The Dissertation Committee for Chang Hoon Lee Certifies that this is the approved version of the following dissertation:

SPLICING OF HUMAN GABAB RECEPTOR SUBUNIT 1 (GABAB1) IN NON-ALCOHOLIC AND ALCOHOLIC BRAINS

Committee:

R. Adron Harris, Supervisor

R. Dayne Mayfield, Co-Supervisor

Nigel S. Atkinson

Vishwanath R. Iyer

Jonathan T. Pierce-Shimomura

Scott P. Hunicke-Smith

SPLICING OF HUMAN GABAB RECEPTOR SUBUNIT 1 (GABAB1) IN NON-ALCOHOLIC AND ALCOHOLIC BRAINS

by

Chang Hoon Lee, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2011

Acknowledgements

I would like to thank to my supervisor, Dr. R. Adron Harris, and co-supervisor, Dr. R. Dayne Mayfield, for their time, guidance, generosity, encouragement, and patience. I especially appreciate their input to improve my writing, presentation skills, and scientific understanding. Many times they gave me direction when I was confused and encouraged me to go on when I was frustrated. They are undoubtedly great supervisors.

My thanks also go to all my committee members, Drs. Nigel Atkinson, Vishwanath Iyer, Jonathan Pierce-Shimomura, and Scott Hunicke-Smith, for constructive criticism, invaluable discussion, and recommendations. I am especially appreciative of Drs. Vishwanath Iyer and Scott Hunicke-Smith for their advice and technical support on next generation sequencing experiments.

I would like to thank Dr. Claus Wilke, Dr. Shyamal Mitra, Dhivya Arasappan, and Dr. Gyoungju Nah for helping me to develop bioinformatics knowledge or allowing me to audit their classes. I would like to extend a special thanks to Dr. Claus Wilke for detailed computational and statistical assistance.

I also thank my father and mother, Choong Yul Lee and Sang Sook Lee, for their love and support, and my brother, Dong Hoon Lee, for his friendship.

I am grateful to all members of the Waggoner Center for Alcohol and Addiction Research for their scientific, technical, and personal support. In particular, I thank Debbie

James, Dr. Igor Ponomarev, Dr. Cecilia Borghese, Dr. Yuri Blednov, Jana Cormack, Dr. Giorgio Gorini, Dr. David Johnson II, Dr. Rebecca Howard, Dr. Yury Núñez, Dr. Jody Mayfield, Lindsay McCracken, Mandy McCracken, Dana Most, Jay Truitt, Elizabeth Osterndorf-Kahanek, Olga Ponomareva, Marsha Berkman, Gayatri Tiwari, and Jill Benavidez.

SPLICING OF HUMAN GABAB RECEPTOR SUBUNIT 1 (GABAB1) IN NON-ALCOHOLIC AND ALCOHOLIC BRAINS

Chang Hoon Lee, Ph.D.

The University of Texas at Austin, 2011

Supervisor: R. Adron Harris

Co-Supervisor: R. Dayne Mayfield

Gamma-aminobutyric acid type B (GABAB) receptor is a G protein coupled receptor (GPCR) that mediates decreased neural activity. It has two subunits, GABAB1 and GABAB2. Previous complementary DNA (cDNA) microarray data showed strong GABAB1 signals from human prefrontal cortex using an intron 4 region probe, and these studies indicated that novel intron 4 containing GABAB1 splicing variants exist. We cloned GABAB1k, l, m, and n including mouse GABAB1j. Expression of these variants are much lower than other major known splicing variants, but GABAB1k, l, m, and n levels are similar across brain tissues. GABAB1l and GABAB1m impair GABAB receptor induced function. To better define GABAB1 splicing in alcoholic brains, whole transcriptome shortgun sequencing (RNA-seq) experiments were proposed. Due to the complexity of GABAB1 splicing, we used gene specific libraries as well as whole transcriptome libraries to maximize GABAB1 specific splicing junction search. The splicing junction search data found that GABAB1 gene is 2 to 3 times longer than the

previous known gene length. Extremely low expression at 5’ end exons was found, and GABAB1 exons were grouped based on expression levels. Chronic alcohol altered exon/intron expression and splicing junctions more than overall gene expression. Decreased exon expression at a GABA binding site, a transmembrane domain (TM), and a microRNA (miRNA) binding site may diminish the normal GABAB1 transcript population and compromise signal transduction following chronic alcohol exposure. This may explain why GABAB receptor agonists have therapeutic benefit in treating alcoholism. During the sequence mapping, read pile-ups and gaps were found from whole transcriptome libraries in known exons. These may prevent single nucleotide polymorphism (SNP) and splicing junction identification and gene expression calculations. Sequence analysis found sequence biases from their mapped reads. The major sequence biases were from RNaseIII RNA fragmentation and T4 polynucleotide kinase (T4PNK) reaction. Heat fragmentation and OptiKinase treatment removed the read pile-ups and gaps including the sequence biases. The identification of RNaseIII target sequences can be incorporated into methods of miRNA gene prediction. These data showed the complexity of GABAB1 receptor splicing and the perturbation of splicing by chronic alcohol abuse demonstrate the power of RNA-seq to provide new insight into gene expression and the role of GABAB receptors in alcoholism. In addition, many other important brain genes may have unexplored splicing variants which will be important for alcoholism and other psychiatric diseases. Also, new RNA-seq library constructions improved the quality of gene expression studies.

Table of Contents

List of Tables

List of Figures

Abbreviations

GABAB1	gamma-aminobutyric acid type B receptor subunit 1
GABAB	gamma-aminobutyric acid type B
GPCR	G protein coupled receptor
cDNA	complementary DNA
RNA-seq	whole transcriptome shortgun sequencing
TM	transmembrane domain
miRNA	microRNA
SNP	single nucleotide polymorphism
T4PNK	T4 polynucleotide kinase
CNS	central nervous system
PNS	peripheral nervous system
FAS	fetal alcohol syndrome
VTA	ventral tegmental area
NAc	nucleus accumbens
WKS	Wernicke-Korsakoff syndrome
GlyR	glycine receptor
AC	adenylyl cyclase
ORF	open reading frame
UTR	untranslated region
ATF4	activating transcription factor 4
GHB	gamma-Hydroxybutyrate
RNAi	RNA interference

HPN	hippocampus pyramidal neuron
IPSP	inhibitory postsynaptic potential
IPSC	inhibitory postsynaptic current
CeA	central amygdala nucleus
snRNP	small nuclear ribonucleoprotein
AE2	anion exchange protein 2
VDR	vitamin D receptor
HNF4 alpha	hepatocyte nuclear factor 4 alpha
SPG4	spastic paraplegia type 4
EST	expressed sequence tag
TRPV1	Transient receptor potential cation channel subfamily V member 1
Sxl	Sex-lethal
BDNF	brain-derived neurotrophic factor
NMDA	N-Methyl-D-aspartate
BK	large conductance Ca^{2+} activated K^{+}
BLA	basal lateral amygdala
bp	base pair
SOLiD	sequencing by oligonucleotide ligation and detection
ePCR	emersion PCR
Mb	megabyte
RT	reverse transcription
rRNA	ribosomal RNA
RACE	rapid amplification of cDNA ends
GIRK	G protein-activated inwardly rectifying K+ channel
aa	amino acids

cRNA	complementary RNA
49mM HK	49 mM high-K^+ solution
SEM	standard error of the mean
HEK-293	human embryonic kidney 293
GUSB	beta-glucuronidase
RefSeq	Reference Sequence
dscDNA	double strand cDNA
sscDNA	single strand cDNA
IGV	Integrative Genomics Viewer
GSL	gene specific library
WTL	whole transcriptome library
UCSC	The University of California, Santa Cruz
RPKM	reads per kilobase of exon model per million mapped reads
RPGM	reads per kilobase of gene model per million mapped reads
RPJM	reads per kilobase of splicing junction model per million uniquely mapped reads
TRC	Tissue Resource Centre
NIAAA	National Institute on Alcohol Abuse and Alcoholism
PMI	postmortem interval
GSAF	Genomic Sequencing and Analysis Facility
tRNA	transfer RNA
TACC	Texas Advanced Computing Center
ssRNA	single stranded RNA
Co-IP	co-immunoprecipitation
ChIP	chromatin immunoprecipitation

EMSA	electrophoretic mobility shift assay
snoRNA	small nucleolar RNA
ChIP-seq	chromatin immunoprecipitation sequencing
HA	hemagglutinin
Myc	anti c-Myc
EtOH	ethanol
min	minute

CHAPTER 1. INTRODUCTION

1.1 ALCOHOL AND ALCOHOLISM

Alcohol was discovered as a fermented beverage as early as 10,000 B.C. and was commonly used in many cultures. Evidence of alcoholic beverage use has been found from all over the world, for example Persia, Egypt, China, India, Babylon, Greece, and Rome. The geographical and sociological conditions affected the types of alcoholic beverages consumed (McGovern, Zhang et al. 2004; McGovern, Mirzoian et al. 2009; McGovern, Christofidou-Solomidou et al. 2010).

Historically, doctors promoted alcoholic beverage drinking for health benefits and most recently for cardiovascular disease protection. Though there is the evidence of cardiovascular benefit from 1 - 2 drinks per day, the benefit is controversial (Sellman, Connor et al. 2009; Sinkiewicz and Weglarz 2009). However, there is a strong correlation between high level of alcohol usage and cardiovascular disease, chronic pancreatitis, alcoholic liver disease, cancer, central nervous system (CNS) damage, and peripheral nervous system (PNS) damage. Long-tern alcohol usage in excessive amounts can damage nearly all organs in the body (Muller, Koch et al. 1985; Testino 2008).

Excessive alcohol intake is one of the leading causes of death in the world today. One study showed that alcohol can be linked to one in every 25 deaths worldwide (Lopez, Mathers et al. 2006). In the United States, deaths resulting from alcohol abuse

reached about 100,000 (McGinnis and Foege 1993), and excessive alcohol usage is currently the third leading cause of death. About $200 billion is spent per year to treat alcohol abuse (Gordis 2000).

There is no cure for alcohol addiction. Currently, there are a few drugs that have been approved for the treatment of alcoholism in the U.S., e.g. disulfiram, acamprosate, and naltrexone (Bouza, Angeles et al. 2004; Kenna, McGeary et al. 2004). Ondansetron, aripiprazole, topiramate, and baclofen also showed promising effects for chronic alcohol use (Bouza, Angeles et al. 2004; Johnson 2004; Kenna, McGeary et al. 2004; Enserink 2011). However, the treatments are not successful for all patients, and molecular mechanisms of alcoholism are still undefined.

1.2 ALCOHOL AND BRAIN DAMAGE

The brain is one of the most sensitive organs to the toxic effects of chronic alcohol consumption. Excessive alcohol use can result in brain development impairment, lesion introducing dementia, cognitive disorder, and depression (Olegard, Sabel et al. 1979; Neiman 1998; Pierucci-Lagha and Derouesne 2003; Heffernan 2008; Fergusson, Boden et al. 2009). Drinking during pregnancy can also impair brain development and cause fetal alcohol syndrome (FAS). Children with FAS have distinct facial features and small brains. The smaller brains have fewer neurons and lead to learning and behavior problems (Olegard, Sabel et al. 1979).

A severe consequence of long term heavy alcohol consumption is also associated with brain lesions. This brain damage is linked to a direct toxic effect of alcohol, alcohol withdrawal, and nutrient deficiency (Neiman 1998). Brain imaging studies showed significant loss of volume in alcoholic brains. Brain white matter was significantly correlated with the amount of alcohol consumption (Jensen and Pakkenberg 1993; Kril and Halliday 1999). One postmortem human brain study showed the reduction of neuronal area in several cortical brain regions, e.g. frontal cortex, motor cortex, and cingulate cortex (Kril and Harper 1989).

Chronic alcohol consumption damages prefrontal cortex more than the other brain region. Selective neuronal loss in the prefrontal cortex was found in alcoholics (Kril and Harper 1989). Chronic alcohol induced cognitive disorder, such as planning ability especially in speed and flexibility, suggested prefrontal cortex damage (Flatscher-Bader, van der Brug et al. 2005; Yizhar, Fenno et al. 2011). The prefrontal cortex is involved in psychopathology (Alexander and Brown 2011; Yizhar, Fenno et al. 2011) and is connected to and regulates mesolimbic dopaminergic pathway. This pathway is involved in reward and is the primary site of the pharmacological action of alcohol and drugs of abuse. It projects from ventral tegmental area (VTA) to nucleus accumbens (NAc), amygdala, and prefrontal cortex. Alcohol and other drugs activate the mesolimbic dopaminergic pathway, and prefrontal cortex and the amygdala send excitatory projections to the VTA (Piomelli 2001; O'Brien and Gardner 2005) (Figure 1.1).

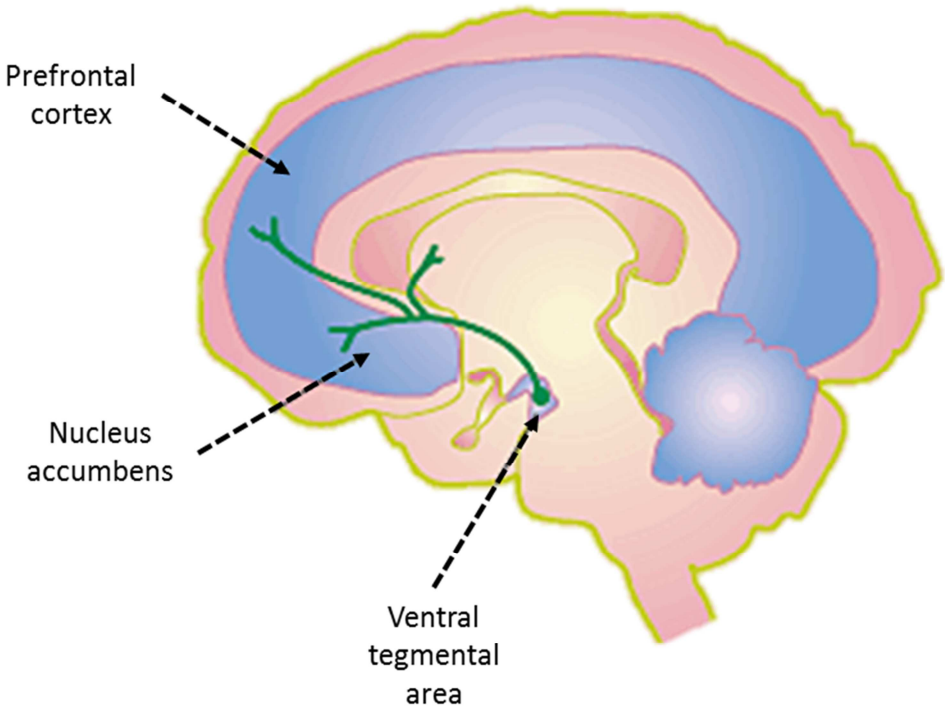


Figure 1.1 Mesolimbic dopaminergic pathway.
This mesolimbic dopaminergic pathway projects from VTA (ventral tegmental area) to NAc (nucleus accumbens) and prefrontal cortex. Green lines indicate connections in the mesolimbic dopaminergic pathway. (modified from O'Brien and Gardner 2005)

1.3 GENE EXPRESSION CHANGES IN ALCOHOLIC PREFRONTAL CORTEX

Genetic and environmental factors almost equally contribute to the development of alcoholism (Chao and Nestler 2004). Gene expression can be regulated by alcohol treatment in neuronal cells (Wilke, Sganga et al. 1994). Previous microarray studies showed gene expression changes in postmortem prefrontal cortex of alcoholics (Liu, Lewohl et al. 2004; Flatscher-Bader, van der Brug et al. 2005). As previously mentioned, prefrontal cortex regulates the mesolimbic dopaminergic pathway that is involved in reward. Therefore, gene expression changes in alcoholic prefrontal cortex are of interest to better understand alcohol addiction. Previous microarray studies compared gene expression in alcoholic and non-alcoholic prefrontal cortices. One study detected 533 alcohol responsive genes (Liu, Lewohl et al. 2006). Among the 533 genes, 27 were selected as candidate genes because they were previously identified as alcohol responsive genes in other microarray studies (Lavoie and Butterworth 1995; Lewohl, Wang et al. 2000; Mayfield, Lewohl et al. 2002; Sokolov, Jiang et al. 2003; Flatscher-Bader, van der Brug et al. 2005). These genes were divided into 12 functional groups. Among the functional groups that were identified, cell adhesion, myelination, neural disease, and neurogenesis groups have more than three genes among the 27 genes. Alcoholism was more related to the neural disease group than the other groups. The neural disease group had Calpain 3, GABAB1, Transferrin, and Transketolase (Liu, Lewohl et al. 2006). Transketolase and GABAB1 were previously implicated in chronic alcoholism by other studies.

Among the candidate genes from the microarray studies, the GABAB receptor was of interest because several GABAB agonists are effective in alcohol withdrawal syndrome and promotion of alcohol abstinence and are currently being used for treating alcoholics (Krupitsky, Burakov et al. 1993; Addolorato, Leggio et al. 2006). Thus, GABAB1 was selected as a research target.

1.4 GABAB RECEPTOR

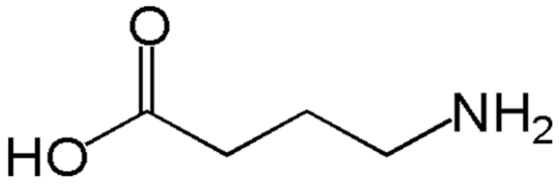
GABA and glycine are the major inhibitory neurotransmitters in vertebrate CNS. Both neurotransmitters bind to their respective receptors to activate chloride-sensitive channels. GABA receptors are also critical drug targets for anticonvulsants, sedatives and anesthetics (Figure 1.2A). GABAA and glycine receptors (GlyR) are highly homologous; however, GABAA receptors are abundant in the cortex and cerebellum while GlyRs predominate in the spinal cord and brain stem (Langosch, 1995; Legendre, 2001). Another type of GABA receptor, the GABAB receptor, is highly expressed in brain and also expressed in heart, small intestine, uterus and the other tissues at lower levels (Calver, Medhurst et al. 2000).

1.4.1 GABAB receptor function

The GABAB receptor is a GPCR which decreases neural activity. The intracellular domain of GABAB1 dimerizes with GABAB2. GABA binds to the extracellular domain of GABAB1 and transfers signals through G proteins. G protein α

subunits are linked with adenylyl cyclase (AC), and G protein βγ subunits alter presynaptic Ca^{2+} channels and postsynaptic K^+ channels. The GABAB receptor decreases the activity of AC and decreases neurotransmitter release by inhibiting Ca^{2+} influx through presynaptic Ca^{2+} channels. At postsynaptic neurons, it activates postsynaptic K^+ channels, and outward K^+ current induces hyperpolarization and K^+ equilibrium potential. Therefore, it prevents Na^+ channel opening and action potential firing (Couve, Moss et al. 2000). Thus, GABAB receptor is considered an inhibitory receptor (Figure 1.2B).

A.



B.

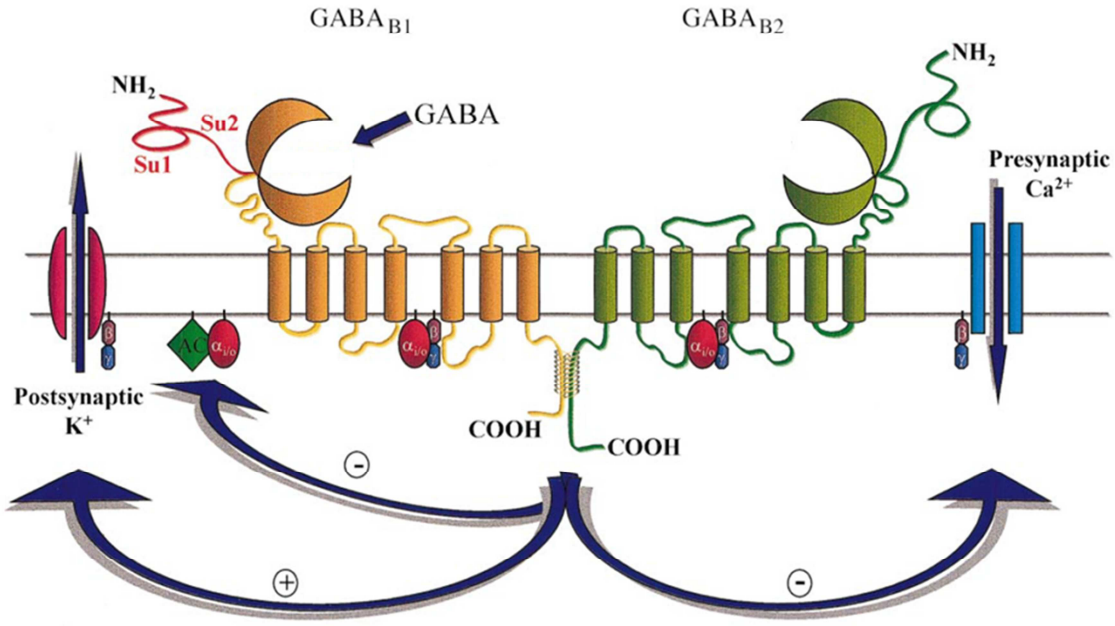


Figure 1.2 GABA and GABAB receptor function.
A. The chemical structure of GABA. B. The function of GABAB receptor. The intracellular domain of GABAB1 dimerizes with GABAB2. It inhibits AC and presynaptic Ca^{2+} channel. It enhances postsynaptic K^{+} channel. Thus, GABAB receptor decreases neural excitability. (modified from Couve et al, 2000)

1.4.2 GABAB1 splicing variants

GABAB1 has two major splicing variants (GABAB1a and b) in human CNS. GABAB1a is the longest splicing variant and has 23 exons (Figure 1.3A). GABAB1b is an N-terminal truncated form of GABAB1a, and its alternative N-terminal is from intron 5 of GABAB1a. GABAB1a and b are the most widely studied. GABAB1a appears to be mainly presynaptic and coupled with Ca^{2+} channels. GABAB1b might be a postsynaptic receptor and linked with K^{+} channels. However, their localization and functions are not clear (Couve, Moss et al. 2000; Billinton, Ige et al. 2001). GABAB1c has no exon 4 of GABAB1a, and GABAB1d has a partial deletion of exon 23. The functions of GABAB1c and d have not been well studied (Isomoto, Kaibara et al. 1998). GABAB1e is mainly expressed in peripheral tissues. Because it has no exon 15, it is a C-terminal truncated form of GABAB1a. It prevents GABAB1a and GABAB2 heterodimerization (Schwarz, Barry et al. 2000). In rats, there is one more GABAB1 isoform, GABAB1j, expressed as highly as GABAB1a and b in CNS. It has only exons 1 through 4 and 3' intron 4 of GABAB1a. Therefore, it is a very small C-terminal truncated GABAB1 isoform. It binds to neuronal membranes, and its purified sushi domains impair the inhibitory effect of the GABAB receptor in whole cell patch clamp experiments (Figure 1.3B) (Tiao, Bradaia et al. 2008).

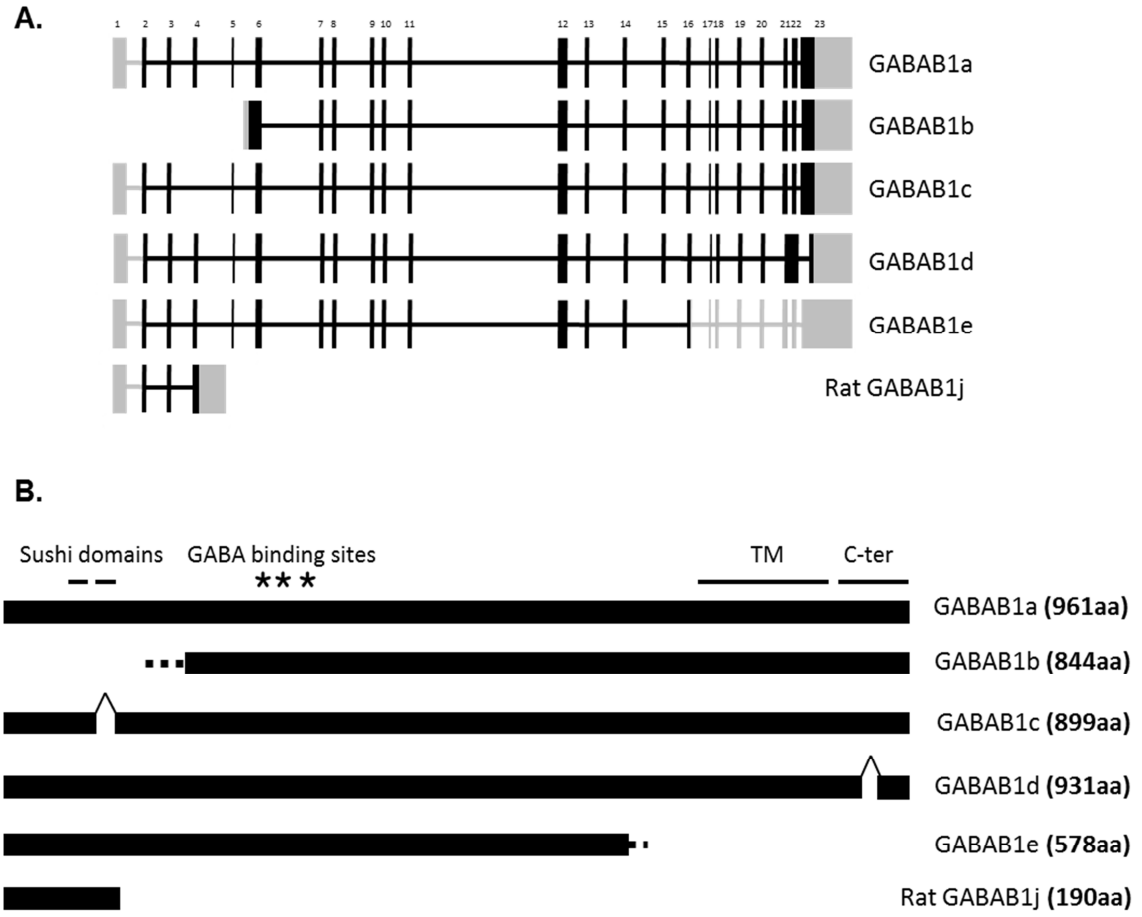


Figure 1.3 GABAB1 splicing variants.
A. Exon and intron structures of GABAB1 splicing variants. Human GABAB1 splicing variants are GABAB1a, b, c, d, and e. GABAB1j exists in rats. Black lines are open reading frames (ORFs), and grey lines indicate untranslated regions (UTRs). Thick and thin lines represent exons and introns, respectively. B. Functional domains of GABAB1 splicing variants. GABAB1 functional domains are sushi domains, GABA binding sites, TMs, and C-terminal intracellular domain. Thick lines are ORFs of GABAB1 splicing variants, and bent thin lines are for splicing outs. Dashed lines represent UTRs.

1.4.3 GABAB1 functional domains

GABAB1 has three major functional domains, e.g. GABA binding sites, TMs, and C-terminal intracellular domain (Figure 1.3B). These are common domains for two major splicing variants, GABAB1a and b. GABAB1 has a long extracellular region. It has three GABA binding sites, Ser359, Tyr479, and Asp584 of GABAB1a, and they are located at exon 7, 10, and 12, respectively. G proteins bind to an intracellular region between TM 3 and TM 4 for signal transduction. C-terminal intracellular domain has a GABAB1 and GABAB2 dimerization region as well as an activating transcription factor 4 (ATF4) interaction region. Two sushi domains are at exon 3 and 4 (Couve, Moss et al. 2000).

Though GABAB1a contains all of the domains, N-terminal truncated GABAB1b does not have sushi domains (Couve, Moss et al. 2000). GABAB1c has only one sushi domain due to exon 4 that has been spliced out. The splicing out of 5' exon 23 provided GABAB1d with a short C-terminal intracellular domain (Isomoto, Kaibara et al. 1998). GABAB1e does not have TM and C-terminal intracellular domains (Schwarz, Barry et al. 2000). Rat GABAB1j has only sushi domains (Tiao, Bradaia et al. 2008) (Figure 1.3B).

1.4.4 GABAB receptor and alcoholism

GABAB receptor studies showed a strong link between the GABAB receptor and alcoholism. The positive allosteric modulator, GS39783, suppresses alcohol drinking and reinforcement in rats (Maccioni, Fantini et al. 2008). A GABAB receptor agonist, gamma-Hydroxybutyrate (GHB), reduces voluntary ethanol drinking and withdrawal symptoms and is used as a treatment for alcoholism. GHB decreases inhibitory synaptic

transmission via GABAB receptors in CA1 of hippocampus (Schweitzer, Roberto et al. 2004; Maccioni, Pes et al. 2008). Baclofen is another GABAB receptor agonist and is also used to treat alcoholism (Enserink 2011). In electrophysiology studies, ethanol potentiation of Cl^- conductance required activation of both GABAA and GABAB receptors based on treatment with the GABAA antagonist, bicuculline, or GABAB antagonists, phaclofen or 2OH-saclofen (Allan, Burnett et al. 1991). This indicated that GABAB receptors are also involved in ethanol's mechanism of action.

From two previous microarray experiments, GABAB1 expression was increased in human prefrontal cortices from alcoholics (Flatscher-Bader, van der Brug et al. 2005; Liu, Lewohl et al. 2006). GABAB1 is also involved in behavioral actions of ethanol. In *Drosophila*, both GABAB1 RNA interference (RNAi) and the GABAB receptor antagonist, CGP 54626, reduced the behavior impairing effects of ethanol (Dzitoyeva, Dimitrijevic et al. 2003).

However, other electrophysiological studies using GABAB receptor antagonists showed that an ethanol effect on GABAB receptors is controversial. In hippocampus pyramidal neurons (HPNs), GABAergic inhibitory postsynaptic potentials (IPSPs) were enhanced by low ethanol concentrations only after application of the GABAB antagonist, CGP-35348 (Wan, Berton et al. 1996). Another GABAB receptor antagonist, CGP 55845A, did not change the ethanol enhancement of IPSPs and inhibitory postsynaptic currents (IPSCs) in central amygdala nucleus (CeA) neurons (Roberto, Madamba et al. 2003). Therefore, GABAB receptor involvement in ethanol actions may vary with different antagonists and different neurons.

1.5 SPLICING

1.5.1 Splicing overview

Splicing is a RNA modification after transcription, which removes introns and joins exons. For many eukaryotic RNAs, most splicings are done in a series of reactions of spliceosome, a complex of small nuclear ribonucleoproteins (snRNPs). Within introns, there are three conserved sites for splicing, e.g. a 5' splice site, a 3' splice site, and a branch site (Figure 1.4). The 5' splice site, a splicing donor, includes a conserved sequence, GU, at the 5' end of introns. The 3' splice site is called a splicing acceptor and is terminated with mostly AG sequence. Upstream from this pyrimidine rich region is the branch point, and its nucleotide sequence is A (Black 2003).

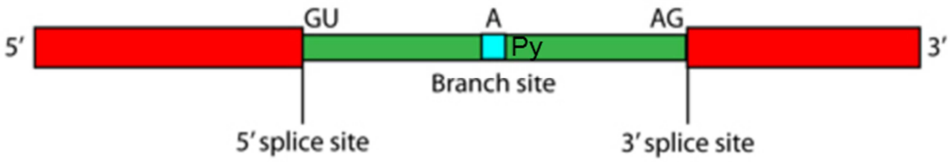


Figure 1.4 Splicing sites.
Spliced introns contain several important and conserved sites for splicing. A 5’ splice site has 5’ GU sequence, a branch site, A, located near a pyrimidine rich region (Py), and a 3’ splice site has 3’ AG sequence.
(modified from http://www.hartnell.edu/tutorials/biology/splicing.html)

1.5.2 Splicing functions

Splicing affects the function of splicing variants. A large number of splicing variants can either have novel functions or complete loss-of-function (Stamm, Ben-Ari et al. 2005).

Splicing also determines the cellular localization of splicing variants. GABAB1a and GABAB1b localize specifically at pre- and post-synaptic locations, respectively (Vigot, Barbieri et al. 2006). Splicing alters membrane interaction and influences signaling. The splicing variants that lack membrane binding properties accumulate in the cytosol or are secreted into the extracellular space (Stamm, Ben-Ari et al. 2005). The newly located splicing variants can have different functions. C-terminal GABAB1 splicing variants, GABAB1e and GABAB1j, lose membrane binding properties since they lack TMs. They inhibit other major GABAB receptor functions in both intracellular and extracellular spaces (Schwarz, Barry et al. 2000; Tiao, Bradaia et al. 2008). This phenomenon has been observed for splicing variants of other secreted genes that lose TMs by alternative splicing (Ezzat, Zheng et al. 2001; Bulanova, Budagian et al. 2007).

Different splicing variants exist in specific tissues or cell types. A C-terminal truncated GABAB1 splicing variant, GABAB1e, is strongly expressed in peripheral tissues though other major GABAB splicing variants mainly exist in CNS (Schwarz, Barry et al. 2000). GABAB1a mostly expresses in granule cells, and an N-terminal truncated splicing variant, GABAB1b, is in Purkinje cells (Billinton, Ige et al. 2001). Like GABAB1 splicing variants, spatially different localization patterns were also found from other N-terminal truncated splicing variants, including Cl^-/HCO_3^- exchanger

isoform anion exchange protein 2 (AE2), vitamin D receptor (VDR; a member of the nuclear hormone receptor superfamily that includes steroid, thyroid, retinoid, and orphan receptors), acetyl-CoA carboxylase-alpha, and P2 promoter-derived hepatocyte nuclear factor 4 alpha (HNF4 alpha) (Wang, Schultheis et al. 1996; Crofts, Hancock et al. 1998; Sunn, Cock et al. 2001; Travers, Vallance et al. 2003; Huang, Levitsky et al. 2009).

The majority of splicing alterations result in loss-of-function. Mutation of a splice site can destroy spliceosome binding site. This prevents normal splicing mechanism and results in loss-of-function. This generates transcripts that contain undesired introns that often compromise gene function. Autosomal dominant spastic paraplegia patients have frequent splice site mutations in spastic paraplegia type 4 (SPG4) gene. The altered splicing causes loss-of-function of SPG4 in dominant spastic paraplegia (Patrono, Casali et al. 2002).

1.5.3 Splicing regulators

Splicing is not always conserved across species and is regulated by evolution. The vast majority of human splicing patterns in the expressed sequence tag (EST) database are not conserved in mouse (Yeo, Van Nostrand et al. 2005). The different splicings provide species specific phenotypes. Transient receptor potential cation channel subfamily V member 1 (TRPV1) of vampire bats shows a species specific splicing pattern in ganglia and generates infrared sensation (Gracheva, Cordero-Morales et al. 2011). For GABAB1 gene splicing, human, mouse, and rat have their own unique splicing variants based on different splicing site sequences at their introns (Billinton, Ige

et al. 2001; Wei, Jia et al. 2001; Tiao, Bradaia et al. 2008; Lee, Mayfield et al. 2010). Minor splicing variants produced by rare splicing events were less conserved in other genomes. The minor splicing variants were lost in one genome after divergence of the two genomes during evolution (Modrek and Lee 2003).

Splicing can be regulated by development. Developmental stages provide unique cellular environments. The splicing of Sex-lethal (sxl) gene in female *Drosohpila* is regulated by developmental stages. The sxl splicing variants generated in the early embryogenesis stage masks a specific sxl gene splicing signal for the second intron in the late embryo resulting in a different splicing variant (SchuLtt, Hilfiker et al. 1998).

Chemical signals also regulate splicings. Anticancer chemotherapeutic agents affect the splicings of Bcl-x gene and other 95 apoptotic genes (Shkreta, Froehlich et al. 2008). Cocaine abuse regulates brain-derived neurotrophic factor (BDNF) splicing and changes its expression in brain (Liu, Lu et al. 2006).

Chronic ethanol treatment is another splicing regulator. An N-Methyl-D-aspartate (NMDA) receptor subunit, NR2, is known as a potential target of alcohol action. Among NR2 splicing variants, chronic ethanol exposure specifically increased NR2B expression, and 18 hour withdrawal brought its expression back to normal (Hardy, Chen et al. 1999). Chronic ethanol exposure also increased 5' splicing variant expression of another NMDA receptor subunit, NR1, though it had no effect on NR1 3' variant expression (Hardy, Chen et al. 1999). L-type voltage-gated Ca^{2+} channel splicing variant expression was increased by chronic ethanol exposure (Walter, McMahon et al. 2000). One of the L-type voltage-gated Ca^{2+} channels, α_{1c}, has two splicing variants, $\alpha_{1c\text{-}1}$ and $\alpha_{1c\text{-}2}$. Chronic

ethanol exposure specifically increased $\alpha_{1c\text{-}1}$ expression without affecting $\alpha_{1c\text{-}2}$ (Walter, McMahon et al. 2000). Large conductance Ca^{2+} activated K^+ (BK) channel is a well-established alcohol target. Different splicing variants of the BK channel have different alcohol sensitivity. Alcohol upregulates mir-9, a miRNA, and destabilizes BK splicing variants that have a mir-9 recognition element at the 3' UTR (Pietrzykowski, Friesen et al. 2008). Therefore, chronic ethanol treatment changes specific alternative splicings resulting in increased or decreased splicing variant populations.

1.5.4 GABAB1 splicing complexity

Two previous cDNA microarray experiments showed the complexity of GABAB1 splicing. Their two probe sequences, NCBI accession numbers N80593 and W07715, were from one single probe, clone 300899. Its expression was increased in alcoholic brains (Figure 1.5A) (Flatscher-Bader, van der Brug et al. 2005; Liu, Lewohl et al. 2006). The probe was aligned to intron 4 of GABAB1a. This is a common intron region of the major human splicing variant isoforms (Figure 1.5B). Frequently, cDNA microarray probes were aligned to an intron region of specific genes and detected unknown splicing variants. The intron 4 probe could suggest the existence of other novel GABAB1 splicing variants. The increased expression detected by this probe after chronic alcoholism raises the possibility of enhanced expression of specific splicing variants of the GABAB1 gene.

The previous microarrays had three GABAB1 microarray probes including clone 300899, and they suggest complexity of GABAB1 gene splicing. Two out of three

GABAB1 probes were generated by unknown splicings. In addition to the clone 300899 from intron 4 of GABAB1, clone 2312175 had an unknown splicing out at exon 23 (Figure 1.5C). Therefore, the probes also suggested the existence of many unknown GABAB1 splicing variants and the overall complexity of GABAB1 splicing. The unknown novel alternative splicing variants could be important to understand brain changes in chronic alcoholism.

Expression changes in brain tissue specific splicing variants also demonstrate the splicing complexity of GABAB1. One cDNA microarray experiment with human alcoholic brain used both prefrontal cortex and NAc (Flatscher-Bader, van der Brug et al. 2005). They found that only 6 % of genes share common expression patterns in both brain regions, and most genes showed different expression levels between the regions. They also found increased GABAB1 expression using clone 300899 in prefrontal cortex but not NAc (Figure 1.5A). Our Illumina oligo microarray data of ILMN_2395373 and ILMN_2395375 detected the expression level of splicing variants that have the 3' end of the GABAB1 gene in four brain regions: prefrontal cortex, basal lateral amygdala (BLA), medial lateral amygdala, and CeA. The expression of ILMN_2395373 was decreased in alcoholic BLA but not in the other tissues (Figure 1.5A, C). ILMN_2395375 was aligned to the splicing out region of clone 2312175 (Figure 1.5C). However, expression of the two probes, ILMN_2395375 and clone 2312175, did not change significantly in any brain region of alcoholics though they target the same exon as ILMN_2395373 (Figure 1.5A). Therefore, in alcoholic brains BLA has less splicing variants containing specific 3' end of

GABAB1 gene than the other brain regions showing the specificity and complexity of GABAB1 splicing.

A.

	Probe	Alignment	Expression in alcoholics	Isoforms detected	Note
* Microarray	clone 300899	intron 4	↑ (Prefrontal cortex)		unknown splicing region
* Microarray	clone 2312175	exon 23	-	1a, 1b, 1c, 1d, 1e	unknown splicing out (339 bp)
* Microarray	clone 298231	exon 23	-	1a, 1b, 1c, 1d, 1e	
* Illumina	ILMN_2395373	exon 23	↓ (BLA)	1a, 1b, 1c, 1d, 1e	3' end
* Illumina	ILMN_2395375	exon 23	-	1a, 1b, 1c, 1d, 1e	aligned to the unknown splicing out region of clone 2312175
* Illumina	ILMN_1658965	intron 5	not detectable	1b	GABAB1b specific

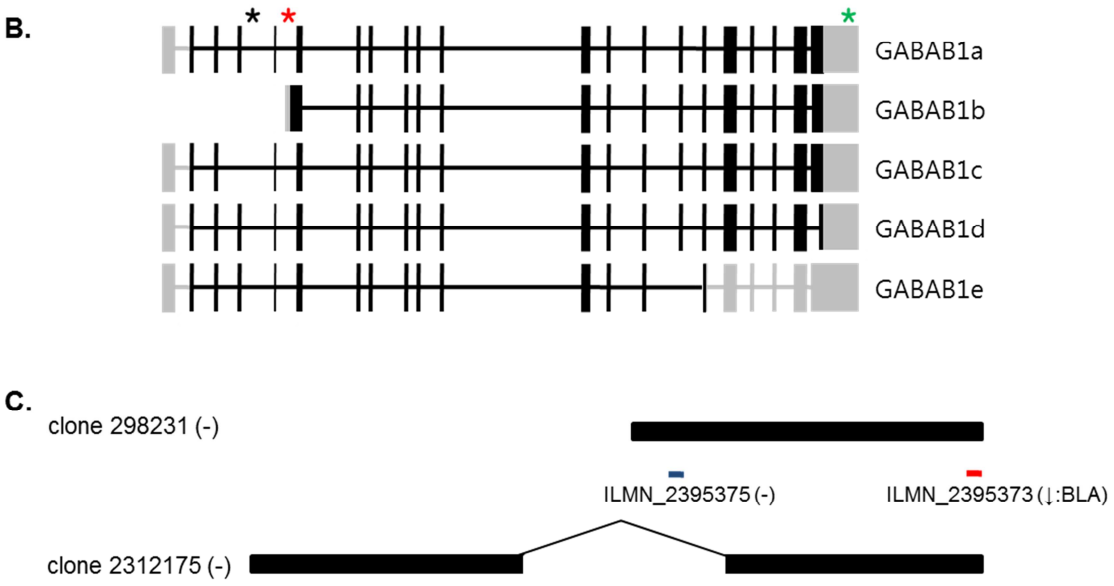


Figure 1.5 GABAB1 probes based on previous cDNA and Illumina oligo microarrays.
A. Summary of previous cDNA and Illumina oligo microarray GABAB1 probes. Among three cDNA microarray probes, clone 2312175 and clone 300899 showed unknown splicing patterns. The expression of clone 300899 was increased in prefrontal cortex only, and the expression of ILMN_2395373 was specifically decreased in BLA. B. Probes in A were marked by colored asterisks based on the locations shown in B. C. The sequence alignment model of previous cDNA microarray probes (clone 298231 and clone 2312175) and Illumina oligo microarray probes (ILMN_239373 and ILMN_2395375) at exon 23 of GABAB1a. Clone 2312175 has unknown 339 base pairs (bp) splicing out. ILMN_2395375 is aligned at the unknown splicing out region, and ILMN_239373 is aligned to 3' UTR end. (-)s represent no expression change of probes in alcoholic brains, and (↓: BLA) is for expression decrease at BLA.

1.5.5 Splicing data limitation

The NCBI database offers data sources to study GABAB1 splicing in brain. However, the database markedly underestimates splicing occurrences of the whole transcriptome (Fodor and Aldrich 2009). Splicing abundantly occurs in brain genes (Stamm, Ben-Ari et al. 2005), and more accurate splicing studies require another database construction using a high throughput sequencing technology, such as RNA-seq.

1.6 RNA-SEQ

1.6.1 RNA-seq overview

RNA-seq is a next generation sequencing technique. This high throughput sequencing technique sequences cDNA to determine RNA content. It is currently replacing microarrays for study of gene expression (Lee, Seo et al. 2011). Unlike microarrays, it has no hybridization step and contains less systematic bias (Marioni, Mason et al. 2008; Lee, Seo et al. 2011). Though microarrays detect gene expression based on specific probes, RNA-seq can measure expression independent of probes, thus detecting all transcripts with less bias. In addition, its deep coverage and base level resolution also provide efficient ways to detect splicing junctions, SNPs, and gene fusions (Maher, Kumar-Sinha et al. 2009; Trapnell, Pachter et al. 2009; Canovas, Rincon et al. 2010; Shen, Catchen et al. 2011).

1.6.2 RNA-seq methods

1.6.2.1 RNA-seq sequencers

For RNA-seq, there are three major sequencers. They are Roche 454 genome sequencer, Illumina (Solexa) genome analyzer, and Sequencing by oligonucleotide ligation and detection (SOLiD) sequencer.

Roche 454 genome sequencer was the first sequencer and was developed by 454 Life Sciences in 2004 (Shendure and Ji 2008). The library can be constructed with cDNA fragments flanked by two adaptors. For clonal amplification, the fragments are amplified by emersion PCR (ePCR), and each amplicon is captured by a single primer coated 28 µm bead (Figure 1.6A). The sequencer has many picoliter volume wells each containing the bead and sequencing enzyme. The sequencing reaction uses luciferase and generates light to detect individual nucleotides. In 2008, the Roche 454 genome sequencer generated 400 - 600 million reads per run of about 400 - 500 bp read length (Table 1.1). This read length is the longest among the three sequencers. However, there are fewer Roche 454 genome sequencer users than Illumina genome analyzer or SOLiD sequencer due to higher cost per megabyte (Mb) (Table 1.1).

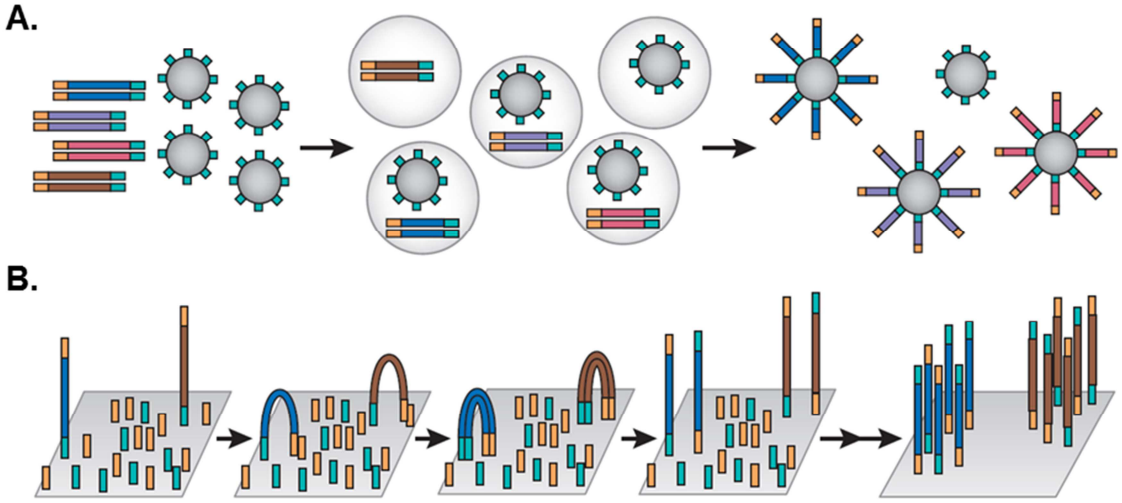


Figure 1.6 Clonal amplifications of three major sequencers.
A. ePCR for Roche 454 genome sequencer and SOLiD sequencers. B. Bridge amplification for Illumina genome analyzer. (from Shendure and Ji 2008)

	Roche 454 genome sequencer	Illumina genome analyzer	SOLiD sequencer
Amplification approach	Emulsion PCR	Bridge amplification	Emulsion PCR
Paired end separation	3 kb	200 bp	3 kb
Mb per run	100 Mb	300 Gb	3000 Mb
Time per paired end run	7 hours	7-14 days	5 days
Read length (update)	250 bp (400 bp)	100 bp (50-100 bp)	35 bp (35-50 bp)
Cost per run	$ 8,438 USD	$ 11,750 USD	$ 17,447 USD
Cost per Mb	$ 84.39 USD	$ 1.00 USD	$ 5.81 USD

Table 1.1 Comparison of RNA-seq among three sequencers.
(modified from Mardis 2008)

Illumina genome analyzer was developed by Solexa in 2006, and it is currently the most popular sequencer for RNA-seq. To form clonal colonies, each adaptor ligated cDNA molecule is attached to a primer on a slide (Shendure and Ji 2008). It is amplified by the bridge amplification method (Figure 1.6B). During amplification, unlabeled nucleotides and enzyme were used. After denature the amplified molecules, first chemistry cycle determine first base using four labeled reversible terminators. After a camera takes fluorescence images from the labeled nucleotides, the reaction is repeated to determine the sequence of bases (Mardis 2008). It generates about 200 million reads of 75 - 100 bp (McPherson 2009).

SOLiD sequencer was developed by Life Technologies in 2007. Like the Roche 454 genome sequencer, ePCR is used for clonal amplification (Figure 1.6A). For its ePCR, 1 µm paramagnetic beads capture amplicons (Shendure and Ji 2008). Each bead contains the same cDNA molecule and is deposited onto a glass slide. The SOLiD sequencer employs sequencing by oligonucleotide ligation. Preferential ligation for its matching sequences generates a nucleotide signal at each position. Unlike the other two sequencers, it generates color signals. As the color signals encode dinucleotides, each base is encoded twice. As shown in the left hand table of Figure 1.7A, each color has 4 possible combinations of the first and the second bases. For example, the combinations of red color are AT, CG, GC, and TA. If the first sequence is A, color signals (red, green, blue, and yellow) encode the red circled dinucleotide. They indicate that the real nucleotide sequence is ATGGA (Figure 1.7B). The SOLiD sequencer can produce 400 million reads of 50 bp length (McPherson 2009).

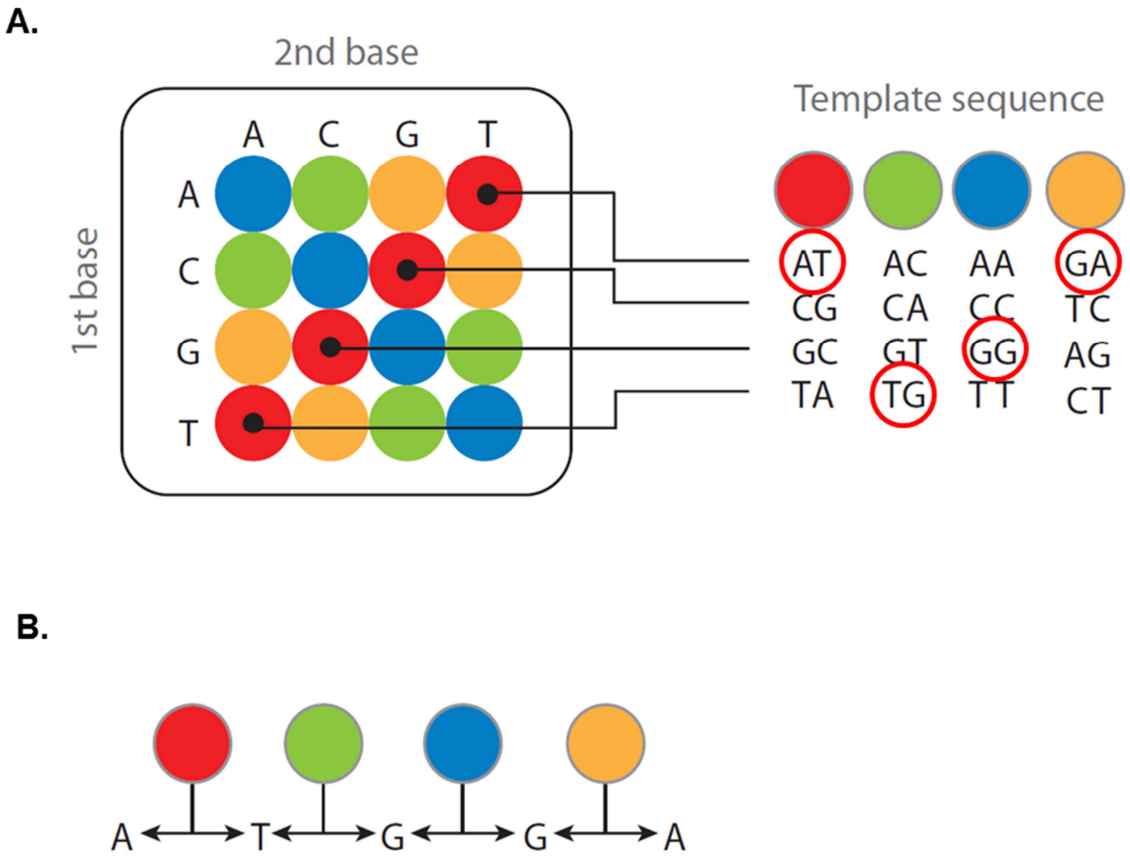


Figure 1.7 Color signals encoding at SOLiD sequencer.
A. Four possible dinucleotides were encoded by each color. If a sequence starts with A, red, green, blue, and yellow color signals encode the red circled dinucleotides. B. The real sequence is ATGGA. (Modified from Mardis 2008)

1.6.2.2 RNA-seq library construction

RNA-seq library construction methods vary for each sequencer. Most methods were generated using random primers for reverse transcription (RT) (Cloonan, Forrest et al. 2008; Hansen, Brenner et al. 2010). Some customized methods were also designed based on research interests. For the RT step, libraries were primed using oligo(dT) primers or target specific primers (Li, Levanon et al. 2009; Yoon and Brem 2010).

The RNA-seq libraries constructed using randomly primed RT can measure whole transcriptome expression. For the Illumina genome analyzer, mRNAs are extracted from total RNA using poly(T) magnetic beads (Morin, Bainbridge et al. 2008). After fragmentation of the mRNAs, cDNAs are synthesized by randomly primed RT. Sequencing is performed after adaptor ligation and amplification (Figure 1.8A). Though this method is the most popular method for RNA-seq library construction, it can lose non-coding RNA information from whole transcriptome expression data. For RNA-seq library construction using the SOLiD sequencer, ribosomal RNA (rRNA) depleted RNAs are prepared and fragmented. After adaptor ligation, cDNAs are generated by randomly primed RT. After size selection, the cDNAs are amplified and sequenced (Figure 1.8B) (Cui, Lin et al. 2010). This library construction can provide both mRNA and non-coding RNA expression data.

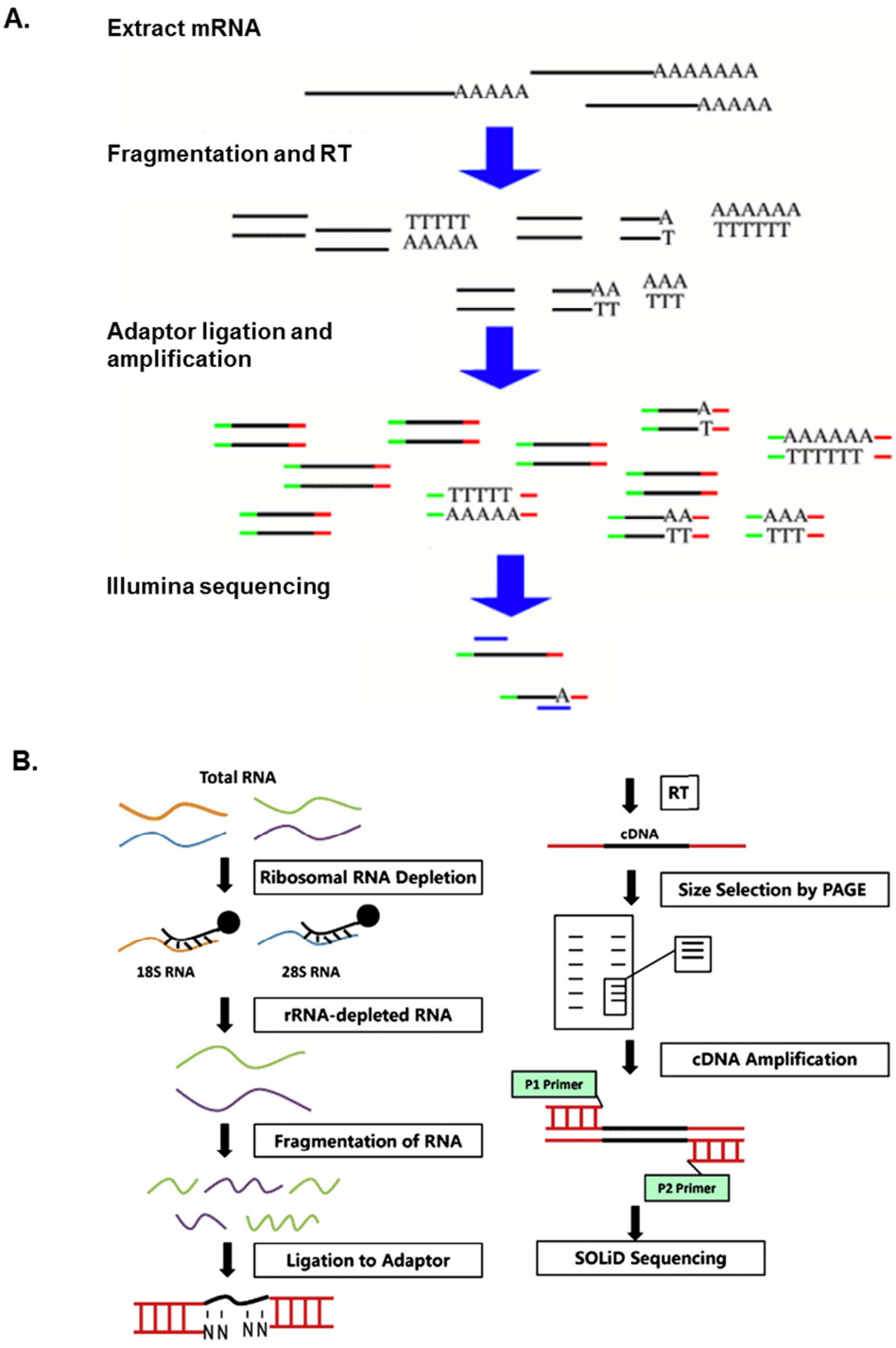


Figure 1.8 RNA-seq library constructions by randomly primed RT.
A. RNA-seq library preparation for Illumina genome analyzer. B. RNA-seq construction for SOLiD sequencer. (Modified from Cui, Lin et al. 2010)

RNA-seq libraries are also constructed to study 3’ and 5’ end sequences of whole transcriptome (Nagalakshmi, Wang et al. 2008; Yoon and Brem 2010). For library construction to generate 3’ end rich sequence data, polyA containing RNAs are selected after total RNA fragmentation. cDNAs are synthesized by oligo(dT) primers. After adaptor ligation and amplification, the library is sequenced (Figure 1.9A). To study 5’ end information of transcriptome, RNA-seq library is designed based on 5’ rapid amplification of cDNA ends (RACE) (Nagalakshmi, Wang et al. 2008).

Target specific primers generate RNA-seq library for specific target transcript regions. For RNA editing study, padlock probes capture target cDNAs, and polymerization and ligation produce a circular cDNA library containing the target cDNAs. After amplification, the specific target region is sequenced (Figure 1.9B) (Li, Levanon et al. 2009). After comparing the sequencing data with genomic DNA sequencing data, RNA editing is identified.

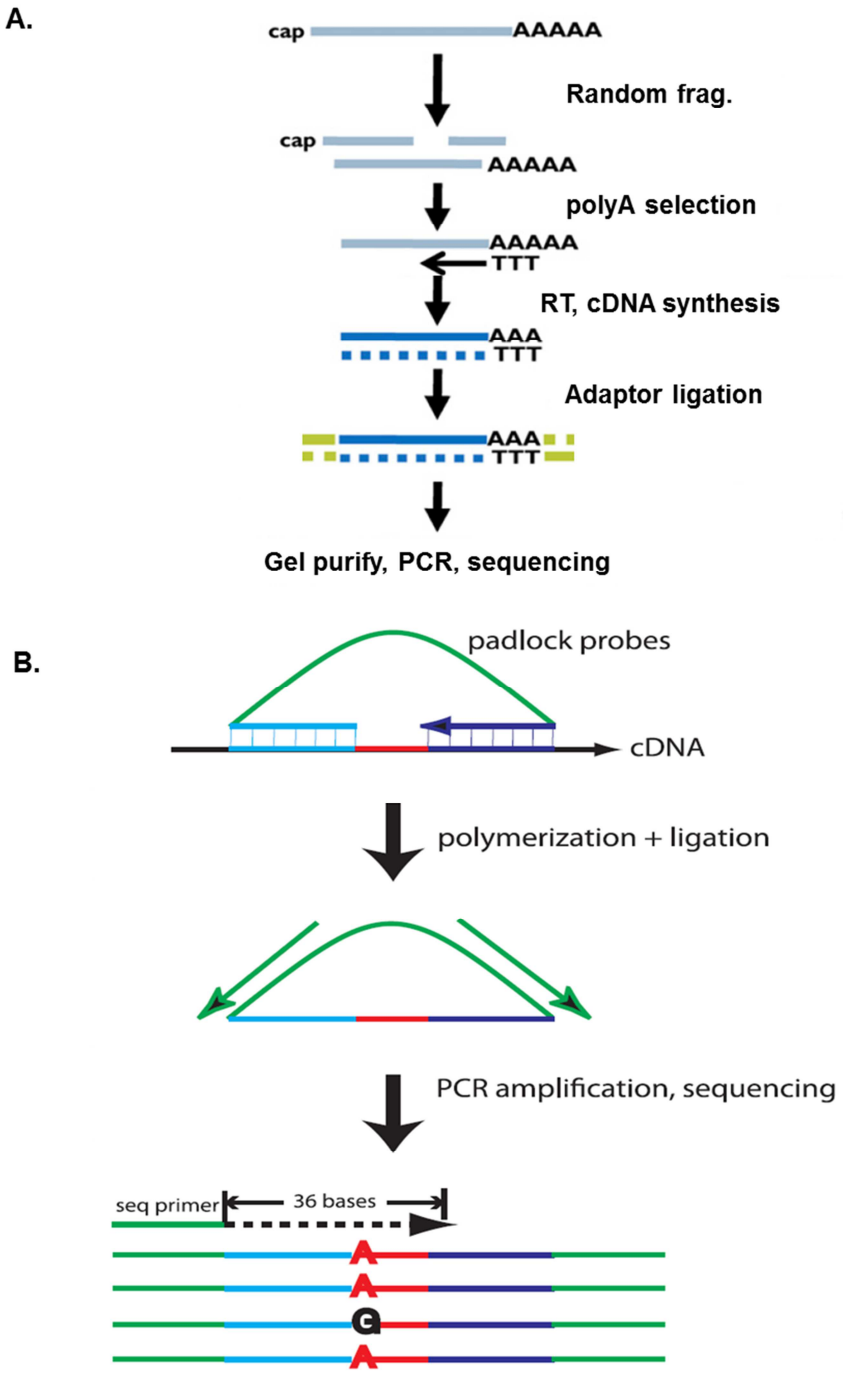


Figure 1.9 RNA-seq library constructions using oligo(dT) and target specific primers. A. RNA-seq library preparation by oligo(dT) primed RT. (Modified from Yoon and Brem 2010) B. RNA-seq library construction using target specific padlock probes for RNA editing. (Modified from Li, Levanon et al. 2009)

1.6.3 RNA-seq biases

Though RNA-Seq has provided genome wide gene expression data, RNA-seq libraries can generate biases during library construction. The biases can distort expression data and prevent detections of SNPs and splicing junctions. RNA fragmentation, RT, and PCR amplification are known bias sources during RNA-seq library construction. For example, biased mapped read distributions are generated by fragmentations and RT primers. Randomly primed RT and PCR amplification result in sequence biases (Romaniuk, McLaughlin et al. 1982; Wang, Gerstein et al. 2009; Hansen, Brenner et al. 2010; Metzker 2010; Willerth, Pedro et al. 2010).

A biased mapped read distribution is generated from RNA-seq library construction using RNA fragmentation. This method uses randomly primed RT from fragmented RNAs (Figure 1.10) (Wang, Gerstein et al. 2009). After RNA fragmentation, about 200 bp RNA fragments are selected for library construction. However, 5' and 3' end containing fragments are mostly shorter than 200 bp and have less chance to be involved in the RNA-seq library construction. Therefore, fewer reads are mapped at the 5' and 3' ends. RNA-seq library construction using cDNA fragmentation method also causes biased distribution (Figure 1.10). This method fragments the oligo(dT) primed cDNAs using sonication, and generates greater bias distribution compared to the RNA fragmentation method (Figure 1.10). It has extremely skewed mapped read distribution toward the 3' end of transcripts and very few mapped reads at the 5' end. oligo(dT) primed RT usually cannot reach 5' end of transcripts for long transcripts. cDNAs for 5'

end transcripts are less cDNAs for 3’ end. This skewed cDNA distribution becomes even severe after amplification.

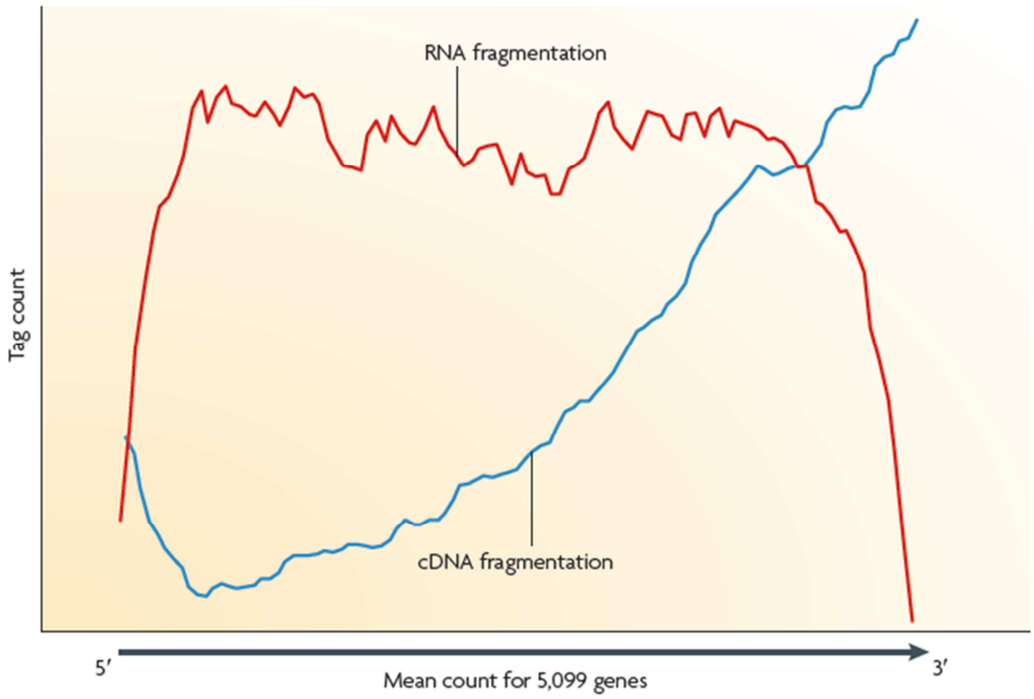


Figure 1.10 Mapped read distribution biases of RNA-seq libraries.
The red line (RNA fragmentation) is for a RNA-seq library prepared using randomly primed RT after RNA fragmentation. The blue line (cDNA fragmentation) represents a RNA-seq library prepared using oligo(dT) primed RT. Its oligo(dT) primed cDNAs were fragmented for RNA-seq library construction. Tag count is the average sequencing coverage for 5,099 yeast gene ORFs. (from Wang, Gerstein et al. 2009)

Randomly primed RT also causes a sequence bias during RNA-seq library construction (Hansen, Brenner et al. 2010). All six RNA-seq library preparation methods for the Illumina genome analyzer introduced the same sequence bias (Figure 1.11A). The sequence bias was not related to RNA fragmentation. The random hexamers binding location is the starting 6 nucleotide positions of sequences data. From the starting point, about 10 nucleotide positions have severe sequence bias.

PCR amplification has also been shown to generate sequence bias during RNA-seq library construction (Willerth, Pedro et al. 2010; Aird, Ross et al. 2011). PCR amplified cDNA using target specific primers generates reads containing significant homology with the primer. During library construction for the Illumina genome analyzer, PCR amplification depleted reads that have over 65 % GC ratio and diminished about 90 % of reads that have less than 12 % GC ratio (Figure 1.11B) (Aird, Ross et al. 2011).

To overcome the bias issues during RNA-seq library construction, new library preparation methods have been developed based on molecular biology techniques (Willerth, Pedro et al. 2010; Aird, Ross et al. 2011). Computational analyses also have provided various solutions for the biases (Degner, Marioni et al. 2009; Gao, Fang et al. 2011).

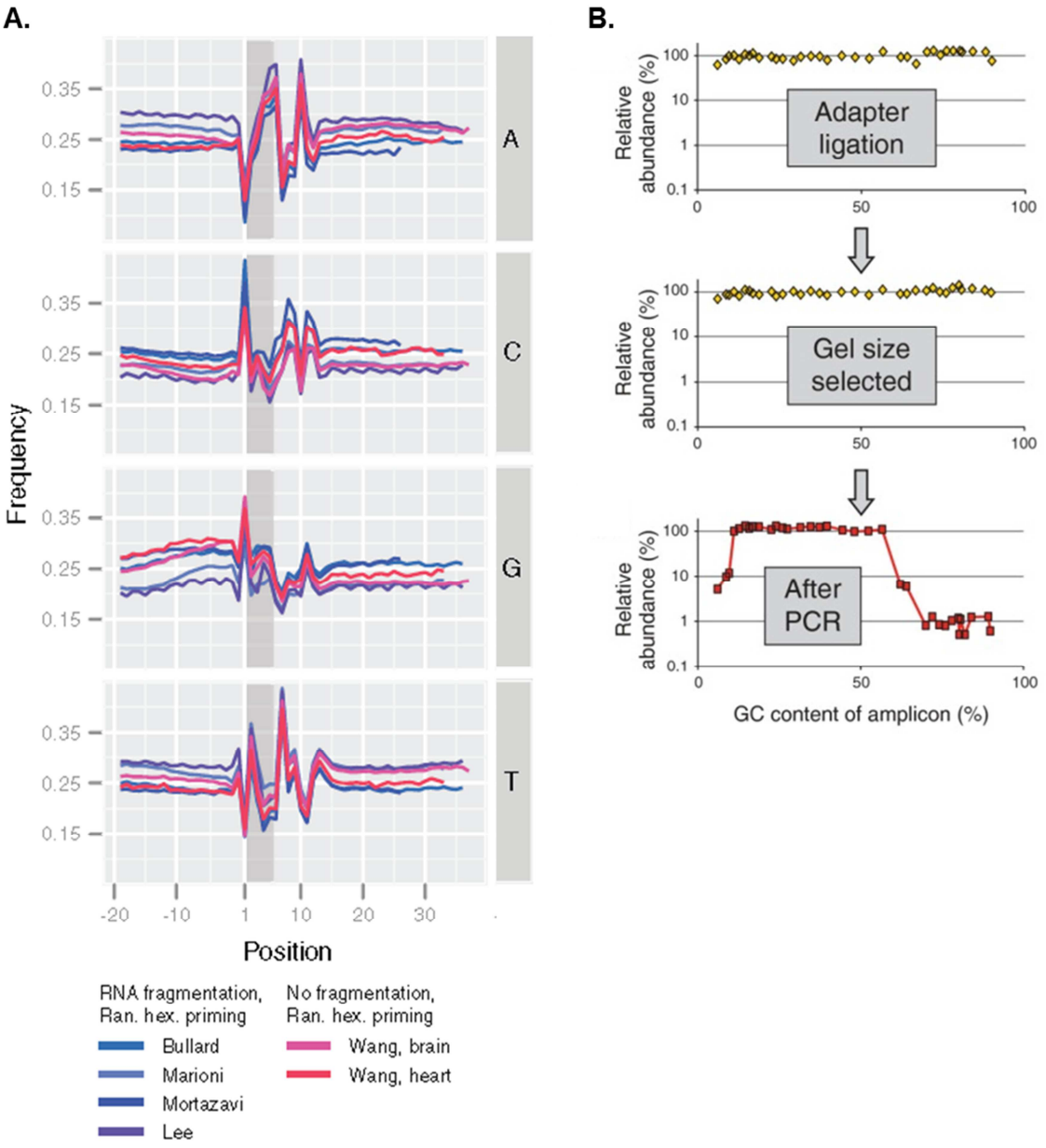


Figure 1.11 Sequence biases.
A. Sequence bias for RNA-seq library using randomly primed RT. Mapped reads were extended upstream and downstream of the first position of the actual read. The first position of the read is 1 at X-axis. Grey section (1 - 6 of X-axis) is the random primer binding location for RT, and six different libraries were compared. (modified from Hansen, Brenner et al. 2010) B. PCR amplification causes GC bias. (modified from Aird, Ross et al. 2011)

1.7 SPECIFIC AIMS

The overall goal of this project is to study the splicing difference of human GABAB1 in non-alcoholic and alcoholic brains. The hypotheses are: 1) there are GABAB1 splicing variants yet to be identified; 2) chronic alcohol abuse alters GABAB1 splicings; 3) less sequence biased RNA-seq libraries improve RNA-seq data quality for the splicing study. These experiments will define GABAB1 splicing in alcoholics and may lead to a better understanding of chronic alcohol action on the GABAB1 receptor. This plan includes three specific aims to be tested:

Aim 1: Intron 4 containing novel GABAB1 isoforms impair GABAB receptor function

Based on the analysis of previous microarray data, the existence of novel GABAB1 isoforms was predicted. After cloning of four new isoforms, GABAB1k, l, m, and n, their expression levels were evaluated. Their expression levels were much less than other known isoforms and similar in prefrontal cortex, frontal cortex, motor cortex, and total brain. Inhibitory actions of GABAB1l and m were identified using two-electrode voltage clamp in *Xenopus laevis* oocytes expression system.

Aim 2: Alcoholism correlates with GABAB1 splicing in human brain

To better define GABAB1 splicing in human brains, gene specific RNA-seq study was performed. It found that the GABAB1 gene is at least 2 to 3 times longer than the known GABAB1 gene and indicated that there are more short splicing variants than just GABAB1a, the longest known major transcript containing all known exons. Chronic alcohol altered exon/intron and splicing junction expression and decreased the population

of normal GABAB1 transcripts. This suggested the therapeutic benefit of GABAB receptor agonists used in treating alcoholism.

Aim 3: Use new RNA-seq library construction to solve RNaseIII and T4PNK sequence biases

During the RNA-seq study, we found that unknown RNaseIII sequence biases that introduced duplicated mapped read pile-ups and gaps in known exons. To remove the biases, RNaseIII fragmentation was replaced with heat fragmentation using T4PNK and OptiKinase. The heat fragmentation removed the RNaseIII sequence biases and significantly reduced the pile-ups and gaps. OptiKinase minimized unexpected T4PNK sequence biases and removed all pile-ups and gaps. These also found unexpected sequence biases of RNaseIII and T4PNK.

CHAPTER 2. INTRON 4 CONTAINING NOVEL GABAB1 ISOFORMS IMPAIR GABAB RECEPTOR FUNCTION

2.1 INTRODUCTION

The GABAB receptor is a metabotropic receptor that is highly expressed in brain and weakly expressed in heart, small intestine, uterus and other tissues (Calver, Medhurst et al. 2000). The functional receptor is a hetero-oligomer of GABAB1 and GABAB2 subunits, where the intracellular domain of GABAB1 dimerizes with GABAB2. GABA binds to the extracellular domain of GABAB1 and transfers signals through G proteins. G protein α subunits are linked with AC, and G protein $\beta\gamma$ subunits alter Ca^{2+} channels and G protein-activated inwardly rectifying K+ channel (GIRK) channels. The GABAB receptor decreases the activity of AC and decreases neurotransmitter release by inhibiting Ca^{2+} influx through presynaptic Ca^{2+} channels. At postsynaptic neurons, it activates K^+ channels, and an outward K^+ current induces hyperpolarization preventing Na^+ channel opening and action potential firing (Couve, Moss et al. 2000). Thus, the GABAB receptor mediates an inhibitory neurotransmission.

GABAB1 has two major splicing variants, GABAB1a and GABAB1b, in the human CNS. GABAB1a is the longest form and has 23 exons (Figure 2.1A). GABAB1b is an N-terminal truncated form of GABAB1a and has an alternative exon from intron 5 resulting in a unique N-terminal without sushi domains (Figure 2.1A, B). GABAB1a and

b are the most widely studied isoforms to date and show some distinct expression patterns in brain as GABAB1a is more presynaptic in localization, and GABAB1b is more postsynaptic (Billinton, Ige et al. 2001; Bettler, Kaupmann et al. 2004; Gassmann, Haller et al. 2005). However, the differences in their localization and function are still not known precisely (Billinton, Ige et al. 2001). Additional splice variants of GABAB1 have been reported although some were found only in one species, and most do not have defined functions (Bettler, Kaupmann et al. 2004; Holter, Davies et al. 2005; Tiao, Bradaia et al. 2008). Human GABAB1c lacks exon 4 (Figure 2.1A), and its functions have not been studied in detail (Billinton, Ige et al. 2001). Human GABAB1e is mainly expressed in peripheral tissues. Because it lacks exon15, GABAB1e is a C-terminal truncated form of GABAB1a and does not have seven TMs, a G-protein coupling region, nor a C-terminal intracellular region (Figure 2.1A, B). However, it is functionally active and prevents GABAB1a and GABAB2 heterodimerization (Schwarz, Barry et al. 2000). In rat, the GABAB1j isoform was recently discovered and found to be expressed as highly as GABAB1a and b in brain. It contains exon 1 through 4 and the 5' part of intron 4 and is thus a small C-terminal truncated GABAB1 isoform (Figure 2.1A). The GABAB1j is a soluble isoform with sushi domains (Figure 2.1B), which selectively impair the function of presynaptic (but not postsynaptic) GABAB receptors (Tiao, Bradaia et al. 2008). Thus, alternative splicing provides a large diversity of structural and functional variation in GABAB1 receptors.

Figure 2.1 Schematic summary of GABAB1 isoforms.

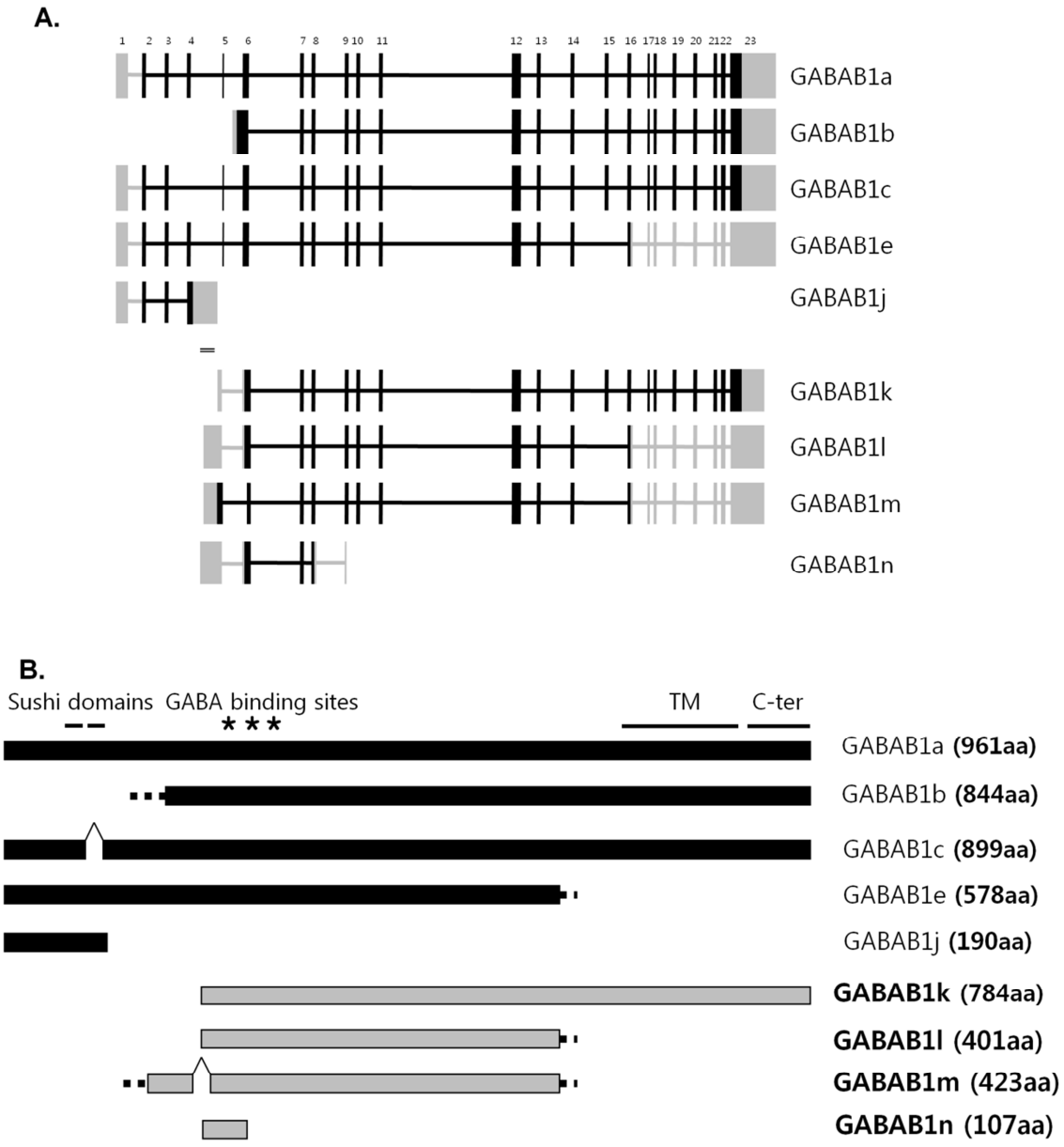


Figure 2.1 Schematic summary of GABAB1 isoforms.
A. Introns and exons of known and novel GABAB1 isoforms were compared. The numbers listed at the top are exon numbers of GABAB1a. Blocks represent exons, and lines are introns of the GABAB1 gene. Black lines and blocks are ORFs, and grey lines and blocks are UTRs. Double line is cDNA microarray probe, clone 300899, which aligns to the intron 4 of GABAB1a. Human GABAB1a, b, c, e, and rat GABAB1j are previously identified isoforms. After cloning, the following novel isoforms were found in human: GABAB1k, GABAB1l, GABAB1m, and GABAB1n. Human GABAB1j was predicted. Mouse GABAB1j and GABAB1k were also cloned. GABAB1k contains the 3' part of intron 4 through exon 23. GABAB1l has additional splicing out of exon 15, and GABAB1m has similar splicing patterns as GABAB1l except for additional splicing out at 5' part of exon 6. GABAB1n has an SNP on exon 8. B. ORFs of GABAB1 isoforms are shown with important functional domains. Black bars are previously known isoforms, and grey bars represent novel isoforms. Dotted lines stand for unique sequence of each isoform. GABAB1k contains GABA binding sites, seven TMs, a G-protein coupling region, and a C-terminal intracellular region. GABAB1l and GABAB1m do not have TMs, a G-protein coupling region, or C-terminal intracellular regions (same as GABAB1e). GABAB1m has a longer N-terminal than GABAB1l. GABAB1n contains only a partial GABA binding site.

The diversity of GABAB1 receptors is indicated by these previous studies, and it is possible that additional unknown isoforms exist. Two previous cDNA microarray studies found strong GABAB1 signals in human brain mRNA (Flatscher-Bader, van der Brug et al. 2005; Liu, Lewohl et al. 2006). The two probe sequences used in these studies, NCBI accession numbers N80593 and W07715, were obtained from one single probe with clone ID 300899. The probe aligns to intron 4 of GABAB1a. This is a common intron region of the major human GABAB1 splice variants (Figure 2.1A) and has been noted as a potentially important location for alternative splicing in rat (Holter, Davies et al. 2005). Because the probe was made from a cDNA library, the probe region should be an alternative exon in the GABAB1 gene suggesting the existence of novel GABAB1 isoforms. The goals of this study were to identify the novel isoforms and determine their functions.

2.2 RESULTS

2.2.1 Microarray probe sequence analysis

The expression of clone 300899 increased significantly in human prefrontal cortices in two independent microarray studies (Flatscher-Bader, van der Brug et al. 2005; Liu, Lewohl et al. 2006), and the clone aligned to the intron 4 region (Figure 2.1A). Sequencing of clone 300899 showed multiple stop codons from all six translated frames and likely represented UTRs. Thus, two possible isoform models, GABAB1j and GABAB1k, were proposed for human allowing us to study novel GABAB1 splicing

variant isoforms. Human GABAB1j and GABAB1k contain clone 300899 at their 3' and 5' UTR, respectively. In human brain, GABAB1k was expected to have intron 4 and exon 5 through 23, and GABAB1j was also proposed to contain exon 1 through 4 and intron 4 like rat GABAB1j.

2.2.2 GABAB1j: Cloning and mRNA levels in human and mouse brain

GABAB1j was previously cloned in rat, but not in human or mouse. Rat GABAB1j ORF is 229 amino acids (aa) (Figure 2.2A) (Tiao, Bradaia et al. 2008). Though the predicted human GABAB1j has some different aa sequences and a shorter ORF of 190 aa, it shares similar motifs with rat GABAB1j such as C-terminal truncated splicing and lack of important functional domains of common GABAB1 isoforms. Although human GABAB1a has 23 exons, it has only the first 4 exons and partial 5' intron 4 (Figure 2.1A). Two novel clones were found from the mouse cDNA microarray probe set after comparing sequences with the possible human GABAB1j (Ponomarev, Maiya et al. 2006). The lengths of the two clones are different because their UTR lengths are not the same, but they shared identical ORF. They have the same exon composition and ORF length as rat GABAB1j. Sequencing alignment also showed that mouse GABAB1j shares the similar ORF except a few aa differences (Figure 2.2A).

The intron 4 sequence of the predicted human GABAB1j is different from mouse and rat GABAB1j. Therefore, sequence alignment showed that it has a very unique C-terminal and a 39 aa shorter ORF than mouse and rat GABAB1j (Figure 2.2A). To do quantitative real-time PCRs, primer and probe sets for general GABAB1 isoform (GB1)

were designed between the last two exons of GABAB1a, exon 22 and 23, and they were specific for most known GABAB1 isoforms except GABAB1j. GABAB1j specific primer and probe sets were also prepared. Using quantitative real-time PCR, the existence of the predicted human GABAB1j was verified. Relative mouse GABAB1j mRNA expression level was much higher than human GABAB1j when compared to GB1 expression (Figure 2.2B). GB1 sets and the northern blot probe of the rat GABAB1j paper (Tiao, Bradaia et al. 2008) were supposed to detect the same major known isoforms. Though rat GABAB1j expression level in brain were as high as GB1, human and mouse GABAB1j expression were less than GB1 expression (Figure 2.2B).

GABAB1j function in human and rodent might not be identical because of their expression and sequence differences. Rat GABAB1j expression is as strong as major GABAB1 isoforms (Tiao, Bradaia et al. 2008). Though mouse GABAB1j expression is only 5.5 fold lower than them (Figure 2.2B), human GABAB1j showed much lower expression (136 fold) (Figure 2.2C). If the fold differences were compared, the relative human GABAB1j expression level to GB1 was also dramatically less (25 times) than mouse GABAB1j expression though identical detection regions were used for both expression studies. Thus, GABAB1j expressions are not consistent across species. Interestingly, the C-terminal sequence of human GABAB1j did not share a common pattern with mouse and rat GABAB1j (Figure 2.2A). From previous research, purified sushi domains of rat GABAB1j impaired the inhibitory effect of the GABAB receptor, but the function of rat GABAB1j full ORF had not been studied in detail (Tiao, Bradaia

et al. 2008). So, it is not clear whether GABAB1j functions could be conserved across species.

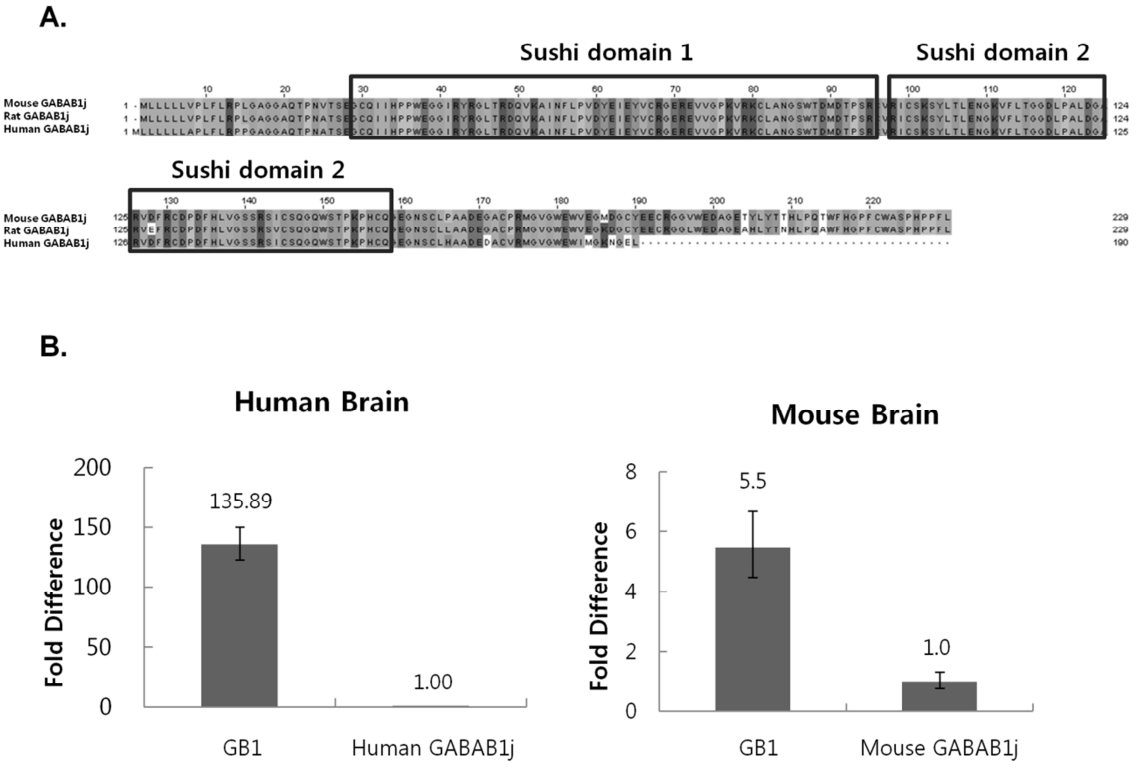


Figure 2.2 Sequences for human, mouse, and rat GABAB1j isoforms and relative expression levels.

A. Sequence alignment of human, mouse, and rat GABAB1j showed a different C-terminal pattern in human GABAB1j. B. Human and mouse GABAB1j mRNA expression levels were measured with quantitative real-time PCR. GB1 primer and probe sets were used to detect most known major GABAB1 isoforms except GABAB1j. Human and mouse GABAB1j primer and probe sets were specific for only GABAB1j. GB1 expressions were shown as fold differences from human and mouse GABAB1j. Although a previous rat GABAB1j study indicated that GB1 and GABAB1j expression levels were similar, human and mouse GABAB1j expressions were lower than GB1. Mouse GABAB1j expression was 5.5 fold lower than GB1 expression, and human GABAB1j showed 135.89 fold lower expression. If the fold differences were compared, human GABAB1j expression level was much less (25 times) than mouse GABAB1j expression. Therefore, GABAB1j expressions vary across species.

2.2.3 GABAB1k sequence analysis in human, mouse, and rat

The predicted human GABAB1k is an N-terminal truncated splicing variant. It contains 3' part of intron 4 and exons 5 through 23, and the ORF was 784 aa (Figure 2.1A). It had important functional domains, such as GABA binding sites (Ser70, Tyr190, and Asp295), seven TMs (415 - 677 aa), a G-protein coupling region (linker region between TM 3 and 4), and a C-terminal intracellular region (678 - 784 aa: dimerize with GABAB2) (Figure 2.1B). It shared most domains with GABAB1b, a functional receptor subunit that had no the sushi domains (Steiger, Bandyopadhyay et al. 2004) (Figure 2.1B). Thus, GABAB1k might also be a functional receptor subunit with greater physiological effect than human GABAB1j, given that GABAB1j only contains the sushi domains (without any important domains of GABAB1b) and exhibits much lower relative expression in human than in mouse and rat. Therefore, we decided to focus on the longer splice variant GABAB1k instead of the very short splice variant GABAB1j.

Before cloning the predicted GABAB1k, its existence was verified in human, mouse, and rat. First, mRNAs containing microarray probe region, clone 300899, were detected across species. After purifying total RNA from human cultured cell lines, mouse midbrain, and rat hippocampus, DNase treatment removed possible genomic DNA contamination. After making mRNA specific cDNA with oligo(dT) primer, PCRs were performed within the probe regions. Three different clones (#2, #3, and #4) were isolated from human (Figure 2.3). After probe homologous regions in mouse and rat were found with Blast N Search, #11 and #18 were cloned in mouse and rat, respectively (Figure 2.3). The previous microarray probe sequence analysis proved that the microarray probe

was a part of intron 4 and either 3' or 5' UTR of novel isoforms. Because there was no known human GABAB1 isoform that had the intron 4 region as an exon, the two possible clones, GABAB1j and GABAB1k, were predicted to contain the intron 4 regions as 3' or 5' UTR, respectively. GABAB1j and GABAB1k were also only possible isoform models that contain the microarray probe in intron 4. Thus, #2, #3, #4, #11, and #18 can be partial GABAB1j or GABAB1k, and mRNAs containing the probe regions exist in human, mouse, and rat.

To confirm the predicted GABAB1k existence and to remove the possibility of genomic DNA contamination, PCRs were performed between the probe region and other exons. Partial GABAB1k forms containing exon-intron junctions were also cloned from human cultured cell lines, mouse midbrain, and rat hippocampus. There were six different N-terminal partial clones, #5-1, #5-2, #6-1, #6-2, #12, and 17&19 (Figure 2.3). They share the partial ORF sequences of other known isoforms in the N-terminal and also have common 5' UTRs. There is no possibility of genomic DNA contamination because introns are spliced out. Thus, the N-terminal partial clones suggested that GABAB1k exists. They are not long enough to have other functional domains, such as TMs, a G-protein coupling region, and a C-terminal intracellular region. It was, thus, necessary to obtain the full ORF for functional study of GABAB1k.

Figure 2.3 Strategy for cloning GABAB1k.

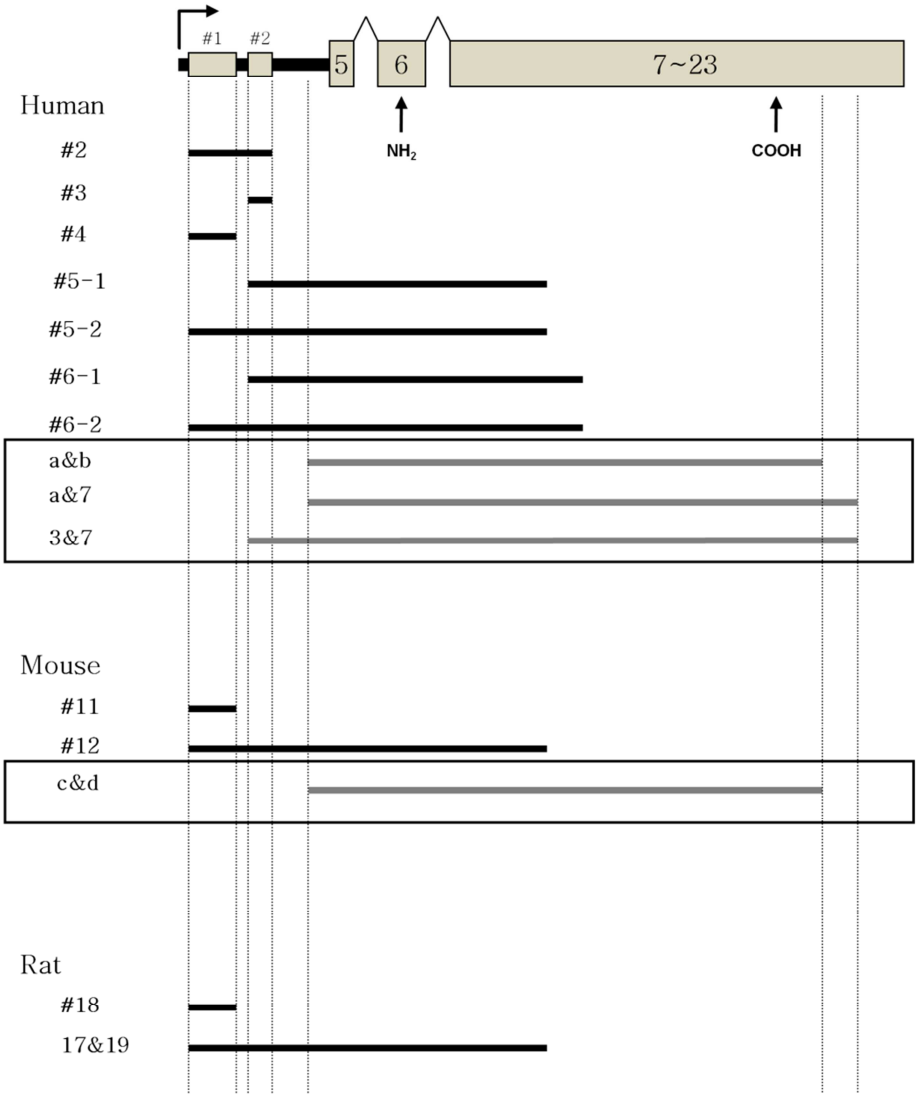


Figure 2.3 Strategy for cloning GABAB1k.
The top figure represents exon-intron composition of the proposed GABAB1k. Numbered rectangular boxes represent known exons, and lines between boxes are introns. Two arrows on the bottom of the boxes represent ORF. Black bar is intron 4, and the two probe sequence data, NCBI accession numbers W07715 and N80593, are indicated as #1 and #2 above small rectangular boxes, respectively. They are 5' and 3' sequences of the microarray probe, clone 300899. A arrow above the black bar is a possible transcription initiation site. Horizontal lines indicate sequences detected after PCR. Based on microarray probe, clone 300899, the following clones were detected in human cultured cells, mouse midbrain, and rat hippocampus: #2, #3, #4, #11, and #18. From the previous microarray probe sequence analysis, the microarray probe aligned to intron 4 and was either 3' or 5' UTR of novel isoforms. Because there was no known human GABAB1 isoform that had intron 4 as an exon, the two possible isoform models, GABAB1j and GABAB1k, were proposed. The predicted GABAB1j and GAGAG1k contain intron 4 as 3' or 5' UTR, respectively. Because GABAB1j and GABAB1k are only two possible isoform models that contain clone 300899, #2, #3, #4, #11, and #18 can be partial GABAB1j or GABAB1k and show their existences. From the same RNA sources six different N-terminal partial clones, #5-1, #5-2, #6-1, #6-2, #12, and 17&19, indicated the existence of GABAB1k. They share common 5' UTRs and partial N-terminal ORFs of GABAB1k and suggest that GABAB1k exists in human, mouse, and rat. To clone the full ORF of GABAB1k, additional primers were designed at its possible 3' UTR region based on other known isoform sequence analyses. Full ORF containing clones were cloned from human brain and mouse midbrain. Bottom two boxes show clones which contain the full ORF (a&b, a&7, 3&7 in human and c&d in mouse).

2.2.4 Cloning GABAB1k, GABAB1l, GABAB1m, and GABAB1n

To clone the full ORF, additional primers were designed at the possible 3' UTR of the predicted GABAB1k based on known isoform sequence analysis. From human whole brain and mouse midbrain, several clones were detected by PCR. They were a&b, a&7, 3&7, and c&d (Figure 2.3). The expected human GABAB1k and mouse GABAB1k were cloned. They have the 3' part of intron 4 through exon 23 (Figure 2.1A). They are an N-terminal truncated splicing variant isoforms, and their ORFs are 784 aa. They contain functional domains, such as GABA binding sites, seven TMs, a G-protein coupling region, and a C-terminal intracellular region (Figure 2.1B).

In addition to GABAB1k, two other forms, GABAB1l and m, were also cloned in human. GABAB1l has additional splicing out of exon 15 and is both N-terminal and C-terminal truncated (Figure 2.1A, B). The C-terminal truncated pattern is the same as GABAB1e (Schwarz, Barry et al. 2000), and its ORF is 401 aa. Though it contains GABA binding sites, it does not have the other functional domains (Figure 2.1B). GABAB1m has almost the same splicing pattern as GABAB1l except for an additional missing at 5' part of exon 6 (Figure 2.1A). Thus, its ORF is longer (423 aa) than human GABAB1l and starts at intron 4 whereas the ORFs of GABAB1k begin at exon 6 (Figure 2.1B).

During cloning, one more isoform, GABAB1n, was also found. It has a SNP at exon 8 and so has a very small ORF, 107 aa. It does not have any important functional domains except a partial GABA binding site (Ser70) (Figure 2.1A, B).

2.2.5 GABAB1k, GABAB1l, GABAB1m, and GABAB1n mRNA levels in human and mouse brains

After cloning the novel GABAB1 forms, GABAB1k, GABAB1l, GABAB1m, and GABAB1n, their mRNA expression levels were compared with other known isoforms in brain.

Previous human cDNA microarrays had three GABAB1 probes, clones 2312175, 298231, and 300899 (Flatscher-Bader, van der Brug et al. 2005; Liu, Lewohl et al. 2006). Clones 2312175 and 298231 aligned to exon 23, and clone 300899 aligned to intron 4. Our mouse cDNA microarray probe set had six different GABAB1 probes (Ponomarev, Maiya et al. 2006). Three probes aligned to 3' exons of the GABAB1 gene including exon 23, and the other three probes contained intron 4. Therefore, the human and mouse probe sets are two types that include exon 23 or intron 4 of GABAB1 gene. Based on the probe sequence information, GB1 primer and probe sets of the previous GABAB1j expression measurement experiments were used for quantitative real-time PCR. GB2 and GB3 primer and probe sets were designed (Figure 2.4A). GB1 sets targeting exon 22 and 23 junction represented human and mouse cDNA microarray probes that aligned to 3' exons of GABAB1 gene. They detect all known GABAB1 isoforms except GABAB1j. Since they identify the two major isoforms, GABAB1a and b, their expressions were expected to be strong. Because GB1 probe sets were designed at the exon junction, there is no possibility of false signals due to genomic contamination. Among commercial primer and probe sets, GB1 sets also best represented the microarray probes that contain exon 23. GB2 sets were specific for other microarray probes containing intron 4. The

probes were human cDNA microarray clone 300899 and its mouse homologs. GB3 sets were designed within intron 4 and exon 6. The two primer sets, GB2 and GB3, were used to measure mRNA levels of multiple overlapping splice variants (e.g. primer set GB2 could detect expression of variants GABAB1j, k, l, m, and n while primer set GB3 could detect GABAB1k, l, m, and n except GABAB1j). Therefore, GB2 expression could include GB3.

Quantitative real-time PCR experiments showed different expression patterns between human and mouse brain. In both species, GB2 and GB3 expression was less than GB1 (Figure 2.4B, C). After comparing GB2 and GB3 expression with GB1, the relative GB2 and GB3 expression in human was less than in mouse (Figure 2.4B, C). The relative GB3 expression level difference between human and mouse was 4.2 fold which was less than the relative GABAB1j expression difference (25 fold) as shown in Figure 2.2B.

Prefrontal cortex, frontal cortex, and motor cortex were also used to characterize GB3 expression in human; GB3 expression levels were not significantly different across the brain regions (Figure 2.4D).

A.

Primer and probe set	Region	Isoforms detected	Note
GB1	exon 22-23	**GABAB1a**, **1b**, 1c, 1e, 1k, 1l, 1m, and 1n	Most isoforms except GABAB1j
GB2	intron 4	GABAB1j, 1k, 1l, 1m, and 1n	New isoforms including GABAB1j
GB3	intron 4 - exon 6	GABAB1k, 1l, 1m, and 1n	New isoforms except GABAB1j

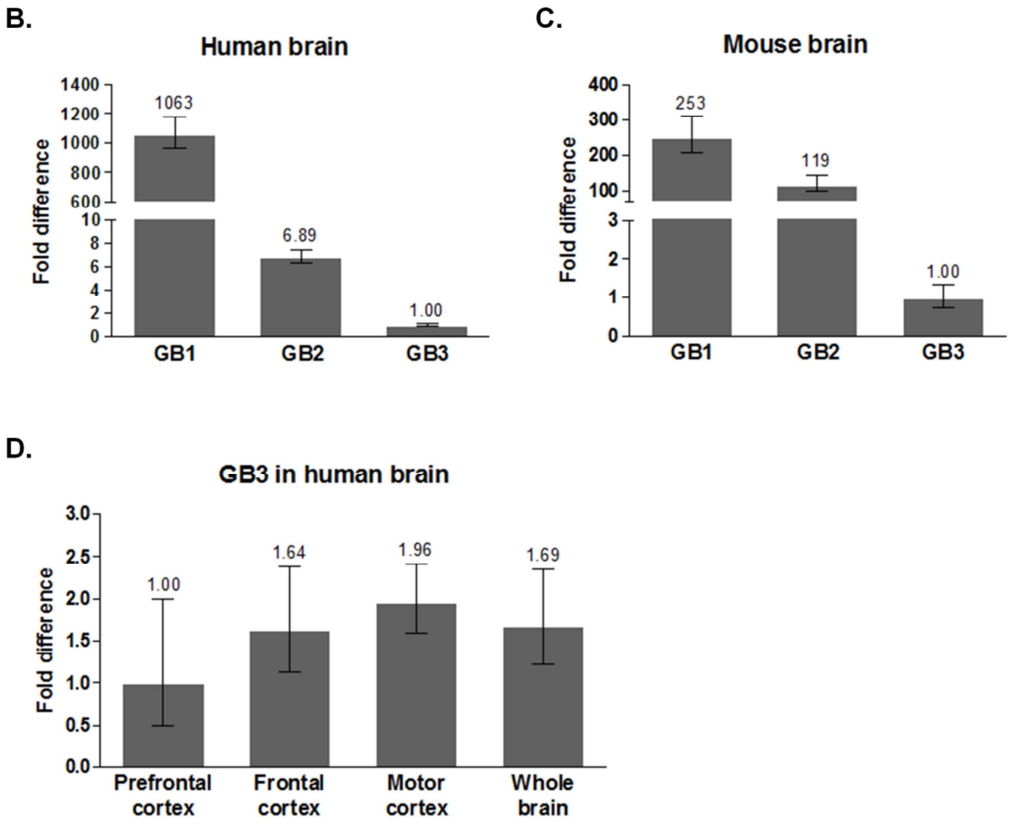


Figure 2.4 Expression of GABAB1 isoforms.
A. For quantitative real-time PCR, primer and probe sets were designed from human and mouse. GB1 sets were designed by the last two exons, exon 22 and 23, and were used for detecting most known GABAB1 isoforms except GABAB1j. They also represented the microarray probes that aligned to 5' exons of GABAB1 gene. GB2 sets were for detection of GABAB1j, k, l, m, and n. They aligned to intron 4 and designed for human cDNA microarray probe, clone 300899, and its mouse homologous probes. GB3 sets targeted intron 4 through exon 6 and were specific for GABAB1k, l, m, and n. B. In human brain, GB1 showed much more dominant expression than GB2 and GB3. C. GB1 expression was also stronger than GB2 and GB3 in mouse brain. However, GB2 and GB3 expression levels in mouse brain were much higher than human brain if they were compared with GB1 expressions. D. In human, GB3 expressions were similar in prefrontal cortex, frontal cortex, and motor cortex.

2.2.6 GABAB1k, GABAB1l, and GABAB1m function

The GABAB receptor is a heterodimeric (GABAB1 + GABAB2) GPCR. GABA binding to GABAB1 transfers a signal through G proteins whose functions are linked to several effectors, including GIRK. Thus, GIRK currents were measured using two-electrode voltage clamp to evaluate the functions of the new GABAB1 isoforms.

GABAB receptor responses were first measured in oocytes expressing GABAB1a and GABAB2 along with GIRK1 and GIRK2 with a range of GABA concentrations. 1mM GABA was the maximum response concentration. The EC_{50} was 1.9 uM, and Hill slope was 0.59 (Figure 2.5A).

For functional characterization, GABAB1k, l, and m were chosen because they have more functional domains than GABAB1n. Various amounts of each isoform cDNA were injected into oocytes together with GABAB2, GIRK1, and GIRK2. GIRK currents were measured following the previous GABAB1a and GABAB2 response measurement protocol (Figure 2.5B). However, no oocyte showed GABA-evoked GIRK currents.

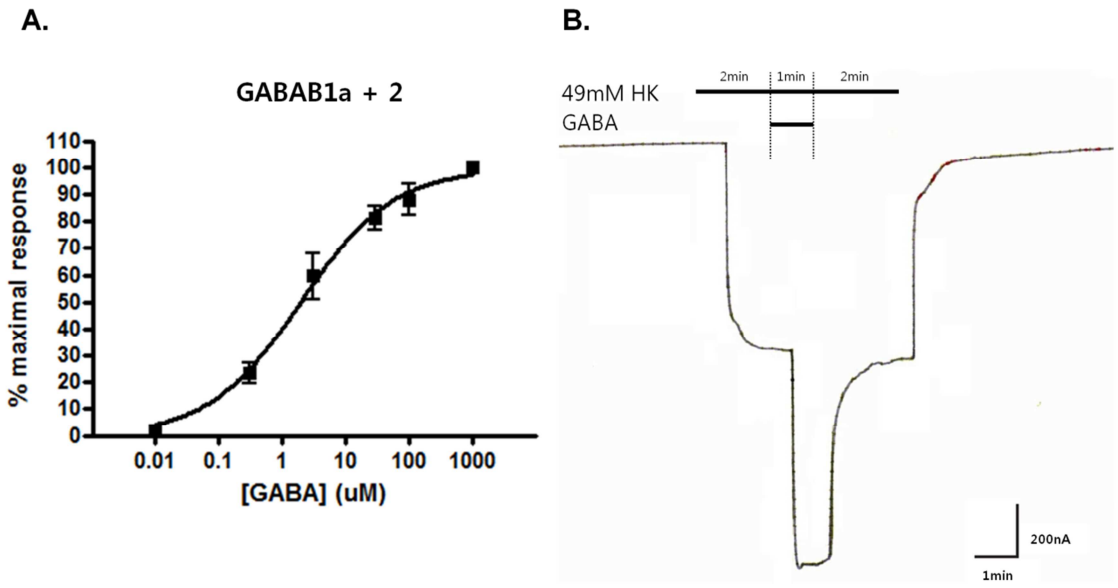


Figure 2.5 GABAB1a/2 concentration response curve.
A. cDNA and complementary RNA (cRNA) double injections were performed. For cDNAs, both GABAB1a and GABAB2 were injected into the nucleus of *Xenopus* oocytes. GIRK1 and GIRK2 cRNAs were also injected into the cytosol of the same oocytes. Current response of GABAB receptor and GIRK expressing oocytes were measured by two-electrode voltage clamp. 1mM GABA was the maximum response concentration; EC_{50} was 1.9 uM; Hill slope was 0.59. B. For GABAB1a/2 concentration response curve measurement, GABA dissolved in 49 mM high-K^+ solution (49mM HK) was applied for 1 min following a 2 min 49mM HK application. After 2 additional min of 49mM HK application, a 20 min wash with ND96 was followed.

Although GABAB1k, l, and m combined with GABAB2 did not induce GIRK currents, we asked if they might inhibit GABAB receptor function. Each isoform cDNA was injected along with GABAB1a, GABAB2, GIRK1, and GIRK2 into oocytes. For vector controls, the same amount of vector (instead of the new isoform) was injected along with GABAB1a, GABAB2, GIRK1, and GIRK2 into the control oocytes. The new isoform- and vector-injected oocytes were paired into groups based on injection amount. Injection amounts varied from 0 to 15 times the amount of GABAB1a injection (Figure 2.6). After 5 days of incubation, GABA-evoked GIRK currents were measured. GABA, 3 mM - an approximate EC_{100} for GABAB1a and GABAB2 receptors, was used. For each batch of oocytes, individual GIRK currents from the isoform-expressing group were normalized as a percentage of the average current of the vector control group. For nonparametric test, $\log_{10}$ values were calculated for the individual percentage values. One-way ANOVA and Dunnett's tests showed significant effects of GABAB1l and m ($p = 0.006$ and $p = 0.004$, respectively); however, GABAB1k did not show significant effects (Figure 2.6A, B, C). For GABAB1l and m data, one-tailed unpaired Student's t-tests were performed between no isoform injection data (0 x) and the other individual data (0.1 x ~ 15 x). GABAB1l significantly decreased GIRK currents at 5 x and 6 x of the GABAB1a concentration (**$p = 0.0014$ and *$p = 0.023$, respectively; Figure 2.6B). GABAB1m also evoked significant decreases at 11 x and 12 x of the GABAB1a concentration (***$p = 0.0003$ and **$p = 0.0064$, respectively; Figure 2.6C). Among 9 to 12 different oocyte batches, the inhibitory patterns were consistently observed. Except for the GABAB1l and m injection ranges that decreased GIRK currents, other injection

amounts did not show the inhibitory effects. This pattern was also observed at additional injection amounts (8 x of GABAB1k, 4.5 x and 5.5 x of GABAB1l, and 8 x of GABAB1m). In the GABAB1l inhibition range (5 x ~ 6 x of the GABAB1a concentration), 5.5 x also decreased GIRK current. The other injection amounts did not produce inhibition (data not shown), as expected from the inhibitory range shown in Figure 2.6.

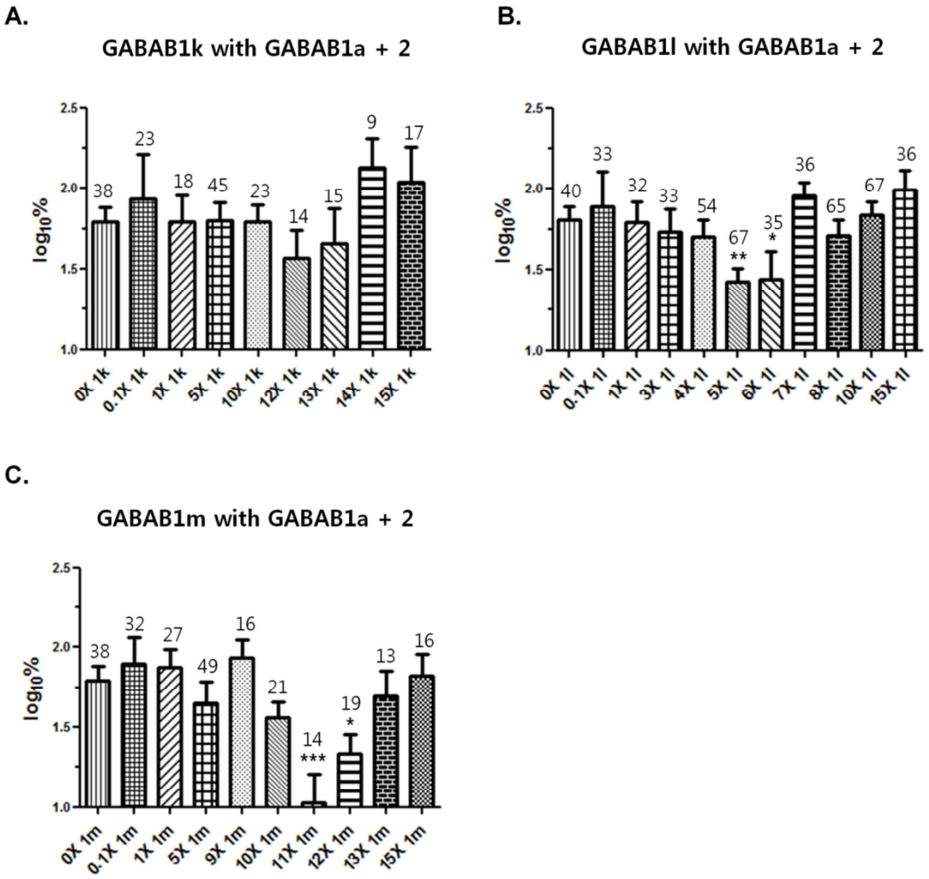


Figure 2.6 Functional effects of three novel human isoforms, GABAB1k, l, and m. Inhibitory effects were measured in *Xenopus* oocytes 5 days after injecting GABAB1a, GABAB2 cDNA and GIRK1, GIRK2 cRNA along with GABAB1k, l, or m cDNA. For vector controls, vectors were injected in the same amounts as used for the isoforms. Isoform and vector injection amounts varied from 0 to 15 times of GABAB1a injection amount. Used GABA concentration was 3 mM, an approximate maximum response concentration. For each batch, individual GIRK currents from isoform injected oocytes were normalized to percentage by the average vector control current. For nonparametric tests, the individual percent values were changed to $\log_{10}$. After one-way ANOVA and Dunnett's tests, one-tailed unpaired Student's t-tests were performed between no isoform injected groups and the isoform injected groups. A. GABAB1k had no inhibitory effect. B. GABAB1l data showed significant differences by ANOVA and Dunnett's tests ($p = 0.0058$), and caused significantly decreased GIRK currents at 5 x and 6 x GABAB1l (**$p = 0.0014$ and *$p = 0.0229$, respectively). C. ANOVA and Dunnett's tests showed the significant differences of GABAB1m data ($p = 0.0040$). GABAB1m showed significant decreases at 11 x and 12 x GABAB1m (***$p = 0.0003$ and **$p = 0.0064$, respectively). The numbers of recorded oocytes are shown above the bars by each group. Data are expressed as the mean ± standard error of the mean (SEM).

2.3 DISCUSSION

In this study, novel GABAB1 splicing variants containing the intron 4 region were cloned from human and mouse confirming and extending the previous suggestion that this intron is a frequent alternative splicing spot in the rat (Wei, Jia et al. 2001; Holter, Davies et al. 2005; Tiao, Bradaia et al. 2008). Previous studies cloned GABAB1g, i, and j from rat, all of which have a portion of intron 4 (Wei, Jia et al. 2001; Holter, Davies et al. 2005; Tiao, Bradaia et al. 2008). However, intron 4 containing GABAB1 splicing variants were not previously reported from human or mouse. Our quantitative real-time PCR data implied that intron 4 is also an alternative exon in human and mouse. Therefore, mouse GABAB1j clones were found from the cDNA microarray probe set. Though human GABAB1j was not cloned in this study, its existence was confirmed by quantitative real time PCR. From human and mouse, four more intron 4 containing clones were identified. GABAB1k, l, m, and n were cloned from human, and another GABAB1k was also cloned from mouse. Our study further emphasizes that intron 4 is a frequent splicing spot of the GABAB1 gene in multiple species.

Though expression levels of these novel isoforms in total brain were lower than other abundant isoforms (GABAB1a and b), they may be preferentially expressed in specific cell types just as GABAB1a and GABAB1b show specificity for pre- and postsynaptic localization (Vigot, Barbieri et al. 2006) and are localized in the different cells using their own unique promoters (Bettler, Kaupmann et al. 2004; Gassmann, Haller et al. 2005). In cerebellum, GABAB1a is predominantly expressed in granule cells, and GABAB1b is abundant in Purkinje cells (Billinton, Ige et al. 2001). Like GABAB1a and

GABAB1b, there are many other precedents for N-terminal truncated splicing variants that have their own promoters and show spatially different expression patterns (Wang, Schultheis et al. 1996; Crofts, Hancock et al. 1998; Sunn, Cock et al. 2001; Travers, Vallance et al. 2003; Huang, Levitsky et al. 2009). The novel isoforms are also N-terminal truncated splicing variants and probably use specific promoters for their expression. After identifying GABAB1 gene putative transcription factor binding sites from the PROMO website, possible promoter regions were found on exon 4 and the 5' part of intron 4 (data not shown) (Messeguer, Escudero et al. 2002; Farre, Roset et al. 2003). Thus, the novel isoforms may be restricted to specific cell types and inhibit GABAB receptor function.

GABAB1l and m might be secreted like GABAB1e because they also lack TMs and have the same C-terminal truncated patterns as GABAB1e (Schwarz, Barry et al. 2000). This phenomenon has been observed for splicing variants of Interleukin-15 receptor α, Fibroblast growth factor receptor 4, human GABAB1e, and rat GABAB1j which lose TMs by alternative splicing and are secreted (Schwarz, Barry et al. 2000; Ezzat, Zheng et al. 2001; Bulanova, Budagian et al. 2007; Tiao, Bradaia et al. 2008). Our functional studies used continuous perfusion of the oocytes, and it is possible that GABAB1l and m were secreted and washed away resulting in low intracellular and extracellular concentrations and obscuring some functional effects. However, many secreted splicing variants including the C-terminal truncated GABAB1 isoforms inhibit the functions of their other membrane-bound isoforms (Mosley, Beckmann et al. 1989; Schwarz, Barry et al. 2000; Ezzat, Zheng et al. 2001; Bulanova, Budagian et al. 2007;

Tiao, Bradaia et al. 2008). It is possible that secreted GABAB1l, m, and n could also inhibit GABAB receptor function.

The inhibitory effects of GABAB1l and m were found only within a narrow concentration window of injected cDNA. The reasons are not clear, but it is not likely due to nonspecific artifact or toxicity with the highest amounts injected. Therefore, above the optimal concentration window, their effects were diminished. The vector controls also had the same injection amounts, but their GABA responses were constant and stable. Also, the inhibitory effects were seen consistently among all 9 - 12 different oocyte batches. Some genes require optimal expression levels to display their functional properties (Lamond and Travers 1985; Voliva, Aronheim et al. 1992; Park, Lee et al. 2005), and GABAB1m and n may also need optimal expression ranges to inhibit function.

GABAB1e binds GABAB2 though a dimerization motif (166 - 439 aa of GABAB1e) and appears to dominantly bind GABAB2 as well as competing with GABAB1a (Schwarz, Barry et al. 2000). Its dimerization motif is not sushi domains but the rest of the N-terminal extracellular domain. GABAB1k, l, m, n, but not GABAB1j contain the dimerization motif and would be expected to bind GABAB2 dominantly instead of GABAB1a and even reduce receptor function.

In summary, alternative splicings at intron 4 generate novel isoforms, GABAB1j, k, l, m, and n (Figure 2.1A). The relative mRNA expression patterns of GABAB1k, l, m, and n are more conserved than GABAB1j across species when compared to general GABAB1 isoform expression. However, the expression of novel N-terminal splicing

variants, GABAB1k, l, m, and n, is low in whole brain (Figure 2.4D). They may have restricted expression at specific cell types because they have putative promoter sites. Though GABAB1k, l, m, and n are expected to dominantly dimerize with GABAB2 like GABAB1e, only GABAB1l and m were observed to disrupt the function of GABAB1a and 2. GABAB1e has only C-terminal truncation (Schwarz, Barry et al. 2000), but the GABAB1l and m have both C-terminal and N-terminal truncated patterns (Figure 2.1B). Therefore, this additional N-terminal truncated pattern may increase their inhibitory functions.

This work expands evidence for the diversity of GABAB1 splice variants in brain. It is important to note that variants were found in multiple species, and it has been noted that splice variants which are likely to contribute to evolution and have functional value are conserved across species (Lareau, Green et al. 2004). In considering possible functional roles of these splice variants, we should note that GABAB receptors provide diverse neuronal signaling through G proteins and may have additional functions as they form complexes with other proteins (Chang, Tu et al. 2007).

2.4 MATERIALS AND METHODS

2.4.1 Sequence analysis

To predict two possible isoforms, GABAB1j and GABAB1k, SOURCE (http://smd.stanford.edu/cgi-bin/source/sourceSearch) was used to search the sequence of the microarray probe, clone 300899. The sequence was converted to a FASTA format by

EditSeq programs of DNASTAR (DNASTAR, Inc., Madison, WI, USA). Blast N Search (www.ncbi.nlm.nih.gov/blast) showed where the probe aligned in GABAB1 gene. Its multiple stop codons were founded from six different translated frames using MapDraw programs of DNASTAR.

To compare GABAB1 isoform sequences, their sequences and exon-intron structure information were found from NCBI data base (http://www.ncbi.nlm.nih.gov). Their sequences were aligned with MegAlign programs of DNASTAR, and their conserved domains were searched by ExPASy Proteomics Server (http://expasy.org) and CD-Search (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). To align GABAB1j sequences, ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html) were used.

2.4.2 Identification of new GABAB1 isoform, GABAB1k

To determine the existence of novel GABAB1 isoforms, GABAB1j and GABAB1k, across species, the microarray probe, clone 300899, was cloned from human, and its homologous regions were also cloned from mouse and rat based on Blast N Search. For further examination, partial GABAB1k clones were identified. The partial GABAB1k clones contain the microarray probe and GABAB1k unique region, 3' intron 4 of GABAB1a. For these experiments, total RNA was isolated from human cultured cells, including human embryonic kidney 293 (HEK-293) (ATCC, Manassas, VA, USA), SH-SY5Y (ATCC, Manassas, VA, USA), and SK-N-SH (ACTT, Manassas, VA, USA). Total RNA was also isolated from mouse midbrain and rat hippocampus. After treating DNase (Qiagen, Valencia, CA, USA), cDNAs were made using oligo(dT) primer

(Invitrogen, Carlsbad, CA, USA). To clone the microarray probe, primers were designed within the probe region using Primer3 (http://frodo.wi.mit.edu/primer3/) and PrimerSelect program of DNASTAR. For partial GABAB1k cloning, the probe region and other 3' exons were targeted to design primers. PCRs were performed using the primers and the prepared human, mouse, and rat cDNAs.

PCR primers (5' to 3'):

- Human

Not1-hGABBR1iso-probe-S:
AAGGAAAAAAGCGGCCGCCCTTCTTCAGGTTTAAATACTCCA

Sal1-hGABBR1iso-probe-A:
ACGCGTCGACCCATCGAGCCCTCTATATTATTAG

Not1-hGABBR1iso-#1probe-S1:
AAGGAAAAAAGCGGCCGCCTTGGGCTTTACTTTCCTCACAT

Sal1-hGABBR1iso-#2probe-A1:
ACGCGTCGACGAATCTACCAGGGTAGAGGAAGT

Sal1-hGABBR1-A: ACGCGTCGACCTTGATAGGGTCGTTGTAGAGC

Sal1-hGABBR1-A1: ACGCGTCGACAGAGTCGAAGTGAAGACCTCAG

- Mouse

Cla1-mGABBR1iso-probe-S:
CCATCGATTAGACTTAGAACATCCTCTGTATCA

Cla1-mGABBR1iso-probe-A: CCATCGATAGAGTTTGGGTGCCCAGAAAAG

Cla1-mGABBR1-A: CCATCGATGGATAAGCTTGAGCTCGTAGTC

- Rat

Not1-rGABBR1iso-probe-S:

AAGGAAAAAAGCGGCCGCCCCTCTCTGTAGATGTAGAACATT

Sal1-rGABBR1iso-probe-A:

ACGCGTCGACTTCTGGTCATTCTGCCCCAACT

Sal1-rGABBR1-A: ACGCGTCGACGGATAAGCTTGAGCTCGTAGTC

2.4.3 Cloning novel isoforms, GABAB1k, GABAB1l, GABAB1m, and GABAB1n

To clone full ORFs of the novel isoforms, cDNAs were generated with oligo(dT) primer from human brain total RNA (Ambion, Austin, TX, USA) and mouse midbrain total RNA. Primers were designed based on intron 4 of GABAB1a and common 3' UTR of known isoforms, and PCRs were performed using the human and mouse brain cDNAs.

PCR primers (5' to 3'):

- Human

Not1-hGABBR1iso-#1probe-S1:

AAGGAAAAAAGCGGCCGCCTTGGGCTTTACTTTCCTCACAT

Sal1-hGABBR1-A2: ACGCGTCGACGGTAACTAGAAGAGGGTGTTGC

Not1-hGABBR1cl-CD-S:

AAGGAAAAAAGCGGCCGCTCTTCTCTGATCCCCGTCTTT

Sal1-hGABBR1cl-CD-A: ACGCGTCGACTACTGGCCTGTCCTCCCTC

- Mouse

Cla1-mGABBR1cl-CD-S: CCATCGATTTCTCTGATCCCCGTCTTTC

Cla1-mGABBR1cl-CD-A: CCATCGATTGCTGGCCTCATCCTTCTC

2.4.4 Quantitative real-time PCR

TaqMan® Gene Expression Assays and Custom TaqMan® Gene Expression Assays (Applied Biosystems, Foster City, CA, USA) were used to detect GABAB1 isoform expressions. cDNA microarray probes are two different kinds containing exon 23 or intron 4. The probes containing exon 23 detect all known isoform expressions except GABAB1j. The exon 22 - 23 junction was targeted using predesigned TaqMan® Gene Expression Assays. Custom TaqMan® Gene Expression Assays were used for intron 4 containing probes, and thus were able to detect GABAB1j, GABAB1k, GABAB1m, GABAB1l, and GABAB1n. Using the same method, the unique alternative splicing region of GABAB1k, GABAB1m, GABAB1l, and GABAB1n were targeted, and GABAB1j specific expression was also detected. For endogenous control gene, beta-glucuronidase (*GUSB*) was used.

After making cDNAs with random primers (Invitrogen, Carlsbad, CA, USA), quantitative real-time PCR was performed. Isoform expressions were compared between human and mouse whole brain RNAs (Ambion, Austin, TX, USA). To measure isoform expression differences in select human brain regions, RNAs from whole brain, prefrontal cortex, frontal cortex (Ambion, Austin, TX, USA), and motor cortex (Liu, Lewohl et al. 2004) were used.

2.4.5 cDNA and cRNA preparation

To do injections into *Xenopus* oocytes for electrophysiological recordings, cDNAs and cRNAs were prepared. For cDNA injection, human GABAB1a clone in pcDNA3.1(-) vector (Invitrogen, Carlsbad, CA, USA) and human GABAB2 clone in pcDNA3 vector (Invitrogen, Carlsbad, CA, USA) were kindly provided by Dr. Uezono (Nagasaki University, Nagasaki, Japan) with the permission of GlaxoSmithKline (Brentford, Middlesex, UK). The cDNAs encoding novel human GABAB1 isoforms, GABAB1k, GABAB1l, and GABAB1m, were also subcloned into pcDNA3.1(-) vector. The cRNAs for GIRK1 and GIRK2 were synthesized as described previously for cRNA injections (Lewohl, Wilson et al. 1999).

2.4.6 Electrophysiological recording

For electrophysiological recordings, oocytes from *Xenopus laevis* frog were used. The frog was used according to the National Institutes of Health guide for the care and use of laboratory animals.

To form functional receptors, nuclear injection of GABAB1a and GABAB2 cDNAs were performed into *Xenopus* oocytes with 0.75 ng and 0.63 ng, respectively. GIRK1 (0.2 ng) and GIRK2 (0.2 ng) cRNAs were also injected into the cytosol of the same oocytes in order to measure GABAB receptor-induced currents. Total injection volumes of cDNAs and cRNAs were 60 nl. Electrophysiological recordings were made with the oocytes after 5 to 7 day incubation at 19 °C. Using two-electrode voltage clamp, the oocytes were clamped at -60 mV while being perfused (2 ml/min) with ND96 buffer

(in mM: 96 NaCl, 2 KCl, 1 $CaCl_2$, 1 $MgCl_2$, 5 HEPES, pH7.5). A 49mM HK (in mM: 48 NaCl, 49 KCl, 1 $CaCl_2$, 1 $MgCl_2$, 5 HEPES, pH7.5) was applied to the oocytes for 2 min (Uezono, Kanaide et al. 2006). GABA (Sigma, St. Louis, MO, USA), ranging in concentration from 10 nM to 1 mM, was dissolved in 49mM HK and applied for 1 min. ND96 was applied for 20 min following the 2 min application of 49mM HK.

The responses of GABAB1a and GABAB2 to the range of GABA concentrations were used to generate GABA concentration response curves with GraphPad Prism, Version 5.01 (GraphPad Software, San Diego, CA, USA). EC_{50} and Hill coefficient values were calculated from the GABA concentration response curves. A maximally effective GABA concentration was also determined from it and was used to test inhibitory effects of the isoforms.

To test if the isoforms can substitute for GABAB1a as a functional receptor subunit, 0.63 ng GABAB2, 0.2 ng GIRK1, and 0.2 ng GIRK2 were injected into oocytes with various amounts (0.001 ng to 20 ng) of isoform cDNAs. Total injection volumes were 60 nl. As in the GABA concentration response experiment, GABA-evoked GIRK currents were measured to determine receptor function.

To determine the inhibitory functions of isoforms, GABAB1a was co-injected with each isoform. Except the isoforms, the injection amounts of GABAB1a, GABAB2, GIRK1, and GIRK2 were the same as the GABA concentration response experiment. For vector controls, the same amount of vector, pcDNA3.1(-), was injected into control oocytes instead of the isoforms. Total injection volumes were 60 nl. The isoform and vector injection amounts varied from 0 to 15 times the GABAB1a injection amount (0.75

ng). After a 5 day incubation at 19 °C, GABA-evoked GIRK currents were measured. The GABA concentration was 3 mM which was the maximal concentration for GABA responses in GABAB1a and GABAB2 expressing oocytes. Nine batches of oocyte were used for both GABAB1k and m experiments, and 12 batches were for GABAB1l. For each batch of oocytes, individual GIRK currents in isoform injected oocytes were expressed as a percentage of average current in vector control oocytes. For nonparametric tests, $\log_{10}$ data were calculated from the values.

2.4.7 Statistical analysis

For statistical analysis and graphing, GraphPad Prism and Excel (Microsoft, Redmond, WA, USA) were used. For electrophysiology data, one-way ANOVA tests and unpaired Student's t-tests were used to define statistical significances. Dunnett's tests were also used for post hoc tests.

CHAPTER 3. ALCOHOLISM CORRELATS WITH GABAB1 SPLICING IN HUMAN BRAIN

3.1 INTRODUCTION

GABAB receptors have been implicated in regulation of alcohol drinking. The positive allosteric modulator, GS39783, suppresses alcohol drinking and reinforcement in rat models (Maccioni, Fantini et al. 2008). A GABAB receptor agonist, GHB, can reduce voluntary ethanol drinking and withdrawal symptoms in humans and is a treatment for alcoholism (Cammalleri, Brancucci et al. 2002; Schweitzer, Roberto et al. 2004). Another GABAB receptor agonist, baclofen, may also be effective in treating the disease (Enserink 2011).

Previous microarray studies have shown strong GABAB1 expression in prefrontal cortices from human alcoholics (Flatscher-Bader, van der Brug et al. 2005; Liu, Lewohl et al. 2006). These experiments suggested complexity of GABAB1 gene splicing, which we confirmed by cloning of new splice variants from human brain (Isomoto, Kaibara et al. 1998; Schwarz, Barry et al. 2000; Tiao, Bradaia et al. 2008; Lee, Mayfield et al. 2010). The microarray experiments used a cDNA probe set, and two out of three GABAB1 probes were generated by unknown splicings. A GABAB1 probe, clone 300899, was aligned to intron 4 of GABAB1. The intron 4 is a common intron region from Reference Sequence (RefSeq) Genes model and major GABAB1 splicing variants

(Lee, Mayfield et al. 2010). Another probe, clone 2312175 targeted an unknown splicing out at exon 23 of GABAB1, and a new variant, GABAB1m, has unexpected splicing out at exon 6 of GABAB1 in brain (Lee, Mayfield et al. 2010). Five different splicing variants were cloned as containing clone 300899 (Tiao, Bradaia et al. 2008; Lee, Mayfield et al. 2010). These also showed many unknown GABAB1 splicings and the overall complexity of GABAB1 splicing.

Among the three GABAB1 microarray probes, only the intron 4 containing probe, clone 300899, showed increased expression in alcoholic brain(Flatscher-Bader, van der Brug et al. 2005; Liu, Lewohl et al. 2006). This suggested a specific GABAB1 splicing change resulting from chronic alcohol exposure. Signal intensities of the three probes were different, and the difference was not explicable with current known splicing variants. Due to this complexity of GABAB1 splicing, it is likely that previously unknown GABAB1 splicing variants are differentially expressed in alcoholic brains. Other studies have also shown that chronic ethanol treatments change specific splicing patterns of channel or receptor genes in brain. Chronic ethanol exposure and withdrawal increased 5' splicing variant expression of the NMDA receptor subunit, NR1, without NR1 3' splicing variant expression changes (Hardy, Chen et al. 1999). The L-type voltage-gated Ca^{2+} channel, α_{1c}, has two splicing variants, $\alpha 1_{c\text{-}1}$ and $\alpha 1_{c\text{-}2}$. Chronic ethanol treatment enhanced specific splicings to increase only the $\alpha 1_{c\text{-}1}$ population (Walter, McMahon et al. 2000).

To identify novel splicing junctions of GABAB1, the NCBI sequence database is one possible option. However, the actual splicing variant number of the whole genome is

far higher than the database (Fodor and Aldrich 2009). Therefore, many unknown splicings of the GABAB1 gene likely exist, and we used RNA-seq to detect new variants of GABAB1 and to quantitate splicing differences in prefrontal cortex of brain tissue from alcoholic and non-alcoholic subjects.

3.2 RESULTS

3.2.1 Gene specific library construction to maximize GABAB1 splicing junction detection

A gene specific library was designed for RNA-seq to maximize splicing junction identification in the GABAB1 gene. Current approaches to RNA-seq libraries construct often utilize random primers or oligo(dT) primers for reverse transcription. However, this approach does not generate the required number of mapped reads from single RNA-seq run to find rare splicing junctions in the GABAB1 gene. Therefore, we used a different approach by generating a gene specific library for our splice junction study in addition to the whole transcriptome library. The gene specific library was constructed using gene specific primers instead of random primers for the reverse transcription; thus, it provided a large population of gene specific reads and maximized detection of gene specific splice junctions.

rRNA constitutes 90-95% of total RNA. If we use total RNA for gene specific library construction, it must generates more nonspecific sequence data from rRNAs than GABAB1 gene specific data. Therefore, we used rRNA depleted total RNA (Figure

3.1).Two gene specific primers were designed at the common 3’ UTR of the GABAB1 transcripts. After reverse transcription using one gene specific primer, the template switching method generated double strand cDNAs (dscDNAs). For amplification, we used the other gene specific primer designed closer to the ORF at the 3’ UTR. Gel extraction removed unbound primers, and the dscDNAs were fragmented. After adaptor ligation, SOLiD sequencing was performed (Figure 3.1). The whole transcriptome library was also prepared for SOLiD sequencing in order to compare the gene specific library with the whole transcriptome library.

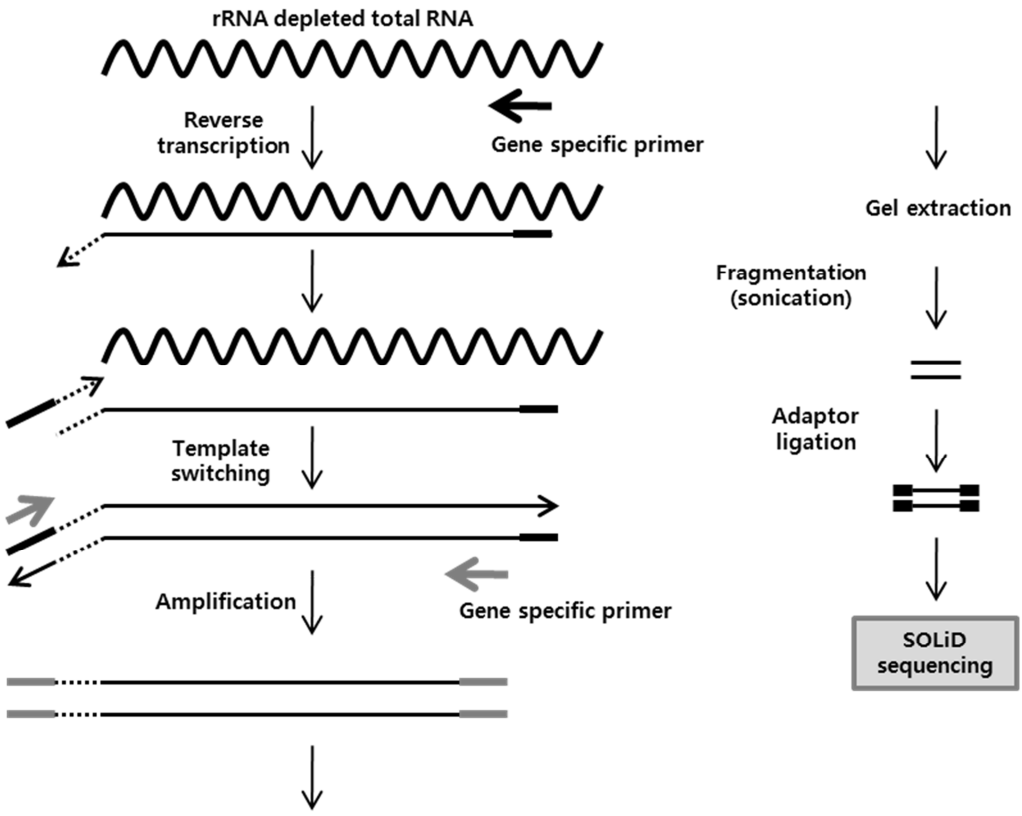


Figure 3.1 Gene specific library construction.
To prepare the gene specific library, rRNA was depleted from total RNA. Single stranded cDNA (sscDNA) was generated by reverse transcription using a gene specific primer. By template switching dscDNA was synthesized, and another gene specific primer was used for amplification. After gel extraction, size selected dscDNAs were sonicated, and adaptors were ligated. The gene specific library was sequenced by SOLiD system.

3.2.2 Comparison between result derived from gene specific and whole transcriptome libraries

The gene specific and the whole genome libraries were compared in several different ways: 1) number of GABAB1 specific mapped reads, 2) splicing junctions, 3) mapping patterns, and 4) possible transcription initiation sites. As expected, the gene specific library had more GABAB1 specific mapped reads than the whole transcriptome library (Figure 3.2). The gene specific library had 134-fold more mapped reads at GABAB1 than the whole transcriptome library. More splicing junctions were also identified in the gene specific library than the whole transcriptome library as visualized by the Integrative Genomics Viewer (IGV) (Figure 3.2).

The gene specific library contained more mapped reads on the 3' end of GABAB1 while mapped read numbers were decreased at the 5' end of GABAB1 (Figure 3.2A). Due to the uneven read distribution of the gene specific library, 5' exons (from exon 1 to exon 5) had no mapped reads while the whole transcriptome library did (Figure 3.2). The biased distribution likely resulted from gene specific reverse transcription and amplification based on the GABAB1 3' UTR. Similar biased distributions were also found from another 3' UTR targeting whole transcriptome libraries using oligo(dT) primed reverse transcription (Wang, Gerstein et al. 2009).

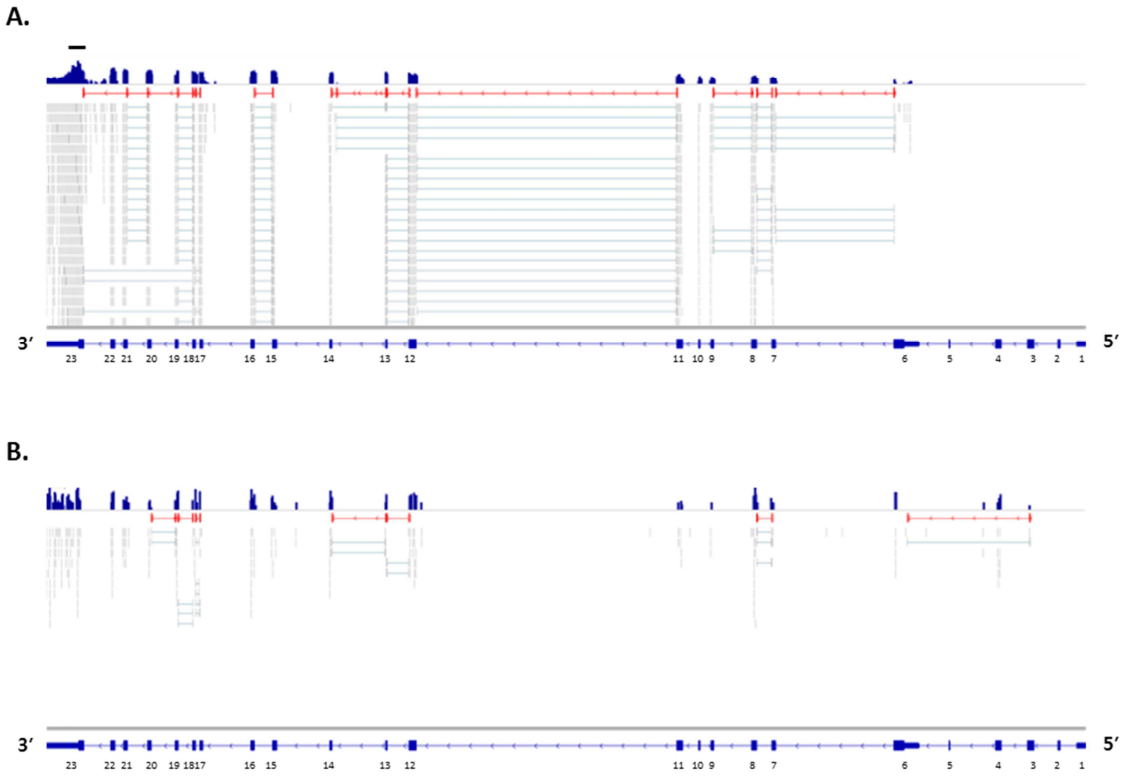


Figure 3.2 Compared result from between gene specific library and whole transcriptome library.

All TopHat mapping files, .wig, .bed, and .sam, were loaded on the IGV and showed at the GABAB1 gene in blue, red, and grey, respectively. The .wig files (blue) contain the number of mapped reads at each nucleotide position. Splicing junctions (red) were uniquely represented using the .bed files. The .sam (grey) files contained individual mapped reads. For the .bed and .sam files, the thick lines represented mapped reads, and the thin lines represented splicing junctions. Mapped read numbers were compared in log space using the .wig files. With the three files, the gene specific library and whole transcriptome library results were shown at A and B, respectively. Because GABAB1 is negatively oriented at chromosome 6, the 5' and 3' ends of GABAB1 are located on the right and left during IGV visualization, respectively. For the gene specific library, only the upper portion of the .sam file data is shown due to space limitations. The gene specific library results had more GABAB1 specific mapped reads and splicing junctions than the whole transcriptome library. The black bar on top of figure A represents the location of both RT and amplification primers. Exon numbers are labeled below the gene model of A and B.

3.2.3 RNA-seq data from alcoholic and control human subjects

To compare GABAB1 splicing in prefrontal cortex alcoholics and matched controls, the data were processed to optimize the discovery of splice junctions and minimize false negative. The comparison of the gene specific and the whole transcriptome libraries showed that the gene specific library was better suited for splicing junction identification because the gene specific library could found more GABAB1 specific splicing junctions than the whole transcriptome library. Moreover, TopHat mapping using additional junction information can identify additional splice junctions (http://tophat.cbcb.umd.edu/manual.html).

To maximize the search for splice junctions, an additional junction search step was added prior to the final mapping (Figure 3.3). For the junction search step, we used both the gene specific and whole transcriptome libraries. Specically, the gene specific libraries provided a large population of GABAB1 gene specific splicing junctions including rare splicing junctions (Figure 3.2A). After TopHat mapping of individual libraries, all splice junction data were collected (Figure 3.3), and unique junction information was used for the final mapping of the whole transcriptome library data from brain tissues of 14 alcoholic and 15 control human subjects. Splice junction information was collected again from the individual mappings. Based on the splicing junction information, minimum GABAB1 gene length was defined and further data analyses were performed including visualization.

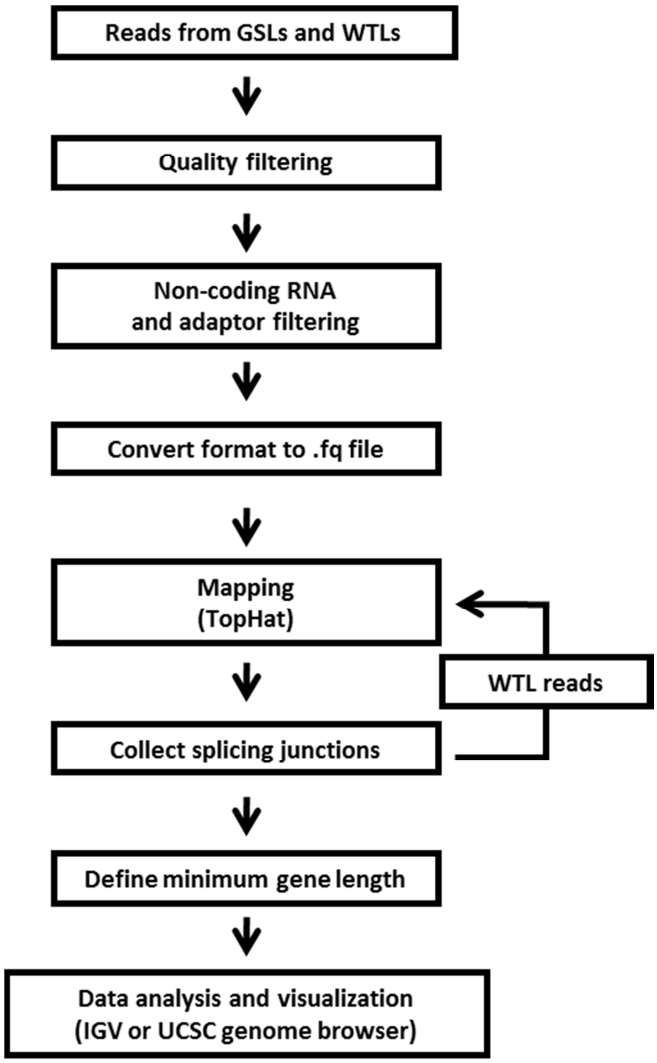


Figure 3.3 Data analysis to maximize splice junction detection.
To maximize the identification of splice junctions using TopHat, a junction search step was added to the data analysis pipeline. We used gene specific libraries (GSLs) and whole transcriptome libraries (WTLs) prepared from commercial and postmortem human prefrontal cortices. After filtering low quality reads and undesired sequence containing reads, the data were converted to .fq files. After TopHat mapping, splice junction information was collected for the remapping step. We used the whole transcriptome library data from 14 alcoholic and 15 control human brain samples for the remapping step. After recollecting splicing junctions, we defined the minimum GABAB1 gene length.

For each step, the remaining read number and its ratio compared to raw reads were calculated (Table 3.1). Control and alcoholic samples had similar ratios from each step though there were some individual variations. Based on the quality filter, about 20 % of raw reads were filtered out, and the second filter for non-coding RNAs and adaptors left approximately 55 % of raw reads. Format change caused about an additional 2 ~ 3 % loss. After TopHat mapping, the unique mapped reads showed about 41 % mappability from its mapping inputs.

Table 3.1 Mapping statistics of 15 control and 14 alcoholic human subjecs.

Control samples	cSFC209	cSFC228	cSFC232	cSFC257	cSFC295	cSFC329	cSFC339	cSFC366	cSFC377	cSFC442	cSFC451	cSFC459	cSFC463	cSFC474	cSFC477
Raw reads	33,094,843	35,343,212	18,581,068	38,888,111	36,224,909	83,654,681	36,663,941	22,449,165	11,911,746	32,833,007	35,882,930	40,277,699	48,561,631	25,737,539	34,263,844
Qaulity filter	26,632,535	24,225,746	15,668,505	32,749,659	28,681,117	70,114,358	29,285,931	17,992,194	10,075,364	26,096,695	28,305,869	33,316,350	40,821,155	21,519,617	25,889,478
	80.47%	68.54%	84.33%	84.22%	79.18%	83.81%	79.88%	80.15%	84.58%	79.48%	78.88%	82.72%	84.06%	83.61%	75.56%
ncRNA and adaptor filter	18,850,924	12,975,759	12,232,580	25,464,797	18,601,606	59,450,301	22,684,693	10,918,514	7,237,675	18,699,278	22,277,760	26,803,061	25,251,036	17,005,900	11,290,576
	56.96%	36.71%	65.83%	65.48%	51.35%	71.07%	61.87%	48.64%	60.76%	56.95%	62.08%	66.55%	52.00%	66.07%	32.95%
Format change (a)	18,267,859	12,356,032	11,638,800	24,390,013	17,750,765	57,245,003	21,713,129	10,383,532	6,886,818	17,876,790	21,322,125	25,701,722	24,003,408	16,197,374	10,611,480
	55.20%	34.96%	62.64%	62.72%	49.00%	68.43%	59.22%	46.25%	57.82%	54.45%	59.42%	63.81%	49.43%	62.93%	30.97%
Unique mapped reads (b)	7,866,682	4,165,394	4,867,740	10,981,783	7,290,526	27,434,260	8,444,624	4,349,063	3,087,588	7,190,915	7,682,991	12,082,894	11,115,208	6,969,258	4,992,016
	23.77%	11.79%	26.20%	28.24%	20.13%	32.79%	23.03%	19.37%	25.92%	21.90%	21.41%	30.00%	22.89%	27.08%	14.57%
(b)/(a)	43.06%	33.71%	41.82%	45.03%	41.07%	47.92%	38.89%	41.88%	44.83%	40.22%	36.03%	47.01%	46.31%	43.03%	47.04%

Alcoholic samples	aSFC204	aSFC210	aSFC217	aSFC234	aSFC260	aSFC303	aSFC313	aSFC315	aSFC327	aSFC330	aSFC332	aSFC351	aSFC430	aSFC476
Raw reads	27,821,439	26,047,363	14,566,308	21,984,169	126,588,851	36,030,714	52,006,412	34,642,762	90,749,179	32,384,057	35,020,905	38,226,079	29,939,449	25,688,439
Qaulity filter	21,406,081	21,840,200	12,404,367	18,367,384	99,273,661	27,099,685	37,856,724	27,081,823	75,869,348	26,756,321	27,865,908	25,866,170	23,554,695	21,218,728
	76.94%	83.85%	85.16%	83.55%	78.42%	75.21%	72.79%	78.17%	83.60%	82.62%	79.57%	67.67%	78.67%	82.60%
ncRNA and adaptor filter	14,260,967	15,525,194	9,607,281	13,366,925	74,456,408	13,603,479	25,701,777	19,371,738	46,758,853	19,897,024	18,548,839	20,354,921	18,089,543	16,388,660
	51.26%	59.60%	65.96%	60.80%	58.82%	37.76%	49.42%	55.92%	51.53%	61.44%	52.97%	53.25%	60.42%	63.80%
Format change (a)	13,609,570	14,751,500	9,144,836	12,710,443	71,410,465	12,952,198	24,532,333	18,495,269	45,739,200	19,094,716	17,724,874	19,449,128	17,347,506	15,598,453
	48.92%	56.63%	62.78%	57.82%	56.41%	35.95%	47.17%	53.39%	50.40%	58.96%	50.61%	50.88%	57.94%	60.72%
Unique mapped reads (b)	5,364,440	6,617,476	4,022,523	5,315,184	32,701,183	5,051,370	4,546,158	6,862,433	12,282,055	8,108,307	7,473,312	6,860,571	7,612,834	7,027,911
	19.28%	25.41%	27.62%	24.18%	25.83%	14.02%	8.74%	19.81%	13.53%	25.04%	21.34%	17.95%	25.43%	27.36%
(b)/(a)	39.42%	44.86%	43.99%	41.82%	45.79%	39.00%	18.53%	37.10%	26.85%	42.46%	42.16%	35.27%	43.88%	45.06%

Under the read numbers, ratios were calculated from raw reads. (b)/(a) represented the ratio of unique mapped reads (b) to format change (a).

3.2.4 GABAB1 splicing junctions and gene lengths

After the final mapping, unique splicing junction data were visualized using the University of California, Santa Cruz (UCSC) genome browser (Figure 3.4A). 20 splicing junctions were found among the 22 known splicing junctions from the RefSeq Genes model. An additional 43 novel splicing junctions were also found; therefore, there are at least 65 splice junctions associated with GABAB1 (visualized in Figure 3.4A). The 43 splicing junctions are rare splicing junctions. Rare splicing junctions are also identified from other intensive junction discovery based on computational approach (Pickrell, Pai et al. 2010). They found 23 rare splicing junctions out of total 33 splicing junctions from HERPUD1 gene. Their rare junction finding ratio was almost the same as out data. Also, TopHat false discovery rate is lower than 1 % for rare splicing junction (Dimon, Sorber et al. 2010).

Based on this splicing junction information, the minimum GABAB1 gene length is 72,333 bp (chr6: 29,653,437 - 29,725,769) on the human reference genome (hg18) (Figure 3.4A). This GABAB1 gene length is about 41 Kbp longer than the known GABAB1 length of 30,958 bp (chr6: 29,677,984 - 29,708,941). The minimum GABAB1 gene length from UCSC databases (i.e., human mRNAs from GeneBank and human ESTs that have been spliced) is 100,266 bp (chr6: 29,631,387 - 29,731,652). These findings provide strong evidence that the GABAB1 gene is larger than previously reported.

We categorized the 65 GABAB1 splicing junctions into four groups: known, known-alternative, partial-novel, and novel splicing junctions (Figure 3.4B). The known

splicing junctions represent those found in the RefSeq Genes model. The known-alternative splicing junctions represent different splicing site combinations of the known splicing junctions. The partial-novel splicing junctions represent junctions with one known splicing site and another unknown site. The novel splicing junctions represent junctions with both unknown splicing sites. For GABAB1, there were 22 known, 5 known-alternative, 8 partial-novel, and 30 novel splicing junctions. Thus, the number of known-alternative, partial-novel, and novel splicing junctions are almost twice the number of known splicing junctions.

Figure 3.4 GABAB1 splicing junctions and gene boundaries.

A.

B.

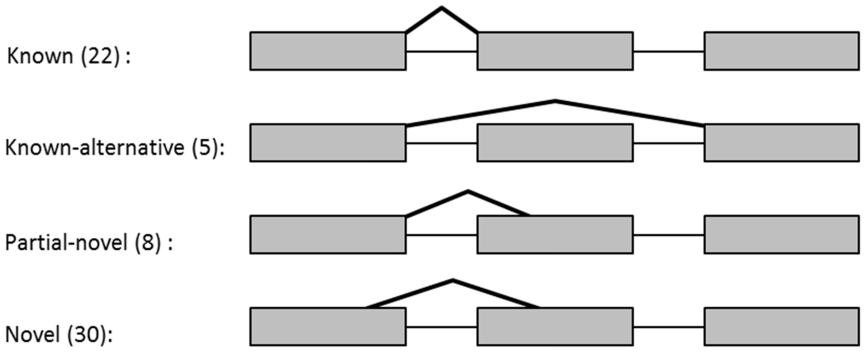


Figure 3.4 GABAB1 splicing junctions and gene boundaries.
After combining splicing junctions from the gene specific and the whole transcriptome library, we found that there are at least 65 splicing junctions associated with the GABAB1 gene. A. The 65 splice junctions were visualized with the UCSC genome browser using a .juncs file and hg18. From the splice junction information, we determined that the GABAB1 gene is at least 72,333 bp (chr6: 29,653,437 - 29,725,769) long which is about 41 Kbp larger than the expected size of 30,958 bp (chr6: 29,677,984 - 29,708,941) based on the RefSeq Genes model. From two databases of the UCSC genome browser (Human mRNAs from GeneBank and Human ESTs that Have Been Spliced), we estimated a minimum GABAB1 gene length of 100,266 bp (chr6: 29,631,387 - 29,731,652). From top to bottom the tracts include a user supplied track (.juncs file loaded track), RefSeq Genes, Human mRNAs from GeneBank and Human ESTs that Have been spliced. GABAB1 gene is at the center of RefSeq Genes tract, and its 5' end exons are located to the right side of 3' end exons. Red arrows point to clones that showed extended GABAB1 gene length from the two UCSC genome browser databases. B. The 65 splicing junctions were one of four types (known, known-alternative, partial-novel, and novel splicing junctions). Known" represents the known splicing junctions from the RefSeq Genes model. Known-alternative" represents splicing junctions that have different splicing site combinations of the known splicing junctions. Partial-novel" represents splicing junctions that have only one known splicing site and an unknown splicing site. Novel" represents splicing junctions that do not have any known splicing sites. The numbers in parentheses are counts of each type. The known splicing junctions are about one third of the other GABAB1 splicing junctions.

3.2.5 GABAB1 gene expression differences in alcoholic versus control subjects

We compared GABAB1 gene expression levels between 14 alcoholics and 15 controls (Figure 3.5). Reads per kilobase of exon model per million mapped reads (RPKM) is most commonly used for current RNA-seq expression studies and is calculated from all known exons (Mortazavi, Williams et al. 2008). However, we found that unknown splicing junctions were found at introns and intergenic regions and many reads were also mapped there. Reads per kilobase of gene model per million mapped reads (RPGM) was introduced to consider the intron and intergenic regions. RPGM allowed calculation of expression levels from each gene as well as any interesting region within gene model in genome wide study.

RPGMs were calculated from the three different GABAB1 gene boundaries: 1) known GABAB1 gene boundary from the RefSeq Genes model, 2) minimum GABAB1 gene boundary from RNA-seq, and 3) the other GABAB1 gene boundary from UCSC genome browser databases (Figure 3.5). However, the expression levels of these three gene boundaries did not significantly change in alcoholics compared to controls.

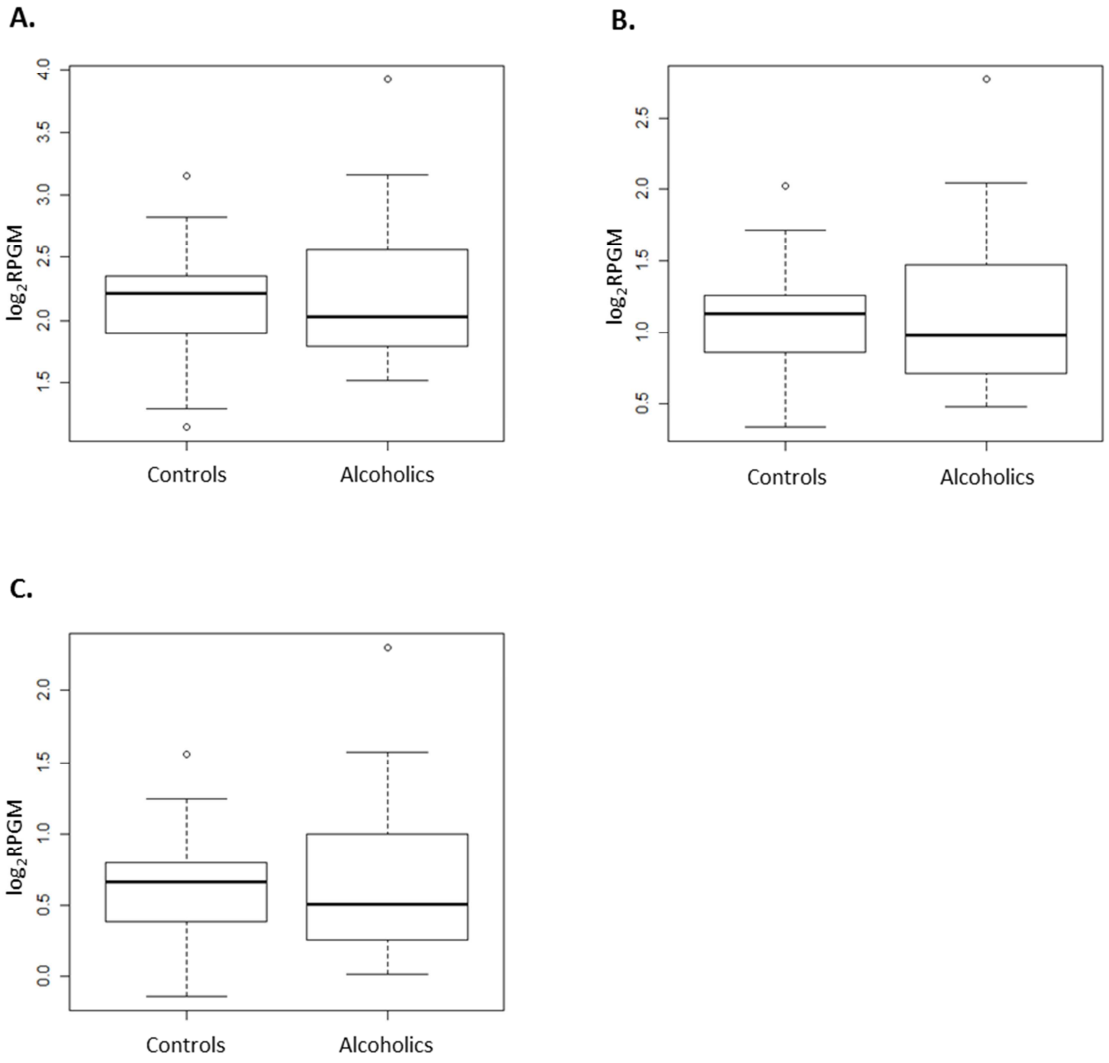


Figure 3.5 GABAB1 gene expression comparison in alcoholics and controls.
GABAB1 expression was compared in 14 alcoholics and 15 controls. A. At the known GABAB1 gene boundary of the RefSeq Genes model, expression levels were compared between alcoholics and controls. After log2 transformation of RPGM values, unpaired student's t-tests were carried out to define statistically significant differences. Y-axis represents log2RPGM. From top to bottom, horizontal lines of each boxplots represent the largest non-outlier, upper quartile, median, lower quartile, and the smallest non-outlier values. B. GABAB1 gene expression levels were compared based on the minimum GABAB1 gene boundary from RNA-seq data. C. Gene expression levels were assessed based on the minimum GABAB1 gene boundary obtained from the UCSC genome browser database.

To confirm GABAB1 gene expression level changes in alcoholics, RPGM values were calculated for all genes from RefSeq Genes model and analyzed using another statistical analysis, DEGseq package (Wang, Feng et al. 2010). Among differentially expressed gene in alcoholics, GABAB1 gene was not found. As shown in Figure 3.5A, the expression level at the known GABAB1 gene length, 30,958 bp, was not changed significantly.

DEGseq was developed for RPKM values to analyze transcript expression data. RPKM and RPGM generate different values for their expressions. To verify RPGM values are also good data input for DEGseq, we compared their transcripts expression data analyses using DEGseq (Figure 3.6). Both RPGM and RPKM showed similar MA plot data. They also found similar number of significantly changed transcripts in alcoholics. Their ratios to total transcripts were 0.74 % and 0.99 % at RPGM and RPKM (p-value <0.001). Based on p-values, most significantly changed transcripts were also conserved.

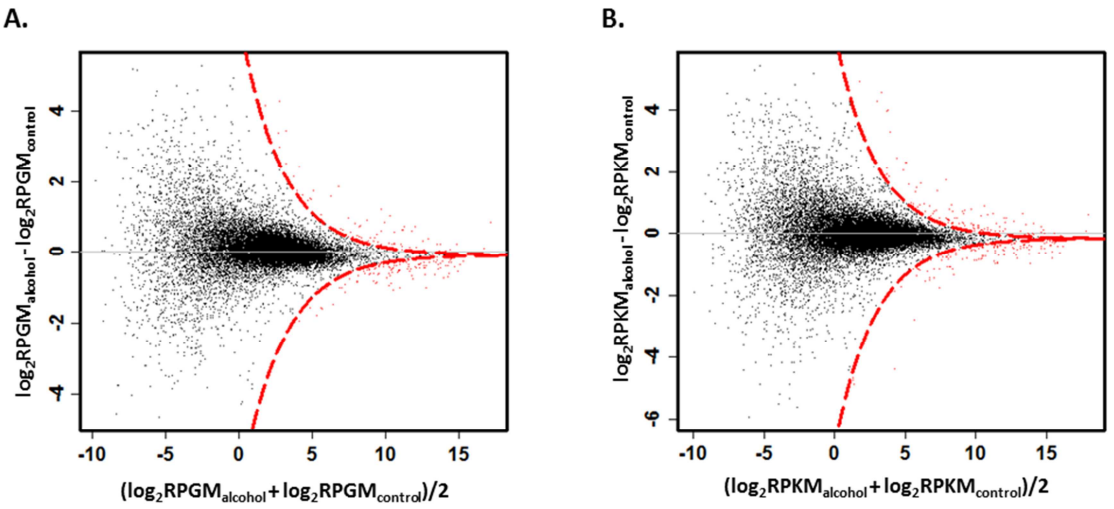


Figure 3.6 Genome wide transcripts expression change comparison in alcoholics using RPGM and RPKM values.

A. Genome wide transcripts expression changes were compared in alcoholics using RPGM. From RPGM values of all transcripts based on the RefSeq Genes model, we compared transcripts expression levels between alcoholics and controls. B. Based on RPKM, genome wide transcripts expression changes were also compared in alcoholics. Like RPGM calculation, RPKM values were calculated for all transcripts based on the RefSeq Genes model, and their expression changes in alcoholics were examined. Both RPGM and RPKM data showed similar patterns from MA plot analyses using DEGseq package. The significantly changed transcripts were 0.74 % and 0.99 % at A and B, respectively (p-values < 0.001). X-axes represent the averages of log_2RPGM or log_2RPKM values for alcoholic and control groups. Y-axes are for differences between the values. Red dots represent genes, exons, introns, or splicing junctions that changed in alcoholics, and the red dotted lines are p-value 0.001 lines.

3.2.6 GABAB1 exon and intron expression changes in alcoholics

We next asked if expression levels of individual exons and introns of the GABAB1 gene might be different in non-alcoholic and alcoholic brain. GABAB1 has 23 exons and 22 introns based on the RefSeq Genes model (Figure 3.4A). Among RPGM values of the exons and introns from all alcoholic and control samples, intron RPGM values were much smaller than exon expression levels although exon RPGM values vary among individual exons (Figure 3.7). RPGM values from exon 1 to 6 were much less than other exon expression levels. This indicates that GABAB1a splicing variant that contains all 23 exons cannot represent the overall GABAB1 gene expression though it is regarded as a major GABAB1 transcript and used for most GABAB1 functional studies (Schwarz, Barry et al. 2000; Lee, Mayfield et al. 2010). Various exon combinations can yield numerous short GABAB1 transcripts, and their combined expression levels possibly represent the overall GABAB1 expression.

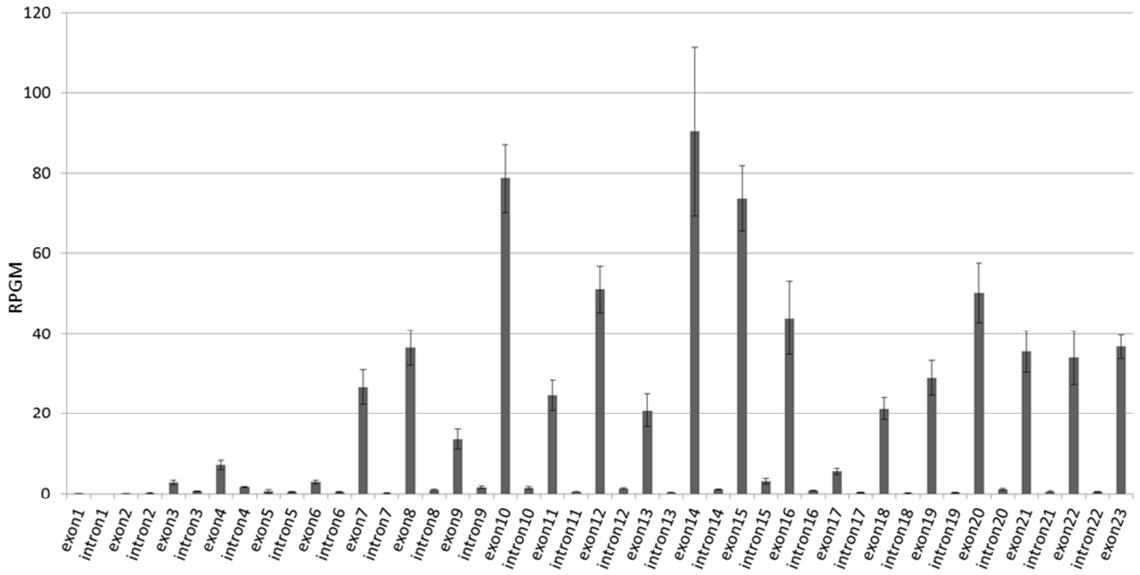


Figure 3.7 GABAB1 exon and intron expressions.
RPGM values were calculated for each GABAB1 exon and intron from15 alcoholic and 14 control samples. 5' end GABAB1 exons showed extremely lower expression and indicated the abundant existence of small splicing variants instead of splicing variants containing all 23 exons. Y-axis represents RPGM values. Data are expressed as the mean ± SEM.

For each GABAB1 exon and intron, we assessed expression level changes in alcoholics. From RPGMs for all exons and introns based on the RefSeq Genes model, differentially expressed exons and introns were identified in alcoholics using the DEGseq package (p-value < 0.05). Among them, GABAB1 data were selected (Table 3.2). Among 23 GABAB1 exons, the expression of 9 exons increased, and 2 exon expressions decreased in alcoholics. Based on the expression level and direction, the exons were grouped (exons 7 - 8, exon 10, exons 11 - 12, exons 14 - 16, and exon 17). Within groups, expression levels (Figure 3.7) and directions (Table 3.2, Figure 3.8) were similar. The groups were separated by discontinuous exon expression patterns.

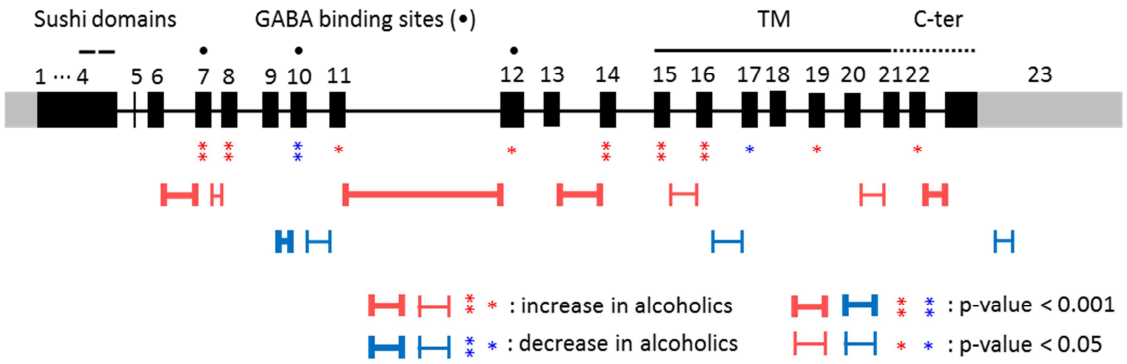


Figure 3.8 GABAB1 exon, intron, and splicing junction changes in alcoholics.
Exon and intron expression and splicing junction existence changes were compared between alcoholics and controls. Based on the RefSeq Genes model, RPGM values were calculated for all exons and introns. For all splicing junctions found from TopHat mapping, we calculated reads per kilobase of splicing junction model per million uniquely mapped reads (RPJMs). Among significant RPGM and RPJM changes, GABAB1 specific data were selected and summarized. The expressions of 11 exons were changed in alcoholics. In alcoholics, exon 10 and 17 expressions decreased, and expressions increased at 9 exons, e.g. exon 7, 8, 11, 12, 14, 15, 16, 19, and 22. Exon 7 - 8, exon 10, exon 11 - 12, exon 14 - 16, and exon 17 were grouped based on expression directions and levels at Figure 3.7. Most the splicing junctions well explained why exon expressions were changed in alcoholics. Two more splicing junction changes could not alter neighboring exon expressions and showed the complexity of GABAB1splicing. Red and blue stars indicate significant exon expression increases and decreases in alcoholics, respectively. Double stars indicate p-values below 0.001, and single stars represent p-values less than 0.05. Red and blue bars represent splicing junctions that significantly increased and decreased in alcoholics, respectively. Thick bars represent p-value below 0.001, and thin bars represent represented p-value lower than 0.05.

Exon / intron	Fold change	p-value	q-value(Benjamini et al. 1995)
exon 7	1.47	9.39E-08	5.38E-06
exon 8	1.38	2.30E-07	1.18E-05
exon 10	0.79	2.33E-08	1.55E-06
exon 11	1.27	1.68E-03	2.15E-02
exon 12	1.14	9.58E-03	8.28E-02
exon 14	1.31	2.99E-12	4.55E-10
exon 15	1.32	2.30E-10	2.43E-08
exon 16	1.47	9.02E-12	1.24E-09
exon 17	0.63	3.72E-03	4.02E-02
exon 19	1.22	3.76E-03	4.06E-02
exon 22	1.16	1.96E-02	1.41E-01

Table 3.2 GABAB1 exon and intron expression changes in alcoholics.

3.2.7 GABAB1 splicing junction changes in alcoholics

To study GABAB1 splice junction changes in alcoholics, we calculated RPJM for each junction. The RPJM calculation also considered the introns and intergenic regions because many unknown splicing junctions and mapped reads were often found at introns and intergenic regions.

Among RPJMs of all splice junctions from TopHat mappings, we identified junctions that significantly change in alcoholics (p-value < 0.05). Among these, 11 differentially expressed GABAB1 junctions were found (Figure 3.8). They were ten known and one novel splice junctions (Table 3.3).

Location	Type	Exon / intron	Fold change	p-value	q-value(Benjamini et al. 1995)	q-value(Storey et al. 2003)
chr6:29,699,781-29,703,241	Known	exon 6 - exon 7	1.71	5.42E-05	7.27E-04	3.67E-04
chr6:29,699,232-29,699,645	Known	exon 7 - exon 8	1.52	2.20E-03	1.39E-02	7.03E-03
chr6:29,697,574-29,697,859	Known	exon 9 - exon 10	0.71	6.90E-05	8.86E-04	4.48E-04
chr6:29,697,049-29,697,507	Known	exon 10 - exon 11	0.72	9.83E-03	4.26E-02	2.15E-02
chr6:29,689,242-29,696,856	Known	exon 11 - exon 12	1.36	1.76E-04	1.91E-03	9.65E-04
chr6:29,686,758-29,688,306	Known	exon 13 - exon 14	1.51	2.31E-14	2.73E-12	1.38E-12
chr6:29,684,490-29,684,984	Known	exon 15 - exon 16	1.29	3.31E-03	1.91E-02	9.64E-03
chr6:29,682,975-29,684,356	Known	exon 16 - exon 17	0.19	8.63E-03	3.88E-02	1.96E-02
chr6:29,680,745-29,681,324	Known	exon 20 - exon 21	6.89	8.20E-03	3.73E-02	1.89E-02
chr6:29,679,473-29,680,249	Known	exon 22 - exon 23	2.40	7.35E-05	9.34E-04	4.72E-04
chr6:29,679,072-29,679,179	Novel	exon 23	0.27	4.13E-02	1.20E-01	6.05E-02

Table 3.3 GABAB1 splicing junction existence changes in alcoholics.

The changes of splicing junctions could explain most exon and intron changes in alcoholics. Splicing junctions increased at exon 6 - 7 and exon 7 - 8 probably altered their neighboring exons 7 and 8 expression and explained why their expression increased. Splicing junctions decreased at exon 9 - 10 and exon 10 - 11 showing that exon 10 spliced in transcripts were less abundant in alcoholics. The exon 11 - 12 splicing junction increase potentially explained why exon 11 and 12 expression increased in alcoholics. Splicing junctions increased at exon 13 - 14 as suggesting expression increase of transcripts that contained exon 14. Lastly, splicing junctions increased at exon 15 - 16 and deceased at exon 16 - 17 indicating that the expression of transcripts containing exon 15 and 16 increased without exon 17 (Figure 3.8).

Changes in splicing junctions did not always change neighboring exon expression levels (Figure 3.8). A known junction at exon 20 - 21 and a novel splicing junction at exon 23 changed without affecting surrounding exon expression levels. Exon 20 and 21 were the third and fourth left end exons of GABAB1 at RefSeq Genes tract (Figure 3.4A). Between the two exons many unknown novel splicing junctions were loaded at User Supplied Track from our RNA-seq junction search step. They possibly diminished exon 20 and 21 expression though their splicing junction increased in alcoholics. A novel splicing junction spliced out 108 bp at the 3' UTR and decreased existence in alcoholics. This novel splicing junction was not as abundant as other known junctions (Pickrell, Pai et al. 2010; Sorber, Dimon et al. 2011), and its length was much shorter than 1489 bp exon 23. Therefore, this splicing junction could not change exon 23 expression level.

However, miRBase search (http://www.mirbase.org/) showed that the splicing junction is a possible target of two miRNAs, e.g. has-miR-3916 and has-miR-2060.

3.3 DISCUSSION

In this study, our first gene specific RNA-seq study demonstrated the power of its deep sequencing to detect unexpected rare splicing junctions and splicing complexity from a single gene, GABAB1. The gene specific library data were provided to query whole transcriptome library data for junctions and exon/intron usage changes in a human psychiatric disease. Specifically, we showed that chronic alcohol abuse alters the production of variants of the GABAB1 gene without altering the overall abundance of transcripts. Our results from the GABAB1 gene raise the possibility that chronic alcohol abuse may affect the splicing of many genes in brain and other tissues. This is supported by our preliminary genome wide analysis showing many changes in exons/introns and splicing junctions in the brain samples used for the analysis of GABAR1 (data not shown).

Exons were grouped based on similar expression patterns and large RPGM changes of their neighboring exons. These groups showed increased or decreased expression in alcoholics, and their RPGM changes were supported by RPJM changes. However, the expression changes seen with individual exon groups were apparently not large enough to change overall GABAB1 expression. Reasons for this include the possibility of transcript lengths much shorter than the whole GABAB1 gene and the

finding that different exons did not follow the same expression direction. For all 23 GABAB1 exons, 9 exons increased in alcoholics. 2 exons decreased, and 12 exons had no change. This variable expression pattern probably prevented an overall GABAB1 expression change in alcoholics.

Quantitative real-time PCR and microarray data validated the variable GABAB1 splicing changes. Our RNA-seq study showed that a known splicing junction at exon 22 - 23 increased RPJM value in alcoholics and a known-alternative splicing junction between exon 14 and 16 did not show RPJM change (Figure 3.8). Quantitative real-time PCR data also generated the same patterns as the RNA-seq splicing junctions and showed the variable splicing changes (data not shown). Previous microarray studies using cDNA and oligonucleotide probes also validated most of our RNA-seq data. Although the increase in one cDNA microarray probe signal at intron 4 was not found in our RNA-seq data, four probes targeting exon 23 and another probe for intron 5 showed no GABAB1 expression changes in agreement with our RNA-seq data (Flatscher-Bader, van der Brug et al. 2005; Liu, Lewohl et al. 2006; Ponomarev, Wang et al. 2011).

It is of interest to note the functional importance of exons 10 and 17 which showed decreased expressions in alcoholics. Exon 10 participates in the binding of GABA on the receptors (Couve, Moss et al. 2000; Lee, Mayfield et al. 2010). A loss of GABA binding sites would likely diminish the response of GABA, and signal transduction of the GABAB receptor would be less efficient. Exon 17 contains the entire TM 3 as well as intracellular and extracellular regions between neighboring TMs. G proteins bind to the intracellular region between TM 3 and 4 (Lee, Mayfield et al. 2010).

Without exon 17, GABAB receptors are not able to transfer signals through G proteins. At exon 23, alcoholics showed a RPJM decrease of a novel splicing junction that is a potential target for miRNAs at 3' UTR. Therefore, splicing out of the miRNA target site decreases resulting in more of this target site in alcoholics allowing the miRNAs to destabilize GABAB1 transcripts that contain exon 23. The exon 23 containing transcripts include the most major GABAB1 splicing variants as well as some short splicing variants. Because overall GABAB1 gene expression does not change in alcoholics, an increase in non-functional variants would decrease the fraction of functional GABAB1 receptors in alcoholic brain. Taken together, these results raise the possibility that altered splicing may reduce GABAB receptor function in the alcoholic brain and GABAB agonists may produce their beneficial effects in alcoholics because they enhance receptor signaling (Schweitzer, Roberto et al. 2004; Enserink 2011).

Most RNA-seq libraries have been constructed using random primers and oligo(dT) primers for reverse transcription. The random primer method generates sequencing data from all existing RNAs based on expression levels, and the oligo(dT) method provides mostly 3' UTR sequencing information for all individual transcripts.. To maximize splicing junction detection for the GABAB1 gene, we prepared a gene specific library using gene specific primers for reverse transcription. This reverse transcription step captured GABAB1 specific transcripts and provided high resolution data for GABAB1 gene specific splicing studies. Although it is possible to capture cDNA from specific genes by modifying current DNA capture techniques (Hodges, Xuan et al. 2007; Ng, Turner et al. 2009; Fisher, Barry et al. 2011), it requires large amounts of cDNAs for

RNA-seq library construction. Human postmortem samples often provide very limited RNA and may not allow use of the capture method for RNA-seq library construction.

To study gene expression and splicing junction existence changes, we employed two new strategies, RPGM and RPJM. From RNA-seq mapping data, we found numerous mapped reads at known intron regions of the RefSeq Genes model, and deep splicing junction searches using a gene specific library also identified many novel splicing junctions at known intronic and intergenic regions. RPKM, which is commonly used to study gene expression levels, is calculated based only on exon models and excludes the intron and intergenic regions (Mortazavi, Williams et al. 2008). For high resolution gene specific splicing study, we included sequence information from intronic and intergenic regions. Therefore, we modified the standard RPKM to consider these regions and defined new terminology, RPGM. Although RPKM is used for full-length transcript expression level calculations, RPGM can be used from gene level to small regions of the gene, such as exons and introns. If gene lengths are larger than the RefSeq Genes model as observed for GABAB1, RPGM can be calculated based on the new gene lengths. For RPJM calculation, we considered all of the exon, intron, and intergenic regions as modifying the RPGM calculation method. Previous research only identified splicing junctions without comparing splicing junctions quantitatively between groups (Pickrell, Pai et al. 2010; Sorber, Dimon et al. 2011). However, RPJM allows such quantitative comparisons among experimental groups.

Among 65 splicing junctions found in human brain GABAB1 transcripts, about 45 were rare and most were unknown junctions. Similar to this result, other studies have

also identified rare splicing from different genes and organisms (Pickrell, Pai et al. 2010; Sorber, Dimon et al. 2011). Splicing can provide new splicing variants with unique functions, e.g. down-regulating other major splicing variants (Mosley, Beckmann et al. 1989; Schwarz, Barry et al. 2000; Ezzat, Zheng et al. 2001; Bulanova, Budagian et al. 2007; Tiao, Bradaia et al. 2008; Lee, Mayfield et al. 2010), controlling transcript intracellular localization (Bell, Miyashiro et al. 2008; Jang, Park et al. 2010), or providing novel functions (Bell, Miyashiro et al. 2010; Gracheva, Cordero-Morales et al. 2011).

There are potential implications of changes in GABAB1 splicing and function in prefrontal cortex of alcoholics. First, optimal function of this brain region is required for decision making and regulation of reward pathways, and reduced prefrontal cortex function is linked with dependence on alcohol and other drugs (Piomelli 2001; Flatscher-Bader, van der Brug et al. 2005; O'Brien and Gardner 2005). Second, drugs acting on GABAB receptors have been used for treatment of alcoholism. In particular, baclofen, a GABAB agonist, may be effective at high doses (Enserink 2011) but was not beneficial at lower doses (Garbutt, Kampov-Polevoy et al. 2010). Thus, it is possible that remodeling of the GABAB receptor by chronic alcohol abuse may reduce the effectiveness of baclofen requiring higher than expected doses for successful treatment.

In summary, the complexity of GABAB1 receptor splicing and the perturbation of splicing by chronic alcohol abuse demonstrate the power of RNA-seq to provide new insight into gene expression and suggest that many other important brain genes may have unexplored splicing variants which will be important for alcoholism and other psychiatric diseases.

3.4 MATERIALS AND METHODS

3.4.1 rRNA depletion from total RNA

15 control and 14 alcoholic postmortem prefrontal cortices were obtained from the Tissue Resource Centre (TRC) at the University of Sydney in Australia (http://www.braindonors.org/). TRC was funded by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and provide brain tissues for alcoholic research. Prefrontal cortex has important roles in reward circuitry to develop alcoholism and is the primary site of the pharmacological action of alcohol drug abuse (Alexander and Brown 2011; Yizhar, Fenno et al. 2011). Samples were matched by gender, age, brain pH, and post-mortem interval (PMI) as closely as possible. Detailed sample information was described at our previous microarray research that used the same samples (Ponomarev, Wang et al. 2011). After RNA extraction, DNase (Ambion, Austin, TX, USA) was treating to their RNAs and commercial human prefrontal cortex total RNAs (Ambion, Austin, TX, USA). rRNAs were depleted using RiboMinus kit (Invitrogen, Carlsbad, CA, USA).

3.4.2 RNA-seq library construction and sequencing

To study GABAB1 specific splicing pattern changes, GABAB1 splicing junction identification was maximized using gene specific libraries prepared from the rRNA depleted commercial and postmortem prefrontal cortex total RNAs. During library

constructions, we used the commercial RNAs as internal standards. Unlike other common libraries that used random primers or oligo(dT) primers for reverse transcription, the gene specific library used gene specific primers. Because most known GABAB1 splicing variants share a common 3' UTR, the gene specific primers were designed at 3’ UTR of human GABAB1. The primers were designed using Primer3 (http://frodo.wi.mit.edu/primer3/) and tested for their ability to bind the GABAB1 gene specifically with PrimerSelect (DNASTAR, Inc., Madison, WI, USA). One gene specific primer, GABBR1-C-ter-GSP-2, 5’ - AGAGACACCACAGTGTGAAAGG - 3’, was optimized for reverse transcription at 42 °C using SMARTer PCR cDNA Synthesis Kit (Clontech, Mountain View, CA, USA). Using a template switching method, dscDNAs were generated. For the amplification step, SMARTer2A-22mer, 5’ - AAGCAGTGGTATCAACGCAGAG - 3’ was designed based on the provided sequence of a primer, SMARTer II A Oligonucleotide, 5’ - AAGCAGTGGTATCAACGCAGAGTACXXXXX - 3’, which was bound at the 5’ end of transcripts during template switching step (X is a closed base in the proprietary sequence). After amplification with SMARTer2A-22mer and another gene specific primer, GABBR1-C-ter-GSP-1, 5’ - AGGTCCATCTGTCTATCCCAAC - 3’, dscDNA generation was confirmed with gel electrophoresis. Using quantitative real-time PCR, an amplification cycle that had the least non-specific PCR products was selected among cycles generating maximum GABAB1 specific PCR products. The amplification cycles were from 25 to 35 cycles, and they vary whenever gene specific libraries were prepared.

Gel extraction removed unbound primers including small amplification products. Following the SOLiD™ 3 System Library Preparation Guide (Applied Biosystems, Carlsbad, CA, USA), the remaining dscDNAs were fragmented into about 200 bp, and gene specific libraries were prepared.

The 3’ end exon of the GABAB1 gene is a long 3' UTR. The exon length is about 1315 bp and about one third of the longest known transcript, GABAB1a. Because the gene specific library has a 3' bias, alternative gene specific primers were also designed at the 5’ end of the 3’ UTR to find more splicing junctions from ORF. For the alternative gene specific library, the primers for reverse transcription and amplification were substituted to GB1-Cter-1st-4, 5' - TCCCAGAGGTATGAG - 3', and GB1-Cter-2nd-1, 5' - CTACTGGCCTGTCCTCCCTCA - 3', instead of GABBR1-C-ter-GSP-2 and GABBR1-C-ter-GSP-1, respectively.

For whole transcriptome libraries, we used 15 control and 14 alcoholic postmortem prefrontal cortex total RNAs and commercial human prefrontal cortex total RNA. The whole transcriptome libraries were prepared using SOLiD™ Total RNA-Seq Kit (Applied Biosystems, Foster City, CA, USA) according to the manufacturer’s protocol.

The gene specific and whole transcriptome libraries were sequenced at the Genomic Sequencing and Analysis Facility (GSAF) at the University of Texas at Austin using SOLiD system. For whole transcriptome libraries, some sequencing runs were repeated if they generated less than 10 million raw reads. For a few whole transcriptome

libraries, sequencing runs were maximized and generated approximately 100 million reads. The all reads that we used for our study were 50 mer single end reads.

3.4.3 Comparison of gene specific and whole transcriptome libraries

Total raw reads were filtered by sequencing quality. For every read, the number of quality values that were below 8 was counted. If the numbers were larger than 13 for a 50 mer read, the reads were filtered out. For 35 mer reads, the reads were removed if the numbers were over 9. Non-coding RNAs (mostly rRNAs and transfer RNAs (tRNAs)) and adaptors were filtered out using Mapreads mapper (Applied Biosystems, Carlsbad, CA, USA). Two mismatches out of 20 mer reads were allowed for the mapper. TopHat, a splicing junction mapper, was used to detect splicing junctions (Trapnell, Pachter et al. 2009). Most TopHat applications use .fq files as input files, and its test data set for installation was also the files as sequence inputs and we also used .fq files for our studies. To generate .fq files, our filtered sequencing data were converted to Illumina raw data format, .fastq files, as changing color base to base space. We modified to .fq files using fq_all2std.pl program (http://maq.sourceforge.net/fq_all2std.pl) (Trapnell, Pachter et al. 2009). During format conversion, we removed reads that contain quality values, -1. For each sample, mapping was performed against hg18 using default option of TopHat. To visualize the mapping data, .sam files were modified using igvtools (http://www.broadinstitute.org/software/igv/download). The modified files were visualized using IGV (http://www.broadinstitute.org/software/igv/) including the other TopHat outputs, .wig and .bed files.

3.4.4 Detection of splicing junctions

To maximize splicing junction mining from individual sample mappings, the data process used to compare gene specific and whole transcriptome libraries was modified, and splicing junction data were subjected to TopHat mapping. To prepare splicing junction data, we added splicing junction search step before the final mapping. The step is individual TopHat mapping for the gene specific libraries and the whole transcriptome libraries from all commercial and postmortem brain samples. After filtering based on quality values and non-desired sequences (non-coding RNA and adaptor), their read formats were converted to .fq files. After the splicing junction search step, we collected splicing junctions from .bed files of the individual mapping results. To use only unique splicing junctions for next mapping, the .bed files were combined and converted to a .juncs file. Using the .juncs file, whole transcriptome library data of 14 alcoholics and 15 controls were remapped.

After recollecting splicing junction data from the second mapping, we generated a new .juncs file that contained unique splicing junction information. For visualization of the file, the splicing junction data was loaded at User Supplied Track of the UCSC genome browser. To compare our RNA-seq splicing junction data with other information at UCSC genome browser, RefSeq Genes, Human mRNAs from GeneBank, and Human ESTs That Have Been Spliced were also visualized. New minimum GABAB1 gene lengths were estimated based on the RNA-seq splicing junction data and mapped reads as well as the UCSC genome browser data.

3.4.5 Expression analyses using RPGM

Using supercomputers at the Texas Advanced Computing Center (TACC) of the University of Texas at Austin, RPGMs were calculated for gene, exon, and introns based on the formula below:

$$RPGM = \frac{mapped\ read\ \#}{\frac{total\ mapped\ read\ \#}{1{,}000{,}000} \times \frac{gene\ length\ (bp)}{1{,}000}}$$

For the three different GABAB1 gene lengths, mapped reads were counted from 14 alcoholic and 15 control whole transcriptome libraries. RPGMs were calculated by dividing the mapped reads by the total million unique mapped reads to whole human genome and by kilobase gene length. After $\log_2$ transformation of the RPGM values, their differences in alcoholics were compared with two-tailed unpaired Student's t-tests using R.

We also calculated RPGMs for all genes, exons, and introns based on RefSeq Genes model. The RPGM value differences between alcoholics and controls were tested using MARS of DEGseq R package. For multiple testing correction, the package calculated p-values adjusted to q-values by two alternative strategies (Wang, Feng et al. 2010). GABAB1 specific data were selected from the analyses.

To identify GABAB1 exon and intron expression level differences, RPGM values of the exons and introns were also calculated from 14 alcoholic and 15 control samples. Their data were compared based on mean ± SEM.

3.4.6 Splicing junctions in alcoholic samples

To study GABAB1 splicing junction changes in alcoholics, we calculated RPJMs for all splicing junctions found from TopHat mapping using the supercomputers at TACC. The below formula was used for the RPJM calculation.

$$RPJM = \frac{mapped\ read\ \#}{\frac{total\ mapped\ read\ \#}{1{,}000{,}000} \times \frac{splicing\ junction\ area\ length\ (bp)}{1{,}000}}$$

Mapped reads were counted for each splicing junction from whole transcriptome libraries of 14 alcoholics and 15 controls. The mapped reads were normalized per million unique mapped reads (total mapped read #) of the whole human reference genome. The reads were also divided by kilobase of splicing junction area length where mapped reads can reach the splicing junction. The splicing junction area length is (read length -1) × 2. Therefore, the splicing junction area length for our 50 mer read was 98.

After calculating RPJM for individual samples, RPJM value differences between alcoholic and control groups were also analyzed using DEGseq package. The analysis did multiple test correction as adjusting p-values using two alternative q-values. Among splicing junctions that significantly change in alcoholics, GABAB1 splicing junctions were selected if they were found from multiple samples.

CHAPTER 4. USE NEW RNA-SEQ LIBRARY CONSTRUCTION TO SOLVE RNASEIII AND T4PNK SEQUENCE BIASES

4.1 INTRODUCTION

Next generation sequencing is revolutionizing biological data acquisition. It can be used instead of many existing specialized measurement approaches. For example, one of the next generation sequencing techniques, RNA-seq, may ultimately replace microarrays. Microarrays detect gene expression levels based on known probes. By contrast, RNA-seq does not have this limitation and can measure expression levels at any region. In addition, it also provides information about SNP and the location of splicing junctions (Trapnell, Pachter et al. 2009; Canovas, Rincon et al. 2010; Shen, Catchen et al. 2011).

However, one disadvantage of RNA-seq is that its sequence data can be biased during library construction. For example, mapped read counts can over or underestimate true RNA abundances based on a variety of library construction steps, such as RT, adaptor ligation, or amplification (Romaniuk, McLaughlin et al. 1982; Hansen, Brenner et al. 2010; Metzker 2010; Willerth, Pedro et al. 2010). In a comparison of multiple library preparation methods for RNA-seq, all methods introduced their own bias and showed different expression patterns from the same sample (Linsen, de Wit et al. 2009). Here, we developed three new RNA-seq library preparation methods and compared them

with the current RNA-seq library construction method for the Applied Biosystems SOLiD system. We identified the step in library preparation that caused the most pronounced bias and outlined alternative preparation techniques that can virtually eliminate this concern.

4.2 RESULTS

4.2.1 RNA-seq library construction

RNA-seq library construction methods vary among sequencing instruments. ABI SOLiD, Illumina HiSeq, and Roche 454 are three major RNA-seq sequencers. They have their own library preparation methods, and other companies also developed library construction methods for the sequencers. Furthermore, customized methods were designed based on researchers' interests, for example, RNA-editing or poly(A) tail studies (Li, Levanon et al. 2009; Morabito, Ulbricht et al. 2010; Yoon and Brem 2010).

In previous studies, random primers or oligo(dT) primers were used for RT of library preparation (Cloonan, Forrest et al. 2008; Hansen, Brenner et al. 2010; Yoon and Brem 2010). However, we used gene specific primers to develop a gene specific library for GABAB1 (Figure 4.1). The gene specific primers were designed for the 3' UTR since most human GABAB1 transcripts share a common 3' UTR (Lee, Mayfield et al. 2010). They were used for both RT and amplification (Figure 4.1). For our study, the gene specific library and a whole transcriptome library were prepared for SOLiD system and sequenced.

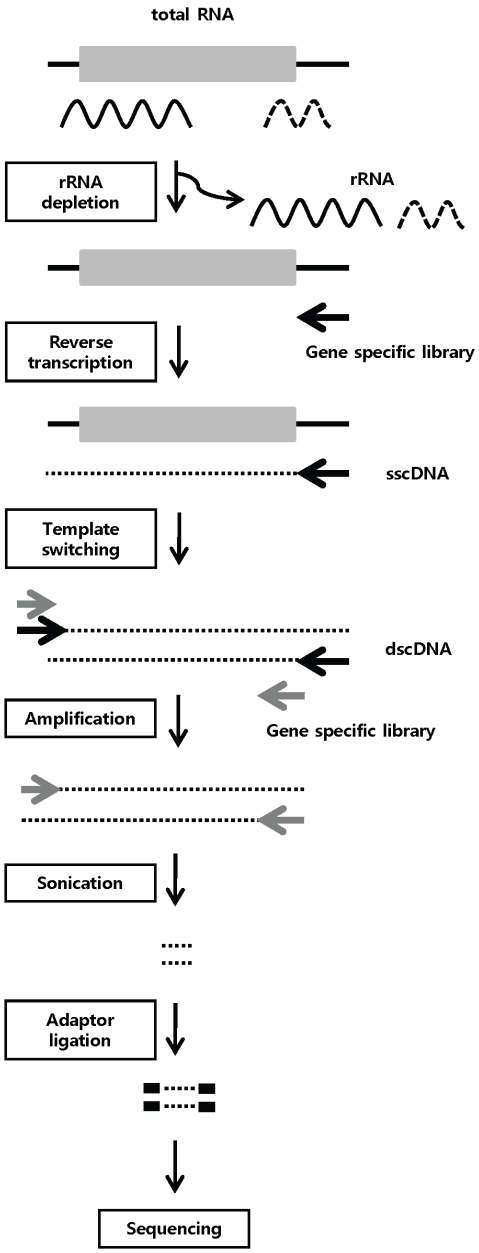


Figure 4.1 Preparation of gene specific library.
mRNA is visualized as a grey box flanked by black lines. The grey box represents the ORF, and the black lines represent UTRs. We obtained total RNA and depleted rRNA from it. RT was carried out using a GABAB1 gene specific primer. The resulting sscDNAs were converted to dscDNAs via template switching. After amplification with another gene specific primer, dscDNAs were sonicated. Adaptor ligated dscDNAs were sequenced with the SOLiD system.

4.2.2 Sequence mapping and visualization

SOLiD system generates color space raw data that are expressed by numbers. Because two sequential bases are encoded by one number, the color space data was converted to base space sequence data for sequence analysis. Before converting the format, raw sequence reads were filtered by sequence quality values. Non-coding RNAs and adaptor sequences were also removed. The non-coding RNAs were primarily rRNAs and tRNAs. The filtered sequences were converted from color base .csfasta files to base space .fastq files that are used as Illumina sequence data format. After further file modification, the sequence data were mapped and visualized (Figure 4.2).

Figure 4.2A shows the mapped reads that result from the gene specific library. Figure 4.2B illustrates the mapped reads resulting from the whole transcriptome library for the same chromosome location. The mapping patterns between these two libraries were dramatically different. Even though the chromosome region was a known exon and should be covered with continuous reads, the mapped reads of the whole transcriptome library showed gaps and pile-ups. (We use the term *pile-ups* to refer to a substantial number of reads having exactly the same sequence as seen in Figure 4.2B.) This uneven coverage of exons with mapped reads demonstrates that the whole transcriptome library has substantial bias.

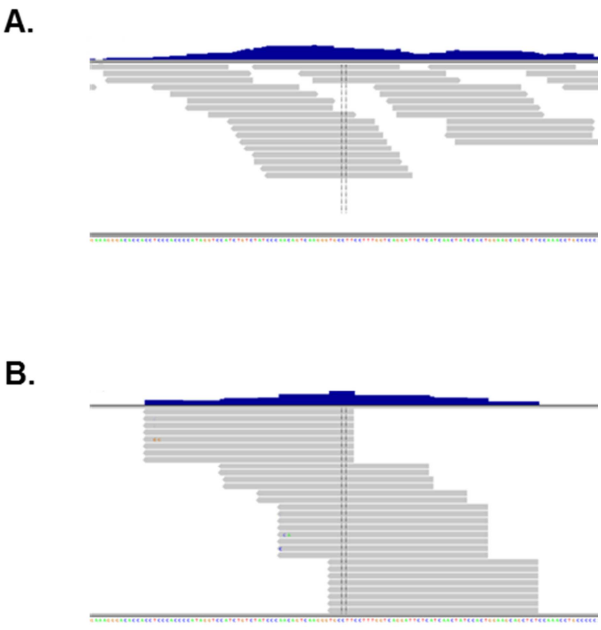


Figure 4.2 Mapping patterns of the gene specific library and the whole transcriptome library.
A. Mapped reads for the gene specific library were visualized at IGV. B. For the same genomic region as in A, mapped reads for the whole transcriptome library were visualized. Compared to the gene specific library, the whole transcriptome library showed read pile-ups as well as gaps. At the gaps, no read was mapped at a known exon even though the gene specific library reads were mapped to these sites.

4.2.3 Comparison of the two sequencing data sets

The pile-ups shown in Figure 4.2B might be the result of sequence bias at the 5' end of the read. To examine this possible sequence bias, we collected all mapped reads and calculated sequence logos (using WebLogo) and entropy at each nucleotide position (Figure 4.3). The height of letters in the sequence logo is based on entropy. Thus, our entropy profiles mimic the sequence logo profiles. We found that the 5’ ends of reads from the whole transcriptome library have a specific sequence pattern. The sequence logo results showed a predominant pattern of AA at the 5’ end of the whole transcriptome library reads (Figure 4.3A). In contrast, reads from the gene specific library did not display this sequence pattern (Figure 4.3B). The strong bias in the whole transcriptome library was reflected in reduced entropy at the 5’ end of reads (Figure 4.3C). In contrast, entropy for the gene specific library showed only a minor fluctuation without the 5’ end drop (Figure 4.3D). For controls of each library, we generated random exome sequences *in silico* based on mapped read locations. These controls showed almost no bias whatsoever (Figure 4.3, bottom rows in A, B, C, and D). The 5’ end sequence bias of the whole transcriptome library was also confirmed from GC ratio and nucleotide frequency calculations at each nucleotide position (data not shown).

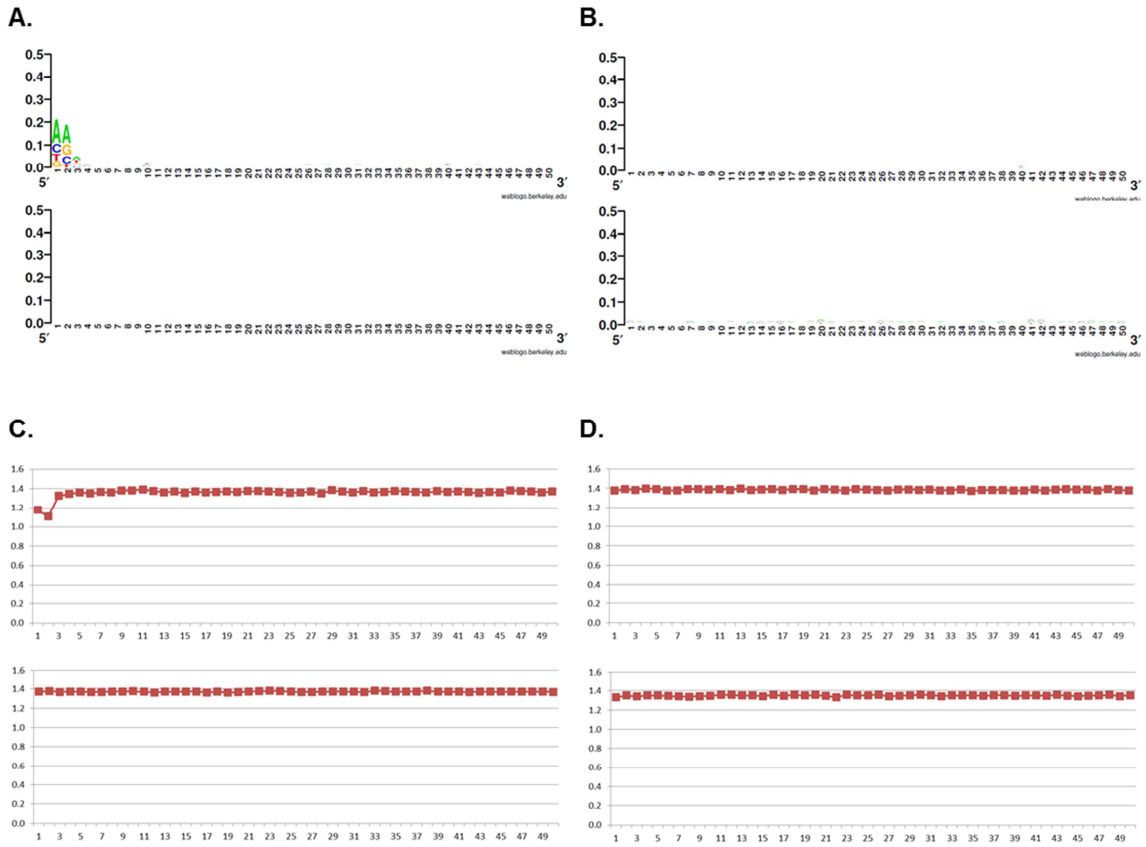


Figure 4.3 Sequence logo and entropy analysis of mapped reads.
We analyzed mapped reads for deviations from randomness using sequence logo and entropy. (Note that the height of letters in the sequence logo is given by the reduction in entropy from random expectation. Thus, large letters in the sequence logo correspond to depressions in the entropy numbers.) Subgraphs, A and C, show sequence logo and entropy data for the whole transcriptome library, respectively. Subgraphs, B and D, show the same quantities for the gene specific library. In each subgraph, the top graph shows actual results, and the bottom graph shows a computational control obtained by randomly generated sequences from human exome based on mapped read locations of each library. In all cases, we analyzed a window of 50 bp starting with the 5' ends of the sequence fragments. The gene specific library showed almost no deviations from randomness. By contrast, the whole transcriptome library had a strong bias at the first 2 bases at the 5' end.

4.2.4 Comparison of the two library preparation methods

The gene specific library and the whole transcriptome library were constructed very differently (Figure 4.4). To identify the source of the bias, we carefully reviewed all steps at which library construction differed. Most importantly, the bias could have arisen from fragmentation or its following steps. For the gene specific library, dscDNA fragmentation was performed with sonication, and adaptors were ligated to the fragmented dscDNA using DNA ligase (Figure 4.4A). Sonication and DNA ligase possibly introduced the extremely minor noise (Figure 4.3B, D). However, these two steps were not known to cause sequence bias, and reads in our experiment were not sufficiently biased to introduce read pile-ups and mapping gaps (Figure 4.2A). By contrast, the whole transcriptome library construction method used RNaseIII fragmentation at the very beginning of the protocol (Figure 4.4B). Perhaps RNaseIII fragmentation causes a bias that has not been previously evaluated. However, the three steps following RNaseIII fragmentation (adaptor ligation, RT, and amplification) (Figure 4.4B) were already known to have sequence biases (Romaniuk, McLaughlin et al. 1982; Hansen, Brenner et al. 2010; Willerth, Pedro et al. 2010). Therefore, we first inspected these three steps for the potential problems.

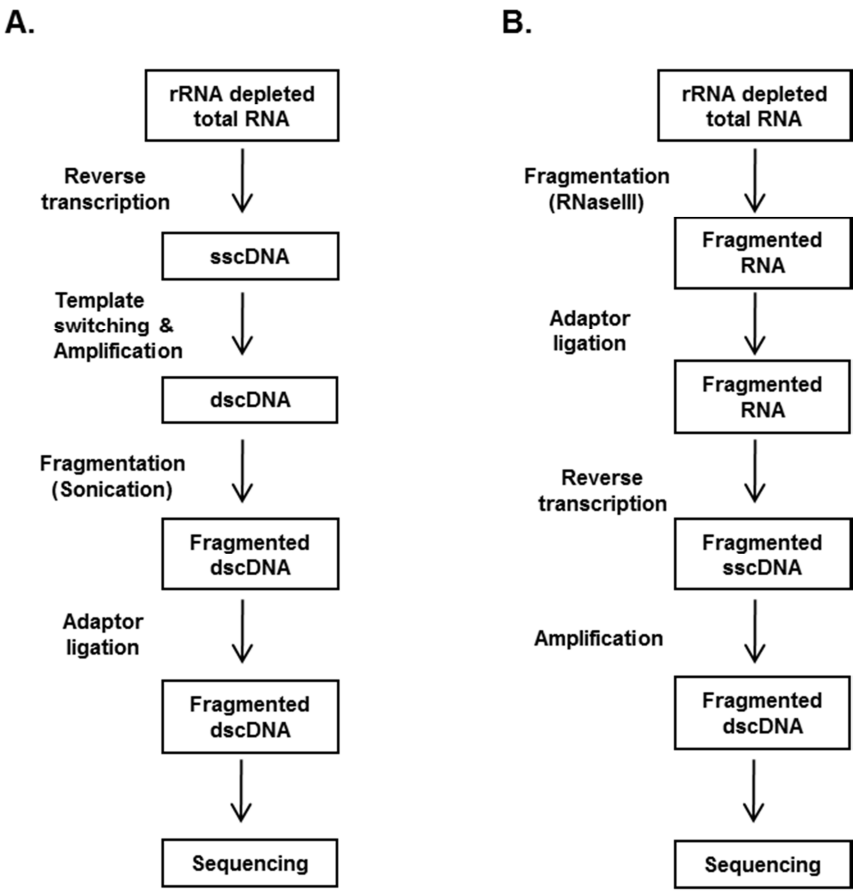


Figure 4.4 Simplified construction methods of the gene specific library and the whole transcriptome library.
A. The construction method of the gene specific library was simplified from Figure 4.1.
B. The whole transcriptome library construction method was simplified.

We used RNA ligase for adaptor ligation. RNA ligase prefers to bind to certain phosphate donors, but preference patterns have been studied only for short oligomers (Hansen, Brenner et al. 2010). In our case, phosphate donors were the fragmented single stranded RNAs (ssRNAs) of 200 bp in length. Figure 4.3A did not show the same sequence pattern as the short oligomer case (Hansen, Brenner et al. 2010). Therefore, its sequence bias either appeared in a different way at the fragmented ssRNAs or was too small to be detected from whole transcriptome library data. We concluded that the sequence bias we saw was likely not caused by RNA ligase.

The bias for RT has been previously studied for the Illumina RNA-seq system (Hansen, Brenner et al. 2010). The randomly primed RT of Illumina RNA-seq causes substantial sequence biases at the 5' and 3' ends. However, our RT method was different from the Illumina method, and their respective biases are likely different. While the random primers of Illumina RNA-seq can hybridize on any locations of mRNAs, the primers in our method are the overhangs of adaptors and can bind only to ends of fragmented ssRNAs. The overhangs are 4 bp random sequence and are used for RT. Because of this targeted priming, the sequence bias of the whole transcriptome library was totally different from the Illumina random primer case (Figure 4.3A) (Hansen, Brenner et al. 2010). Therefore, we concluded that it was unlikely that RT was a major source of bias in our library preparation.

Finally, amplification is known to have GC bias (Willerth, Pedro et al. 2010). This bias could affect some transcript regions that might appear more or less highly expressed, but this could not generate 5' end sequence bias.

Therefore, we excluded the three steps (adaptor ligation, RT, and amplification) as major sources of 5' end sequence bias of the whole transcriptome library. We next focused our efforts on RNaseIII. RNaseIII can digest ssRNA of preribosomal RNA and bacteriophage T4 (Dunn 1976; Elela, Igel et al. 1996). However, we removed all ribosomal RNAs and used only human RNA samples. Though we used only about 200 bp fragmented RNAs, their ssRNA digestion products were not the length. During eukaryotic dsRNA metabolism, RNaseIII is generally considered to be a random cutter when it digests RNAs into 200 bp. It recognizes double stranded RNA structures and cleaves them (Elela, Igel et al. 1996). Cleavage usually produces both short and long fragments from RNA. In yeast, the short fragments are about 28 - 32 bp in length and have unique sequence (AGNN) in the middle as containing hairpin structures (Lamontagne, Larose et al. 2001). In our system, the long fragments were around 200bp and used for RNA-seq. The short fragments were too short for RNA-seq.

Because RNaseIII specifically recognizes RNA secondary structure and leaves the unique sequence in the short fragments, it is not likely to be a perfect random cutter. Therefore, the long fragments could have an RNaseIII specific sequence pattern as well. On the basis of this reason, we replaced RNaseIII fragmentation by an alternative method using heat.

4.2.5 Alternative library construction

We fragmented RNA at 95 °C when it was denatured. In denatured RNA, the hydrogen bonds are broken, and only covalent bonds remain. Denatured RNA has no

secondary structure. Therefore, each base can be attacked equally by heat, and we don't expect that heat fragmentation can introduce 5' end bias.

However, heat fragmented RNA could not be used directly for adaptor ligation. The 5' and 3' ends of the heat fragmented RNA needed to be modified. We used T4PNK for the modification (Figure 4.5B). Its kinase activity adds phosphate groups at 5' fragment ends. Its phosphatase activity removes phosphate from 3' fragment ends and leaves hydroxyl groups.

However, T4PNK is known to have sequence bias in its kinase activity according to the OptiKinase product information of Amersham Biosciences (http://www.gelifesciences.com/aptrix/upp00919.nsf/Content/2B2B93BBD26B478FC1257628001D0D0C/$file/63005432.pdf). Therefore, we also tested an alternative preparation method where we applied OptiKinase before T4PNK treatment (Figure 4.5C). OptiKinase has been shown not to have the sequence bias of T4PNK, but it does not have phosphatase activity at the 3' end. Here, we assumed that T4PNK has negligible bias in its phosphatase activity and used it following OptiKinase treatment for its phosphatase activity. (Our results showed that this assumption is valid, as shown below.) Note that for kinase activity, T4PNK treatment required the optimization for enzyme tittering and treatment time. Therefore, OptiKinase was clearly a better choice for future RNA-seq studies.

To assess the sequence bias of these two newly proposed heat fragmentation methods, we prepared two new whole transcriptome libraries, together with a second iteration of the whole transcriptome library prepared according to the standard protocol

based on RNaseIII (Figure 4.5A). To reduce batch-to-batch sequencing variation, we sequenced all three libraries simultaneously in the same SOLiD system.

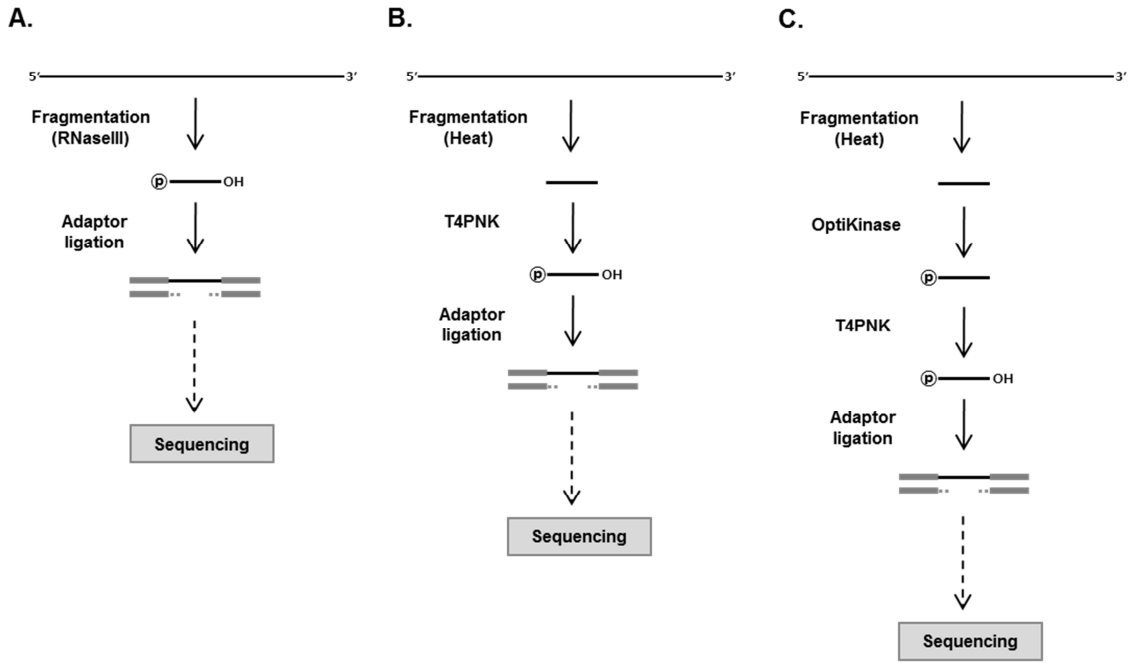


Figure 4.5 Alternative fragmentation methods.
A. RNaseIII was used for the current fragmentation method. B. Heat fragmentation method was used as an alternative fragmentation method. After heat fragmentation, T4PNK was applied to modify RNA fragment ends for adaptor ligation. T4PNK adds phosphate and hydroxyl groups at their 5' and 3' ends, respectively. C. Modified heat fragmentation method was designed using OptiKinase. OptiKinase reduces T4PNK bias introduced during the RNA fragment modification. Because OptiKinase has only kinase activity, T4PNK was used for phosphatase activity after OptiKinase treatment.

4.2.6 Mapping results of novel libraries

We processed paired-end reads instead of previous single-end reads and visualized both 5' and 3' read mapping results (Figure 4.6). Ctl refers to the library prepared with RNaseIII. Heat refers to the library generated with heat fragmentation and T4PNK. Heat + OptiK refers to the library that was generated with an additional OptiKinase step right after the heat fragmentation.

As described before for Figure 4.2B, read pile-up is a indicator of the fragment end sequence bias. Ctl showed clear pile-ups for 5' read mapping results as expected (Figure 4.6A). By contrast, Heat had much fewer pile-ups, and Heat + OptiK had virtually no pile-ups. Also, Heat + OptiK had the most even and smoothly connected distribution of the mapped reads.

Figure 4.6B shows mapping results for 3' reads. Ctl showed pile-up patterns just like in the 5' case. The pile-up patterns were weaker using Heat and largely disappeared using Heat + OptiK.

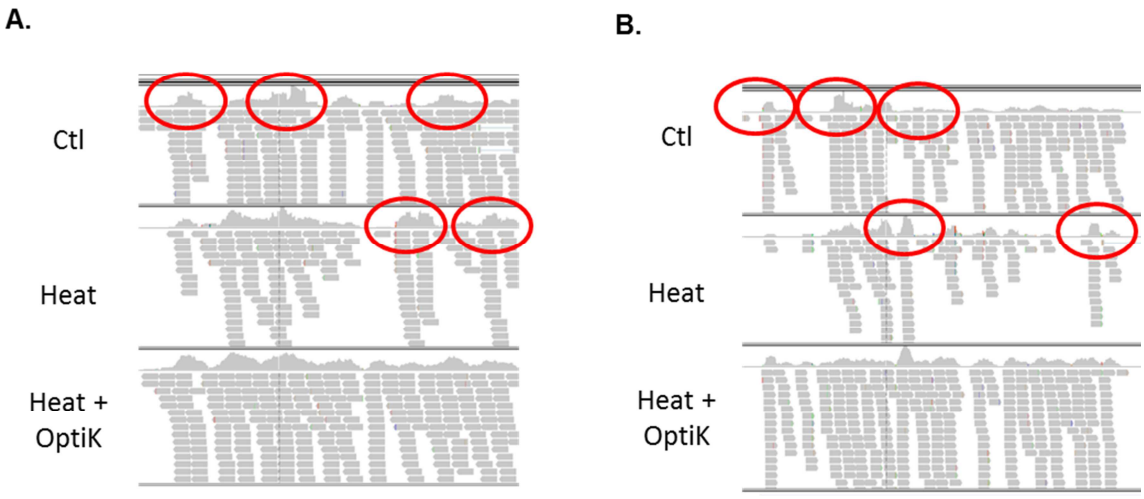


Figure 4.6 Mapping patterns of the three whole transcriptome libraries.
A. 5' read mapping results of the three whole transcriptome libraries were visualized at IGV. Ctl refers to the library constructed using RNaseIII. Heat refers to the library constructed using the heat fragmentation method. Heat + OptiK refers to the library constructed using the modified heat fragmentation method. The pile-ups of duplicated reads are indicators for fragment end sequence bias. The red circles represent the pile-ups. Ctl showed pile-ups. Heat had fewer numbers of pile-ups, and no pile-ups were found at Heat + OptiK. B. 3' read mapping results show the same patterns as the 5' read mapping.

4.2.7 Identifying the detailed sequence bias patterns of the three libraries, Ctl, Heat, and Heat + OptiK

Next, we carried out sequence logo and entropy analyses for the three libraries (Figure 4.7). Figure 4.7A and C represent results for 5' reads, and Figure 4.7B and D represent results for 3' reads.

As described before for Figure 4.3A, Ctl had a strong bias of AA at the 5’ end of reads and showed minor bias at the other positions (Figure 4.7A). Heat had less bias right at the 5’ end but had some bias throughout the entire 50 bp window. T4PNK is biased though its bias is somewhat less than RNaseIII bias at the 5’ end. Finally, Heat + OptiK had almost no bias throughout the entire 50 bp window. Even though its negligible bias shares the same sequences with Heat, OptiKinase decreases the T4PNK sequence bias. For 5’ reads, Heat + OptiK was the least biased method overall.

For 3' reads, biases of Ctl were generally more severe than for the 5’ reads (Figure 4.3B), but the overall bias pattern followed that of the 5’ reads. Ctl showed the strongest bias at the beginning of reads with a preferred sequence of GGTNTA. Heat was biased towards sequences starting with A or T and had small bias ubiquitously throughout the sequence. The bias pattern was very different from the one found in Ctl. Thus, we believe that this bias was caused by the phosphatase activity of T4PNK. Most enzymes acting on nucleic acids are generally affected by nucleotide sequences near their reaction sites (Romaniuk, McLaughlin et al. 1982). Surprisingly, Heat + OptiK showed a much smaller sequence bias for 3' reads than Heat. Thus, it seems that pretreatment with OptiKinase

weakens the bias of T4PNK phosphatase activity because T4PNK mostly has phosphatase activity rather than kinase activity.

As expected, the entropy results generally mirrored the sequence logo results (Figure 4.7C, D). The entropy plots show clearly that the entropy values for Heat + OptiK fall consistently above the entropy values for Ctl and Heat for both 5' and 3' end reads. Therefore, Heat + OptiK is the least biased method for both 5’ and 3’ end reads.

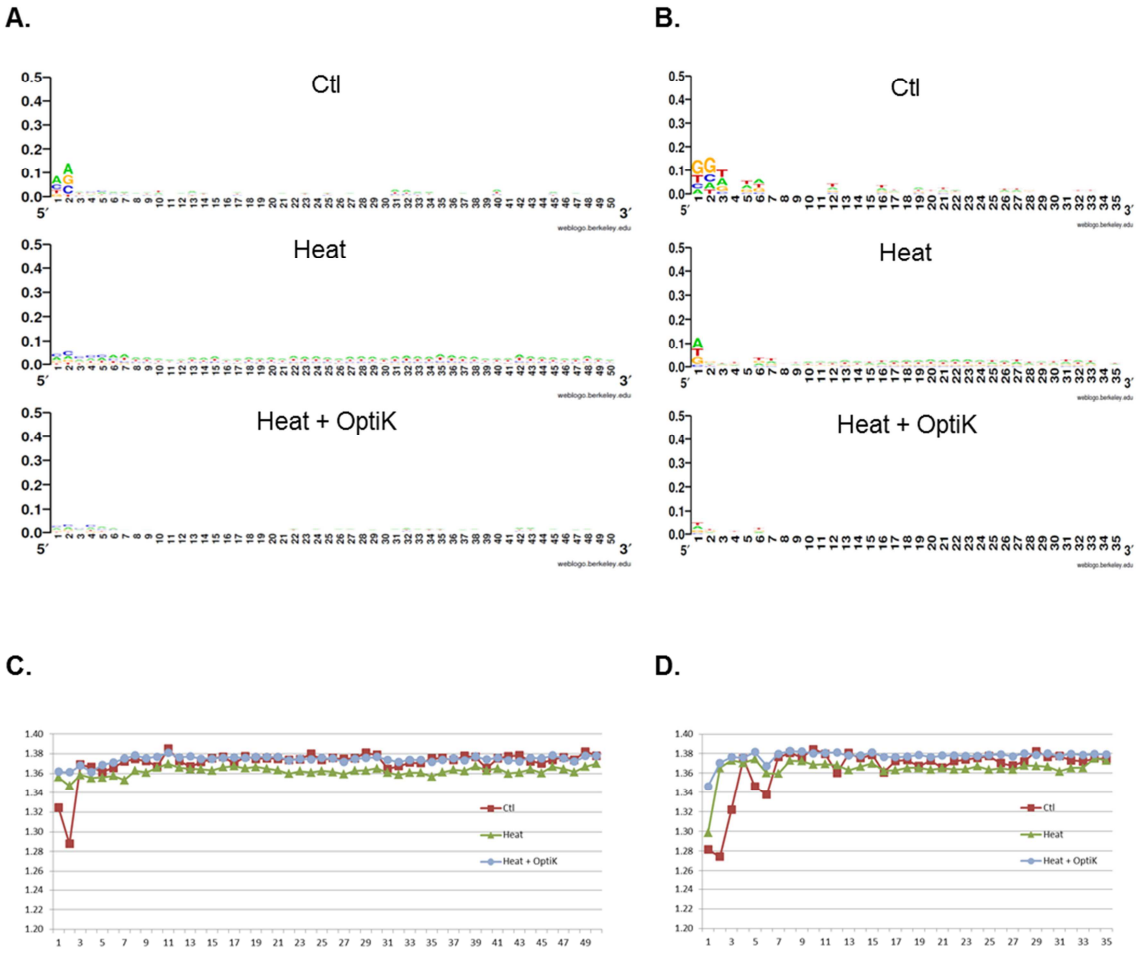


Figure 4.7 Sequence logo and entropy analysis of mapped reads.
A. We compared sequence logo results for 10,000 randomly selected mapped 5' reads of Ctl, Heat, and Heat + OptiK. B. Sequence logo results of 3' reads were also compared as in A. Ctl showed strong biases near the beginning of reads both for 5' and 3' reads. Heat had less biases at 5' and 3' reads but showed persistent low level biases at all nucleotide positions. The biases did not have the same sequence bias patterns as Ctl. Therefore, RNaseIII generates sequence biases at both 5' and 3' ends. The 3' end sequence bias, A or T, indicated the sequence bias of T4PNK phosphatase activity. Though Heat + OptiK had a similar sequence bias pattern as Heat, it was much smaller throughout the entire nucleotide positions. Thus, OptiKinase weakens T4PNK sequence biases, and Heat + OptiK is the best method. C and D show entropy results for 5' and 3' end reads, respectively. Lower entropy values correspond to greater sequence bias.

As a result of our study, we have obtained the sequence biases of RNaseIII and T4PNK. T4PNK kinase activity has a minor bias at the 5’ end of fragmented ssRNA. T4PNK phosphatase activity is biased mostly towards sequences with an A or T nucleotide at the fragment end of 3’ reads. Its sequence biases make less severe pile-ups than the broad RNaseIII sequence biases (Figure 4.6B).

RNaseIII has biases at both the 5' and 3' ends (summarized in Figure 4.8A). Its sequence pattern was mostly 5' - AA··· ···TANACC - 3' (We reverse complemented the 3’ sequence here). Based on the known RNaseIII reaction mechanism, the two 5’ and 3’ biased sequences were separated by the short fragmented RNAs (Figure 4.8B). Thus, we have shown that RNaseIII introduces sequence biases into the long fragments that it produces.

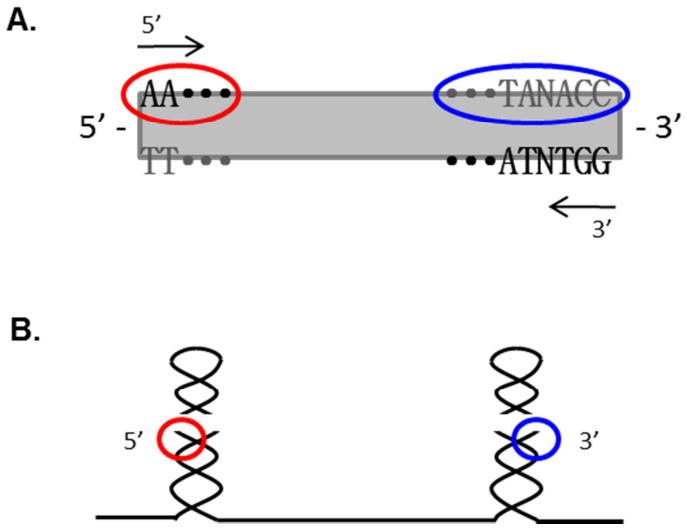


Figure 4.8 RNaseIII specific cutting sites.
A. Sequence biases of RNaseIII were summarized. RNaseIII specific cutting sites were mostly 5' - AA··· ···TANACC - 3'. (The 3' end reads are shown as reverse complement.) The red circle represents the 5' end cutting sites, and the blue circle represents the 3' end cutting site. B. The figure shows how these cutting sites map onto the helical regions cut by RNaseIII. The sites are disconnected by short fragment RNAs. The RNaseIII specific cutting sites are at the long fragment ends.

4.3 DISCUSSION

We have found that the standard ABI RNA-seq protocol, SOLiD™ Total RNA-Seq Kit, has substantial sequence biases and pile-ups in mapped reads. The biases could be tracked to RNaseIII activity which preferentially cuts at specific locations. Other steps in library preparation, specifically RNA ligation, RT, and amplification seemed to be free of bias in the RNA-seq preparation method. By changing fragmentation from RNaseIII to heat, the sequence bias was largely eliminated, and the vast majority of the sequence pile-ups disappeared. Therefore, heat fragmentation was superior to commercial RNaseIII methods when preparing RNA-seq.

The pile-ups and gaps produced by the sequence biases cause erroneous results in the identification of SNPs and splicing junctions (which might be missed) and in the measurement of expression levels. Even though overall gene expression levels and finding were probably not significantly affected by the pile-ups and gaps, a more precise expression analysis, e.g. a comparison of expression patterns among short exons, could be significantly affected by these biases. Also, genome wide splicing junction detection numbers were not significantly different among the three different library preparations (data not shown). During RNA-seq study, most genes were not highly expressed, and only few reads were mapped the genes. Therefore, read pile-up and gap cannot be issue at the less expressed gene and genome wide junction finding probably could not show significant difference. Also, whole transcriptome library could not show any failure to find splicing junctions from GABAB1 and GAPDH gene (data not shown). GABAB1 gene is strongly expressed in brains, and GAPDH gene is an endogenous control for

molecular biology experiments to measure RNA and protein. Therefore, splicing junction finding generally was not affected by the pile-up and gap issues except extremely high expressed genes. The issues probably can mostly fail in SNP finding or precise gene expression calculation.

Our RNA-seq data showed the power of next generation sequencing to find unexpected sequence biases from enzymes and revealed previously unknown enzyme reaction sites. A huge population of RNA-seq reads allowed us to identify previously unknown sequence biases and targeting sites of two enzymes, RNaseIII and T4PNK.

RNaseIII was previously considered to be a random cutter for dsRNAs. However, our study has shown that it indeed has preferred cutting sites. RNaseIII recognizes RNA secondary structure and digests the middle of double stranded RNA (Elela, Igel et al. 1996; Lamontagne, Larose et al. 2001). Yeast RNaseIIIs recognizes the AGNN loop of the double stranded RNA (Lamontagne, Larose et al. 2001). Thus, it is not surprising that its cutting site should be non-random even if it had never before been identified. Our high resolution study of RNA-seq found that RNaseIII caused pile-up mapping at RNA-seq and had specific digesting sites to provide fragment ends with conserved sequences. Even though one recent study had identified conserved sequences at 5' and 3' end reads from several RNA-seq library construction methods (Roberts, Trapnell et al. 2011), it did not identify RNaseIII as the source of the sequence bias. It also did not provide a solution to remove the bias during library preparation. Importantly, we have been able to propose RNA-seq library construction methods that solve the bias issue and result in precise expression analysis.

Similarly, whether T4PNK phosphatase activity had a sequence bias was not previously known even though the bias in its kinase activity was known for short oligomers. We found a bias in its phosphatase activity and a unique minor bias in its kinase activity at ssRNA fragments. However, the amount of bias was strongly reduced when we treated RNA fragments with OptiKinase first. Apparently, when T4PNK is used for both kinase and phosphatase activities, its phosphatase activity is more strongly biased than when it is used only for phosphatase activity.

Among the various methods of library preparation we studied, Heat + OptiK was overall the least biased. However, while vastly improved over the manufacturer's protocol, it is not a completely unbiased method. Entropy calculations and sequence logos showed small amounts of bias near the beginning of both 5' and 3' reads. Clearly, some of the remaining biases at fragment ends seem to be caused by T4PNK. The origins of any other minor deviations from complete randomness are not clear. They may be due to any combination of the other steps in library preparation, i.e. RNA ligation, RT, PCR amplification, and so on. Unfortunately, it is impossible to exclude all of the possibly biased steps from RNA-seq library preparation due to current technical limitations. Therefore, the individual steps should be optimized or substituted for less biased steps.

If all RNaseIII families share the same cutting sites, our identification of the RNaseIII specific cutting sites may provide useful applications in the prediction of miRNA genes. Drosha and Dicer belong to the RNaseIII family and produce miRNAs (Nam, Shin et al. 2005). Drosha recognizes hairpin structures in primary transcripts and produces pre-miRNAs after digestion in the nucleus. Dicer detects the hairpin structures

of the pre-miRNAs and carries out one more cleavage. Current prediction of miRNA genes is mostly based on RNA secondary structure. The mature miRNA sequences are mostly predicted by more than one hairpin precursor locus (Griffiths-Jones, Saini et al. 2008). However, prediction accuracy tends to be low (Nam, Shin et al. 2005). If the RNaseIII specific cutting sites that we have identified are conserved for other members of the RNaseIII family including Drosha and Dicer, the RNaseIII specific cutting sites can be incorporated into methods of miRNA gene prediction and improve the prediction accuracy.

4.4 MATERIALS AND METHODS

4.4.1 Whole transcriptome library construction

Total RNA from human prefrontal cortex was obtained from Ambion (Austin, TX, USA). RNA quality was evaluated using the Agilent 2100 Bioanalyzer RNA Nano Chip (Agilent Technologies, Santa Clara, CA, USA). After treating samples with DNase (Ambion, Austin, TX, USA), rRNA was depleted from total RNA using the RiboMinus kit (Invitrogen, Carlsbad, CA, USA). rRNA depletion was verified by the Agilent 2100 Bioanalyzer RNA Nano Chip. Whole transcriptome libraries were prepared using the SOLiD™ Total RNA-Seq Kit (Applied Biosystems, Carlsbad, CA, USA) according to the manufacturer's protocol.

4.4.2 Gene specific library construction

To prepare a gene specific library, gene specific primers for RT and amplification were designed at the 3' UTR of human GABAB1. The primers were designed using Primer3 (http://frodo.wi.mit.edu/primer3/) and tested for binding specificity with PrimerSelect (DNASTAR, Inc., Madison, WI, USA). After treating the human prefrontal cortex total RNA with DNase, GABBR1-C-ter-GSP-2, 5' - AGAGACACCACAGTGTGAAAGG - 3', was used for RT with the SMARTer PCR cDNA Synthesis Kit (Clontech, Mountain View, CA, USA). Using a template switching method, dscDNA was generated (Cloonan, Forrest et al. 2008). For the amplification step, SMARTer2A-22mer, 5' - AAGCAGTGGTATCAACGCAGAG - 3', was designed based on the sequence of a provided primer, SMARTer 2 A Oligonucleotide, that was bound at the 5' ends of transcripts during the template switching step. After amplification with SMARTer2A-22mer and GABBR1-C-ter-GSP-1, 5' - AGGTCCATCTGTCTATCCCAAC - 3', the dscDNA generation was confirmed with gel electrophoresis.

From the gel electrophoresis data, we optimized the number of amplification cycle based on band intensity. As soon as the band intensity reached maximum, an optimal amplification cycle was selected. The chosen cycle had the lowest level of non-specific PCR products among the amplification cycles containing maximum levels of GABAB1 gene specific PCR products. The optimal amplification cycle was verified with quantitative real-time PCR experiments using TaqMan® Gene Expression Assays (Applied Biosystems, Carlsbad, CA, USA). Based on a previous GABAB1 study (Lee,

Mayfield et al. 2010), the primer and probe set targeting the exon 22 - 23 junction of GABAB1 was used to measure gene specific amplification level. 18S and *GUSB* were used for endogenous controls. For each amplification cycle, gene specific amplification levels were calculated from both endogenous controls. Among amplification cycles, a cycle was selected as soon as both gene specific amplification levels reached plateau. The selected cycle was the same as the optimal amplification cycle chosen from the gel electrophoresis data.

Gel extraction removed small amplification products and the primers for RT, template switching, and amplification. According to the SOLiD™ 3 System Library Preparation Guide (Applied Biosystems, Carlsbad, CA, USA), the remaining dscDNA was sonicated to about 200bp length, and the gene specific library was prepared for SOLiD sequencing.

4.4.3 Sequence mapping and visualization

The libraries were sequenced using the SOLiD system, and TopHat was used for sequence mapping (Trapnell, Pachter et al. 2009). During mapping, reads that had low quality values were difficult to map and increased total mapping time. To eliminate these issues, we set up quality filter parameters to remove the reads. For every read, we counted the number of quality values that were below 8 (higher than 15% probability of incorrect base call). If the numbers were larger than 13 at 50 mer sequence reads, the reads were filtered out. For 35 mer sequence reads, reads were removed if the numbers were over 9. Non-coding RNAs (mostly rRNA and tRNA) and adaptors were filtered out

using the Mapreads mapper (Applied Biosystems, Carlsbad, CA, USA). Two mismatches out of 20 mer reads were allowed for the non-coding RNAs and adaptors mappings. After filtering, the remaining reads were converted to the .fastq file format. For TopHat mapping, the .fastq files were also converted to .fq files using fq_all2std.pl (http://maq.sourceforge.net/fq_all2std.pl). To visualize the mapping data, .sam files were processed by igvtools (http://www.broadinstitute.org/software/igv/download) and visualized with the IGV (http://www.broadinstitute.org/igv/).

4.4.4 Sequence analysis

All mapped reads were selected from the .fq files. From all the mapped reads we calculated entropy ($H = -\sum_{i}^{A,T,G,C} P_i \log P_i$) at each nucleotide position. Here, P_A, P_T, P_G, and P_C represent the relative abundances of each of the 4 bases (A, T, G, and C). Among them, a random sample of 10,000 reads was used as input for sequence logo (a graphical representation of a conserved sequence pattern; http://WebLogo.berkeley.edu) (Crooks, Hon et al. 2004). For the gene specific library, we removed reads that contained the primer sequences used for amplification, and the remaining mapped reads were used for entropy calculation and sequence logo. We used the same number of reads as the remaining reads for the whole transcriptome library. As controls, for each library we computationally generated random exome sequences based on mapped read locations.

4.4.5 Heat fragmentation library construction

We evaluated two alternative heat fragmentation methods. Under the first method, 300 ng rRNA depleted total RNA was added into nuclease free water (Ambion, Austin, TX, USA) up to 4.5 ul. Using a PCR machine, the RNA containing 0.2 ml tube was incubated at 95 °C for 80 minutes. Fragmented RNAs were verified using the Agilent 2100 Bioanalyzer RNA Nano Chip. The size of RNA fragments was around 200bp. For the second method, we performed heat fragmentation in a buffer containing divalent ions following a previous study (Cloonan, Forrest et al. 2008) and other on-line protocols (http://grcf.jhmi.edu/hts/protocols/mRNA-Seq_SamplePrep_1004898_D.pdf; http://www.neb.com/nebecomm/ManualFiles/manualE6114.pdf). Its incubation time was shorter (10 minutes) than in the nuclease free water. Agilent 2100 Bioanalyzer RNA Nano Chip data verified that the second fragmentation method produced RNA fragments of comparable size to the ones produced by the first fragmentation method. The incubation time for both methods was optimized based on PCR machines, incubation temperature, solutions, RNA sources, and tube shapes.

After heat fragmentation in a buffer containing divalent ions, RNA ends had to be modified to allow adaptor ligation. We used T4PNK (USB, Cleveland, OH, USA) to add phosphate groups to 5’ ends and hydroxyl groups to 3’ ends. Adaptor ligation and subsequent steps of library construction were identical to the manufacturer’s protocol of the SOLiD™ Total RNA-Seq Kit (Applied Biosystems, Carlsbad, CA, USA).

4.4.6 Updated heat fragmentation library construction

To minimize known sequence bias of T4PNK, we introduced an additional treatment step to the heat fragmentation library construction. After heat fragmentation, we used OptiKinase (USB, Cleveland, OH, USA) to add phosphate groups to the 5’ ends of fragmented RNAs. Subsequently, we proceeded with T4PNK as before to leave hydroxyl groups to the 3’ ends of fragmented RNAs.

CHAPTER 5. DISCUSSION

5.1 SUMMARY

GABAB receptors are composed of two subunits, GABAB1 and GABAB2. Splicing provides GABAB1 with structural and functional diversity. cDNA microarrays showed strong signals from human prefrontal cortex using GABAB1 intron 4 region probes, suggesting the existence of novel splice variants. Based on the probe sequence analysis, we proposed two possible splice variants, GABAB1j and GABAB1k. The existence of human GABAB1j was verified by quantitative real-time PCR, and mouse GABAB1j was found from a microarray probe set based on the human GABAB1j sequence. GABAB1j ORFs and expression patterns were not conserved across species, and they did not contain any important functional domains except sushi domains. We next focused our attention on GABAB1k.

After obtaining PCR evidence for the presence of GABAB1k in human, mouse, and rat, it was cloned from human and mouse by PCR along with three additional isoforms, GABAB1l, GABAB1m, and GABAB1n. Their expression levels measured by quantitative real-time PCR were relatively low in brain although they may be expressed more robustly in specific cell types. GABAB1l and GABAB1m inhibited GABAB receptor induced GIRK currents in *Xenopus* oocytes measured using the two-electrode voltage clamp system.

This study supports previous suggestions that intron 4 of the GABAB1 gene is a frequent splicing spot shared across species. Like GABAB1e, GABAB1l and GABAB1m

do not have TMs but have a dimerization motif. So, they also could be secreted and bind GABAB2 dominantly instead of GABAB1a. Only GABAB1l and GABAB1m are N- and C-terminal truncated splicing variants, and they impaired receptor function. This suggests that the intron 4 containing N-terminal truncation is necessary for the inhibitory action of the new splice variants.

Given the complexity of these GABAB1 splicing variants, we sequenced gene specific libraries to maximize identification of GABAB1 specific splicing junctions. We obtained splicing junction data from sequencing data of gene specific and whole transcriptome libraries from human prefrontal cortex. We found that the GABAB1 gene is at least 2 to 3 times longer than the known GABAB1 gene.

Mapped reads and splicing junctions were frequently found from intronic and intergenic regions of the RefSeq Genes model. To consider these regions, we introduced new strategies, RPGM and RPJM, to calculate gene expression and splicing junction levels, respectively. GABAB1 exon and intron RPGM data showed extremely low expression at the 5’ end exons and exon grouping pattern. This indicated that there are other short splicing variants besides GABAB1a, the longest known major transcript containing all exons.

Previous microarray studies have shown increased expression of GABAB1 in alcoholic brains although this was dependent upon the probes used to measure expression. Chronic alcohol also changed expression of specific GABAB1 splicing variants. In our current studies, chronic alcohol altered expression of exons/introns and splicing junctions in human brain more than whole transcriptome expression including

GABAB1 gene expression. Most of the splicing junction changes could be explained based on changes in exons and introns.

Decreased exon expression at a GABA biding site, a TM, and a miRNA binding site probably decrease the population of normal GABAB1 transcripts and thereby decrease normal signal transduction. This suggests that the therapeutic benefit of GABAB receptor agonists used in treating alcoholism may be due to a normalization of the impaired GABAB function produced by splicing in alcoholic brains.

In these studies, we found GABAB1 splicing complexity and change in alcoholic brains. Its splicing complexity allowed us to clone novel intron 4 containing GABAB1 splicing variants that impair GABAB receptor function and to find many rare GABAB1 splicing junctions using gene specific RNA-seq approaches. Because chronic alcohol treatment targets GABAB1 splicing rather than gene expression, RNA-seq study found splicing junction and exon expression changes in alcoholic brains without gene expression changes. Additionally, our RNA-seq study showed the power of next generation sequencing using a huge population of reads. It found rare splicing junctions from GABAB1 gene and unexpected RNA-seq sequence biases including previously unknown enzyme reaction sites.

During this RNA-seq study, we noticed that the whole transcriptome library displayed numerous read duplications (pile-ups) and gaps in known exons. The pile-ups and gaps of the whole transcriptome library caused a loss of SNP and splicing junction information and reduced the quality of gene expression results. Further, we found clear sequence biases for both 5’ and 3’ end reads in the whole transcriptome library.

To remove the biases, RNaseIII fragmentation was replaced with heat fragmentation. For adaptor ligation, T4PNK was used following heat fragmentation. However, its kinase and phosphatase activities introduced additional sequence biases. To minimize them, we used OptiKinase before T4PNK. Our results suggest that the heat fragmentation removed the RNaseIII sequence bias and significantly reduced the pile-ups and gaps. OptiKinase minimized the T4PNK sequence biases and removed all remaining pile-ups and gaps thus maximizing the quality of RNA-seq data.

In addition to the sequence bias solutions, our study further revealed the specific target sequences of RNaseIII and T4PNK. This is the first application of RNA-seq to discover unknown enzyme target sequences. The identification of RNaseIII target sequences could improve miRNA gene prediction accuracy if other members of the RNaseIII family, including Drosha and Dicer, also have specific target sequences.

5.2 FUTURE DIRECTIONS

5.2.1 Further studies of intron 4 containing splicing variant

5.2.1.1 Further characterization of GABAB1k, l, m, and n

To initially characterize novel GABAB1 splicing variants, we performed some electrophysiological experiments with four of the splice variants. Though additional GABAB1 splicing variant functions were predicted based on domain homology, their functions have not been studied. Therefore, further characterization studies are warranted.

GABAB1l and m do not contain TM and C-terminal intracellular domains but contain the dimerization motif (Figure 5.1A). Because they have the same C-terminal pattern as GABAB1e (Schwarz, Barry et al. 2000), they are probably secreted and thus inhibit the function of other major splicing variants (Mosley, Beckmann et al. 1989; Schwarz, Barry et al. 2000; Ezzat, Zheng et al. 2001; Bulanova, Budagian et al. 2007; Tiao, Bradaia et al. 2008) (Figure 5.1B). Further studies must be conducted to determine their cellular localization and function. For example, additional electrophysiology experiments should be performed using purified GABAB1l and m. To have inhibitory function in cytosol, their dimerization motif would probably dimerize with GABAB2 in a dominant manner as reported for GABAB1e (Schwarz, Barry et al. 2000). Co-immunoprecipitation (Co-IP) experiments could be performed to confirm the binding.

Like GABAB1l and m, GABAB1k and n also need further characterization to find their functions. GABAB1k has exon 15 as a common splicing pattern in CNS (Schwarz, Barry et al. 2000) and possibly mediates a CNS specific function (Figure 5.1A). Though GABAB1n has only a short extracellular region of GABAB1a, it contains one GABAB binding site and a part of the dimerization motif (Figure 5.1A). Therefore, it might be able to inhibit other major GABAB1 splicing variants like GABAB1l and m.

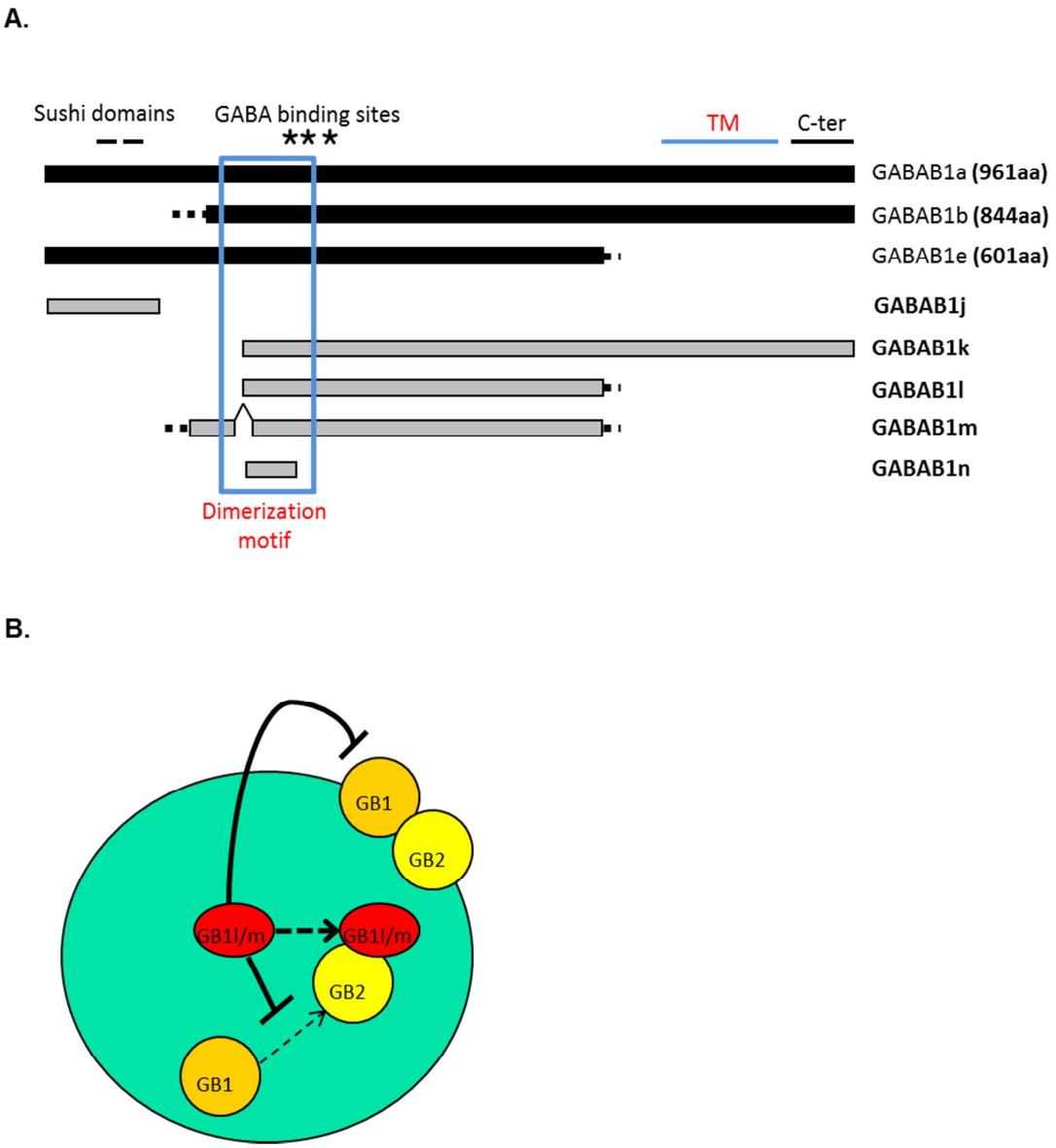


Figure 5.1 GABAB1 splicing variant domains and function.
A. Functional domains of GABAB1 splicing variants. Thick black lines represent known splicing variants, and thick grey bars represent intron 4 containing splicing variants. TM and dimerization motif are emphasized by light blue lines. B. Predicted mechanisms of GABAB1l and m (GB1l/m). Their secreted portion probably inhibits the function of major GABAB1 splicing variants (GB1). In the intracellular region, GB1l/m will dominantly bind to the GABAB2 subunit (GB2) excluding GB1.

GABAB1k, l, m, and n are good candidates for alcohol targets, and their alcohol effects needs to be studied. Alcohol actions on the known major GABAB1 isoforms (GABAB1a and b) are controversial (Dzitoyeva, Dimitrijevic et al. 2003; Roberto, Madamba et al. 2003), and we could not find an acute alcohol effect on GABAB1a mediated currents in oocytes (Figure A.1). However, GABAB receptor agonists are used in the treatment for alcoholism and alter acute behavioral and electrophysiological effects of alcohol (Allan, Burnett et al. 1991; Schweitzer, Roberto et al. 2004; Maccioni, Fantini et al. 2008; Maccioni, Pes et al. 2008; Enserink 2011). The previous microarray data showed expression increase at intron 4. Therefore, alcohol effects of the intron 4 containing isoforms, GABAB1k, l, m, and n, should be examined using the two electrode voltage clamp system in *Xenopus* oocytes to find alcohol targets.

5.2.1.2 Unknown intron 4 containing splicing variant studies based on GABAB1 splicing complexity

RNA-seq and quantitative real-time PCR studies did not duplicate some of the previous cDNA microarray data. Though intron 4 targeting probes of the previous microarrays showed expression increases in alcoholic prefrontal cortices, RNA-seq experiment could not detect the expression increase at intron 4 (Figure 3.8). For quantitative real-time PCR experiments, GB2 was designed based on the microarray probe. However, it did not show increased expression at intron 4 (Figure A.2). Technical differences in the methods combined with the GABAB1 splicing complexity could result in differences in expression levels. RNA-seq and quantitative real-time PCR detect only

short portions (about 50 bp) from highly amplified short cDNA amplicons. However, the cDNA microarrays had about 1 Kbp detection resolution. Their experiments were designed based on hybridization between cDNA probes and cDNAs prepared using amplified total RNAs. Therefore, RNA-seq and quantitative real-time PCR experiments are not always the best methods to verify microarray data. To better repeat the previous microarray data, we should prepare identical cDNAs used for the previous microarrays and carry out Southern blot analyses.

In view of the GABAB1 splicing complexity, there are likely additional intron 4 containing splicing variants that change expression after chronic alcohol exposure. To verify their existences and expressions, northern blot experiments comparing non-alcoholic and alcoholic brains are required. To clone the variants, we can screen prefrontal cortex cDNA libraries using colony hybridization experiments based on the intron 4 probe of the cDNA microarray. The roles of these splice variants and the effect of chronic alcohol on their expressions and functions are of future interest given the promising role of GABAB agonists in treating alcoholism.

N-terminal truncated splicing variants have their own promoters and unique cellular distributions (Wang, Schultheis et al. 1996; Crofts, Hancock et al. 1998; Sunn, Cock et al. 2001; Travers, Vallance et al. 2003; Huang, Levitsky et al. 2009). The N-terminal truncated intron 4 containing GABAB1 splicing variants probably follow the same patterns. During the promoter search using the PROMO website (Farre, Roset et al. 2003), we found potential transcription factor binding sites based on sequence homology. To validate the potential promoters, we need to find the transcription initiation sites using

3’ RACE. Reporter assays will confirm transcription factor binding and promoters upstream from their transcription initiation sites. Chromatin immunoprecipitation (ChIP) and electrophoretic mobility shift assay (EMSA) will provide exact binding sites for specific transcription factors predicted by the PROMO search. Though their ORFs do not contain unique peptide sequences to design antibodies, they have unique transcript sequences. For cellular distribution studies, the transcript sequences will allow *in situ* hybridization or quantitative real-time PCR experiments.

5.2.2 Further GABAB1 splicing and whole transcriptome studies

5.2.2.1 Validation of splicing junction changes found from RNA-seq

Our RNA-seq study found 65 GABAB1 splicing junctions. Some splicing junction changes in the RNA-seq data were also observed in alcoholics using quantitative real-time PCR (Figure A.2, A.3). GB1 targeted a known splicing junction at exon 22 - 23 and showed increased expression in alcoholic brains (Figure A.2). Using quantitative real-time PCR, GBc targeted known-alternative splicing junction between exon 14 and 16 and showed similar results to the RNA-seq study (Figure A.3). We used human prefrontal cortex samples like the previous microarray studies (Liu, Lewohl et al. 2006). However, the quantitative real-time PCR experiments were performed for only two splicing junctions out of the 65 splicing junctions found from RNA-seq study. Therefore, more splicing junction data should be confirmed using quantitative real-time PCR.

The splicing junction changes can be measured in two different ways. We can measure splicing out or in from transcripts. RPJM calculation of RNA-seq junction study

counted mapped reads that have splicing out. TopHat mapper has multiple junction finding steps that validate found splicing out each time to reduce false discovery rates. Though its default setting for maximum intron length is 500,000 Kbp, to diminish false discovery rate split segments of individual reads can be aligned much further and do not miss splicing out that is longer than 500,000 Kbp. Therefore, splicing junction finding and RPJM calculation for splicing out are highly accurate (Dimon, Sorber et al. 2010). However, for rare splicing junctions RPJM values are extremely low. RNA-seq may have difficulty in detecting rare splicing junctions for multiple samples and may not provide enough power for statistical tests. If we calculate another value counting mapped reads that are spliced in, it will provide high values and higher power for rare splicing junctions. Therefore, it could a solution for rare splicing junction study. Mapped reads containing splicing in have short mapping extension into splicing junction regions. Read mapping always allows mismatches. The short mapping extension can have higher chance for false mapping if it has mismatches. Therefore, quantitative real-time PCR validation is strongly required for splicing in validation. In Figure A.2 and A.3, primer and probe sets, GB2, 3, 4, a, b, d, and e, specifically detected splicing ins, and their primers or probes were designed based on splicing junction sequences to detect them. For further study of rare splicing junctions, we can design primer and probe sets like them to perform quantitative real-time PCR experiments.

5.2.2.2 Characterization of GABAB1 splicing junctions

After RNA-seq data analysis, we predicted GABAB1 splicing functions in alcoholic brains based on functional domains. To confirm the predicted functions, we need to characterize particular splicing variants. Using splicing variants without exon 10 (exon 10 contains one GABA binding site), a GABA binding affinity test will help identify the function of this exon. Also, functional measurements (e.g., electrophysiology or cAMP production) can prove functional failure of the splicing variants due to less GABA binding sites. Splicing variants lacking exon 17, and thus lacking TM3, will not be localized on the membrane as functional receptors. Perhaps these variants would mimic the function of other secreted GABAB1 splicing variants and act to inhibit other major GABAB1 isoform functions. For exon 17 studies, we can follow the GABAB receptor inhibitory tests used for GABAB1k, l, and m outlined in chapter 2. Even if the abnormal TMs allow localization to the membrane, the receptors could not transfer GABA signals because a functional G protein binding site located at the intracellular region between TM 3 and 4 would not exist. Altered structural conformation based on the missing TM 3 would be expected to alter their locations and functions. Co-IP experiments of the splicing variants and G proteins would test their interaction.

At exon 23, one miRNA target site is less spliced out and exposes more at 3' UTR in alcoholic brains (Figure 3.8). Our RNA-seq data showed that the splicing site is a novel splicing junction, and the splicing site was a possible target site of 2 miRNAs. To confirm if the two miRNAs downregulate GABAB1 transcript expressions, miRNA luciferase reporter assay is required. For this assay, vectors need to be constructed using

GABAB1 3' UTR with/without the splicing site. After co-transfection with one target miRNA and each vector construct, the assay will verify if the miRNA specifically targets the junction. Chronic alcohol also can alter miRNA expression levels (Avissar, McClean et al. 2009; Lewohl, Nunez et al. 2011). Therefore, expression studies of the two miRNAs should be tested in alcoholic brains. If they did not express in alcoholic brains, the miRNA effects will not likely play a critical role in alcoholics.

5.2.2.3 Genome wide RNA-seq data analyses

Chronic alcohol may target splicing machinery globally in alcoholic brains. Altered splicing junction ratio in whole transcriptome is about 8 times higher than global gene expression change ratios in our RNA-seq study (Figure A.4). Splicing factors can affect splicing junctions (Grosso, Gomes et al. 2008), and neuronal splicing factors alter splicings in a neuron specific manner (Allen, Darnell et al. 2010). Splicing factor expressions were modulated by cellular and environmental signals (Pihlajamaki, Lerin et al. 2011). Therefore, it is feasible that chronic alcohol could influence splicing factor expressions in brain. Among all genes from the RefSeq Genes model, a preliminary analysis suggests that the most significantly changed genes found in the MA plot analysis (p-value < 0.001) were small nucleolar RNAs (snoRNAs) that can also regulate splicing like splicing factors (Table A.1) (Kishore and Stamm 2006). The functional study of snoRNAs and chronic alcohol will help determine if chronic alcohol preferentially targets the splicing mechanism compared to overall gene expression changes.

Like the longer GABAB1 gene length that we identified from its splicing junctions, other genes also can have longer gene lengths than their known gene lengths of the RefSeq Genes model based on their splicing junctions. The current gene model is not always accurate (Bernal, Crammer et al. 2007), and gene predictions have lower accuracy in finding the beginning and end of genes. Longer 3' UTRs were frequently found in mammalian RNA-seq data (Mortazavi, Williams et al. 2008; Mangone, Manoharan et al. 2010). Therefore, gene prediction models need to be updated to improve their accuracy (Guigo, Agarwal et al. 2000; Schweikert, Zien et al. 2009). Accurate gene lengths mean that we can calculate gene expression levels with higher accuracy. Splicing junction data of our RNA-seq study helped update only GABAB gene length. To update all known gene lengths in genome wide, we also need to collect and analyze more splicing junction data for each gene from RNA-seq data.

Though we found 291,553 splicing junctions from the whole transcriptome, only 65 GABAB1 junctions were used for our gene specific study. We thus need to analyze genome wide splicing junction data. After updating all gene lengths based on the RefSeq Genes model, we can assign splicing junctions to individual genes. Using gene and splicing junction expression levels, we can study their relationship like the correlation study between chromatin immunoprecipitation sequencing (ChIP-seq) and RNA-seq data (Pepke, Wold et al. 2009). Our splicing junction data showed that significantly changed junctions are mostly from the mitochondria chromosome (Table A.2). Among the top 100 splicing junctions based on p-values, 85 junctions are from the mitochondria chromosome. Though the RefSeq Genes model is the most commonly used gene model

for RNA-seq, it does not have gene information for the mitochondria chromosome, and we have not calculated gene expression data for the mitochondria chromosome due to lack of a gene model. However, the mitochondria genes and splicings should be further characterized given the chronic alcohol effect that we observed.

5.2.2.4 Case analysis to study splicing variation of individual samples

RPJM values can be altered based on individual sample variations. Including alcoholism, their age, gender, brain pH, PMI, and smoking probably affect RPJM values. For our RNA-seq study, their RPJM fold changes in alcoholics vary among splicing junctions (Figure A.5). Among 65 GABAB1 splicing junctions, some selected GABAB1 splicing junctions at Table A.3 had various mapped reads across samples. Some novel or partial_novel splicing junctions only existed in specific samples (Table A.3). These variations could not be simply grouped and explained only by alcoholism. Therefore, other variables (genetic polymorphisms, age, gender, brain pH, PMI, and smoking) for each sample should be also considered to understand the expression variations.

5.2.3 Further RNA-seq bias studies

5.2.3.1 Further library constructions to diminish RNA-seq biases

In our RNA-seq study using alcoholic brains, we prepared whole transcriptome libraries by RNaseIII fragmentations. To remove RNaseIII bias, a Heat + OptiK method should be used for RNA fragmentation of RNA-seq library reconstructions instead of

RNaseIII. The new libraries will allow more accurate gene expression level calculations and identify more splicing junctions and SNPs.

To remove additional T4PNK sequence bias of the Heat + OptiK method (Figure 4.7B, D), phosphatase could be used instead of T4PNK (Figure 5.2). Because no phosphatase sequence bias has been reported, it probably has less chance to have sequence bias than T4PNK. Therefore, this will be a good trial to remove the known T4PNK bias.

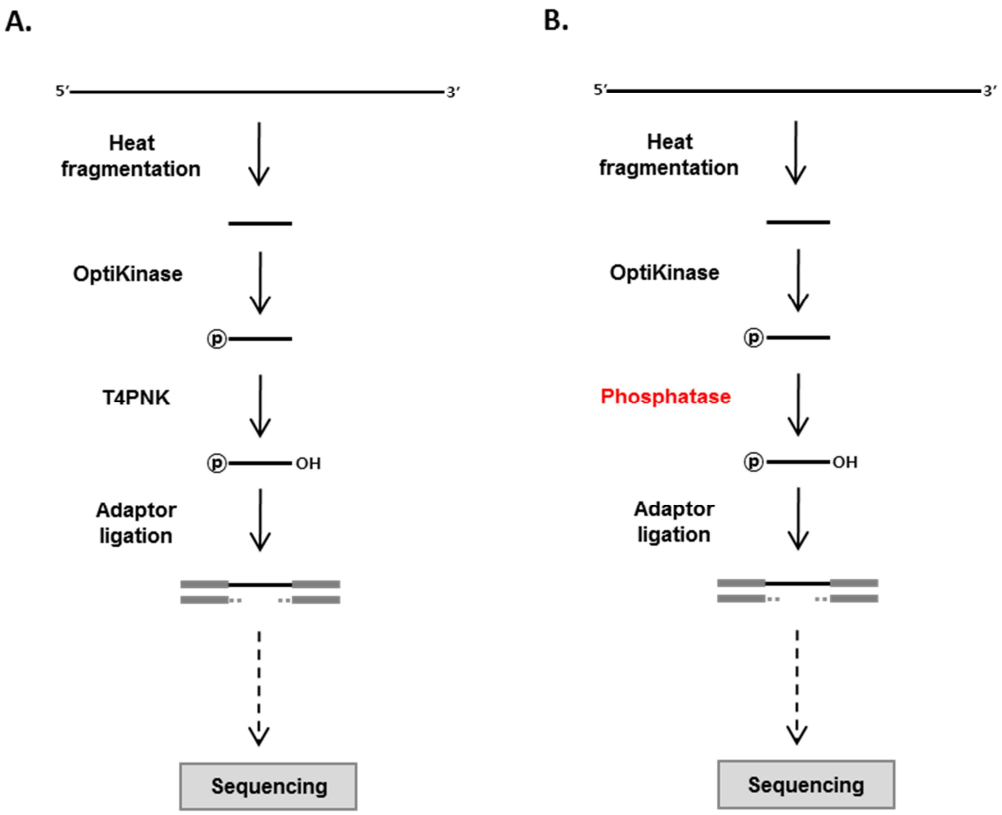


Figure 5.2 Possible whole transcriptome library construction to remove T4PNK sequence bias.
A. Heat + OptiK method uses T4PNK after OptiKinase treatment. B. Updated Heat + OptiK method uses phosphatase instead of T4PNK to remove known T4PNK sequence bias.

5.2.3.2 Minor RNA-seq bias finding

Our study is the first sequence bias study using the SOLiD sequencer. Most RNA-seq biases were found from the Illumina genome analyzer. The biases are from PCR amplification, GC bias, and randomly primed RT (Hansen, Brenner et al. 2010; Aird, Ross et al. 2011). Our preliminary study showed that whole transcriptome libraries do not have GC bias (Figure A.6A). WebLogo sequence analyses using our whole transcriptome libraries did not show the same sequence biases found from the RNA-seq libraries that used randomly primed cDNA for the Illumina genome analyzer (Figure 1.11A, 4.3, 4.7) (Hansen, Brenner et al. 2010). The GC bias studies for the Illumina genome analyzer compared relative read abundances at each GC ratio (Figure 1.11B) (Aird, Ross et al. 2011), but we used absolute read populations (Figure A.6). For our sequence bias study, we measured entropy instead of the nucleic acid frequency that used for the RNA-seq libraries for the Illumina genome analyzer (Figure 1.11A, 4.3, 4.7). The different measurements for the biases might not provide identical results. Nucleic acid frequency test in Figure 1.11A was more sensitive than our entropy calculations. Small sequence biases can result in large signals during nucleic acid frequency calculation. Therefore, we should also measure the biases for our sequencing data following their methods. It will find more minor sequence bias patterns and improve our RNA-seq data quality.

5.3 LIMITATION OF STUDY

As discussed previously, most human postmortem samples provided very limited RNA amounts and did not always allow further experimental designs including data validation. Though we proposed library reconstruction to diminish sequence biases of current RNA-seq data and quantitative real-time PCR to validate expression changes as further studies, they could not be performed using identical samples due to the material limitation.

Though intron 4 containing GABAB1 splicing variant transcripts have unique nucleotide sequences, the peptide sequences are not likely to be sufficiently different to allow design splice-specific antibodies. Therefore, epitope tags (His, Flag, hemagglutinin (HA), or anti c-Myc (Myc) Tags) should be used to find their binding partners using Co-IP experiments, but one must consider unexpected effects of epitope tags for data interpretation. Also, *in situ* hybridization could be designed based on their unique transcript sequences for their cellular localization studies. However, the *in situ* hybridization experiment can show only mRNA location.

RNA-seq study collects reads generated from fragmented RNAs and needs further analysis steps for transcriptome reconstruction to find full length splicing variants. Though there are some software packages, such as Scripture and Cufflinks, for the transcriptome reconstruction (Garber, Grabherr et al. 2011), they could not always construct known splicing variant transcripts. Transcriptome reconstruction is inaccurate for genes that have high splicing complexity like GABAB1. Though RPGM calculation

is currently recommended for gene expression study, it is impossible to study individual transcript expressions.

Due to the difficulty of transcriptome reconstruction, the 65 GABAB1 splicing junctions found from our RNA-seq study cannot be characterized completely. Though we have junction information, we cannot clearly predict the combinations that will constitute all the individual transcripts. We can characterize specific junction containing GABAB1 transcripts based on known splicing transcript sequence, but there will be still questions for their functions.

As previously mentioned, it is currently impossible to design non-biased RNA-seq library due to technical limitation. All individual library preparation steps possibly contain unknown biases. Though computational approaches and new library constructions tried to overcome the bias issue, there is no perfect solution. Therefore, we should use the best methods to solve the biases among them and also consider their limitations.

5.4 CONCLUSIONS

In this study, intron 4 containing novel splicing variant cloning and our first gene specific RNA-seq experiments showed the splicing complexity of GABAB1 gene. They detected unexpected novel splicings from GABAB1 gene. The novel GABAB1 splicing variant cloning confirmed that intron 4 is a frequent alternative splicing spot (Wei, Jia et al. 2001; Holter, Davies et al. 2005; Tiao, Bradaia et al. 2008). The gene specific RNA-

seq detected about 40 unknown splicing junctions from GABAB1 genes and demonstrates the power of deep sequencing. The splicing junctions provided strong evidence to extend GABAB1 gene length and indicated that GABAB1 gene area is overlapped with neighboring genes. Therefore, GABAB1 gene has extreme complexity in splicing.

Chronic alcohol abuse alters GABAB splicing rather than overall gene expression in prefrontal cortex. Though our RNA-seq data showed no GABAB1 gene expression change in alcoholics, GABAB1 splicings were affected by chronic alcohol. Ten splicing junctions out of 22 known GABAB1 junctions were changed, and 11 exons among total 22 known exons were also altered their expressions in alcoholics. Genome wide RNA-seq data analysis also found that chronic alcohol changes splicing and exon/intron expressions more than gene expressions (Figure A.4) and suggested that splicing as a mechanism of brain and transcriptome changes from chronic alcohol. Microarray experiments showed expression increases with only one intron 4 targeting probe among 6 different GABAB1 probes (Flatscher-Bader, van der Brug et al. 2005; Liu, Lewohl et al. 2006; Ponomarev, Wang et al. 2011). Our quantitative real-time PCR experiments showed an expression increase only at exon 22 - 23 junction among 11 target regions at GABAB1 gene (Figure A.2, A.3). These data suggest production of splice variants that will have impaired function with no increase in total GABAB receptors. This could have important implications for use of GABAB drugs in treatment of alcoholism. However, the critical and difficult goal for the future is defining the functional changes produced by the diverse GABAB splicing changes found in alcoholics. This should provide a

paradigm for addressing the broader question of the role of splicing in other genes and other psychiatric diseases.

APPENDIX

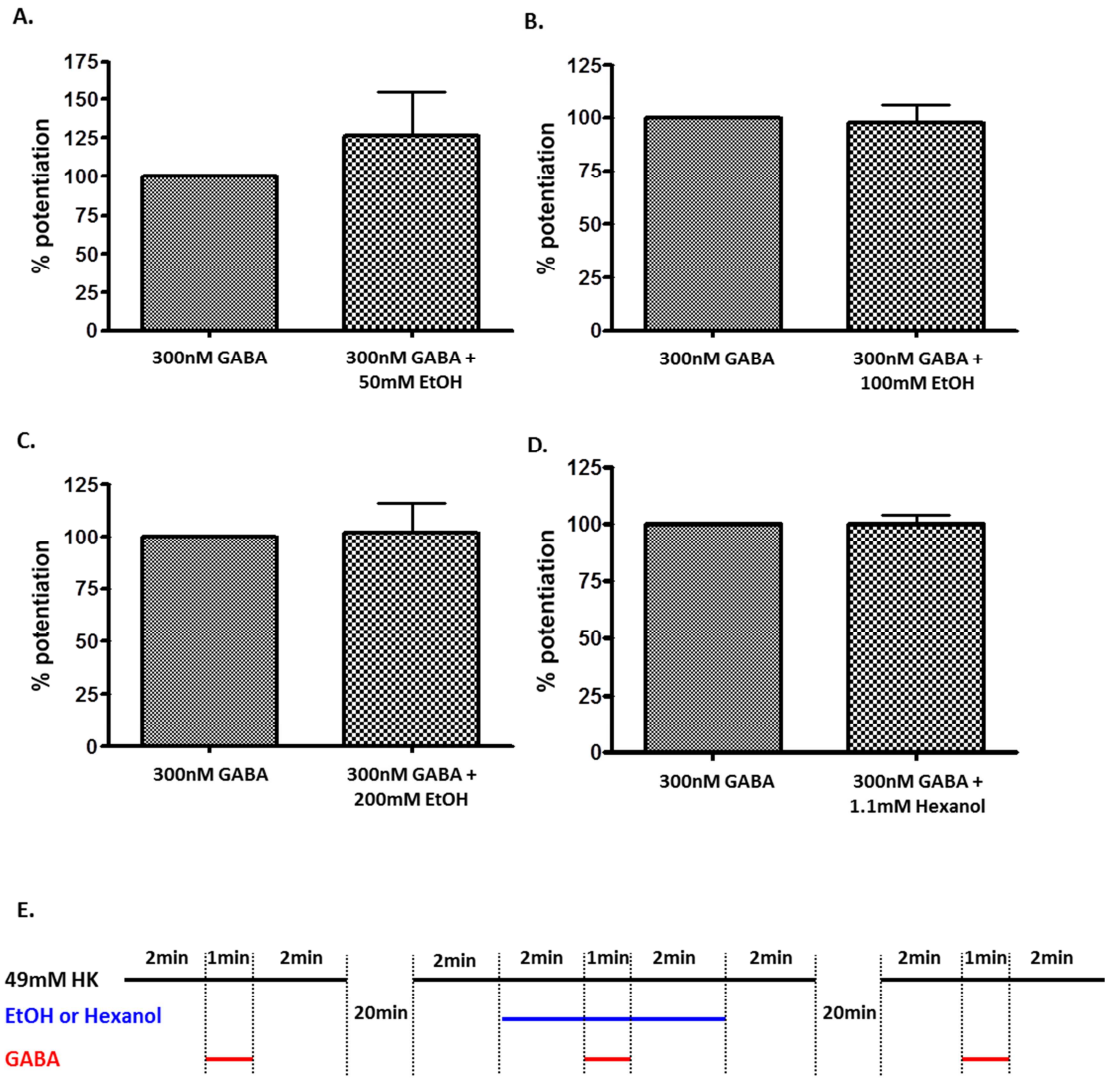
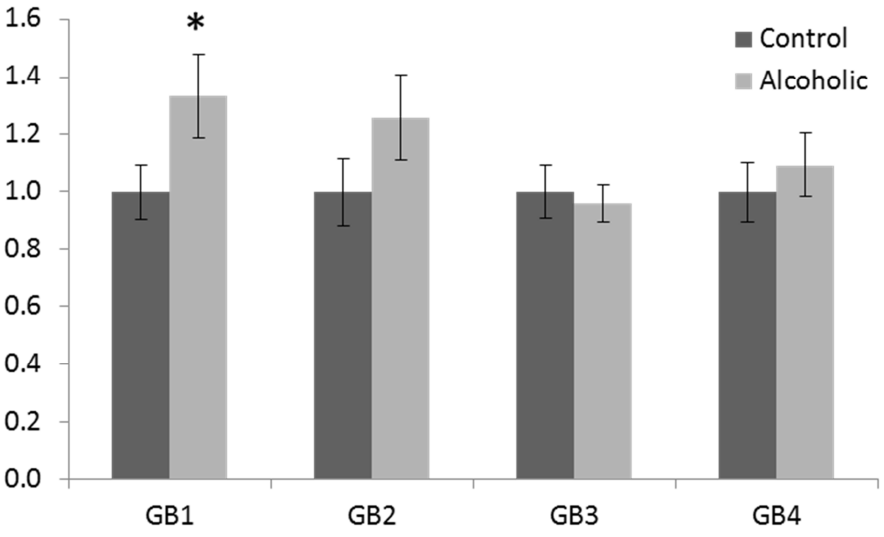


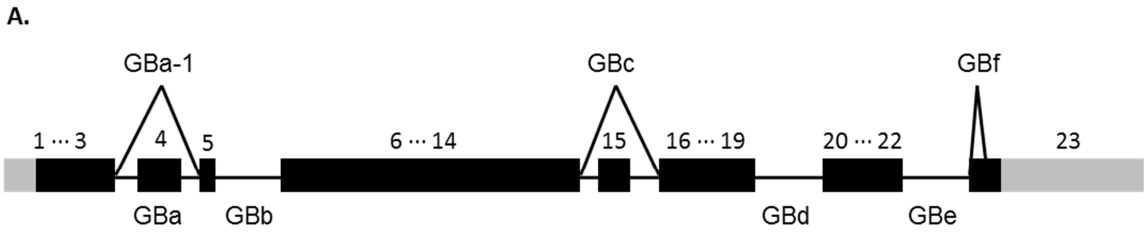
Figure A.1 Alcohol effect on GABAB1a mediated currents in oocytes.
GABAB1a cRNAs were injected into *Xenopus* oocytes including GABAB2 cRNAs and GIRK1/2 cDNAs. GABAB1a mediated currents were measured using the two electrode voltage clamp system. A, B, and C. Acute ethanol (EtOH) effects on GABAB1a mediated currents were tested using 50, 100, and 200 mM EtOH, respectively. D. Effect of 1.1 mM hexanol on GABAB1a mediated currents. E. Experimental design for ethanol and hexanol experiments. 49mM HK, EtOH or Hexanol, and GABA were represented by black, blue, and red, respectively. Each line showed treatment time. 20 minutes (mins) were for washing using ND96 buffer.



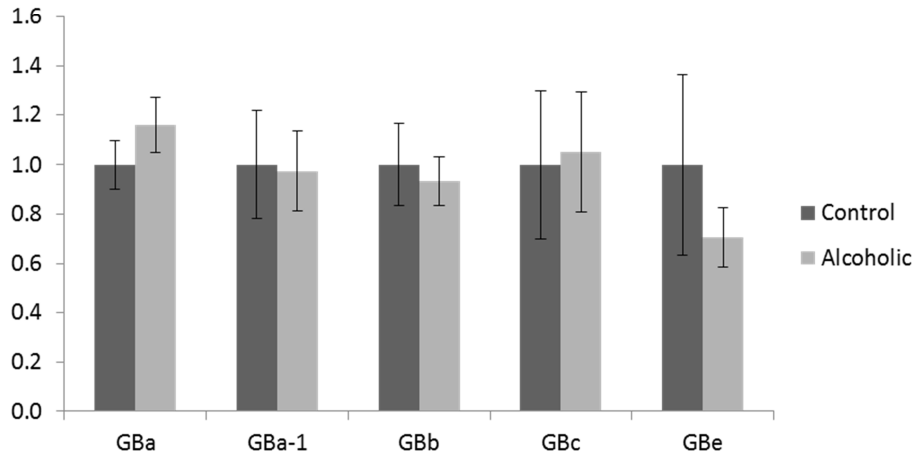
Fold change	1.33	1.26	0.96	1.09
p-value	0.0309	0.0903	0.3674	0.2733

Figure A.2 Quantitative real-time PCR data for previous microarray probes and intron 4 containing splicing variants.

Based on the cDNA microarray data, quantitative real-time PCR was performed. The primer and probe sets, GB1 and GB2, were designed based on previous microarray probes. GB1 targets a splicing junction between exon 22 and 23. GB2 was a microarray probe from intron 4. GB3 and GB4 target splicing variants that have intron 4 in their transcripts. GB3 represents GABAB1k, l, m, and n, and GB4 targets GABAB1j. Though a GB2 expression increase was expected in alcoholic brains based on the microarray data, GB2 expression did not increase significantly here. Only GB1 expression was increased.



B.



	GBa	GBa-1	GBb	GBc	GBe
Fold change	1.16	0.97	0.93	1.05	0.71
p value	0.1433	0.3740	0.4634	0.4486	0.2419

Figure A.3 GABAB1 splicing junction studies using quantitative real-time PCR.
A. After searching known GABAB1 splicing junction information from human, mouse, and rat, quantitative real-time PCR primer and probe sets were designed for the splicing junctions. GBa-1 and GBc represent know-alternative splicing junctions, and GBf represents a spliced out site at exon 23. GBa targets splicing variants containing exon 4. GBb, GBd, and GBe detect splicing variants containing introns that have been spliced in. B. None of the primer and probe sets showed significant difference between alcoholics and controls, but no expression change of GBc was similar to our RNA-seq data. GBd (shown in A) does not exist in human, and the GBf signal was too low to detect.

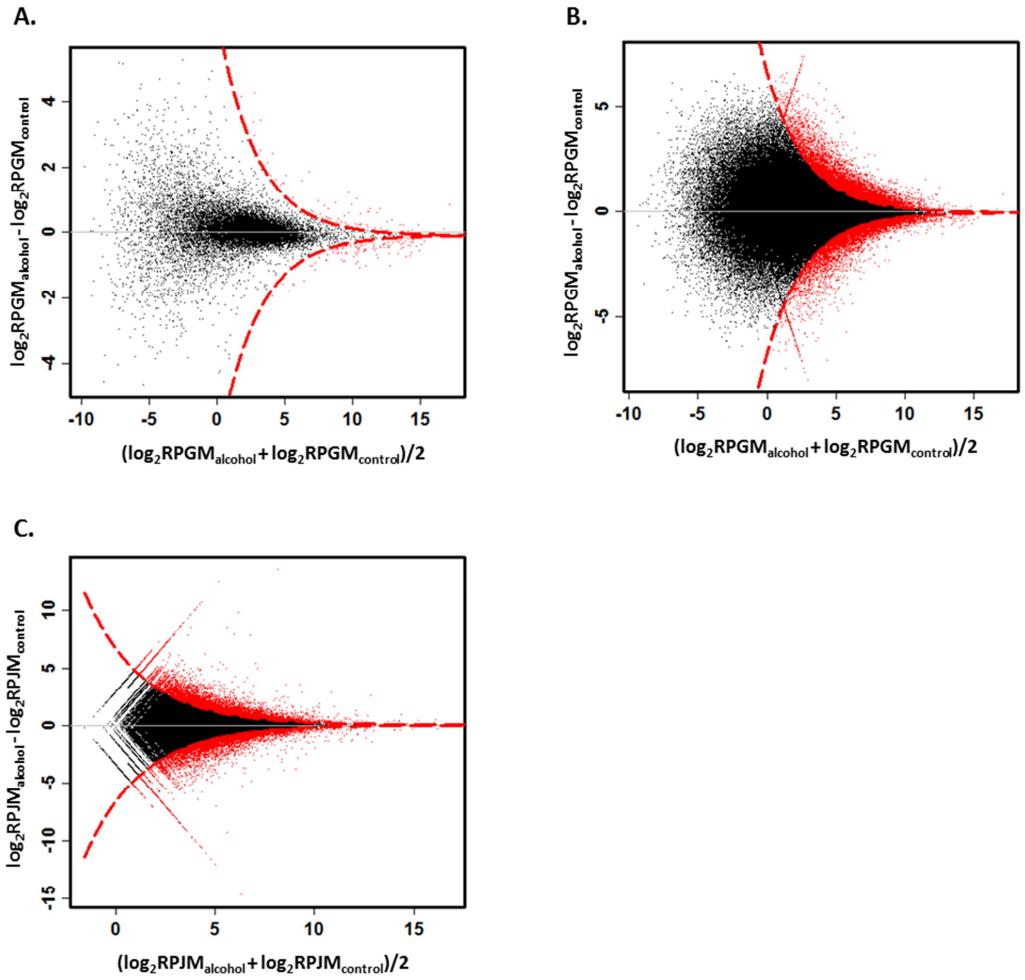


Figure A.4 Genome wide gene, exon, intron, and splicing junction changes in alcoholics. A. Genome wide gene expression changes were compared in alcoholics. From RPGM values of all genes based on the RefSeq Genes model, we compared gene expression levels between alcoholics and controls. B. Exon and intron expressions were compared between alcoholics and controls. Based on the RefSeq Genes model, RPGM values were also calculated for all exons and introns, and significant RPGM changes were assessed. C. For all splicing junctions found from RNA-seq study, we assessed RPJM changes in alcoholics. Significantly different RPGM and RPJM values in alcoholics were identified using the DEGseq package. Significant gene expression changes by alcoholism were about 1.8 % of total number, but exons/introns and splicing junctions were about 14 % and 15 %, respectively (p-values < 0.001). X-axes represent the averages of log_2RPGM or log_2RPJM values for alcoholic and control groups. Y-axes are for differences between the values. Red dots represent genes, exons, introns, or splicing junctions that changed in alcoholics (p-values < 0.001), and the red dotted lines are p-value 0.001 lines.

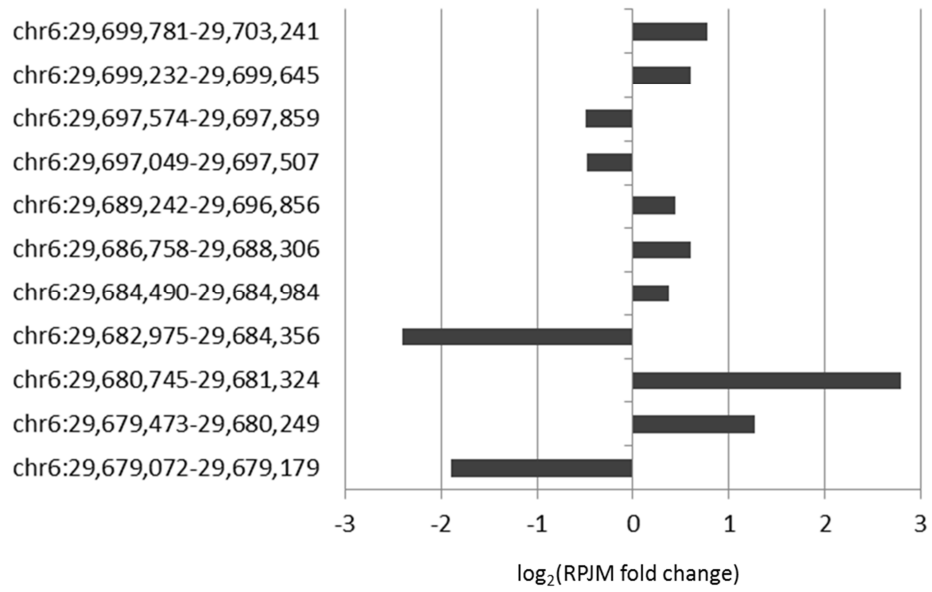


Figure A.5 Fold changes in alcoholics for GABAB1 splicing junctions.
We calculated $\log_2$(fold change) for each splicing junction that significantly changed RPJM in alcoholics. X-axis represents $\log_2$(fold change), and Y-axis is for splicing junction location.

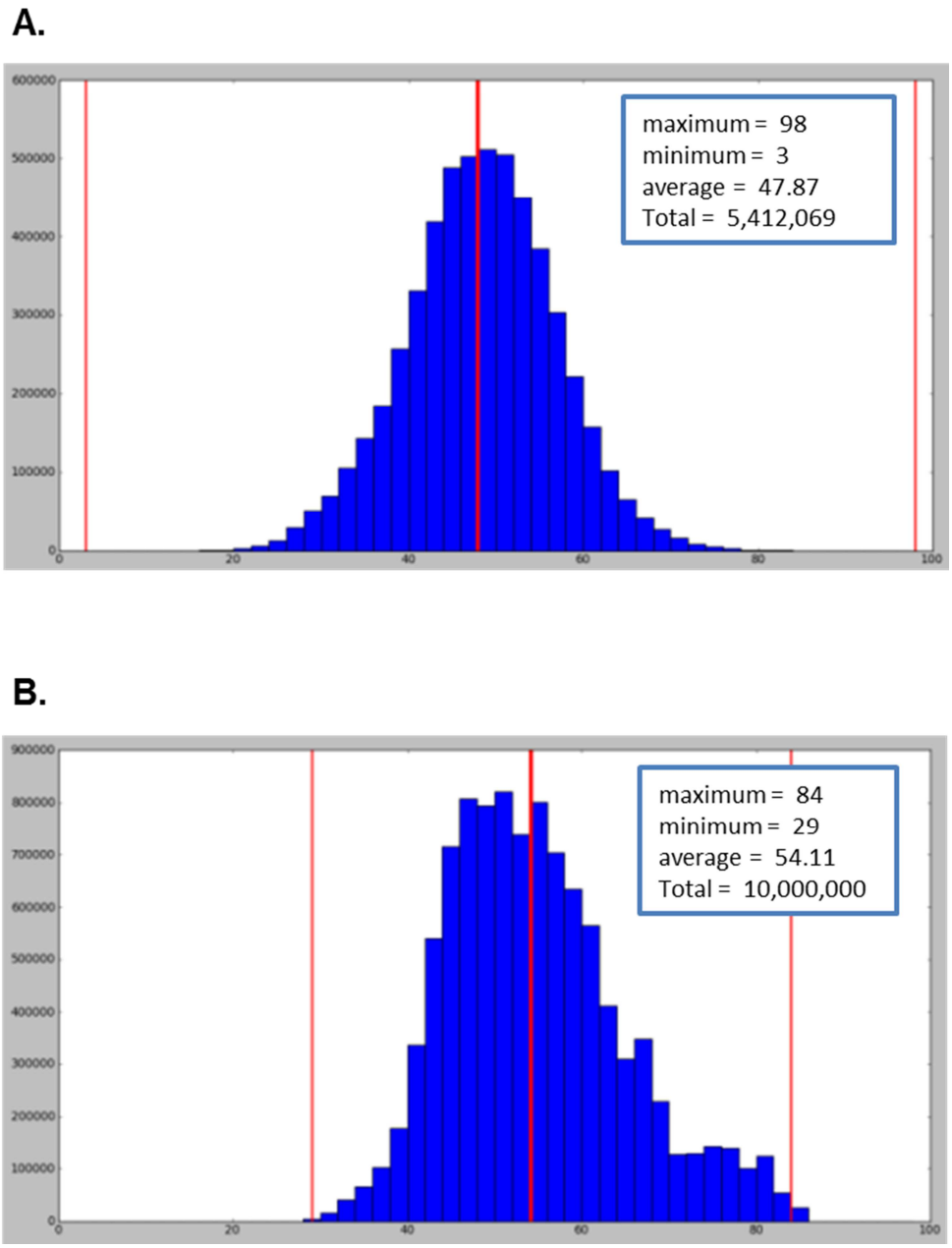


Figure A.6 GC distributions of whole transcriptome library and random sequences.
A. GC distribution was calculated based on read population at each GC ratio. B. GC distribution calculation using random sequences. X-axis is for GC ratio, and Y-axis represents read number. The left, the middle, and the right red lines represent minimum, average, and maximum GC ratios, respectively.

Num	gene ID	type	fold change	p-value	q-value(Benjamini et al. 1995)	q-value(Storey et al. 2003)
1	SNORA42	non-coding RNA	1.90	0.00E+00	0.00E+00	0.00E+00
2	SNORD27	non-coding RNA	0.65	0.00E+00	0.00E+00	0.00E+00
3	RNU6ATAC	non-coding RNA	0.63	0.00E+00	0.00E+00	0.00E+00
4	SNORA81	non-coding RNA	0.72	0.00E+00	0.00E+00	0.00E+00
5	SNORD47	non-coding RNA	0.52	0.00E+00	0.00E+00	0.00E+00
6	SNORD115-10_dup3	non-coding RNA	0.69	0.00E+00	0.00E+00	0.00E+00
7	SNORD115-5_dup1	non-coding RNA	1.13	2.49E-177	6.73E-174	6.73E-174
8	SNORD115-39	non-coding RNA	0.78	4.11E-176	9.72E-173	9.72E-173
9	SNORD76	non-coding RNA	0.68	1.80E-171	3.78E-168	3.78E-168
10	SNORD115-34	non-coding RNA	0.81	6.29E-164	1.19E-160	1.19E-160
11	SNORD100	non-coding RNA	1.24	3.73E-151	6.41E-148	6.41E-148
12	CDR1	coding RNA	1.75	1.36E-148	2.14E-145	2.14E-145
13	SNORD116-8	non-coding RNA	0.68	5.20E-111	7.57E-108	7.57E-108
14	SNORD114-12	non-coding RNA	0.76	1.61E-106	2.17E-103	2.17E-103
15	SNORD115-30	non-coding RNA	0.80	3.38E-102	4.26E-99	4.26E-99
16	SNORD115-33	non-coding RNA	0.78	1.65E-99	1.95E-96	1.95E-96
17	SNORD116-18	non-coding RNA	0.62	1.65E-98	1.84E-95	1.84E-95
18	SNORD115-42	non-coding RNA	0.81	4.67E-98	4.65E-95	4.42E-95
19	SNORD115-15_dup2	non-coding RNA	0.81	4.67E-98	4.65E-95	4.42E-95
20	SNORD115-10_dup5	non-coding RNA	0.81	4.67E-98	4.65E-95	4.42E-95
21	SNORD44	non-coding RNA	0.53	2.11E-92	1.90E-89	1.90E-89
22	SNORA84	non-coding RNA	2.18	7.52E-88	6.47E-85	6.47E-85
23	SNORA67	non-coding RNA	1.59	4.33E-86	3.56E-83	3.56E-83
24	SNORA52	non-coding RNA	1.74	9.46E-84	7.46E-81	7.46E-81
25	MALAT1	non-coding RNA	1.36	6.94E-83	5.25E-80	5.25E-80
26	SNORD113-6	non-coding RNA	0.50	3.03E-78	2.20E-75	2.20E-75
27	SNORD115-23	non-coding RNA	0.88	1.04E-76	7.28E-74	7.28E-74
28	SCARNA20	non-coding RNA	2.44	1.93E-76	1.30E-73	1.30E-73
29	SNORD33	non-coding RNA	1.24	4.03E-73	2.63E-70	2.63E-70
30	SNORD117	non-coding RNA	0.66	1.61E-68	1.02E-65	1.02E-65
31	SNORA10	non-coding RNA	1.63	5.53E-65	3.37E-62	3.37E-62
32	LOC85389	non-coding RNA	1.24	2.42E-62	1.43E-59	1.43E-59
33	SNORA80	non-coding RNA	1.65	1.47E-61	8.44E-59	8.44E-59
34	SNORD115-38	non-coding RNA	0.87	1.04E-54	5.79E-52	5.79E-52
35	SNORD115-14	non-coding RNA	0.76	1.03E-53	5.55E-51	5.55E-51
36	SNORA54	non-coding RNA	0.82	6.62E-53	3.48E-50	3.48E-50
37	SNORD84	non-coding RNA	0.76	1.53E-51	7.80E-49	7.80E-49
38	SNORD95	non-coding RNA	1.31	1.39E-49	6.94E-47	6.94E-47
39	SNORD112	non-coding RNA	0.55	2.75E-49	1.33E-46	1.33E-46
40	SNORA28	non-coding RNA	0.57	3.71E-49	1.75E-46	1.75E-46
41	SNORD115-32	non-coding RNA	0.87	7.05E-49	3.25E-46	3.25E-46
42	SNORD61	non-coding RNA	0.84	1.44E-48	6.47E-46	6.47E-46
43	KIAA0907	coding RNA	1.88	1.57E-43	6.92E-41	6.92E-41
44	SNORD116-22	non-coding RNA	0.76	6.48E-43	2.79E-40	2.79E-40
45	SNORD58A	non-coding RNA	0.73	7.75E-37	3.26E-34	3.26E-34
46	SNORA70	non-coding RNA	1.57	6.52E-36	2.68E-33	2.68E-33
47	SNORD97	non-coding RNA	0.81	7.39E-35	2.97E-32	2.97E-32
48	SNORA63	non-coding RNA	0.77	7.81E-34	3.08E-31	3.08E-31
49	SNORD114-1	non-coding RNA	0.81	8.68E-33	3.35E-30	3.35E-30
50	SNORD116-29	non-coding RNA	0.75	7.84E-32	2.97E-29	2.97E-29
51	SNORD26	non-coding RNA	0.90	7.88E-31	2.92E-28	2.92E-28
52	SNORD115-41	non-coding RNA	0.88	2.54E-30	9.24E-28	9.24E-28
53	RPPH1	non-coding RNA	1.11	1.98E-29	7.06E-27	7.06E-27

Num	gene ID	type	fold change	p-value	q-value(Benjamini et al. 1995)	q-value(Storey et al. 2003)
54	SNORD58B	non-coding RNA	0.79	3.08E-28	1.08E-25	1.08E-25
55	SNORD115-40	non-coding RNA	0.86	3.13E-28	1.08E-25	1.08E-25
56	SNORD12C	non-coding RNA	0.63	2.52E-27	8.51E-25	8.51E-25
57	SNORD31	non-coding RNA	0.75	1.03E-26	3.43E-24	3.43E-24
58	SNORD59B	non-coding RNA	0.71	2.43E-25	7.91E-23	7.91E-23
59	SNORD83A	non-coding RNA	3.85	1.62E-24	5.20E-22	5.20E-22
60	SNORD115-4	non-coding RNA	0.67	3.89E-24	1.23E-21	1.23E-21
61	GFAP	coding RNA	2.00	1.19E-23	3.70E-21	3.70E-21
62	SNORA20	non-coding RNA	0.82	1.64E-23	5.01E-21	5.01E-21
63	SNORA53	non-coding RNA	0.57	1.86E-23	5.59E-21	5.59E-21
64	SNORD46	non-coding RNA	0.82	3.00E-23	8.87E-21	8.87E-21
65	SCARNA1	non-coding RNA	0.59	4.20E-23	1.22E-20	1.22E-20
66	SNORA75	non-coding RNA	0.73	8.75E-23	2.51E-20	2.51E-20
67	SNORD50B	non-coding RNA	0.74	4.55E-22	1.28E-19	1.28E-19
68	SNORA14B	non-coding RNA	0.86	4.93E-22	1.37E-19	1.37E-19
69	SNORA2A	non-coding RNA	1.58	1.91E-21	5.22E-19	5.22E-19
70	SNORA8	non-coding RNA	0.79	1.02E-20	2.76E-18	2.76E-18
71	SNORA41	non-coding RNA	0.74	2.29E-20	6.10E-18	6.10E-18
72	SNORD42B	non-coding RNA	0.71	4.00E-20	1.05E-17	1.05E-17
73	SNORD8	non-coding RNA	0.85	8.47E-20	2.19E-17	2.19E-17
74	SNORD115-15_dup1	non-coding RNA	0.92	2.33E-19	5.96E-17	5.96E-17
75	PAR4	non-coding RNA	0.88	3.09E-19	7.78E-17	7.78E-17
76	SNORD90	non-coding RNA	0.70	3.30E-19	8.21E-17	8.21E-17
77	SNORD116-16	non-coding RNA	0.84	7.64E-19	1.88E-16	1.88E-16
78	SNORD10	non-coding RNA	0.60	1.82E-18	4.40E-16	4.40E-16
79	SNORD9	non-coding RNA	0.90	1.48E-16	3.54E-14	3.54E-14
80	SNORD115-2	non-coding RNA	0.73	4.82E-16	1.14E-13	1.14E-13
81	SNORD113-9	non-coding RNA	0.82	4.00E-15	9.35E-13	9.35E-13
82	SNORA46	non-coding RNA	0.67	8.11E-14	1.87E-11	1.87E-11
83	SNORD116-1	non-coding RNA	0.92	9.92E-14	2.26E-11	2.26E-11
84	SNORD113-7	non-coding RNA	0.75	1.02E-13	2.30E-11	2.30E-11
85	SNORA7B	non-coding RNA	1.50	1.16E-13	2.58E-11	2.58E-11
86	IFITM1	coding RNA	3.19	2.21E-13	4.85E-11	4.85E-11
87	SNORD5	non-coding RNA	0.77	4.20E-13	9.13E-11	9.13E-11
88	SNORD104	non-coding RNA	1.08	6.62E-13	1.42E-10	1.42E-10
89	SNORD105	non-coding RNA	0.69	7.55E-13	1.61E-10	1.61E-10
90	SCARNA5	non-coding RNA	1.09	1.76E-12	3.69E-10	3.69E-10
91	SNORA71D	non-coding RNA	2.08	3.84E-12	7.98E-10	7.98E-10
92	SNORD41	non-coding RNA	1.24	3.99E-12	8.21E-10	8.21E-10
93	SNORA47	non-coding RNA	1.75	6.99E-12	1.42E-09	1.42E-09
94	SNORD38B	non-coding RNA	0.76	9.27E-12	1.87E-09	1.87E-09
95	NCRNA00084	non-coding RNA	1.85	1.01E-11	2.00E-09	2.00E-09
96	SNORD125	non-coding RNA	1.16	1.08E-11	2.12E-09	2.12E-09
97	SNORD116-2_dup1	non-coding RNA	0.93	2.33E-11	4.54E-09	4.54E-09
98	SNORA26	non-coding RNA	0.81	5.52E-11	1.07E-08	1.07E-08
99	SNORD113-4	non-coding RNA	0.68	6.16E-11	1.18E-08	1.18E-08
100	SNORA37	non-coding RNA	1.26	6.49E-11	1.23E-08	1.23E-08
101	SNORD113-5	non-coding RNA	0.80	1.47E-10	2.75E-08	2.75E-08
102	SNORD121B	non-coding RNA	0.48	1.60E-10	2.97E-08	2.97E-08
103	EIF4A2	coding RNA	0.81	2.16E-10	3.97E-08	3.97E-08
104	SNORD115-37	non-coding RNA	0.38	2.19E-10	3.99E-08	3.99E-08
105	SNORA74A	non-coding RNA	0.58	3.05E-10	5.50E-08	5.50E-08
106	LOC85390	non-coding RNA	0.68	1.20E-09	2.14E-07	2.14E-07

Num	gene ID	type	fold change	p-value	q-value(Benjamini et al. 1995)	q-value(Storey et al. 2003)
107	GAS5	non-coding RNA	0.69	2.15E-09	3.80E-07	3.80E-07
108	SNORD96A	non-coding RNA	1.16	3.80E-09	6.65E-07	6.63E-07
109	SNORA35	non-coding RNA	0.60	3.82E-09	6.63E-07	6.63E-07
110	SNORA49	non-coding RNA	1.11	3.93E-09	6.76E-07	6.76E-07
111	SNORD114-14	non-coding RNA	1.19	4.62E-09	7.88E-07	7.88E-07
112	ACTB	coding RNA	1.27	6.61E-09	1.12E-06	1.12E-06
113	SNORD91A	non-coding RNA	0.79	7.47E-09	1.25E-06	1.25E-06
114	SNORD116-6	non-coding RNA	0.92	1.24E-08	2.04E-06	2.04E-06
115	SNORD116-2_dup2	non-coding RNA	0.92	1.24E-08	2.04E-06	2.04E-06
116	SNORD115-44	non-coding RNA	1.23	1.49E-08	2.43E-06	2.43E-06
117	NPAS4	coding RNA	20.37	3.08E-08	4.98E-06	4.98E-06
118	SNORD54	non-coding RNA	0.85	5.65E-08	9.05E-06	9.05E-06
119	SNORD116-21	non-coding RNA	0.87	6.09E-08	9.68E-06	9.68E-06
120	SNORA64	non-coding RNA	1.23	7.39E-08	1.17E-05	1.17E-05
121	SNORD12B	non-coding RNA	0.35	1.25E-07	1.95E-05	1.95E-05
122	SERPINA3	coding RNA	11.45	1.35E-07	2.09E-05	2.09E-05
123	APLNR	coding RNA	3.56	1.53E-07	2.36E-05	2.36E-05
124	SNORD123	non-coding RNA	0.31	1.76E-07	2.68E-05	2.68E-05
125	CLU	coding RNA	1.39	1.89E-07	2.86E-05	2.86E-05
126	SNORD119	non-coding RNA	0.53	1.90E-07	2.86E-05	2.86E-05
127	SNHG1	non-coding RNA	0.79	2.30E-07	3.43E-05	3.43E-05
128	SNORD115-3	non-coding RNA	5.72	2.43E-07	3.60E-05	3.60E-05
129	SNORD63	non-coding RNA	0.69	2.81E-07	4.12E-05	4.12E-05
130	SNORD109B_dup1	non-coding RNA	0.36	3.33E-07	4.83E-05	4.81E-05
131	SNORD109A_dup1	non-coding RNA	0.36	3.33E-07	4.83E-05	4.81E-05
132	SCARNA18	non-coding RNA	0.56	4.53E-07	6.50E-05	6.50E-05
133	SNORD12	non-coding RNA	0.75	5.09E-07	7.24E-05	7.24E-05
134	SNORD93	non-coding RNA	0.52	8.28E-07	1.17E-04	1.17E-04
135	SNORA32	non-coding RNA	0.76	1.01E-06	1.41E-04	1.41E-04
136	SNORD107	non-coding RNA	0.78	1.20E-06	1.66E-04	1.66E-04
137	SNORD114-26	non-coding RNA	0.88	1.35E-06	1.86E-04	1.86E-04
138	SNORD79	non-coding RNA	0.67	1.48E-06	2.03E-04	2.03E-04
139	HIST1H4E	coding RNA	1.29	1.81E-06	2.46E-04	2.46E-04
140	SNORD52	non-coding RNA	0.92	1.86E-06	2.52E-04	2.52E-04
141	SNORD115-35	non-coding RNA	0.89	1.94E-06	2.61E-04	2.61E-04
142	SNORA38B	non-coding RNA	1.32	2.88E-06	3.84E-04	3.84E-04
143	SNORD127	non-coding RNA	0.62	2.94E-06	3.88E-04	3.88E-04
144	SNORA17	non-coding RNA	0.87	3.35E-06	4.40E-04	4.40E-04
145	SNORD75	non-coding RNA	0.74	3.82E-06	4.99E-04	4.99E-04
146	SNORD56	non-coding RNA	0.85	4.21E-06	5.46E-04	5.46E-04
147	SNORD24	non-coding RNA	0.83	4.80E-06	6.17E-04	6.17E-04
148	SNORA68	non-coding RNA	0.75	5.54E-06	7.08E-04	7.08E-04
149	TXNIP	coding RNA	1.58	6.01E-06	7.63E-04	7.63E-04
150	SNORD29	non-coding RNA	0.72	6.50E-06	8.20E-04	8.20E-04
151	SNORD17	non-coding RNA	1.24	6.68E-06	8.37E-04	8.37E-04
152	SNORA3	non-coding RNA	0.74	7.43E-06	9.24E-04	9.24E-04
153	SNORA65	non-coding RNA	1.44	9.51E-06	1.18E-03	1.18E-03
154	SCARNA2	non-coding RNA	1.46	9.97E-06	1.23E-03	1.23E-03
155	SNORD114-23	non-coding RNA	0.80	1.24E-05	1.52E-03	1.52E-03
156	SNORD91B	non-coding RNA	0.94	1.26E-05	1.52E-03	1.52E-03
157	SCARNA6	non-coding RNA	0.82	1.55E-05	1.87E-03	1.87E-03
158	SNORD115-16	non-coding RNA	1.05	1.69E-05	2.02E-03	2.02E-03
159	SNORD121A	non-coding RNA	11.02	1.69E-05	2.02E-03	2.02E-03

Num	gene ID	type	fold change	p-value	q-value(Benjamini et al. 1995)	q-value(Storey et al. 2003)
160	SNORD88C	non-coding RNA	1.39	2.09E-05	2.47E-03	2.47E-03
161	AQP1	coding RNA	2.95	2.14E-05	2.51E-03	2.51E-03
162	MT1X	coding RNA	3.07	2.25E-05	2.63E-03	2.63E-03
163	SNORD30	non-coding RNA	0.84	2.56E-05	2.97E-03	2.97E-03
164	RNU4ATAC	non-coding RNA	0.79	2.99E-05	3.45E-03	3.45E-03
165	SNORA62	non-coding RNA	0.86	3.83E-05	4.39E-03	4.39E-03
166	C7orf41	coding RNA	1.31	5.50E-05	6.27E-03	6.27E-03
167	SNORD114-18	non-coding RNA	4.66	5.87E-05	6.65E-03	6.65E-03
168	RPLP2	coding RNA	1.62	6.01E-05	6.77E-03	6.77E-03
169	VTRNA1-3	non-coding RNA	8.36	6.86E-05	7.68E-03	7.68E-03
170	NAP1L2	coding RNA	0.73	7.14E-05	7.94E-03	7.93E-03
171	SNORA27	non-coding RNA	0.90	7.17E-05	7.93E-03	7.93E-03
172	SNORD4B	non-coding RNA	1.65	7.72E-05	8.49E-03	8.49E-03
173	SNORD1B	non-coding RNA	18.99	8.27E-05	9.04E-03	9.04E-03
174	MARS2	coding RNA	2.90	1.00E-04	1.09E-02	1.09E-02
175	HBII-52-45	non-coding RNA	6.02	1.06E-04	1.14E-02	1.14E-02
176	SNORD36B	non-coding RNA	0.36	1.09E-04	1.18E-02	1.18E-02
177	IFITM3	coding RNA	2.19	1.20E-04	1.28E-02	1.28E-02
178	HBII-52-28	non-coding RNA	0.28	1.41E-04	1.50E-02	1.50E-02
179	SCARNA14	non-coding RNA	0.75	1.51E-04	1.60E-02	1.60E-02
180	FAM107A	coding RNA	1.38	1.63E-04	1.72E-02	1.72E-02
181	SNORD115-31	non-coding RNA	0.91	1.66E-04	1.73E-02	1.73E-02
182	SNORD114-20	non-coding RNA	0.65	1.67E-04	1.74E-02	1.74E-02
183	C7orf68	coding RNA	3.26	1.70E-04	1.76E-02	1.76E-02
184	HIST1H2AE	coding RNA	0.52	1.86E-04	1.91E-02	1.91E-02
185	SNORA43	non-coding RNA	1.50	2.13E-04	2.17E-02	2.17E-02
186	SNORA61	non-coding RNA	0.64	2.15E-04	2.19E-02	2.19E-02
187	RPS2	coding RNA	1.42	2.18E-04	2.21E-02	2.21E-02
188	SNORD114-13	non-coding RNA	0.53	2.27E-04	2.28E-02	2.28E-02
189	UNC45B	coding RNA	0.56	3.09E-04	3.09E-02	3.09E-02
190	SNORD92	non-coding RNA	0.95	3.17E-04	3.16E-02	3.16E-02
191	ACTG1	coding RNA	1.17	3.26E-04	3.23E-02	3.23E-02
192	SPP1	coding RNA	1.51	3.47E-04	3.42E-02	3.42E-02
193	SNORD65	non-coding RNA	0.65	3.63E-04	3.56E-02	3.56E-02
194	ATP5EP2	non-coding RNA	0.64	4.06E-04	3.96E-02	3.96E-02
195	XIST	non-coding RNA	3.53	4.58E-04	4.44E-02	4.44E-02
196	MBP	coding RNA	1.75	4.84E-04	4.67E-02	4.67E-02
197	NAP1L3	coding RNA	0.79	5.11E-04	4.91E-02	4.91E-02
198	SNORD116-24	non-coding RNA	0.84	5.43E-04	5.18E-02	5.18E-02
199	SNORD21	non-coding RNA	0.90	5.49E-04	5.22E-02	5.22E-02
200	SNORD116-4	non-coding RNA	1.44	6.06E-04	5.73E-02	5.73E-02
201	SNORA21	non-coding RNA	1.11	6.27E-04	5.90E-02	5.90E-02
202	HIST1H4C	coding RNA	0.70	6.44E-04	6.03E-02	6.03E-02
203	SNORD18B	non-coding RNA	0.43	7.10E-04	6.61E-02	6.61E-02
204	SNORD43	non-coding RNA	0.83	8.08E-04	7.49E-02	7.49E-02
205	SNORD16	non-coding RNA	0.73	8.30E-04	7.66E-02	7.66E-02
206	SNORA44	non-coding RNA	0.85	8.52E-04	7.82E-02	7.81E-02
207	SNORD115-43_dup1	non-coding RNA	0.15	8.67E-04	7.87E-02	7.81E-02
208	SNORD115-36_dup1	non-coding RNA	0.15	8.67E-04	7.87E-02	7.81E-02
209	SNORD115-29_dup1	non-coding RNA	0.15	8.67E-04	7.87E-02	7.81E-02
210	SNORD115-11_dup1	non-coding RNA	0.15	8.67E-04	7.87E-02	7.81E-02
211	HIST1H2BH	coding RNA	32.38	9.62E-04	8.63E-02	8.63E-02
212	SNORD6	non-coding RNA	1.24	9.72E-04	8.67E-02	8.67E-02

Table A.1 Significantly changed genes in alcoholic brains.
Among all genes from the RefSeq Genes model, the most significantly changed genes in alcoholics found in the MA plot analysis using the DEGseq package (p-value < 0.001).

Num	junction location	type	fold change	p-value	q-value(Benjamini et al. 1995)	q-value(Storey et al. 2003)
1	chrM:2169-7250	novel	11,849.87	0.00E+00	0.00E+00	0.00E+00
2	chrM:1753-9606	novel	5.73	0.00E+00	0.00E+00	0.00E+00
3	chrM:2169-2908	novel	0.00	0.00E+00	0.00E+00	0.00E+00
4	chrM:2169-2324	novel	0.36	0.00E+00	0.00E+00	0.00E+00
5	chrM:2169-3006	novel	1.83	0.00E+00	0.00E+00	0.00E+00
6	chrM:2883-2987	novel	7.04	0.00E+00	0.00E+00	0.00E+00
7	chrM:3042-3284	novel	1.35	0.00E+00	0.00E+00	0.00E+00
8	chrM:2300-10041	novel	0.80	0.00E+00	0.00E+00	0.00E+00
9	chrM:3027-3307	novel	0.07	0.00E+00	0.00E+00	0.00E+00
10	chrM:1717-10770	novel	11.68	0.00E+00	0.00E+00	0.00E+00
11	chrM:2779-3092	novel	0.46	0.00E+00	0.00E+00	0.00E+00
12	chr11:10480107-10486672	novel	0.37	0.00E+00	0.00E+00	0.00E+00
13	chrM:2169-5996	novel	0.40	0.00E+00	0.00E+00	0.00E+00
14	chrM:2024-3000	novel	860.00	0.00E+00	0.00E+00	0.00E+00
15	chrM:2024-6268	novel	224.74	0.00E+00	0.00E+00	0.00E+00
16	chr17:40341044-40343508	known	2.31	0.00E+00	0.00E+00	0.00E+00
17	chrM:2688-2851	novel	0.35	0.00E+00	0.00E+00	0.00E+00
18	chrM:2692-2851	novel	5,618.06	0.00E+00	0.00E+00	0.00E+00
19	chrM:1896-7324	novel	0.06	2.39E-305	1.52E-301	7.67E-302
20	chrM:1753-1932	novel	7.14	1.17E-283	7.07E-280	3.57E-280
21	chr9:5083804-5084484	novel	3.01	9.59E-280	5.51E-276	2.78E-276
22	chrM:1717-15939	novel	0.17	2.03E-269	1.11E-265	5.62E-266
23	chrM:2169-2223	novel	0.14	2.43E-258	1.28E-254	6.45E-255
24	chrM:2813-2968	novel	0.76	2.44E-255	1.23E-251	6.21E-252
25	chrM:2169-8925	novel	0.00	8.43E-251	4.07E-247	2.06E-247
26	chrM:2974-6806	novel	11.55	7.59E-249	3.52E-245	1.78E-245
27	chrM:3072-6899	novel	0.39	3.75E-247	1.67E-243	8.46E-244
28	chrM:1784-2405	novel	0.00	1.03E-231	4.45E-228	2.25E-228
29	chrM:2297-3048	novel	0.73	3.07E-229	1.28E-225	6.45E-226
30	chrM:1717-2987	novel	3.63	1.96E-225	7.87E-222	3.98E-222
31	chrM:2813-3883	novel	1.26	4.40E-225	1.71E-221	8.65E-222
32	chrM:2961-3179	novel	0.25	5.01E-225	1.89E-221	9.54E-222
33	chrM:2904-3006	novel	1.74	1.39E-223	5.07E-220	2.56E-220
34	chrM:2753-3092	novel	0.42	1.02E-220	3.61E-217	1.82E-217
35	chr17:21947572-21947942	novel	0.71	5.34E-212	1.84E-208	9.30E-209
36	chrM:2796-6785	novel	232.30	2.01E-205	6.72E-202	3.40E-202
37	chrM:2169-2242	novel	0.00	2.36E-202	7.70E-199	3.89E-199
38	chrM:1746-7783	novel	40.93	5.21E-202	1.66E-198	8.36E-199
39	chrM:2796-2849	novel	0.54	1.85E-197	5.73E-194	2.90E-194
40	chrM:1753-2242	novel	0.36	1.12E-191	3.38E-188	1.71E-188
41	chrM:2328-6487	novel	0.10	1.47E-187	4.31E-184	2.18E-184
42	chr8:27517830-27518357	known	1.86	1.59E-186	4.56E-183	2.30E-183
43	chrM:1717-5327	novel	342.81	1.47E-184	4.12E-181	2.08E-181
44	chrM:2675-2905	novel	29.35	8.99E-182	2.46E-178	1.25E-178
45	chrM:2169-2391	novel	0.74	4.43E-175	1.19E-171	6.00E-172
46	chrM:1753-12225	novel	0.07	1.90E-169	4.99E-166	2.52E-166
47	chrM:2800-2922	novel	0.37	1.91E-165	4.90E-162	2.48E-162
48	chr8:27512070-27513213	known	1.77	2.41E-165	6.05E-162	3.06E-162
49	chrM:3013-11952	novel	0.50	8.64E-161	2.13E-157	1.07E-157
50	chrM:1753-6714	novel	0.23	5.26E-159	1.27E-155	6.41E-156

Num	junction location	type	fold change	p-value	q-value(Benjamini et al. 1995)	q-value(Storey et al. 2003)
51	chr5:79876547-79981970	novel	0.58	2.07E-156	4.90E-153	2.48E-153
52	chrM:3013-7230	novel	0.13	2.46E-155	5.72E-152	2.89E-152
53	chrM:2306-2699	novel	0.00	3.44E-155	7.82E-152	3.95E-152
54	chrM:1717-9533	novel	0.03	1.14E-152	2.54E-149	1.28E-149
55	chrM:2328-15005	novel	2.98	4.37E-152	9.58E-149	4.84E-149
56	chrM:2024-6361	novel	1.29	7.86E-145	1.69E-141	8.56E-142
57	chr18:72829857-72830899	known	2.73	5.36E-135	1.13E-131	5.73E-132
58	chrM:2100-2538	novel	0.17	4.67E-128	9.72E-125	4.91E-125
59	chrM:1717-9365	novel	0.02	1.10E-126	2.26E-123	1.14E-123
60	chrM:2798-11562	novel	0.00	1.71E-126	3.44E-123	1.74E-123
61	chrM:721-6739	novel	5.23	9.59E-125	1.90E-121	9.59E-122
62	chrM:2466-15931	novel	0.35	9.09E-124	1.77E-120	8.94E-121
63	chrM:2331-14589	novel	1,611.32	1.67E-123	3.19E-120	1.61E-120
64	chrM:1784-7078	novel	7.39	4.24E-122	7.99E-119	4.04E-119
65	chr2:70777037-70784960	known	2.85	2.39E-121	4.43E-118	2.24E-118
66	chr11:10486130-10486790	novel	0.00	1.10E-120	2.01E-117	1.01E-117
67	chrM:2172-15807	novel	0.56	5.36E-120	9.66E-117	4.88E-117
68	chr17:21947905-21949761	novel	0.33	1.44E-118	2.55E-115	1.29E-115
69	chrM:743-3520	novel	0.56	1.92E-118	3.36E-115	1.70E-115
70	chrM:1890-2925	novel	1.94	2.51E-118	4.33E-115	2.19E-115
71	chrM:1717-2485	novel	5.58	1.34E-116	2.28E-113	1.15E-113
72	chrM:1925-2607	novel	0.15	1.79E-115	3.00E-112	1.52E-112
73	chrM:2800-3006	novel	0.00	1.93E-113	3.19E-110	1.61E-110
74	chrM:2100-5906	novel	1,428.42	1.07E-112	1.75E-109	8.85E-110
75	chrM:1842-2068	novel	1,414.85	7.04E-112	1.13E-108	5.73E-109
76	chrM:2058-2538	novel	0.71	2.10E-110	3.34E-107	1.69E-107
77	chrM:2267-9349	novel	0.24	5.84E-107	9.15E-104	4.62E-104
78	chrM:2100-4408	novel	0.67	5.26E-106	8.14E-103	4.11E-103
79	chrM:1881-2450	novel	0.21	4.04E-100	6.18E-97	3.12E-97
80	chr10:57029251-57030307	novel	5.15	2.59E-98	3.90E-95	1.97E-95
81	chrM:1968-7063	novel	1,136.58	1.29E-94	1.91E-91	9.67E-92
82	chrM:1753-2063	novel	0.01	8.92E-94	1.31E-90	6.63E-91
83	chrM:1784-8951	novel	0.00	9.17E-93	1.33E-89	6.74E-90
84	chrM:1896-3077	novel	0.36	1.56E-91	2.24E-88	1.13E-88
85	chrM:2169-2316	novel	1.11	3.39E-90	4.81E-87	2.43E-87
86	chrM:1717-2626	novel	17.33	4.40E-90	6.17E-87	3.12E-87
87	chrM:2480-6133	novel	0.49	6.89E-88	9.55E-85	4.83E-85
88	chrM:2024-7075	novel	1,025.39	2.08E-87	2.85E-84	1.44E-84
89	chrM:2779-2977	novel	0.05	2.82E-87	3.82E-84	1.93E-84
90	chrM:1862-6275	novel	0.00	3.25E-87	4.35E-84	2.20E-84
91	chrM:1759-3079	novel	980.84	1.84E-84	2.44E-81	1.23E-81
92	chrM:2297-2476	novel	2.91	7.15E-84	9.38E-81	4.74E-81
93	chrM:2675-3092	novel	0.22	3.96E-83	5.14E-80	2.60E-80
94	chrM:2904-3765	novel	1.23	1.39E-81	1.78E-78	9.01E-79
95	chr11:2373327-2373674	known	2.29	3.33E-80	4.23E-77	2.14E-77
96	chrM:2974-6808	novel	10.69	3.77E-78	4.74E-75	2.40E-75
97	chrM:1753-3158	novel	2.55	6.11E-78	7.60E-75	3.84E-75
98	chr21:26276662-26294200	known	2.11	4.98E-77	6.13E-74	3.10E-74
99	chrM:2312-3000	novel	0.00	4.16E-76	5.06E-73	2.56E-73
100	chr17:21947029-21947579	novel	4.91	6.06E-76	7.31E-73	3.69E-73

Table A.2 Top 100 most significantly changed splicing junctions in alcoholic brains. Among all splicing junction found from RNA-seq, the top 100 significantly changed junctions in alcoholics found from statistical analysis using the DEGseq package.

location	Type	cSFC477	cSFC366	cSFC377	cSFC232	cSFC474	cSFC228	cSFC209	cSFC442	cSFC295	cSFC339	cSFC451	cSFC463	cSFC257	cSFC329	cSFC459
chr6:29,681,453-29,682,146	known	5	12	0	10	17	12	0	9	16	5	3	10	2	24	17
chr6:29,682,241-29,682,652	known	1	18	15	8	12	10	13	4	3	4	9	17	18	18	11
chr6:29,688,371-29,688,998	known	0	10	2	1	6	13	1	1	9	4	2	10	22	12	9
chr6:29,697,962-29,699,060	known	7	20	0	0	0	9	0	17	4	0	7	0	25	27	1
chr6:29,688,953-29,699,060	partial_novel	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0
chr6:29,684,501-29,689,002	novel	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0
chr6:29,679,125-29,689,002	novel	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0
chr6:29,678,081-29,678,381	novel	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
total mapped read		6,970,403	5,344,243	3,765,907	5,895,336	8,254,606	5,421,041	9,886,852	8,791,261	8,885,572	10,021,595	9,140,872	13,666,586	13,509,520	32,804,613	14,280,904

location	Type	aSFC217	aSFC327	aSFC210	aSFC234	aSFC476	aSFC204	aSFC303	aSFC313	aSFC332	aSFC315	aSFC351	aSFC260	aSFC330	aSFC430
chr6:29,681,453-29,682,146	known	1	20	9	20	1	4	1	13	7	2	2	52	14	6
chr6:29,682,241-29,682,652	known	8	97	2	9	5	30	0	11	28	3	0	16	13	7
chr6:29,688,371-29,688,998	known	0	61	0	0	5	9	0	9	20	1	3	30	7	0
chr6:29,697,962-29,699,060	known	3	0	7	0	1	0	0	18	0	0	6	49	35	4
chr6:29,688,953-29,699,060	partial_novel	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chr6:29,684,501-29,689,002	novel	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chr6:29,679,125-29,689,002	novel	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chr6:29,678,081-29,678,381	novel	0	0	0	0	0	0	0	0	0	0	3	0	0	0
total mapped read		5,009,950	16,553,222	7,855,559	6,401,077	8,285,665	6,502,796	6,089,416	5,858,693	8,923,080	8,074,811	8,029,271	38,820,126	9,691,588	8,872,790

Table A.3 Mapped read counts for not changed GABAB1 splicing junctions in alcoholics.
Sample names starting from a or c represent alcoholic or control samples, respectively.

REFERENCES

Addolorato, G., L. Leggio, et al. (2006). Baclofen in the treatment of alcohol withdrawal syndrome: a comparative study vs diazepam. Am J Med **119**(3): 276 e213-278.

Aird, D., M. G. Ross, et al. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol **12**(2): R18.

Alexander, W. H. and J. W. Brown (2011). Medial prefrontal cortex as an action-outcome predictor. Nat Neurosci **14**(10): 1338-1344.

Allan, A. M., D. Burnett, et al. (1991). Ethanol-induced changes in chloride flux are mediated by both GABA(A) and GABA(B) receptors. Alcohol Clin Exp Res **15**(2): 233-237.

Allen, S. E., R. B. Darnell, et al. (2010). The neuronal splicing factor Nova controls alternative splicing in N-type and P-type CaV2 calcium channels. Channels (Austin) **4**(6): 483-489.

Avissar, M., M. D. McClean, et al. (2009). MicroRNA expression in head and neck cancer associates with alcohol consumption and survival. Carcinogenesis **30**(12): 2059-2063.

Bell, T. J., K. Y. Miyashiro, et al. (2010). Intron retention facilitates splice variant diversity in calcium-activated big potassium channel populations. Proc Natl Acad Sci U S A **107**(49): 21152-21157.

Bell, T. J., K. Y. Miyashiro, et al. (2008). Cytoplasmic BK(Ca) channel intron-containing mRNAs contribute to the intrinsic excitability of hippocampal neurons. Proc Natl Acad Sci U S A **105**(6): 1901-1906.

Bernal, A., K. Crammer, et al. (2007). Global discriminative learning for higher-accuracy computational gene prediction. PLoS Comput Biol **3**(3): e54.

Bettler, B., K. Kaupmann, et al. (2004). Molecular structure and physiological functions of GABA(B) receptors. Physiol Rev **84**(3): 835-867.

Billinton, A., A. O. Ige, et al. (2001). Advances in the molecular understanding of GABA(B) receptors. Trends Neurosci **24**(5): 277-282.

Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem **72**: 291-336.

Bouza, C., M. Angeles, et al. (2004). Efficacy and safety of naltrexone and acamprosate in the treatment of alcohol dependence: a systematic review. Addiction **99**(7): 811-828.

Bulanova, E., V. Budagian, et al. (2007). Soluble Interleukin IL-15Ralpha is generated by alternative splicing or proteolytic cleavage and forms functional complexes with IL-15. J Biol Chem **282**(18): 13167-13179.

Calver, A. R., A. D. Medhurst, et al. (2000). The expression of GABA(B1) and GABA(B2) receptor subunits in the cNS differs from that in peripheral tissues. Neuroscience **100**(1): 155-170.

Cammalleri, M., A. Brancucci, et al. (2002). Gamma-hydroxybutyrate reduces GABA(A)-mediated inhibitory postsynaptic potentials in the CA1 region of hippocampus. Neuropsychopharmacology **27**(6): 960-969.

Canovas, A., G. Rincon, et al. (2010). SNP discovery in the bovine milk transcriptome using RNA-Seq technology. Mamm Genome **21**(11-12): 592-598.

Chang, W., C. Tu, et al. (2007). Complex formation with the Type B gamma-aminobutyric acid receptor affects the expression and signal transduction of the extracellular calcium-sensing receptor. Studies with HEK-293 cells and neurons. J Biol Chem **282**(34): 25030-25040.

Chao, J. and E. J. Nestler (2004). Molecular neurobiology of drug addiction. Annu Rev Med **55**: 113-132.

Cloonan, N., A. R. Forrest, et al. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods **5**(7): 613-619.

Couve, A., S. J. Moss, et al. (2000). GABAB receptors: a new paradigm in G protein signaling. Mol Cell Neurosci **16**(4): 296-312.

Crofts, L. A., M. S. Hancock, et al. (1998). Multiple promoters direct the tissue-specific expression of novel N-terminal variant human vitamin D receptor gene transcripts. Proc Natl Acad Sci U S A **95**(18): 10529-10534.

Crooks, G. E., G. Hon, et al. (2004). WebLogo: a sequence logo generator. Genome Res **14**(6): 1188-1190.

Cui, P., Q. Lin, et al. (2010). A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. Genomics **96**(5): 259-265.

Degner, J. F., J. C. Marioni, et al. (2009). Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. Bioinformatics **25**(24): 3207-3212.

Dimon, M. T., K. Sorber, et al. (2010). HMMSplicer: a tool for efficient and sensitive discovery of known and novel splice junctions in RNA-Seq data. PLoS One **5**(11): e13875.

Dunn, J. J. (1976). RNase III cleavage of single-stranded RNA. Effect of ionic strength on the fideltiy of cleavage. J Biol Chem **251**(12): 3807-3814.

Dzitoyeva, S., N. Dimitrijevic, et al. (2003). Gamma-aminobutyric acid B receptor 1 mediates behavior-impairing actions of alcohol in Drosophila: adult RNA interference and pharmacological evidence. Proc Natl Acad Sci U S A **100**(9): 5485-5490.

Elela, S. A., H. Igel, et al. (1996). RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. Cell **85**(1): 115-124.

Enserink, M. (2011). Addiction research. Anonymous alcoholic bankrolls trial of controversial therapy. Science **332**(6030): 653.

Ezzat, S., L. Zheng, et al. (2001). A soluble dominant negative fibroblast growth factor receptor 4 isoform in human MCF-7 breast cancer cells. Biochem Biophys Res Commun **287**(1): 60-65.

Farre, D., R. Roset, et al. (2003). Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res **31**(13): 3651-3653.

Fergusson, D. M., J. M. Boden, et al. (2009). Tests of causal links between alcohol abuse or dependence and major depression. Arch Gen Psychiatry **66**(3): 260-266.

Fisher, S., A. Barry, et al. (2011). A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol **12**(1): R1.

Flatscher-Bader, T., M. van der Brug, et al. (2005). Alcohol-responsive genes in the frontal cortex and nucleus accumbens of human alcoholics. J Neurochem **93**(2): 359-370.

Fodor, A. A. and R. W. Aldrich (2009). Convergent evolution of alternative splices at domain boundaries of the BK channel. Annu Rev Physiol **71**: 19-36.

Gao, L., Z. Fang, et al. (2011). Length bias correction for RNA-seq data in gene set analyses. Bioinformatics **27**(5): 662-669.

Garber, M., M. G. Grabherr, et al. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods **8**(6): 469-477.

Garbutt, J. C., A. B. Kampov-Polevoy, et al. (2010). Efficacy and safety of baclofen for alcohol dependence: a randomized, double-blind, placebo-controlled trial. Alcohol Clin Exp Res **34**(11): 1849-1857.

Gassmann, M., C. Haller, et al. (2005). The RXR-type endoplasmic reticulum-retention/retrieval signal of GABAB1 requires distant spacing from the membrane to function. Mol Pharmacol **68**(1): 137-144.

Gordis, E. (2000). Contributions of behavioral science to alcohol research: understanding who is at risk and why. Exp Clin Psychopharmacol **8**(3): 264-270.

Gracheva, E. O., J. F. Cordero-Morales, et al. (2011). Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. Nature **476**(7358): 88-91.

Griffiths-Jones, S., H. K. Saini, et al. (2008). miRBase: tools for microRNA genomics. Nucleic Acids Res **36**(Database issue): D154-158.

Grosso, A. R., A. Q. Gomes, et al. (2008). Tissue-specific splicing factor gene expression signatures. Nucleic Acids Res **36**(15): 4823-4832.

Guigo, R., P. Agarwal, et al. (2000). An assessment of gene prediction accuracy in large DNA sequences. Genome Res **10**(10): 1631-1642.

Hansen, K. D., S. E. Brenner, et al. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res **38**(12): e131.

Hardy, P. A., W. Chen, et al. (1999). Chronic ethanol exposure and withdrawal influence NMDA receptor subunit and splice variant mRNA expression in the rat cerebral cortex. Brain Res **819**(1-2): 33-39.

Heffernan, T. M. (2008). The impact of excessive alcohol use on prospective memory: a brief review. Curr Drug Abuse Rev **1**(1): 36-41.

Hodges, E., Z. Xuan, et al. (2007). Genome-wide in situ exon capture for selective resequencing. Nat Genet **39**(12): 1522-1527.

Holter, J., J. Davies, et al. (2005). Identification of two further splice variants of GABABR1 characterizes the conserved micro-exon 4 as a hot spot for regulated splicing in the rat brain. J Mol Neurosci **26**(1): 99-108.

Huang, J., L. L. Levitsky, et al. (2009). Novel P2 promoter-derived HNF4alpha isoforms with different N-terminus generated by alternate exon insertion. Exp Cell Res **315**(7): 1200-1211.

Isomoto, S., M. Kaibara, et al. (1998). Cloning and tissue distribution of novel splice variants of the rat GABAB receptor. Biochem Biophys Res Commun **253**(1): 10-15.

Jang, D. J., S. W. Park, et al. (2010). N termini of apPDE4 isoforms are responsible for targeting the isoforms to different cellular membranes. Learn Mem **17**(9): 469-479.

Jensen, G. B. and B. Pakkenberg (1993). Do alcoholics drink their neurons away?" Lancet **342**(8881): 1201-1204.

Johnson, B. A. (2004). An overview of the development of medications including novel anticonvulsants for the treatment of alcohol dependence. Expert Opin Pharmacother **5**(9): 1943-1955.

Kenna, G. A., J. E. McGeary, et al. (2004). Pharmacotherapy, pharmacogenomics, and the future of alcohol dependence treatment, part 1. Am J Health Syst Pharm **61**(21): 2272-2279.

Kenna, G. A., J. E. McGeary, et al. (2004). Pharmacotherapy, pharmacogenomics, and the future of alcohol dependence treatment, Part 2. Am J Health Syst Pharm **61**(22): 2380-2388.

Kishore, S. and S. Stamm (2006). The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. Science **311**(5758): 230-232.

Kril, J. J. and G. M. Halliday (1999). Brain shrinkage in alcoholics: a decade on and what have we learned?" Prog Neurobiol **58**(4): 381-387.

Kril, J. J. and C. G. Harper (1989). Neuronal counts from four cortical regions of alcoholic brains. Acta Neuropathol **79**(2): 200-204.

Krupitsky, E. M., A. M. Burakov, et al. (1993). Baclofen administration for the treatment of affective disorders in alcoholic patients. Drug Alcohol Depend **33**(2): 157-163.

Lamond, A. I. and A. A. Travers (1985). Genetically separable functional elements mediate the optimal expression and stringent regulation of a bacterial tRNA gene. Cell **40**(2): 319-326.

Lamontagne, B., S. Larose, et al. (2001). The RNase III family: a conserved structure and expanding functions in eukaryotic dsRNA metabolism. Curr Issues Mol Biol **3**(4): 71-78.

Lareau, L. F., R. E. Green, et al. (2004). The evolving roles of alternative splicing. Curr Opin Struct Biol **14**(3): 273-282.

Lavoie, J. and R. F. Butterworth (1995). Reduced activities of thiamine-dependent enzymes in brains of alcoholics in the absence of Wernicke's encephalopathy. Alcohol Clin Exp Res **19**(4): 1073-1077.

Lee, C., R. D. Mayfield, et al. (2010). Intron 4 containing novel GABAB1 isoforms impair GABAB receptor function. PLoS One **5**(11): e14044.

Lee, S., C. H. Seo, et al. (2011). Accurate quantification of transcriptome from RNA-Seq data by effective length normalization. Nucleic Acids Res **39**(2): e9.

Lewohl, J. M., Y. O. Nunez, et al. (2011). Up-Regulation of MicroRNAs in Brain of Human Alcoholics. Alcohol Clin Exp Res.

Lewohl, J. M., L. Wang, et al. (2000). Gene expression in human alcoholism: microarray analysis of frontal cortex. Alcohol Clin Exp Res **24**(12): 1873-1882.

Lewohl, J. M., W. R. Wilson, et al. (1999). G-protein-coupled inwardly rectifying potassium channels are targets of alcohol action. Nat Neurosci **2**(12): 1084-1090.

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

Linsen, S. E., E. de Wit, et al. (2009). Limitations and possibilities of small RNA digital gene expression profiling. Nat Methods **6**(7): 474-476.

Liu, J., J. M. Lewohl, et al. (2004). Gene expression profiling of individual cases reveals consistent transcriptional changes in alcoholic human brain. J Neurochem **90**(5): 1050-1058.

Liu, J., J. M. Lewohl, et al. (2006). Patterns of gene expression in the frontal cortex discriminate alcoholic from nonalcoholic individuals. Neuropsychopharmacology **31**(7): 1574-1582.

Liu, Q. R., L. Lu, et al. (2006). Rodent BDNF genes, novel promoters, novel splice variants, and regulation by cocaine. Brain Res **1067**(1): 1-12.

Lopez, A. D., C. D. Mathers, et al. (2006). Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet **367**(9524): 1747-1757.

Maccioni, P., N. Fantini, et al. (2008). Specific reduction of alcohol's motivational properties by the positive allosteric modulator of the GABAB receptor, GS39783--comparison with the effect of the GABAB receptor direct agonist, baclofen. Alcohol Clin Exp Res **32**(9): 1558-1564.

Maccioni, P., D. Pes, et al. (2008). gamma-Hydroxybutyric acid (GHB) suppresses alcohol's motivational properties in alcohol-preferring rats. Alcohol **42**(2): 107-113.

Maher, C. A., C. Kumar-Sinha, et al. (2009). Transcriptome sequencing to detect gene fusions in cancer. Nature **458**(7234): 97-101.

Mangone, M., A. P. Manoharan, et al. (2010). The landscape of C. elegans 3'UTRs. Science **329**(5990): 432-435.

Mardis, E. R. (2008). Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet **9**: 387-402.

Marioni, J. C., C. E. Mason, et al. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res **18**(9): 1509-1517.

Mayfield, R. D., J. M. Lewohl, et al. (2002). Patterns of gene expression are altered in the frontal and motor cortices of human alcoholics. J Neurochem **81**(4): 802-813.

McGinnis, J. M. and W. H. Foege (1993). Actual causes of death in the United States. JAMA **270**(18): 2207-2212.

McGovern, P. E., M. Christofidou-Solomidou, et al. (2010). Anticancer activity of botanical compounds in ancient fermented beverages (review). Int J Oncol **37**(1): 5-14.

McGovern, P. E., A. Mirzoian, et al. (2009). Ancient Egyptian herbal wines. Proc Natl Acad Sci U S A **106**(18): 7361-7366.

McGovern, P. E., J. Zhang, et al. (2004). Fermented beverages of pre- and proto-historic China. Proc Natl Acad Sci U S A **101**(51): 17593-17598.

McPherson, J. D. (2009). Next-generation gap. Nat Methods **6**(11 Suppl): S2-5.

Messeguer, X., R. Escudero, et al. (2002). PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics **18**(2): 333-334.

Metzker, M. L. (2010). Sequencing technologies - the next generation. Nat Rev Genet **11**(1): 31-46.

Modrek, B. and C. J. Lee (2003). Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat Genet **34**(2): 177-180.

Morabito, M. V., R. J. Ulbricht, et al. (2010). High-throughput multiplexed transcript analysis yields enhanced resolution of 5-hydroxytryptamine 2C receptor mRNA editing profiles. Mol Pharmacol **77**(6): 895-902.

Morin, R., M. Bainbridge, et al. (2008). Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques **45**(1): 81-94.

Mortazavi, A., B. A. Williams, et al. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods **5**(7): 621-628.

Mosley, B., M. P. Beckmann, et al. (1989). The murine interleukin-4 receptor: molecular cloning and characterization of secreted and membrane bound forms. Cell **59**(2): 335-348.

Muller, D., R. D. Koch, et al. (1985). [Neurophysiologic findings in chronic alcohol abuse]. Psychiatr Neurol Med Psychol (Leipz) **37**(3): 129-132.

Nagalakshmi, U., Z. Wang, et al. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science **320**(5881): 1344-1349.

Nam, J. W., K. R. Shin, et al. (2005). Human microRNA prediction through a probabilistic co-learning model of sequence and structure. Nucleic Acids Res **33**(11): 3570-3581.

Neiman, J. (1998). Alcohol as a risk factor for brain damage: neurologic aspects. Alcohol Clin Exp Res **22**(7 Suppl): 346S-351S.

Ng, S. B., E. H. Turner, et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. Nature **461**(7261): 272-276.

O'Brien, C. P. and E. L. Gardner (2005). Critical assessment of how to study addiction and its treatment: human and non-human animal models. Pharmacol Ther **108**(1): 18-58.

Olegard, R., K. G. Sabel, et al. (1979). Effects on the child of alcohol abuse during pregnancy. Retrospective and prospective studies. Acta Paediatr Scand Suppl **275**: 112-121.

Park, H., J. A. Lee, et al. (2005). An Aplysia type 4 phosphodiesterase homolog localizes at the presynaptic terminals of Aplysia neuron and regulates synaptic facilitation. J Neurosci **25**(39): 9037-9045.

Patrono, C., C. Casali, et al. (2002). Missense and splice site mutations in SPG4 suggest loss-of-function in dominant spastic paraplegia. J Neurol **249**(2): 200-205.

Pepke, S., B. Wold, et al. (2009). Computation for ChIP-seq and RNA-seq studies. Nat Methods **6**(11 Suppl): S22-32.

Pickrell, J. K., A. A. Pai, et al. (2010). Noisy splicing drives mRNA isoform diversity in human cells. PLoS Genet **6**(12): e1001236.

Pierucci-Lagha, A. and C. Derouesne (2003). [Alcoholism and aging. 2. Alcoholic dementia or alcoholic cognitive impairment?]. Psychol Neuropsychiatr Vieil **1**(4): 237-249.

Pietrzykowski, A. Z., R. M. Friesen, et al. (2008). Posttranscriptional regulation of BK channel splice variant stability by miR-9 underlies neuroadaptation to alcohol. Neuron **59**(2): 274-287.

Pihlajamaki, J., C. Lerin, et al. (2011). Expression of the splicing factor gene SFRS10 is reduced in human obesity and contributes to enhanced lipogenesis. Cell Metab **14**(2): 208-218.

Piomelli, D. (2001). Cannabinoid activity curtails cocaine craving. Nat Med **7**(10): 1099-1100.

Ponomarev, I., R. Maiya, et al. (2006). Transcriptional signatures of cellular plasticity in mice lacking the alpha1 subunit of GABAA receptors. J Neurosci **26**(21): 5673-5683.

Ponomarev, I., S. Wang, et al. (2011). Gene co-expression networks in human brain identify epigenetic modifications in alcohol dependence. J Neurosci: In submission.

Roberto, M., S. G. Madamba, et al. (2003). Ethanol increases GABAergic transmission at both pre- and postsynaptic sites in rat central amygdala neurons. Proc Natl Acad Sci U S A **100**(4): 2053-2058.

Roberts, A., C. Trapnell, et al. (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol **12**(3): R22.

Romaniuk, E., L. W. McLaughlin, et al. (1982). The effect of acceptor oligoribonucleotide sequence on the T4 RNA ligase reaction. Eur J Biochem **125**(3): 639-643.

SchuLtt, C., A. Hilfiker, et al. (1998). virilizer regulates Sex-lethal in the germline of Drosophila melanogaster. Development **125**(8): 1501-1507.

Schwarz, D. A., G. Barry, et al. (2000). Characterization of gamma-aminobutyric acid receptor GABAB(1e), a GABAB(1) splice variant encoding a truncated receptor. J Biol Chem **275**(41): 32174-32181.

Schweikert, G., A. Zien, et al. (2009). mGene: accurate SVM-based gene finding with an application to nematode genomes. Genome Res **19**(11): 2133-2143.

Schweitzer, P., M. Roberto, et al. (2004). gamma-hydroxybutyrate increases a potassium current and decreases the H-current in hippocampal neurons via GABAB receptors. J Pharmacol Exp Ther **311**(1): 172-179.

Sellman, D., J. Connor, et al. (2009). Alcohol cardio-protection has been talked up. N Z Med J **122**(1303): 97-101.

Shen, Y., J. Catchen, et al. (2011). Identification of transcriptome SNPs between Xiphophorus lines and species for assessing allele specific gene expression within F(1) interspecies hybrids. Comp Biochem Physiol C Toxicol Pharmacol.

Shendure, J. and H. Ji (2008). Next-generation DNA sequencing. Nat Biotechnol **26**(10): 1135-1145.

Shkreta, L., U. Froehlich, et al. (2008). Anticancer drugs affect the alternative splicing of Bcl-x and other human apoptotic genes. Mol Cancer Ther **7**(6): 1398-1409.

Sinkiewicz, W. and M. Weglarz (2009). [Alcohol and wine and cardiovascular diseases in epidemiologic studies]. Przegl Lek **66**(5): 233-238.

Sokolov, B. P., L. Jiang, et al. (2003). Transcription profiling reveals mitochondrial, ubiquitin and signaling systems abnormalities in postmortem brains from subjects with a history of alcohol abuse or dependence. J Neurosci Res **72**(6): 756-767.

Sorber, K., M. T. Dimon, et al. (2011). RNA-Seq analysis of splicing in Plasmodium falciparum uncovers new splice junctions, alternative splicing and splicing of antisense transcripts. Nucleic Acids Res **39**(9): 3820-3835.

Stamm, S., S. Ben-Ari, et al. (2005). Function of alternative splicing. Gene **344**: 1-20.

Steiger, J. L., S. Bandyopadhyay, et al. (2004). cAMP response element-binding protein, activating transcription factor-4, and upstream stimulatory factor differentially control hippocampal GABABR1a and GABABR1b subunit gene expression through alternative promoters. J Neurosci **24**(27): 6115-6126.

Sunn, K. L., T. A. Cock, et al. (2001). Novel N-terminal variant of human VDR. Mol Endocrinol **15**(9): 1599-1609.

Testino, G. (2008). Alcoholic diseases in hepato-gastroenterology: a point of view. Hepatogastroenterology **55**(82-83): 371-377.

Tiao, J. Y., A. Bradaia, et al. (2008). The sushi domains of secreted GABA(B1) isoforms selectively impair GABA(B) heteroreceptor function. J Biol Chem **283**(45): 31005-31011.

Trapnell, C., L. Pachter, et al. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25**(9): 1105-1111.

Travers, M. T., A. J. Vallance, et al. (2003). Characterisation of an N-terminal variant of acetyl-CoA carboxylase-alpha: expression in human tissues and evolutionary aspects. Biochim Biophys Acta **1634**(3): 97-106.

Uezono, Y., M. Kanaide, et al. (2006). Coupling of GABAB receptor GABAB2 subunit to G proteins: evidence from Xenopus oocyte and baby hamster kidney cell expression system. Am J Physiol Cell Physiol **290**(1): C200-207.

Vigot, R., S. Barbieri, et al. (2006). Differential compartmentalization and distinct functions of GABAB receptor variants. Neuron **50**(4): 589-601.

Voliva, C. F., A. Aronheim, et al. (1992). B-cell factor 1 is required for optimal expression of the DRA promoter in B cells. Mol Cell Biol **12**(5): 2383-2390.

Walter, H. J., T. McMahon, et al. (2000). Ethanol regulates calcium channel subunits by protein kinase C delta -dependent and -independent mechanisms. J Biol Chem **275**(33): 25717-25722.

Wan, F. J., F. Berton, et al. (1996). Low ethanol concentrations enhance GABAergic inhibitory postsynaptic potentials in hippocampal pyramidal neurons only after block of GABAB receptors. Proc Natl Acad Sci U S A **93**(10): 5049-5054.

Wang, L., Z. Feng, et al. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics **26**(1): 136-138.

Wang, Z., M. Gerstein, et al. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet **10**(1): 57-63.

Wang, Z., P. J. Schultheis, et al. (1996). Three N-terminal variants of the AE2 Cl-/HCO3- exchanger are encoded by mRNAs transcribed from alternative promoters. J Biol Chem **271**(13): 7835-7843.

Wei, K., Z. Jia, et al. (2001). Cloning and characterization of a novel variant of rat GABA(B)R1 with a truncated C-terminus. Brain Res Mol Brain Res **89**(1-2): 103-110.

Wilke, N., M. Sganga, et al. (1994). Effects of alcohol on gene expression in neural cells. EXS **71**: 49-59.

Willerth, S. M., H. A. Pedro, et al. (2010). Development of a low bias method for characterizing viral populations using next generation sequencing technology. PLoS One **5**(10): e13564.

Yeo, G. W., E. Van Nostrand, et al. (2005). Identification and analysis of alternative splicing events conserved in human and mouse. Proc Natl Acad Sci U S A **102**(8): 2850-2855.

Yizhar, O., L. E. Fenno, et al. (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. Nature **477**(7363): 171-178.

Yoon, O. K. and R. B. Brem (2010). Noncanonical transcript forms in yeast and their regulation during environmental stress. RNA **16**(6): 1256-1267.

VITA

Chang Hoon Lee was born in Daejeon, Korea on June 30, 1975. He is the son of Choong Yul Lee and Sang Sook Lee. He received a B.S. in Biological Science in February, 2002 from Seoul National University and a M.S. in Program in Neuroscience in February, 2004 from Seoul National University. After further research at Baylor College of Medicine as a research technician II, he began his Ph.D. training in the Neuroscience Graduate Program at the University of Alabama at Birmingham in August, 2005. To take care of his family, he moved to the Institute of Cell and Molecular Biology at the University of Texas at Austin in August, 2006. He joined Dr. Harris' laboratory in the spring of 2007. His graduate training has allowed him to make important scientific contributions in understanding human GABAB1 splicing in alcoholic brains. While a graduate student, he was awarded the F. M. Jones and H. L. Bruce Endowed Graduate Fellowships in Addiction Biology and a Jones Fellowship in Alcohol and Addiction Research Travel Award from the Waggoner Center for Alcohol and Addiction Research. He also received a Professional Development Award from the University of Texas at Austin, an Association of Korean Neuroscientists Research Award from Society of Neuroscience, and a Bionet Award from Korean-American Biomedical Scientists Symposium. He was awarded the degree of Doctor of Philosophy in December, 2011.

Permanent address: 29-5 Yongdu-dong, Jung-gu, Daejeon 301-110, Republic of Korea

This dissertation was typed by Chang Hoon Lee.