binding efficacy of AKAP7 protein is developmentally regulated and disruptions to this regulation may affect its binding efficacy to PKAs; therefore possibly affecting synaptic plasticity of neurons in DLPFC.

The second gene which was found to be alternatively spliced both in microarray and RNAseq analyses was BAI1 associated protein 3 (BAIAP3) gene which encodes a protein that functions as brain-specific angiogenesis inhibitor. This protein also belongs to a group of proteins called the mammalian uncoordinated 13 (Munc13). These proteins are known to regulate synaptic secretion of neurotransmitters. BAIAP3 possesses two munc13 homology domains (MHD) together with two C2 domains. C2 domains are frequently found among proteins involved in signal transduction [79]. Our results revealed that an exon (location: chr16:1,348,379-1,348,640) of BAIAP3 gene is developmentally regulated and differentially used in schizophrenia. It does correspond neither with MHD nor C2 domains according to domains shown in Ensembl web page. Therefore it is difficult to draw conclusions about the effect of alternative splicing of this specific exon. Its effect should be investigated further by functional assays. BAIAP3 is an attractive candidate for further studies and analyses since it is associated with anxiety in women and with substance abuse in men [79].

The common alternatively spliced third gene in microarray and RNAseq analyses was semaphorin-3B (SEMA3B) which encodes a secreted chemorepulsion molecule to inhibit axonal growth at nearby regions. It possess a conserved sema domain at the amino terminal which contributes to repelling function and

immunoglobulin like domain at the carboxyl terminal. Immunoglobulin like domains are found at many proteins and involved in protein-protein interactions [80]. Our results revealed two exons of SEMA3B gene are developmentally regulated and differentially used in schizophrenia. They both are in the region of sema domain according to domains shown in ensembl web page. However, their expression remains to be lower among schizophrenia patients compared to healthy controls. It is known that the expression of semaphorins through adulthood shows their importance in adult synaptic plasticity but the importance of different splice variants are not known yet [81]. Also SEMA3B is associated with schizophrenia and anxiety disorders through genome wide association studies. Therefore to study the alternative splicing of these two exons that we found are important to understand its relation to schizophrenia and normal development of brain.

Although this study may reveal important results, there are some points that are needed to be considered before making further conclusions. First of all, here we compared different data types with each other; for splicing analysis in psychologically healthy samples, exon microarray data was used while splicing analysis in schizophrenia samples was carried out by using RNA-Seq data. In a previous study using both exon microarray and RNA-seq in order to investigate splicing differences between two different brain regions have found a significant correlation between the results obtained by two methods [24]. Although both methods produced similar results, we investigated them by using different statistical method. While Altanalyze software that was used to analyze exon microarray data uses simpler statistical method: it calculates splicing index (SI) values for each exon in each sample by normalizing exon intensities to gene intensities and compare Sis by t-test. However DEXseq that was used to analyze RNA-Seq data uses a more sophisticated statistical method. It first calculates a size factor (S) for each sample by taking medians of all counts (of counting bins) divided by geometric mean of each in all samples. Then calculates normalized counts by dividing raw counts with the calculated S. Then exon counts for each sample is obtained by fitting them by generalized linear models and they were compared with chi-square likelihood ratio test. Despite the fact that there are such differences in data acquisition and analysis methods, we were still able to replicate some of the

results in the literature but in order to further continue with these results, the study should be replicated by using RNA-Seq data and apply the same statistical analysis for both developmental and schizophrenia related splicing analyses. Secondly, in the pursuit of our research question we found the common exons between the results of two analyses. The first part of analysis revealed 1289 developmentally-regulated exons while the second part of the analyses revealed 3340 exons that show expression difference between schizophrenia and healthy samples. Although these numbers are large, 4 exons that are found to be common between them can be purely by chance and be false-positive. Therefore, before making further conclusions, if possible, our results should be verified by measuring the expression of these 4 exons in primary cell lines obtained by postmortem healthy and schizophrenia brain samples.

Furthermore, we did not look at effects of medication status and gender on the expressions of exons that we have found. The effect of medication status on the expressions of exons can be assessed by taking information related to dosage and type of antipsychotics that each patient uses and if there is any effect on exon expression can be investigated by statistical tests. Also investigating gender differences in the contex of schizophrenia disease is important since we know that schizophrenia occurs more commonly and earlier among men compared to women, which may reflect differences in early development since it has been known that men and women develop differently. For example, one study revealed that mRNA amount of schizophrenia-associated ZNF804A gene was expressed higher in men compared to women with schizophrenia but there was no significant differences in splice variants of this gene with respect to two sex [53]. Another factor that needed to be considered is to look at the exon expression differences in first-degree relatives of patients with schizophrenia since they may also show middle expression changes between schizophrenia patients and healthy controls. For example, although carried in blood platelet cells, one study found that the expression of fynT splice variant (containing different exon 7) was significantly lower among schizophrenia patients compared to controls and was lowest in relatives and the expression of fynΔ7 splice ariant (lacking exon 7) was higher among schizophrenia patients and was highest in relatives [82].

Although there are some drawbacks of the study, it should be noted that we tried to implement a different perspective in our analysis. It is known that many of the splicing changes occur before birth but there are still ongoing postnatal changes. Also it is a fact that synaptic changes, which are necessary to adopt to dynamic environmental stimuli and form the basis of cognitive functioning, occur as a result of transcriptional changes. Further accumulating evidence indicates that disruptions of alternative splicing underlie many neurodevelopmental disorders [83]. Therefore to study the effects of alternative splicing in the context of plasticity and neurodevelopmental disorders may be important to develop therapeutic strategies to cure these diseases. Our results similarly reveal that many genes show alternative splicing aberrations in schizophrenia samples. And some of them also show developmental expression changes and not surprisingly they are involved in synaptic plasticity. If their effects on synaptic plasticity and their abnormal expression change during early development are proved, it may be possible to intervene and develop therapeutical approaches.

Chapter 5

Supplementary Material

5.1 Supplementary Figures

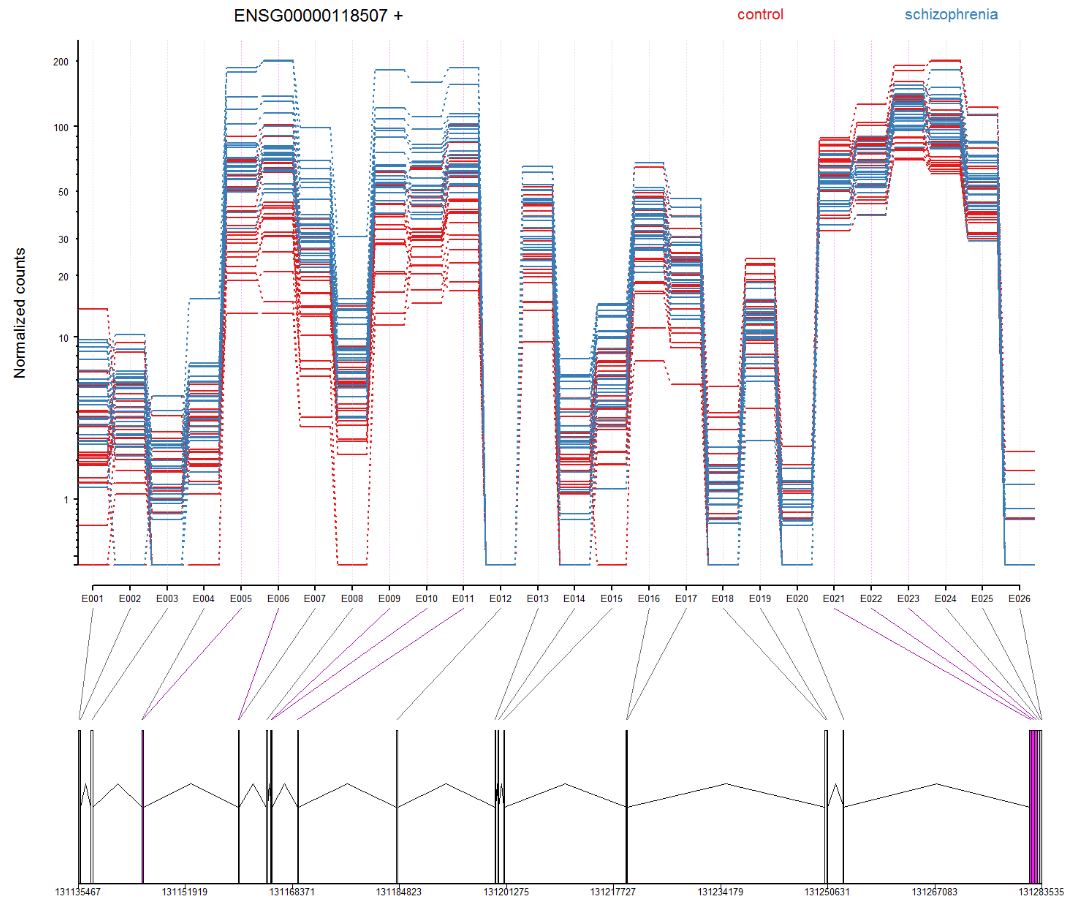


Figure 5.1: **The Normalized DLPFC Counts of AKAP7 Gene.** Expression graph on the top shows normalized expression of DLPFCs of AKAP7 gene for all samples: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all DLPFCs of AKAP7 transcripts and genomic locations of them: pink bars indicate differentially expressed DLPFCs (FDR corrected p value < 0.1) between schizophrenia and healthy groups.

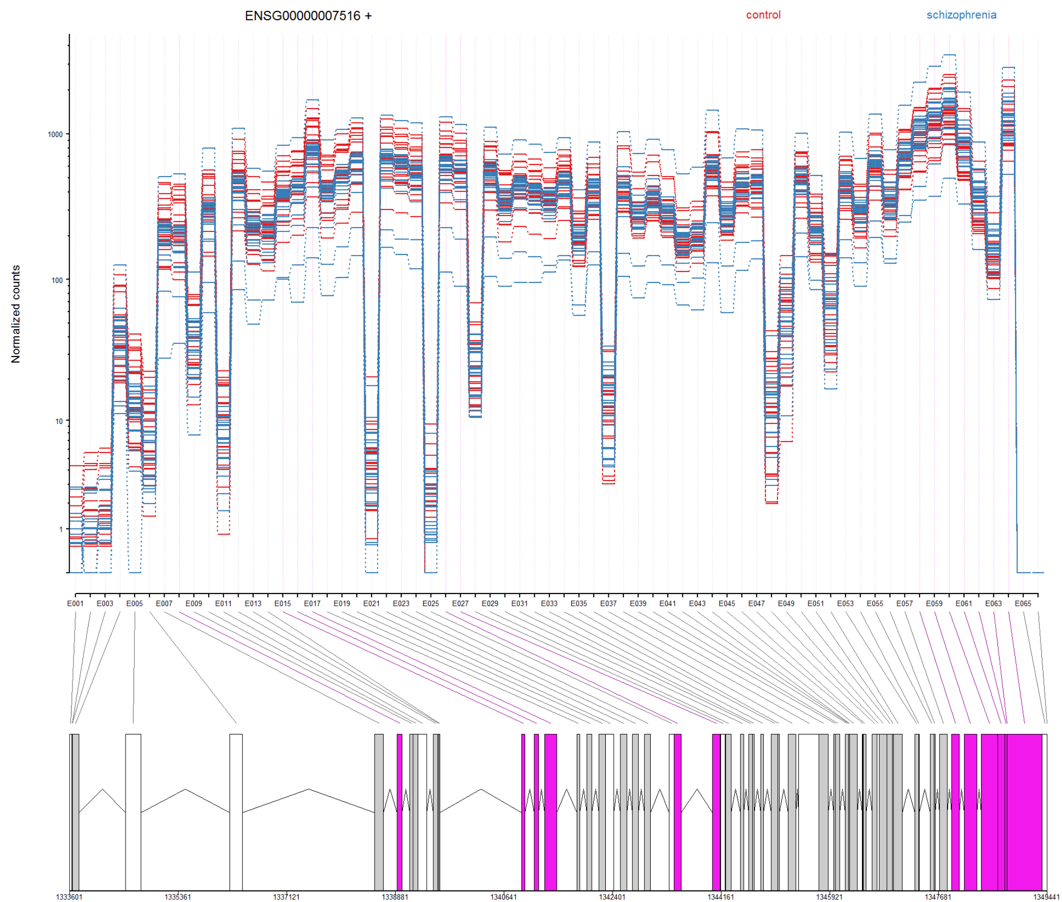


Figure 5.2: **The Normalized DLPFC Counts of BAIAP3 Gene.** Expression graph on the top shows normalized expression of DLPFCs of BAIAP3 gene for all samples: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all DLPFCs of BAIAP3 transcripts and genomic locations of them: pink bars indicate differentially expressed DLPFCs (FDR corrected p value < 0.1) between schizophrenia and healthy groups.

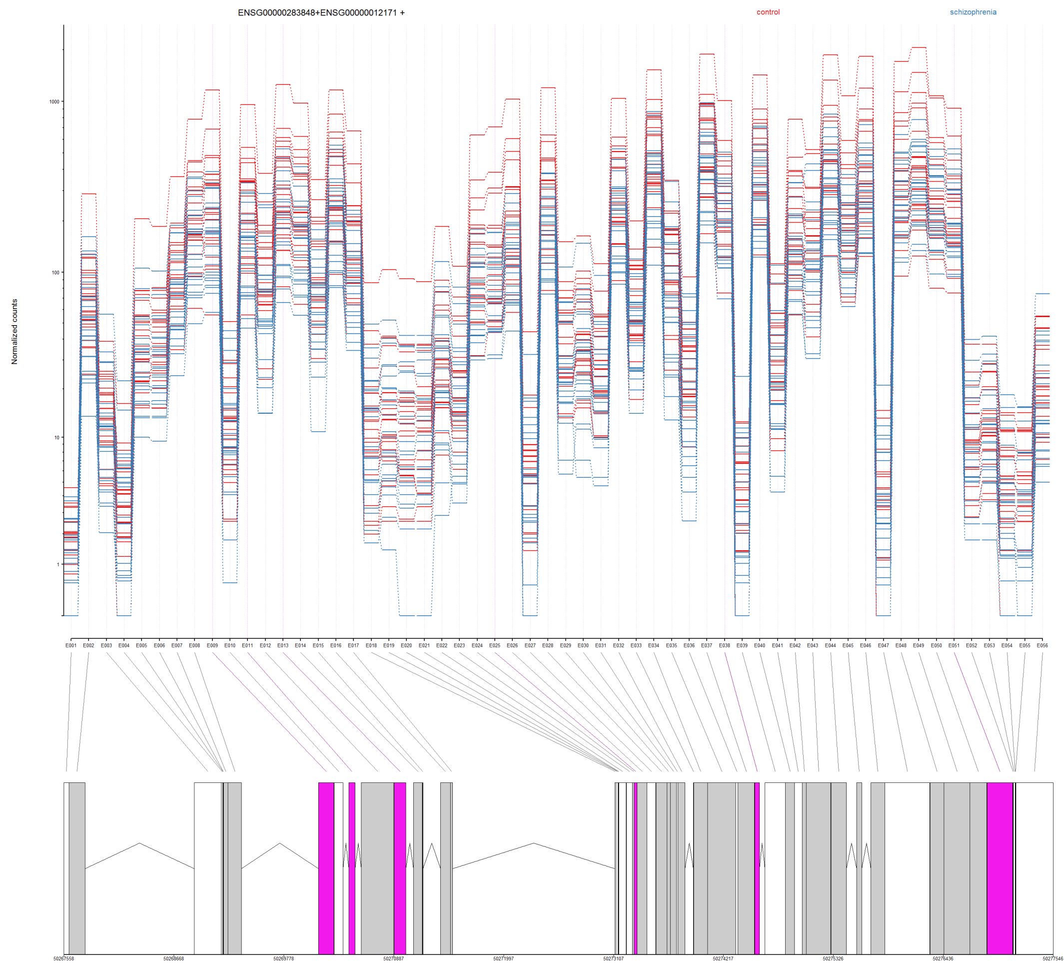
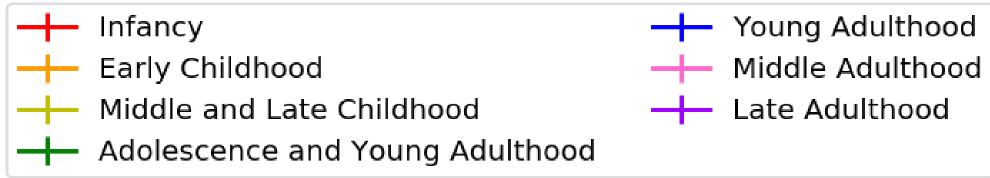


Figure 5.3: **The Normalized DLPFC Counts of SEMA3B Gene.** Expression graph on the top shows normalized expression of DLPFCs of SEMA3B and MIR6872 gene for all samples: blue lines indicate the expression of schizophrenia samples while red lines indicate the expression of the samples of the healthy group. On the bottom, bars show all DLPFCs of SEMA3B transcripts and genomic locations of them: pink bars indicate differentially expressed DLPFCs (FDR corrected p value < 0.1) between schizophrenia and healthy groups.



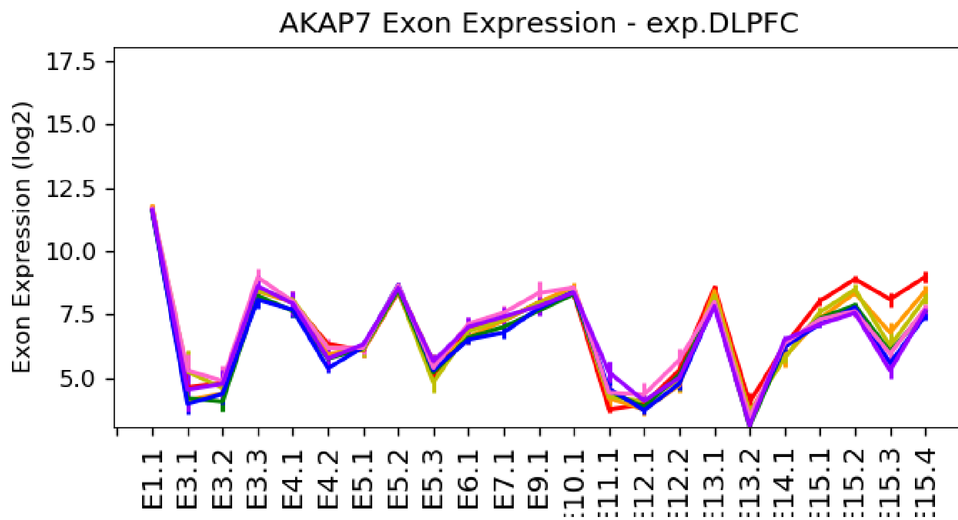
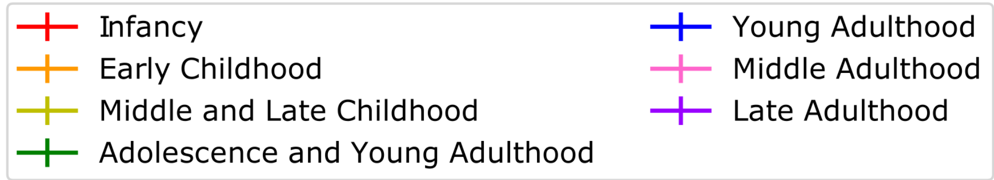


Figure 5.4: **The Expression of DLPFC Counts of AKAP7 Gene for All Developmental Periods.** Log2 expressions of DLPFCs of AKAP7 gene during different developmental periods.



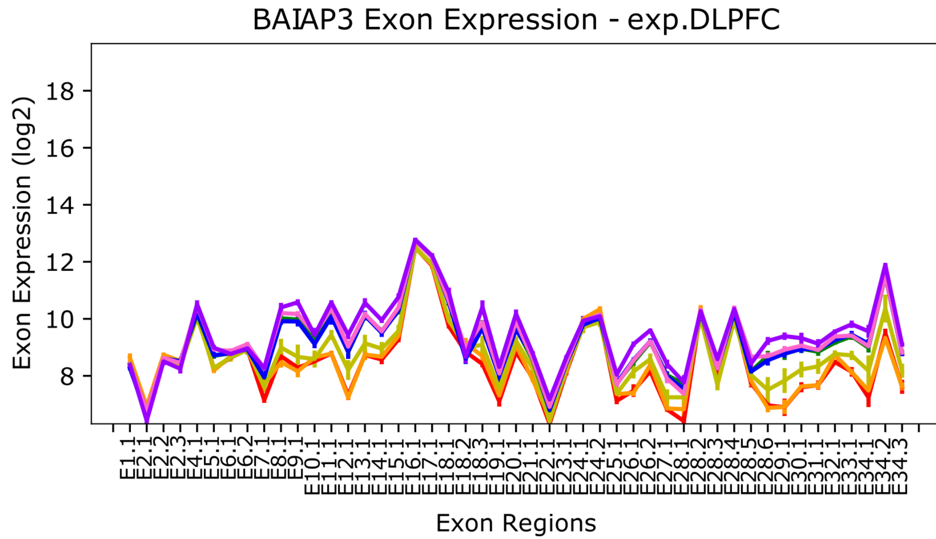
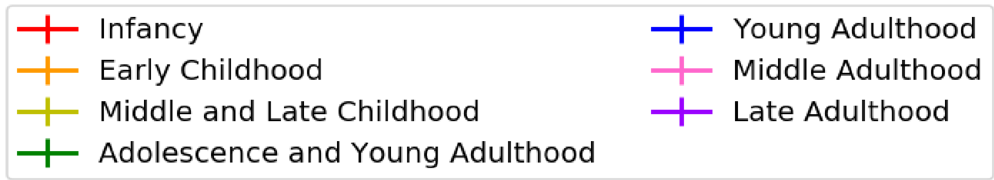


Figure 5.5: **The Expression of DLPFC Counts of BAIAP3 Gene for All Developmental Periods.** Log2 expressions of DLPFCs of BAIAP3 gene during different developmental periods.



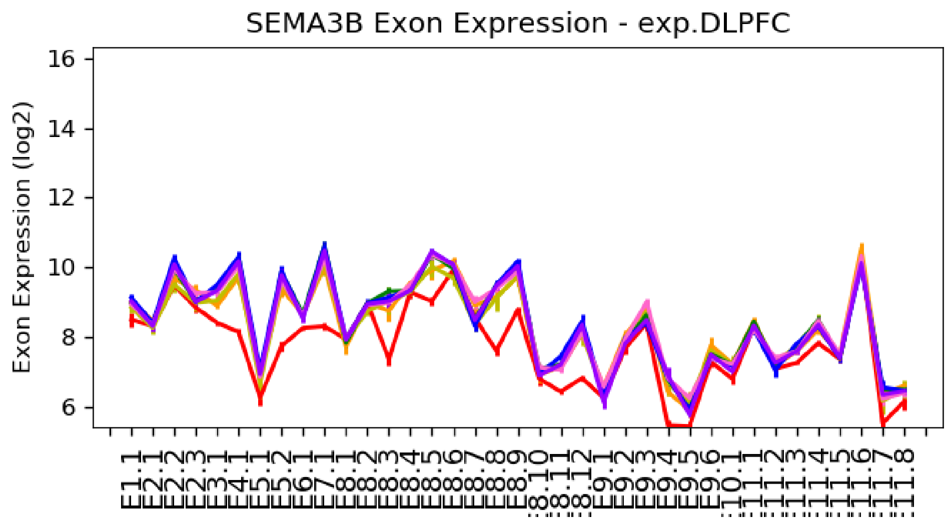


Figure 5.6: **The Expression of DLPFC Counts of SEMA3B Gene for All Developmental Periods.** Log2 expressions of DLPFCs of SEMA3B gene during different developmental periods.

5.2 Supplementary Tables

Table 5.1: **GSM accession codes, developmental group, sex, age, hemisphere and region information for each array included in the study**

Sample	Developmental Group	Sex	Age	Hemisphere	Region
GSM708782	Infancy	M	4 M	L	VFC
GSM708777	Infancy	M	4 M	L	OFC
GSM708776	Infancy	M	4 M	L	MFC
GSM708774	Infancy	M	4 M	L	M1C
GSM708770	Infancy	M	4 M	L	DLPFC
GSM708695	Infancy	M	4 M	R	VFC
GSM708694	Infancy	M	4 M	L	VFC
GSM708685	Infancy	M	4 M	R	OFC
GSM708684	Infancy	M	4 M	L	OFC
GSM708683	Infancy	M	4 M	R	MFC
GSM708682	Infancy	M	4 M	L	MFC
GSM708679	Infancy	M	4 M	R	M1C
GSM708678	Infancy	M	4 M	L	M1C
GSM708671	Infancy	M	4 M	R	DLPFC
GSM708670	Infancy	M	4 M	L	DLPFC
GSM708475	Infancy	M	4 M	R	VFC
GSM708474	Infancy	M	4 M	L	VFC
GSM708465	Infancy	M	4 M	R	OFC
GSM708464	Infancy	M	4 M	L	OFC
GSM708463	Infancy	M	4 M	R	MFC
GSM708462	Infancy	M	4 M	L	MFC
GSM708459	Infancy	M	4 M	R	M1C
GSM708458	Infancy	M	4 M	L	M1C
GSM708451	Infancy	M	4 M	R	DLPFC
GSM708450	Infancy	M	4 M	L	DLPFC
GSM708652	Infancy	F	6 M	L	VFC
GSM708647	Infancy	F	6 M	L	OFC

Sample	Developmental Group	Sex	Age	Hemisphere	Region
GSM708646	Infancy	F	6 M	L	MFC
GSM708644	Infancy	F	6 M	L	M1C
GSM708640	Infancy	F	6 M	L	DLPFC
GSM709159	Infancy	M	10 M	L	VFC
GSM709155	Infancy	M	10 M	L	OFC
GSM709154	Infancy	M	10 M	L	MFC
GSM709150	Infancy	M	10 M	L	DLPFC
GSM708507	Early Childhood	F	1 Y	R	VFC
GSM708506	Early Childhood	F	1 Y	L	VFC
GSM708497	Early Childhood	F	1 Y	R	OFC
GSM708496	Early Childhood	F	1 Y	L	OFC
GSM708495	Early Childhood	F	1 Y	R	MFC
GSM708494	Early Childhood	F	1 Y	L	MFC
GSM708491	Early Childhood	F	1 Y	R	M1C
GSM708490	Early Childhood	F	1 Y	L	M1C
GSM708483	Early Childhood	F	1 Y	R	DLPFC
GSM708482	Early Childhood	F	1 Y	L	DLPFC
GSM708857	Early Childhood	F	2 Y	R	VFC
GSM708856	Early Childhood	F	2 Y	L	VFC
GSM708849	Early Childhood	F	2 Y	R	OFC
GSM708848	Early Childhood	F	2 Y	L	OFC
GSM708847	Early Childhood	F	2 Y	R	MFC
GSM708846	Early Childhood	F	2 Y	L	MFC
GSM708844	Early Childhood	F	2 Y	R	M1C
GSM708843	Early Childhood	F	2 Y	L	M1C
GSM708837	Early Childhood	F	2 Y	R	DLPFC
GSM708836	Early Childhood	F	2 Y	L	DLPFC
GSM709186	Early Childhood	F	3 Y	L	VFC
GSM709180	Early Childhood	F	3 Y	L	M1C
GSM709173	Early Childhood	M	3 Y	L	VFC
GSM709169	Early Childhood	M	3 Y	L	OFC

Sample	Developmental Group	Sex	Age	Hemisphere	Region
GSM709168	Early Childhood	M	3 Y	L	MFC
GSM709167	Early Childhood	M	3 Y	L	M1C
GSM709163	Early Childhood	M	3 Y	L	DLPFC
GSM708428	Early Childhood	M	4 Y	R	VFC
GSM708423	Early Childhood	M	4 Y	R	MFC
GSM708421	Early Childhood	M	4 Y	R	M1C
GSM708417	Early Childhood	M	4 Y	R	DLPFC
GSM709198	Middle and Late Childhood	M	8 Y	L	VFC
GSM709195	Middle and Late Childhood	M	8 Y	L	OFC
GSM709194	Middle and Late Childhood	M	8 Y	L	MFC
GSM709190	Middle and Late Childhood	M	8 Y	L	DLPFC
GSM708798	Middle and Late Childhood	M	8 Y	L	VFC
GSM708793	Middle and Late Childhood	M	8 Y	L	OFC
GSM708792	Middle and Late Childhood	M	8 Y	L	MFC
GSM708790	Middle and Late Childhood	M	8 Y	L	M1C
GSM708786	Middle and Late Childhood	M	8 Y	L	DLPFC
GSM709212	Middle and Late Childhood	F	11 Y	L	VFC
GSM709208	Middle and Late Childhood	F	11 Y	L	OFC
GSM709207	Middle and Late Childhood	F	11 Y	L	MFC
GSM709206	Middle and Late Childhood	F	11 Y	L	M1C
GSM709189	Middle and Late Childhood	F	11 Y	L	DLPFC
GSM708568	Middle and Late Childhood	F	13 Y	R	VFC
GSM708567	Middle and Late Childhood	F	13 Y	L	VFC
GSM708559	Middle and Late Childhood	F	13 Y	R	OFC
GSM708558	Middle and Late Childhood	F	13 Y	L	OFC
GSM708557	Middle and Late Childhood	F	13 Y	R	MFC
GSM708556	Middle and Late Childhood	F	13 Y	L	MFC
GSM708553	Middle and Late Childhood	F	13 Y	R	M1C
GSM708552	Middle and Late Childhood	F	13 Y	L	M1C
GSM708545	Middle and Late Childhood	F	13 Y	R	DLPFC
GSM708544	Middle and Late Childhood	F	13 Y	L	DLPFC

Sample	Developmental Group	Sex	Age	Hemisphere	Region
GSM708432	Adolescence and Young Adulthood	M	15 Y	L	DLPFC
GSM708435	Adolescence and Young Adulthood	M	15 Y	L	M1C
GSM708437	Adolescence and Young Adulthood	M	15 Y	L	MFC
GSM708438	Adolescence and Young Adulthood	M	15 Y	L	OFC
GSM708443	Adolescence and Young Adulthood	M	15 Y	L	VFC
GSM708208	Adolescence and Young Adulthood	M	18 Y	R	DLPFC
GSM708207	Adolescence and Young Adulthood	M	18 Y	L	DLPFC
GSM708216	Adolescence and Young Adulthood	M	18 Y	R	M1C
GSM708215	Adolescence and Young Adulthood	M	18 Y	L	M1C
GSM708218	Adolescence and Young Adulthood	M	18 Y	R	MFC
GSM708217	Adolescence and Young Adulthood	M	18 Y	L	MFC
GSM708220	Adolescence and Young Adulthood	M	18 Y	R	OFC
GSM708219	Adolescence and Young Adulthood	M	18 Y	L	OFC
GSM708228	Adolescence and Young Adulthood	M	18 Y	R	VFC
GSM708227	Adolescence and Young Adulthood	M	18 Y	L	VFC
GSM708608	Adolescence and Young Adulthood	F	19 Y	L	DLPFC
GSM708612	Adolescence and Young Adulthood	F	19 Y	L	M1C
GSM708614	Adolescence and Young Adulthood	F	19 Y	L	MFC
GSM708615	Adolescence and Young Adulthood	F	19 Y	L	OFC
GSM708620	Adolescence and Young Adulthood	F	19 Y	L	VFC
GSM708624	Adolescence and Young Adulthood	F	21 Y	L	DLPFC
GSM708628	Adolescence and Young Adulthood	F	21 Y	L	M1C
GSM708630	Adolescence and Young Adulthood	F	21 Y	L	MFC
GSM708631	Adolescence and Young Adulthood	F	21 Y	L	OFC
GSM708636	Adolescence and Young Adulthood	F	21 Y	L	VFC
GSM708806	Adolescence and Young Adulthood	M	22 Y	R	DLPFC
GSM708805	Adolescence and Young Adulthood	M	22 Y	L	DLPFC
GSM708814	Adolescence and Young Adulthood	M	22 Y	R	M1C
GSM708813	Adolescence and Young Adulthood	M	22 Y	L	M1C
GSM708818	Adolescence and Young Adulthood	M	22 Y	R	MFC
GSM708817	Adolescence and Young Adulthood	M	22 Y	L	MFC

Sample	Developmental Group	Sex	Age	Hemisphere	Region
GSM708820	Adolescence and Young Adulthood	M	22 Y	R	OFC
GSM708819	Adolescence and Young Adulthood	M	22 Y	L	OFC
GSM708830	Adolescence and Young Adulthood	M	22 Y	R	VFC
GSM708829	Adolescence and Young Adulthood	M	22 Y	L	VFC
GSM708742	Adolescence and Young Adulthood	M	23 Y	R	DLPFC
GSM708741	Adolescence and Young Adulthood	M	23 Y	L	DLPFC
GSM708750	Adolescence and Young Adulthood	M	23 Y	R	M1C
GSM708749	Adolescence and Young Adulthood	M	23 Y	L	M1C
GSM708754	Adolescence and Young Adulthood	M	23 Y	R	MFC
GSM708753	Adolescence and Young Adulthood	M	23 Y	L	MFC
GSM708756	Adolescence and Young Adulthood	M	23 Y	R	OFC
GSM708755	Adolescence and Young Adulthood	M	23 Y	L	OFC
GSM708766	Adolescence and Young Adulthood	M	23 Y	R	VFC
GSM708765	Adolescence and Young Adulthood	M	23 Y	L	VFC
GSM708702	Young Adulthood	F	27 Y	L	OFC
GSM708701	Young Adulthood	F	27 Y	L	MFC
GSM708698	Young Adulthood	F	27 Y	L	DLPFC
GSM708575	Young Adulthood	M	28 Y	R	OFC
GSM708574	Young Adulthood	M	28 Y	L	OFC
GSM708573	Young Adulthood	M	28 Y	R	MFC
GSM708572	Young Adulthood	M	28 Y	L	MFC
GSM708604	Young Adulthood	F	30 Y	R	VFC
GSM708603	Young Adulthood	F	30 Y	L	VFC
GSM708594	Young Adulthood	F	30 Y	R	OFC
GSM708593	Young Adulthood	F	30 Y	L	OFC
GSM708592	Young Adulthood	F	30 Y	R	MFC
GSM708591	Young Adulthood	F	30 Y	L	MFC
GSM708588	Young Adulthood	F	30 Y	R	M1C
GSM708587	Young Adulthood	F	30 Y	L	M1C
GSM708581	Young Adulthood	F	30 Y	R	DLPFC
GSM708580	Young Adulthood	F	30 Y	L	DLPFC

Sample	Developmental Group	Sex	Age	Hemisphere	Region
GSM708921	Young Adulthood	M	36 Y	R	VFC
GSM708920	Young Adulthood	M	36 Y	L	VFC
GSM708911	Young Adulthood	M	36 Y	R	OFC
GSM708910	Young Adulthood	M	36 Y	L	OFC
GSM708909	Young Adulthood	M	36 Y	R	MFC
GSM708908	Young Adulthood	M	36 Y	L	MFC
GSM708905	Young Adulthood	M	36 Y	R	M1C
GSM708904	Young Adulthood	M	36 Y	L	M1C
GSM708897	Young Adulthood	M	36 Y	R	DLPFC
GSM708896	Young Adulthood	M	36 Y	L	DLPFC
GSM709308	Young Adulthood	F	37 Y	R	VFC
GSM709307	Young Adulthood	F	37 Y	L	VFC
GSM709298	Young Adulthood	F	37 Y	R	OFC
GSM709297	Young Adulthood	F	37 Y	L	OFC
GSM709296	Young Adulthood	F	37 Y	R	MFC
GSM709295	Young Adulthood	F	37 Y	L	MFC
GSM709292	Young Adulthood	F	37 Y	R	M1C
GSM709291	Young Adulthood	F	37 Y	L	M1C
GSM709284	Young Adulthood	F	37 Y	R	DLPFC
GSM709283	Young Adulthood	F	37 Y	L	DLPFC
GSM708538	Young Adulthood	M	37 Y	R	VFC
GSM708537	Young Adulthood	M	37 Y	L	VFC
GSM708528	Young Adulthood	M	37 Y	R	OFC
GSM708527	Young Adulthood	M	37 Y	L	OFC
GSM708526	Young Adulthood	M	37 Y	R	MFC
GSM708525	Young Adulthood	M	37 Y	L	MFC
GSM708522	Young Adulthood	M	37 Y	R	M1C
GSM708521	Young Adulthood	M	37 Y	L	M1C
GSM708514	Young Adulthood	M	37 Y	R	DLPFC
GSM708513	Young Adulthood	M	37 Y	L	DLPFC

Sample	Developmental Group	Sex	Age	Hemisphere	Region
GSM708889	Middle Adulthood	M	40 Y	R	VFC
GSM708888	Middle Adulthood	M	40 Y	L	VFC
GSM708879	Middle Adulthood	M	40 Y	R	VFC
GSM708878	Middle Adulthood	M	40 Y	L	VFC
GSM708877	Middle Adulthood	M	40 Y	R	VFC
GSM708876	Middle Adulthood	M	40 Y	L	VFC
GSM708873	Middle Adulthood	M	40 Y	R	VFC
GSM708872	Middle Adulthood	M	40 Y	L	VFC
GSM708865	Middle Adulthood	M	40 Y	R	VFC
GSM708864	Middle Adulthood	M	40 Y	L	VFC
GSM708734	Middle Adulthood	M	40 Y	R	VFC
GSM708733	Middle Adulthood	M	40 Y	L	VFC
GSM708724	Middle Adulthood	M	40 Y	R	VFC
GSM708723	Middle Adulthood	M	40 Y	L	VFC
GSM708722	Middle Adulthood	M	40 Y	R	VFC
GSM708721	Middle Adulthood	M	40 Y	L	VFC
GSM708718	Middle Adulthood	M	40 Y	R	VFC
GSM708717	Middle Adulthood	M	40 Y	L	VFC
GSM708711	Middle Adulthood	M	40 Y	R	VFC
GSM708710	Middle Adulthood	M	40 Y	L	VFC
GSM709267	Middle Adulthood	M	42 Y	L	VFC
GSM709262	Middle Adulthood	M	42 Y	L	VFC
GSM709261	Middle Adulthood	M	42 Y	L	VFC
GSM709259	Middle Adulthood	M	42 Y	L	VFC
GSM709255	Middle Adulthood	M	42 Y	L	VFC
GSM709251	Middle Adulthood	M	55 Y	L	VFC
GSM709247	Middle Adulthood	M	55 Y	L	VFC
GSM709246	Middle Adulthood	M	55 Y	L	VFC
GSM709241	Middle Adulthood	M	55 Y	L	VFC

Sample	Developmental Group	Sex	Age	Hemisphere	Region
GSM708260	Late Adulthood	M	64 Y	R	VFC
GSM708259	Late Adulthood	M	64 Y	L	VFC
GSM708250	Late Adulthood	M	64 Y	R	OFC
GSM708249	Late Adulthood	M	64 Y	L	OFC
GSM708248	Late Adulthood	M	64 Y	R	MFC
GSM708247	Late Adulthood	M	64 Y	L	MFC
GSM708244	Late Adulthood	M	64 Y	R	M1C
GSM708243	Late Adulthood	M	64 Y	L	M1C
GSM708236	Late Adulthood	M	64 Y	R	DLPFC
GSM708235	Late Adulthood	M	64 Y	L	DLPFC
GSM708323	Late Adulthood	F	70 Y	R	VFC
GSM708322	Late Adulthood	F	70 Y	L	VFC
GSM708313	Late Adulthood	F	70 Y	R	OFC
GSM708312	Late Adulthood	F	70 Y	L	OFC
GSM708311	Late Adulthood	F	70 Y	R	MFC
GSM708310	Late Adulthood	F	70 Y	L	MFC
GSM708307	Late Adulthood	F	70 Y	R	M1C
GSM708306	Late Adulthood	F	70 Y	L	M1C
GSM708299	Late Adulthood	F	70 Y	R	DLPFC
GSM708298	Late Adulthood	F	70 Y	L	DLPFC
GSM709116	Late Adulthood	F	82 Y	R	VFC
GSM709107	Late Adulthood	F	82 Y	R	OFC
GSM709106	Late Adulthood	F	82 Y	R	MFC
GSM709105	Late Adulthood	F	82 Y	L	MFC
GSM709102	Late Adulthood	F	82 Y	R	M1C
GSM709101	Late Adulthood	F	82 Y	L	M1C
GSM709094	Late Adulthood	F	82 Y	R	DLPFC
GSM709093	Late Adulthood	F	82 Y	L	DLPFC

OFC: Orbital Prefrontal Cortex, MFC: Medial Prefrontal Cortex, DLPFC: Dorsolateral Prefrontal Cortex, VFC: Ventrolateral Prefrontal Cortex, M1C: Primary Motor Cortex; M:male, F: female; R: right hemisphere, L: left hemisphere; M: month, Y:years

Table 5.2: **Exon IDs, Probeset numbers and Genomic Locations of AKAP7 Provided by AltAnalyze**

ExonID	Probeset	Genomic Location
E1-1	2925726	chr6:131,135,726-131,135,772
E3-2	2925730	chr6:131,145,342-131,145,369
E3-3	2925731	chr6:131,145,383-131,145,410
E4-1	2925735	chr6:131,160,041-131,160,144
E4-2	2925736	chr6:131,160,143-131,160,198
E5-1	2925739	chr6:131,164,371-131,164,397
E5-2	2925740	chr6:131,164,397-131,164,437
E5-3	2925741	chr6:131,164,479-131,164,513
E6-1	2925742	chr6:131,165,081-131,165,200
E7-1	2925744	chr6:131,169,112-131,169,264
E9-1	2925748	chr6:131,199,483-131,199,573
E10-1	2925751	chr6:131,200,029-131,200,104
E11-1	2925753	chr6:131,200,848-131,200,897
E12-1	2925758	chr6:131,219,669-131,219,710
E12-2	2925759	chr6:131,219,717-131,219,808
E13-1	2925762	chr6:131,250,366-131,250,540
E14-1	2925764	chr6:131,253,036-131,253,095
E15-1	2925768	chr6:131,281,545-131,281,615
E15-2	2925769	chr6:131,281,800-131,282,019
E15-3	2925770	chr6:131,282,152-131,282,358
E15-4	2925771	chr6:131,282,444-131,283,259

Table 5.3: **Exon IDs, Probeset numbers and Genomic Locations of BA-IAP3 Provided by AltAnalyze**

ExonID	Probeset	Genomic Location
E4-1	3643760	chr16:1,338,552-1,338,655
E5-1	3643761	chr16:1,338,920-1,338,989
E7-1	3643764	chr16:1,339,496-1,339,560
E8-1	3643765	chr16:1,340,921-1,340,980
E9-1	3643766	chr16:1,341,139-1,341,179
E10-1	3643767	chr16:1,341,293-1,341,442
E11-1	3643768	chr16:1,341,823-1,341,858
E12-1	3643769	chr16:1,341,986-1,342,055
E13-1	3643770	chr16:1,342,185-1,342,282
E14-1	3643771	chr16:1,342,540-1,342,605
E15-1	3643772	chr16:1,342,732-1,342,769
E16-1	3643773	chr16:1,342,936-1,342,969
E17-1	3643774	chr16:1,343,392-1,343,440
E18-1	3643775	chr16:1,344,025-1,344,133
E18-3	3643777	chr16:1,344,257-1,344,299
E19-1	3643778	chr16:1,344,468-1,344,523
E20-1	3643780	chr16:1,344,608-1,344,693
E21-1	3643781	chr16:1,344,797-1,344,849
E23-1	3643783	chr16:1,345,268-1,345,371
E24-1	3643784	chr16:1,345,511-1,345,730
E24-2	3643785	chr16:1,345,746-1,345,861
E25-1	3643786	chr16:1,345,988-1,346,055
E26-1	3643787	chr16:1,346,169-1,346,211
E26-2	3643788	chr16:1,346,302-1,346,361
E27-1	3643789	chr16:1,346,441-1,346,510
E28-1	3643790	chr16:1,346,612-1,346,674
E28-2	3643791	chr16:1,346,708-1,346,734
E28-4	3643793	chr16:1,346,810-1,346,846
E28-6	3643795	chr16:1,346,899-1,346,955

ExonID	Probeset	Genomic Location
E29-1	3643796	chr16:1,347,303-1,347,331
E30-1	3643797	chr16:1,347,560-1,347,625
E31-1	3643798	chr16:1,347,709-1,347,766
E32-1	3643799	chr16:1,347,925-1,348,014
E33-1	3643800	chr16:1,348,121-1,348,219
E34-1	3643801	chr16:1,348,421-1,348,480
E34-2	3643802	chr16:1,348,875-1,349,278
E34-3	3643803	chr16:1,349,376-1,349,589

Table 5.4: **Exon IDs, Probeset numbers and Genomic Locations of SEMA3B Provided by AltAnalyze**

ExonID	Probeset	Genomic Location
E1-1	2622699	chr3:50,267,479-50,267,749
E2-1	2622701	chr3:50,269,000-50,269,157
E2-2	2622702	chr3:50,269,181-50,269,206
E3-1	2622705	chr3:50,270,152-50,270,262
E4-1	2622706	chr3:50,270,448-50,270,478
E5-1	2622707	chr3:50,270,671-50,270,794
E5-2	2622708	chr3:50,270,900-50,270,970
E6-1	2622710	chr3:50,271,087-50,271,176
E7-1	2622711	chr3:50,271,372-50,271,465
E8-1	2622713	chr3:50,273,130-50,273,227
E8-2	2622714	chr3:50,273,238-50,273,288
E8-3	2622715	chr3:50,273,317-50,273,441
E8-4	2622716	chr3:50,273,455-50,273,534
E8-5	2622717	chr3:50,273,561-50,273,644
E8-6	2622718	chr3:50,273,647-50,273,706
E8-8	2622720	chr3:50,273,764-50,273,828
E8-9	2622721	chr3:50,273,916-50,274,038
E8-10	2622722	chr3:50,274,067-50,274,105
E8-11	2622723	chr3:50,274,108-50,274,244
E8-12	2622724	chr3:50,274,368-50,274,562
E9-1	2622725	chr3:50,274,736-50,274,826
E9-2	2622726	chr3:50,274,855-50,274,884
E9-4	2622728	chr3:50,275,017-50,275,053
E11-1	2622732	chr3:50,275,729-50,275,823
E11-2	2622733	chr3:50,275,953-50,276,226
E11-6	2622737	chr3:50,276,615-50,276,691
E11-7	2622738	chr3:50,276,808-50,277,116

Table 5.5: **Exon IDs and Genomic Locations of AKAP7 Provided by DEXseq**

ExonID	Genomic Location
E001	chr6:131,135,467-131,135,736
E002	chr6:131,135,737-131,135,782
E003	chr6:131,137,378-131,137,751
E004	chr6:131,145,285-131,145,287
E005	chr6:131,145,288-131,145,416
E006	chr6:131,160,059-131,160,174
E007	chr6:131,160,175-131,160,198
E008	chr6:131,164,336-131,164,532
E009	chr6:131,165,081-131,165,149
E010	chr6:131,165,150-131,165,217
E011	chr6:131,169,113-131,169,273
E012	chr6:131,184,373-131,184,558
E013	chr6:131,199,461-131,199,573
E014	chr6:131,199,991-131,200,062
E015	chr6:131,200,827-131,200,910
E016	chr6:131,219,661-131,219,750
E017	chr6:131,219,751-131,219,808
E018	chr6:131,250,159-131,250,429
E019	chr6:131,250,430-131,250,616
E020	chr6:131,253,037-131,253,105
E021	chr6:131,281,530-131,281,943
E022	chr6:131,281,944-131,282,398
E023	chr6:131,282,399-131,282,844
E024	chr6:131,282,845-131,283,184
E025	chr6:131,283,185-131,283,532
E026	chr6:131,283,533-131,283,535

Table 5.6: **Exon IDs and Genomic Locations of BAIAP3 Provided by DEXseq**

ExonID	Genomic Location
E001	chr16:1,333,601-1,333,637
E002	chr16:1,333,638-1,333,642
E003	chr16:1,333,643-1,333,644
E004	chr16:1,333,645-1,333,749
E005	chr16:1,334,504-1,334,756
E006	chr16:1,336,197-1,336,399
E007	chr16:1,338,540-1,338,680
E008	chr16:1,338,902-1,338,989
E009	chr16:1,339,113-1,339,163
E010	chr16:1,339,164-1,339,244
E011	chr16:1,339,245-1,339,386
E012	chr16:1,339,496-1,339,564
E013	chr16:1,339,565-1,339,579
E014	chr16:1,339,580-1,339,603
E015	chr16:1,340,922-1,340,981
E016	chr16:1,341,129-1,341,195
E017	chr16:1,341,294-1,341,489
E018	chr16:1,341,822-1,341,866
E019	chr16:1,341,986-1,342,063
E020	chr16:1,342,181-1,342,283
E021	chr16:1,342,284-1,342,421
E022	chr16:1,342,527-1,342,634
E023	chr16:1,342,719-1,342,814
E024	chr16:1,342,913-1,343,016
E025	chr16:1,343,318-1,343,392
E026	chr16:1,343,393-1,343,513
E027	chr16:1,344,022-1,344,146
E028	chr16:1,344,147-1,344,226
E029	chr16:1,344,227-1,344,317

ExonID	Genomic Location
E030	chr16:1,344,469-1,344,525
E031	chr16:1,344,601-1,344,667
E032	chr16:1,344,668-1,344,698
E033	chr16:1,344,798-1,344,849
E034	chr16:1,344,969-1,345,073
E035	chr16:1,345,074-1,345,099
E036	chr16:1,345,249-1,345,372
E037	chr16:1,345,418-1,345,746
E038	chr16:1,345,747-1,345,890
E039	chr16:1,345,986-1,346,008
E040	chr16:1,346,009-1,346,078
E041	chr16:1,346,170-1,346,220
E042	chr16:1,346,221-1,346,225
E043	chr16:1,346,226-1,346,237
E044	chr16:1,346,238-1,346,361
E045	chr16:1,346,442-1,346,458
E046	chr16:1,346,459-1,346,510
E047	chr16:1,346,605-1,346,684
E048	chr16:1,346,685-1,346,718
E049	chr16:1,346,719-1,346,846
E050	chr16:1,346,847-1,346,933
E051	chr16:1,346,934-1,346,955
E052	chr16:1,346,956-1,347,091
E053	chr16:1,347,298-1,347,358
E054	chr16:1,347,359-1,347,369
E055	chr16:1,347,545-1,347,610
E056	chr16:1,347,611-1,347,625
E057	chr16:1,347,701-1,347,821
E058	chr16:1,347,894-1,348,017
E059	chr16:1,348,096-1,348,301
E060	chr16:1,348,379-1,348,640

ExonID	Genomic Location
E061	chr16:1,348,641-1,348,755
E062	chr16:1,348,756-1,348,793
E063	chr16:1,348,794-1,348,794
E064	chr16:1,348,795-1,349,365
E065	chr16:1,349,366-1,349,438
E066	chr16:1,349,439-1,349,441

Table 5.7: **Exon IDs and Genomic Locations of SEMA3B Provided by DEXseq**

ExonID	Genomic Location
E001	chr3:50,267,558-50,267,607
E002	chr3:50,267,608-50,267,769
E003	chr3:50,268,874-50,269,146
E004	chr3:50,269,147-50,269,160
E005	chr3:50,269,161-50,269,164
E006	chr3:50,269,165-50,269,168
E007	chr3:50,269,169-50,269,213
E008	chr3:50,269,214-50,269,349
E009	chr3:50,270,127-50,270,284
E010	chr3:50,270,285-50,270,376
E011	chr3:50,270,433-50,270,495
E012	chr3:50,270,561-50,270,889
E013	chr3:50,270,890-50,271,009
E014	chr3:50,271,088-50,271,177
E015	chr3:50,271,178-50,271,181
E016	chr3:50,271,361-50,271,458
E017	chr3:50,271,459-50,271,480
E018	chr3:50,273,119-50,273,129
E019	chr3:50,273,130-50,273,151
E020	chr3:50,273,152-50,273,156
E021	chr3:50,273,157-50,273,159
E022	chr3:50,273,160-50,273,235
E023	chr3:50,273,236-50,273,297
E024	chr3:50,273,298-50,273,314
E025	chr3:50,273,315-50,273,345
E026	chr3:50,273,346-50,273,443
E027	chr3:50,273,444-50,273,534
E028	chr3:50,273,535-50,273,646

Table 5.8: **Exon IDs and Genomic Locations of SEMA3B Provided by DEXseq**

ExonID	Genomic Location
E029	chr3:50,273,647-50,273,680
E030	chr3:50,273,681-50,273,743
E031	chr3:50,273,744-50,273,758
E032	chr3:50,273,759-50,273,828
E033	chr3:50,273,913-50,273,915
E034	chr3:50,273,916-50,274,057
E035	chr3:50,274,058-50,274,340
E036	chr3:50,274,341-50,274,362
E037	chr3:50,274,363-50,274,530
E038	chr3:50,274,531-50,274,582
E039	chr3:50,274,634-50,274,842
E040	chr3:50,274,843-50,274,934
E041	chr3:50,274,935-50,275,011
E042	chr3:50,275,012-50,275,053
E043	chr3:50,275,054-50,275,301
E044	chr3:50,275,302-50,275,459
E045	chr3:50,275,560-50,275,615
E046	chr3:50,275,705-50,275,844
E047	chr3:50,275,845-50,276,301
E048	chr3:50,276,302-50,276,444
E049	chr3:50,276,445-50,276,706
E050	chr3:50,276,707-50,276,876
E051	chr3:50,276,877-50,277,139
E052	chr3:50,277,140-50,277,142
E053	chr3:50,277,143-50,277,161
E054	chr3:50,277,162-50,277,165
E055	chr3:50,277,166-50,277,170
E056	chr3:50,277,171-50,277,546

Bibliography

[1] Y. Lee and D. C. Rio, "Mechanisms and regulation of alternative pre-mrna splicing," *Annual Review of Biochemistry*, vol. 84, no. 1, pp. 291–323, 2015. PMID: 25784052.

[2] C. Will, "Spliceosome structure and function," *The RNA world*, pp. 369–400, 2006.

[3] H. Lodish, A. Berk, C. A. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. Ploegh, P. Matsudaira, *et al.*, *Molecular cell biology.* Macmillan, 2008.

[4] K. Fox-Walsh, Y. Dou, B. Lam, S.-P. Hung, P. Baldi, and K. Hertel, "The architecture of pre-mrnas affects mechanisms of splice-site pairing," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 16176–81, 12 2005.

[5] C. L. Will and R. Lührmann, "Spliceosome structure and function," *Cold Spring Harbor perspectives in biology*, vol. 3, no. 7, p. a003707, 2011.

[6] B.-S. Jo and S. S. Choi, "Introns: the functional benefits of introns in genomes," *Genomics & informatics*, vol. 13, no. 4, p. 112, 2015.

[7] X.-Z. Wang, "Mechanism of alternative splicing and its regulation (review)," *BIOMEDICAL REPORTS*, vol. 3, pp. 152–158, 01 2015.

[8] L. Cartegni, J. Wang, Z. Zhu, M. Q. Zhang, and A. R. Krainer, "Esefinder: a web resource to identify exonic splicing enhancers," *Nucleic acids research*, vol. 31, no. 13, pp. 3568–3571, 2003.

[9] A. J. McCullough and S. M. Berget, “An intronic splicing enhancer binds u1 snrnps to enhance splicing and select 5’ splice sites,” *Molecular and cellular biology*, vol. 20, no. 24, pp. 9225–9235, 2000.

[10] J. Hui, L.-H. Hung, M. Heiner, S. Schreiner, N. Neumüller, G. Reither, S. A. Haas, and A. Bindereif, “Intronic ca-repeat and ca-rich elements: a new class of regulators of mammalian alternative splicing,” *The EMBO journal*, vol. 24, no. 11, pp. 1988–1998, 2005.

[11] J. G. Underwood, P. L. Boutz, J. D. Dougherty, P. Stoilov, and D. L. Black, “Homologues of the caenorhabditis elegans fox-1 protein are neuronal splicing regulators in mammals,” *Molecular and cellular biology*, vol. 25, no. 22, pp. 10005–10016, 2005.

[12] G. W. Yeo, E. L. Van Nostrand, and T. Y. Liang, “Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements,” *PLoS Genet*, vol. 3, no. 5, p. e85, 2007.

[13] C. Siebel, L. Fresco, and D. Rio, “The mechanism of somatic inhibition of drosophila p-element pre-mrna splicing: multiprotein complexes at an exon pseudo-5’splice site control u1 snrnp binding.,” *Genes & Development*, vol. 6, no. 8, pp. 1386–1401, 1992.

[14] R. Marcucci, F. E. Baralle, and M. Romano, “Complex splicing control of the human thrombopoietin gene by intronic g runs,” *Nucleic acids research*, vol. 35, no. 1, pp. 132–142, 2007.

[15] C. D. Chen, R. Kobayashi, and D. M. Helfman, “Binding of hnrnp h to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat β-tropomyosin gene,” *Genes & development*, vol. 13, no. 5, pp. 593–606, 1999.

[16] M. Caputi and A. M. Zahler, “Determination of the rna binding specificity of the heterogeneous nuclear ribonucleoprotein (hnrnp) h/h’/f/2h9 family,” *Journal of Biological Chemistry*, vol. 276, no. 47, pp. 43850–43859, 2001.

[17] C. P. Donahue, C. Muratore, J. Y. Wu, K. S. Kosik, and M. S. Wolfe, "Stabilization of the tau exon 10 stem loop alters pre-mrna splicing," *Journal of Biological Chemistry*, vol. 281, no. 33, pp. 23302–23306, 2006.

[18] Z. Wang and C. B. Burge, "Splicing regulation: from a parts list of regulatory elements to an integrated splicing code," *Rna*, vol. 14, no. 5, pp. 802–813, 2008.

[19] M. de la Mata, C. R. Alonso, S. Kadener, J. P. Fededa, M. Blaustein, F. Pelisch, P. Cramer, D. Bentley, and A. R. Kornblihtt, "A slow rna polymerase ii affects alternative splicing in vivo," *Molecular cell*, vol. 12, no. 2, pp. 525–532, 2003.

[20] H. J. Kang, Y. I. Kawasawa, F. Cheng, Y. Zhu, X. Xu, M. Li, A. M. Sousa, M. Pletikos, K. A. Meyer, G. Sedmak, *et al.*, "Spatio-temporal transcriptome of the human brain," *Nature*, vol. 478, no. 7370, pp. 483–489, 2011.

[21] E. T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth, and C. B. Burge, "Alternative isoform regulation in human tissue transcriptomes," *Nature*, vol. 456, no. 7221, pp. 470–476, 2008.

[22] Q. Pan, O. Shai, L. J. Lee, B. J. Frey, and B. J. Blencowe, "Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing," *Nature genetics*, vol. 40, no. 12, pp. 1413–1415, 2008.

[23] R. Tao, C. Li, E. N. Newburn, T. Ye, B. K. Lipska, M. M. Herman, D. R. Weinberger, J. E. Kleinman, and T. M. Hyde, "Transcript-specific associations of slc12a5 (kcc2) in human prefrontal cortex with development, schizophrenia, and affective disorders," *Journal of Neuroscience*, vol. 32, no. 15, pp. 5216–5222, 2012.

[24] P. Mazin, J. Xiong, X. Liu, Z. Yan, X. Zhang, M. Li, L. He, M. Somel, Y. Yuan, Y.-P. Phoebe Chen, *et al.*, "Widespread splicing changes in human brain development and aging," *Molecular systems biology*, vol. 9, no. 1, p. 633, 2013.

[25] C. Paterson, Y. Wang, J. E. Kleinman, and A. J. Law, “Effects of schizophrenia risk variation in the nrg1 gene on nrg1-iv splicing during fetal and early postnatal human neocortical development,” *American Journal of Psychiatry*, vol. 171, no. 9, pp. 979–989, 2014.

[26] R. Tao, K. Davis, C. Li, J. H. Shin, Y. Gao, A. Jaffe, M. Gondre-Lewis, D. Weinberger, J. Kleinman, and T. Hyde, “Gad1 alternative transcripts and dna methylation in human prefrontal cortex and hippocampus in brain development, schizophrenia,” *Molecular psychiatry*, vol. 23, no. 6, pp. 1496–1505, 2018.

[27] K. Wang, D. Wu, H. Zhang, A. Das, M. Basu, J. Malin, K. Cao, and S. Hannenhalli, “Comprehensive map of age-associated splicing changes across human tissues and their contributions to age-associated diseases,” *Scientific reports*, vol. 8, no. 1, pp. 1–12, 2018.

[28] P. V. Mazin, X. Jiang, N. Fu, D. Han, M. Guo, M. S. Gelfand, and P. Khaitovich, “Conservation, evolution, and regulation of splicing during prefrontal cortex development in humans, chimpanzees, and macaques,” *RNA*, vol. 24, no. 4, pp. 585–596, 2018.

[29] Q. Li, J.-A. Lee, and D. L. Black, “Neuronal regulation of alternative pre-mrna splicing,” *Nature Reviews Neuroscience*, vol. 8, no. 11, pp. 819–831, 2007.

[30] D. Schmucker, J. C. Clemens, H. Shu, C. A. Worby, J. Xiao, M. Muda, J. E. Dixon, and S. L. Zipursky, “Drosophila dscam is an axon guidance receptor exhibiting extraordinary molecular diversity,” *Cell*, vol. 101, no. 6, pp. 671–684, 2000.

[31] C. Schmauss, V. Haroutunian, K. L. Davis, and M. Davidson, “Selective loss of dopamine d3-type receptor mrna expression in parietal and motor cortices of patients with chronic schizophrenia,” *Proceedings of the National Academy of Sciences*, vol. 90, no. 19, pp. 8942–8946, 1993.

[32] C. Zhao, Z. Xu, J.-H. Chen, Z. Yu, K.-L. Tong, W.-S. Lo, F. Pun, S.-K. Ng, S.-Y. Tsang, and H. Xue, “Two isoforms of gaba(a) receptor beta2

subunit with different electrophysiological properties: Differential expression and genotypical correlations in schizophrenia," *Molecular psychiatry*, vol. 11, pp. 1092–105, 01 2007.

[33] C. Zhao, Z. Xu, F. Wang, J. Chen, S.-K. Ng, P.-W. Wong, Z. Yu, F. W. Pun, L. Ren, W.-S. Lo, *et al.*, "Alternative-splicing in the exon-10 region of gaba a receptor β 2 subunit gene: relationships between novel isoforms and psychotic disorders," *PloS one*, vol. 4, no. 9, p. e6977, 2009.

[34] M. M. Huntsman, B.-V. Tran, S. G. Potkin, W. E. Bunney, and E. G. Jones, "Altered ratios of alternatively spliced long and short γ2 subunit mrnas of the γ-amino butyrate type a receptor in prefrontal cortex of schizophrenics," *Proceedings of the National Academy of Sciences*, vol. 95, no. 25, pp. 15066–15071, 1998.

[35] S. M. Clinton, V. Haroutunian, K. L. Davis, and J. H. Meador-Woodruff, "Altered transcript expression of nmda receptor-associated postsynaptic proteins in the thalamus of subjects with schizophrenia," *American Journal of Psychiatry*, vol. 160, no. 6, pp. 1100–1109, 2003.

[36] S. Le Corre, C. G. Harper, P. Lopez, P. Ward, and S. Catts, "Increased levels of expression of an nmdar1 splice variant in the superior temporal gyrus in schizophrenia," *Neuroreport*, vol. 11, no. 5, pp. 983–986, 2000.

[37] L. Ding and A. N. Hegde, "Expression of rgs4 splice variants in dorsolateral prefrontal cortex of schizophrenic and bipolar disorder patients," *Biological psychiatry*, vol. 65, no. 6, pp. 541–545, 2009.

[38] S. Mexal, R. Berger, L. Pearce, A. Barton, J. Logel, C. E. Adams, R. G. Ross, R. Freedman, and S. Leonard, "Regulation of a novel αn-catenin splice variant in schizophrenic smokers," *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, vol. 147, no. 6, pp. 759–768, 2008.

[39] A. Gibbons, E. Thomas, and B. Dean, "Regional and duration of illness differences in the alteration of ncam-180 mrna expression within the cortex of subjects with schizophrenia," *Schizophrenia research*, vol. 112, no. 1-3, pp. 65–71, 2009.

[40] S. J. Huffaker, J. Chen, K. K. Nicodemus, F. Sambataro, F. Yang, V. Mattay, B. K. Lipska, T. M. Hyde, J. Song, D. Rujescu, *et al.*, "A primate-specific, brain isoform of kcnh2 affects cortical physiology, cognition, neuronal repolarization and risk of schizophrenia," *Nature medicine*, vol. 15, no. 5, pp. 509–518, 2009.

[41] A. J. Law, B. K. Lipska, C. S. Weickert, T. M. Hyde, R. E. Straub, R. Hashimoto, P. J. Harrison, J. E. Kleinman, and D. R. Weinberger, "Neuregulin 1 transcripts are differentially expressed in schizophrenia and regulated by 5' snps associated with the disease," *Proceedings of the National Academy of Sciences*, vol. 103, no. 17, pp. 6747–6752, 2006.

[42] R. Hashimoto, R. Straub, C. Weickert, T. Hyde, J. Kleinman, and D. Weinberger, "Expression analysis of neuregulin-1 in the dorsolateral prefrontal cortex in schizophrenia," *Molecular psychiatry*, vol. 9, no. 3, pp. 299–307, 2004.

[43] W. Tan, Y. Wang, B. Gold, J. Chen, M. Dean, P. J. Harrison, D. R. Weinberger, and A. J. Law, "Molecular cloning of a brain-specific, developmentally regulated neuregulin 1 (nrg1) isoform and identification of a functional promoter variant associated with schizophrenia," *Journal of Biological Chemistry*, vol. 282, no. 33, pp. 24343–24351, 2007.

[44] W.-T. Kao, Y. Wang, J. E. Kleinman, B. K. Lipska, T. M. Hyde, D. R. Weinberger, and A. J. Law, "Common genetic variation in neuregulin 3 (nrg3) influences risk for schizophrenia and impacts nrg3 expression in human brain," *Proceedings of the National Academy of Sciences*, vol. 107, no. 35, pp. 15619–15624, 2010.

[45] A. J. Law, J. E. Kleinman, D. R. Weinberger, and C. S. Weickert, "Disease-associated intronic variants in the erbb4 gene are related to altered erbb4 splice-variant expression in the brain in schizophrenia," *Human molecular genetics*, vol. 16, no. 2, pp. 129–141, 2007.

[46] L. J. Sartorius, D. R. Weinberger, T. M. Hyde, P. J. Harrison, J. E. Kleinman, and B. K. Lipska, "Expression of a grm3 splice variant is increased in

the dorsolateral prefrontal cortex of individuals carrying a schizophrenia risk snp," *Neuropsychopharmacology*, vol. 33, no. 11, pp. 2626–2634, 2008.

[47] K. Nakata, B. K. Lipska, T. M. Hyde, T. Ye, E. N. Newburn, Y. Morita, R. Vakkalanka, M. Barenboim, Y. Sei, D. R. Weinberger, *et al.*, "Disc1 splice variants are upregulated in schizophrenia and associated with risk polymorphisms," *Proceedings of the National Academy of Sciences*, vol. 106, no. 37, pp. 15873–15878, 2009.

[48] K. Åberg, P. Saetre, E. Lindholm, B. Ekholm, U. Pettersson, R. Adolfsson, and E. Jazin, "Human qki, a new candidate gene for schizophrenia involved in myelination," *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, vol. 141, no. 1, pp. 84–90, 2006.

[49] K. Aberg, P. Saetre, N. Jareborg, and E. Jazin, "Human qki, a potential regulator of mrna expression of human oligodendrocyte-related genes involved in schizophrenia," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 7482–7, 06 2006.

[50] M. Smith, R. Sharples, G. Evin, C. McLean, B. Dean, G. Pavey, E. Fantino, R. Cotton, K. Imaizumi, C. Masters, R. Cappai, and J. Culvenor, "Expression of truncated presenilin 2 splice variant in alzheimer's disease, bipolar disorder, and schizophrenia brain cortex," *Brain research. Molecular brain research*, vol. 127, pp. 128–35, 09 2004.

[51] G. Habl, A. Schmitt, M. Zink, M. Wilmsdorff, P. Yeganeh-Doost, A. Jatzko, T. Schneider-Axmann, M. Bauer, and P. Falkai, "Decreased reelin expression in the left prefrontal cortex (ba9) in chronic schizophrenia patients," *Neuropsychobiology*, vol. 66, pp. 57–62, 07 2012.

[52] C. Oldmeadow, D. Mossman, T.-J. Evans, E. Holliday, P. Tooney, M. Cairns, J. Wu, V. Carr, J. Attia, and R. Scott, "Combined analysis of exon splicing and genome wide polymorphism data predict schizophrenia risk loci," *Journal of Psychiatric Research*, vol. 52, 05 2014.

[53] R. Tao, H. Cousijn, A. Jaffe, P. Burnet, F. Edwards, S. Eastwood, J. Shin, T. Lane, M. Walker, B. Maher, D. Weinberger, P. Harrison, T. Hyde, and

J. Kleinman, "Expression of znf804a in human brain and alterations in schizophrenia, bipolar disorder, and major depressive disorder," *JAMA psychiatry*, vol. 71, 08 2014.

[54] O. Cohen, S. McCoy, F. Middleton, S. Bialosuknia, Y. Zhang-James, L. Liu, M. Tsuang, S. Faraone, and S. Glatt, "Transcriptomic analysis of postmortem brain identifies dysregulated splicing events in novel candidate genes for schizophrenia," *Schizophrenia research*, vol. 142, 10 2012.

[55] J. Wu, X. Wang, N. Beveridge, P. Tooney, R. Scott, V. Carr, and M. Cairns, "Transcriptome sequencing revealed significant alteration of cortical promoter usage and splicing in schizophrenia," *PloS one*, vol. 7, p. e36351, 04 2012.

[56] A. Uezato, N. Yamamoto, Y. Iwayama, S. Hiraoka, E. Hiraaki, A. Umino, E. Haramo, M. Umino, T. Yoshikawa, and T. Nishikawa, "Reduced cortical expression of a newly identified splicing variant of the dlg1 gene in patients with early-onset schizophrenia," *Translational psychiatry*, vol. 5, p. e654, 10 2015.

[57] O. Cohen, T. Weickert, J. Hess, L. Paish, S. McCoy, D. Rothmond, C. Galletly, D. Liu, D. Weinberg, X.-F. Huang, Q. Xu, Y. Shen, D. Zhang, W. Yue, J. Yan, L. Wang, T. Lu, L. He, Y. Shi, and S. Glatt, "A splicing-regulatory polymorphism in drd2 disrupts zranb2 binding, impairs cognitive functioning and increases risk for schizophrenia in six han chinese samples," *Molecular psychiatry*, vol. 21, 09 2015.

[58] M. Li, A. Jaffe, R. Straub, R. Tao, J. Shin, Y. Wang, Q. Chen, C. Li, Y. Jia, K. Ohi, B. Maher, N. Brandon, A. Cross, J. Chenoweth, D. Hoeppner, H. Wei, T. Hyde, R. McKay, J. Kleinman, and D. Weinberger, "A human-specific as3mt isoform and borcs7 are molecular risk factors in the 10q24.32 schizophrenia-associated locus," *Nature medicine*, vol. 22, 05 2016.

[59] M. Gandal, P. Zhang, E. Hadjimichael, R. Walker, C. Chen, S. Liu, H. Won, H. Bakel, M. Varghese, Y. Wang, A. Shieh, J. Haney, S. Parhami, J. Belmont, M. Kim, P. Moran Losada, Z. Khan, J. Mleczko, Y. Xia, and D. Geschwind,

"Transcriptome-wide isoform-level dysregulation in asd, schizophrenia, and bipolar disorder," *Science*, vol. 362, p. eaat8127, 12 2018.

[60] L. Ma, S. Semick, Q. Chen, C. Li, R. Tao, A. Price, J. Shin, Y. Jia, N. Brandon, A. Cross, T. Hyde, J. Kleinman, A. Jaffe, D. Weinberger, and R. Straub, "Schizophrenia risk variants influence multiple classes of transcripts of sorting nexin 19 (snx19)," *Molecular Psychiatry*, vol. 25, 01 2019.

[61] P. Kim, M. Scott, and J. Meador-Woodruff, "Dysregulation of the unfolded protein response (upr) in the dorsolateral prefrontal cortex in elderly patients with schizophrenia," *Molecular Psychiatry*, 10 2019.

[62] L. Ma, A. Shcherbina, and S. Chetty, "Variations and expression features of cyp2d6 contribute to schizophrenia risk," *Molecular Psychiatry*, 02 2020.

[63] C. Moody, A. Funk, E. Devine, R. Homan, D. Boison, R. Mccullumsmith, and S. O'Donovan, "Adenosine kinase expression in the frontal cortex in schizophrenia," *Schizophrenia bulletin*, vol. 46, pp. 690–698, 04 2020.

[64] S. M. Singh, C. A. Castellani, and K. A. Hill, "Postzygotic somatic mutations in the human brain expand the threshold-liability model of schizophrenia," *Frontiers in Psychiatry*, vol. 11, 2020.

[65] S. Ripke, B. Neale, A. Corvin, J. Walters, K.-H. Farh, P. Holmans, P. Lee, B. Bulik-Sullivan, D. Collier, H. Huang, T. Pers, I. Agartz, E. Agerbo, M. Albus, M. Alexander, F. Amin, S. Bacanu, M. Begemann, R. Belliveau, and M. O'Donovan, "Biological insights from 108 schizophrenia-associated genetic loci," *Nature*, vol. 511, 07 2014.

[66] B. Blencowe, "Alternative splicing: New insights from global analyses," *Cell*, vol. 126, pp. 37–47, 08 2006.

[67] G. Yeo, D. Holste, G. Kreiman, and C. Burge, "Variation in alternative splicing across human tissues," *Genome biology*, vol. 5, p. R74, 02 2004.

[68] X. Ding, S. Liu, M. Tian, W. Zhang, T. Zhu, D. Li, J. Wu, H. Deng, Y. Jia, W. Xie, H. Xie, and J.-S. Guan, "Activity-induced histone modifications

govern neurexin-1 mrna splicing and memory preservation," *Nature neuroscience*, vol. 20, 03 2017.

[69] N. Salomonis, J. Baumbach, T. Lengauer, B. Conklin, and M. Albrecht, "Altanalyze and domaingraph: Analyzing and visualizing exon expression data," *Nucleic acids research*, vol. 38, pp. W755–62, 07 2010.

[70] J. Gertz, K. Varley, N. Davis, B. Baas, I. Goryshin, R. Vaidyanathan, S. Kuersten, and R. Myers, "Transposase mediated construction of rna-seq libraries," *Genome research*, vol. 22, pp. 134–41, 11 2011.

[71] X. Wang and M. Cairns, "Understanding complex transcriptome dynamics in schizophrenia and other neurological diseases using rna sequencing," *International review of neurobiology*, vol. 116C, pp. 127–152, 08 2014.

[72] S. Anders, A. Reyes, and W. Huber, "Detecting differential usage of exons from rna-seq data," *Genome research*, vol. 22, pp. 2008–17, 06 2012.

[73] J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, G. Walters, F. Garcia, N. Young, *et al.*, "The genotype-tissue expression (gtex) project," *Nature genetics*, vol. 45, no. 6, pp. 580–585, 2013.

[74] B. Chilian, H. Abdollahpour, T. Bierhals, I. Haltrich, G. Fekete, I. Nagel, G. Rosenberger, and K. Kutsche, "Dysfunction of shank2 and chrna7 in a patient with intellectual disability and language impairment supports genetic epistasis of the two loci," *Clinical genetics*, vol. 84, 01 2013.

[75] O. Senormanci, S. Celik, E. Valipour, V. Dogan, and G. Senormanci, "Determination of candidate genes involved in schizophrenia using the whole-exome sequencing," *Bratislava Medical Journal*, vol. 119, pp. 572–576, 01 2018.

[76] M. Caracci, L. Fuentealba, and M.-P. Marzolo, "Golgi complex dynamics and its implication in prevalent neurological disorders," *Frontiers in Cell and Developmental Biology*, vol. 7, p. 75, 05 2019.

[77] B. Jones, J. Deem, T. Younts, M. Weisenhaus, C. Sanford, M. Slack, J. Chin, D. Nachmanson, A. McKennon, P. Castillo, and G. McKnight, "Targeted

deletion of akap7 in dentate granule cells impairs spatial discrimination," *eLife*, vol. 5, 12 2016.

[78] K. Trotter, I. Fraser, G. Scott, J. Stutts, J. Scott, and S. Milgram, "Alternative splicing regulates the subcellular localization of a-kinase anchoring protein 18 isoforms," *The Journal of cell biology*, vol. 147, pp. 1481–92, 12 1999.

[79] S. M. Wojcik, M. Tantra, B. Stepniak, K. nok M. Man, K. Muller-Ribbe, M. Begemann, A. Ju, S. Papiol, A. Ronnenberg, A. Gurvich, Y. Shin, I. Augustin, N. Brose, and H. Ehrenreich, "Genetic markers of a munc13 protein family member, BAIAP3, are gender specifically associated with anxiety and benzodiazepine abuse in mice and humans," *Molecular Medicine*, vol. 19, pp. 135–148, jan 2013.

[80] K. Ochi, T. Mori, Y. Toyama, Y. Nakamura, and H. Arakawa, "Identification of semaphorin3b as a direct target of p53," *Neoplasia : An International Journal for Oncology Research*, vol. 4, 01 2002.

[81] F. Winter, J. Kwok, J. Fawcett, T. Vo, D. Carulli, and J. Verhaagen, "The chemorepulsive protein semaphorin 3a and perineuronal net-mediated plasticity," *Neural Plasticity*, vol. 2016, pp. 1–14, 01 2016.

[82] K. Hattori, H. Fukuzako, T. Hashiguchi, S. Hamada, Y. Murata, T. Isosaka, S. Yuasa, and T. Yagi, "Decreased expression of fyn protein and disbalanced alternative splicing patterns in platelets from patients with schizophrenia," *Psychiatry research*, vol. 168, pp. 119–28, 08 2009.

[83] R. Porter, F. Jaamour, and S. Iwase, "Neuron-specific alternative splicing of transcriptional machineries: Implications for neurodevelopmental disorders," *Molecular and Cellular Neuroscience*, vol. 87, 12 2017.