12-16-2016

Assessing Epigenetic Features of GABA(A) Receptor Genes in IPSCs, IPSC-Derived Neuroepithelial Cells, and IPSC-Derived Neural Cell Cultures

Maegan Watson
University of Connecticut, maegan.gross@uconn.edu

Assessing Epigenetic Features of $GABA_A$ Receptor Genes in IPSCs, IPSC-Derived Neuroepithelial Cells, and IPSC-Derived Neural Cell Cultures

Maegan Watson

University of Connecticut School of Medicine

Department of Neuroscience

2017

Abstract

Alcohol use disorders (AUDs) affect approximately 13.9% of the population in the United States. AUDs result from a combination of biological and environmental factors, and have a strong genetic component with heritability rates estimated to be 50-60%. Alcohol consumption has been shown to enhance $GABA_A$ receptor activity, and genetic studies have shown an association between AUDs and single nucleotide polymorphisms (SNPs) within $GABA_A$ receptor subunit genes. Many groups have found a correlation between AUDs and a synonymous SNP in exon 5 of the *GABRA2* gene (rs279858 T to C; C allele associated with AUDs). However, the biological effects of this and other SNPs in this region are unknown.

Our lab has been using iPSC technology for the past few years to study AUDs. The lab has shown, using qPCR of mRNA, that a cluster of $GABA_A$ receptor subunit genes on chromosome 4p12 is expressed at minimal levels in neural cells derived from half of the iPSC lines (both controls and alcoholics) studied. Expression had a high correlation to genotype; lines with minimal expression tended to have at least one C allele of the rs279858 SNP. However, subunits for $GABA_A$ receptor subunit genes

located on other chromosomes showed robust expression in all lines. Data from RNA-seq of 12-week old iPSC-derived neural cultures showed similar results.

Results generated in this thesis project from chromatin immunoprecipitation experiments suggest that genetic factors linked to the rs279858 tag-SNP may moderate the developmental expression of the chromosome 4p12 $GABA_A$ receptor subunit gene cluster by altering epigenetic marks in the promoters of these genes. IPSCs from expresser lines had epigenetic marks typical of a bivalent promoter (both histone H3K4Me3 and H3K27Me3 marks present) for *GABRA2*. During neural development, the active H3K4Me3 modification remained with a reduction in levels of H3K27Me3 in neuroepithelial cells and differentiated neural cultures in expresser lines. There were significantly lower levels of the H3K4Me3 mark in non-expresser iPSCs and neural cell cultures. In addition, DNA methylation assays looking at a CpG island in the *GABRA2* promoter showed that fibroblasts of both expresser and non-expresser lines have very low CpG DNA methylation. However, after reprogramming, iPSCs from non-expresser lines show significant amounts of CpG island DNA methylation compared with iPSC lines which generate chromosome 4p12 $GABA_A$ receptor subunit gene expressing neural cells.

These data suggests that an AUD-linked functional genetic variation may have a direct influence on epigenetic state of the chromosome 4p12 $GABA_A$ receptor subunit gene cluster and that this effect, while appears very strong in generation of iPSCs, may have variable effects on epigenetic processes during specific times and brain regions during neural development.

Assessing Epigenetic Features of $GABA_A$ Receptor Genes in IPSCs, IPSC-Derived Neuroepithelial Cells, and IPSC-Derived Neural Cell Cultures

Maegan Watson

B.S., Franklin Pierce University, 2007

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut
2017

Approval Page

Doctor of Philosophy Dissertation

Assessing Epigenetic Features of $GABA_A$ Receptor Genes in IPSCs, IPSC-Derived Neuroepithelial Cells, and IPSC-Derived Neural Cell Cultures

Presented by:

Maegan Watson

Major Advisor _____
Jonathan Covault, M.D., Ph.D.

Associate Advisor _____
Richard Mains, Ph.D.

Associate Advisor _____
Stormy Chamberlain, Ph.D.

Associate Advisor _____
Justin Cotney, Ph.D.

University of Connecticut

2017

Acknowledgements

First off I would like to thank my advisor, Dr. Jonathan Covault, for being a terrific mentor. When I was looking for a new lab he was open and allowed me to jump in on a project and took a chance on me. For that I will be forever grateful.

Second, I would like to thank all of my friends from graduate school, especially Dr. Richard Lieberman and Kaitlin Clinton from the Covault lab for listening to my daily struggles and triumphs. And my friends Kasey Johnson, Kyle Denton, and John Wizeman for always being there to help me out during the many stresses that arose on my journey throughout graduate school.

Thirdly, I must thank my parents for their love, support, and constant encouragement. They have given me the means to get to where I am today and have instilled in me a passion for learning. My Mom, Boni Gross, is my best friend and has always believed in me every step of the way. My Dad, Stan Gross, has always shown a great interest in my research and is always excited to hear about experiments and my little “iPS’rs.” And my new Mom, Nancy Watson, who has cheered me on every step of the way.

Finally, I would like to thank my husband, Justin Watson, for keeping with me and believing in me even when I didn’t want to be around myself. You have been my greatest teacher.

I would also be remiss if I did not include our cat, Mowgli, as he has provided much joy and comfort, and has stayed up with me late into the night helping me achieve my goal.

Table of Contents

Chapter 1.

Introduction

Alcohol Use Disorders

Alcohol use in the United States of America is common as more than 60% of US adults consume alcoholic beverages regularly. The National Institute of Alcohol Abuse and Alcoholism (NIAAA) defines moderate drinking as seven drinks per week and no more than three drinks in a day for women, and 14 drinks per week and no more than four drinks consumed in a day for men. This moderate alcohol consumption may have potential beneficial effects as studies have shown that these levels of alcohol use are associated with reduced the risk of heart disease and stroke (O'Keefe, Bybee et al. 2007), type-2 diabetes (Conigrave, Hu et al. 2001, Djousse, Biggs et al. 2007), and gallstones (Leitzmann, Giovannucci et al. 1999). However, consuming alcohol in excess (drinking five or more drinks in a day more than five times in 30 days) can become hazardous to health and can cause an increase in blood pressure (Maheswaran, Gill et al. 1991, Husain, Ansari et al. 2014), liver disease (O'Shea, Dasarathy et al. 2010), and can also increase the risk for various cancers (Boffetta, Hashibe et al. 2006). In addition, excessive drinking increases the risk of developing an alcohol use disorder (AUD). The prevalence of AUDs in the US is estimated to be 13.9% (Grant, Goldstein et al. 2015), meaning approximately one in six adults who drink have some level of an AUD. This number is slightly higher in men (17.6% 12-month prevalence) than women (10.4% 12-month prevalence) (Grant, Goldstein et al. 2015). Diagnosis of AUDs is given according to the criteria in the Diagnostic Statistics Manual version 5 (DSM-5). To be

diagnosed with an AUD, the patient must meet two or more criteria provided within a 12-month period. Briefly, the criteria include: consuming larger amounts of alcohol than intended, spending a great deal of time drinking, or recovering from drinking, craving alcohol (a new addition to the DSM-5), failing to fulfill obligations, a persistent desire to cut down on alcohol intake, giving up activities, continuing the use of alcohol despite interpersonal problems, acquired tolerance, and withdrawal symptoms. AUD severity ranges from mild (only meeting two criteria) to severe (meeting six or more criteria). In addition, AUDs are multifactorial and result from a combination of environmental factors and complex genetic factors complicating efforts to develop effective treatments even further. Comparison studies of monozygotic twins and dizygotic twins indicate that heritability rates of AUDs are 50-60% (Devor and Cloninger 1989, Prescott and Kendler 1999, Goldman, Oroszi et al. 2005, Gelernter and Kranzler 2009). For initiation of use and early drinking it is thought that environmental effects outweigh genetic risks; an environment where alcohol is readily available may promote initiation of drinking, whereas an individual who is never exposed to alcohol will not initiate drinking. However, genetic effects play a greater role in the transition from moderate drinking to excessive drinking, alcohol dependence, and AUDs (Gelernter and Kranzler 2009). Because of the varying levels of severity of AUDs and the complex genetic and environmental factors, one size fits all treatments have shown limited relief in many alcoholic subjects (Litten, Egli et al. 2012). In order to develop more effective treatments, more complete understanding about the biology of AUDs is an important goal. Because there is a strong genetic component associated with AUDs, it seems a logical place to explore possible abnormalities and risk factors associated with AUDs.

The Genetics of AUDs

Genes directly involved with the metabolism of alcohol, including the genes coding for alcohol dehydrogenase, *ADH1B,* and aldehyde dehydrogenase, *ALDH2,* have been extensively studied in their role of AUD risk (Edenberg 2007). Mutations in these genes produce an accumulation of acetaldehyde, which results in the drinker feeling sick. Because of this effect, these mutations are considered to be protective against AUDs, as carriers of these mutations would generally avoid drinking alcohol and the associated adverse sensations generated by acetaldehyde.

In addition to these genes, genome-wide association studies (GWAS) looking at control subjects and alcoholic subjects have provided potential risk genes, including genes for subunits of $GABA_A$ receptors. However, the mechanism behind these variants is largely unknown. By elucidating the way in which these mutations act, we can provide more effective and personalized treatments.

$GABA_A$ Receptors and AUDs

While the exact mechanism is unknown, acute alcohol exposure is known to potentiate $GABA_A$ receptor activity in the brain (Aguayo 1990). $GABA_A$ receptors are ligand-gated ion channels permeable to chloride and facilitate fast synaptic inhibition. GABA is the major inhibitory neurotransmitter in the brain and after binding to the $GABA_A$ receptor, induces the potentiation of the ion channel. These ion channels are pentamers generated by various compositions of 19 different $GABA_A$ subunits including α1-6, β1-3, γ1-3, δ, ε, θ, π, and ρ1-3. The genes encoding the most common subunits are located on five different chromosomes; chromosome 4p12 harbors γ1, α2, α4, and

β1; chromosome 5q34 (β2, α6, α1, γ2, and π); chromosome 15q12 (β3, α5, and γ3); chromosome Xq28 (ε, α3, and θ); and chromosome 1p36 (δ). The structure of the $GABA_A$ receptor and the layout of genes across human chromosomes is shown in **Figure 1.1** (adapted from (Jacob, Moss et al. 2008)). Most commonly, the receptors are composed of two α subunits, two β subunits, and a γ subunit. However, the subunit composition for $GABA_A$ receptors varies by brain region and differences in subunit composition create unique electrophysiological and pharmacological properties (Barnard, Skolnick et al. 1998, Picton and Fisher 2007). The three most common combinations in the adult brain are α1β2γ2 (60%), α2β3γ2 (15%), and α3β3γ2 (10%) (Enoch 2008). During development, the subunit composition changes. In rats, primates, and humans, α1 subunit expression increases with development and α2 expression decreases showing an importance of the α2 subunit in early development (Laurie, Wisden et al. 1992, Hashimoto, Nguyen et al. 2009, Fillman, Duncan et al. 2010); a feature to remember in later discussion.

Numerous studies have shown associations of SNPs in a 140kb region spanning the 3' end of *GABRA2* and the intergenic region between *GABRA2* and *GABRG1* to AUDs (Covault, Gelernter et al. 2004, Edenberg, Dick et al. 2004, Lappalainen, Krupitsky et al. 2005, Fehr, Sander et al. 2006, Covault, Gelernter et al. 2008, Enoch, Hodgkinson et al. 2009, Olfson and Bierut 2012). Haplomaps of this area show high correlations of SNPs within the 140kb haplotype block to expression status of chromosome 4 $GABA_A$ receptor subunit genes in iPSC-derived neural cell culture versus SNPs outside the haplotype block, which are not associated (**Figure 1.2**). The most reproducible studies (see above) have reported an association of AUDs with a

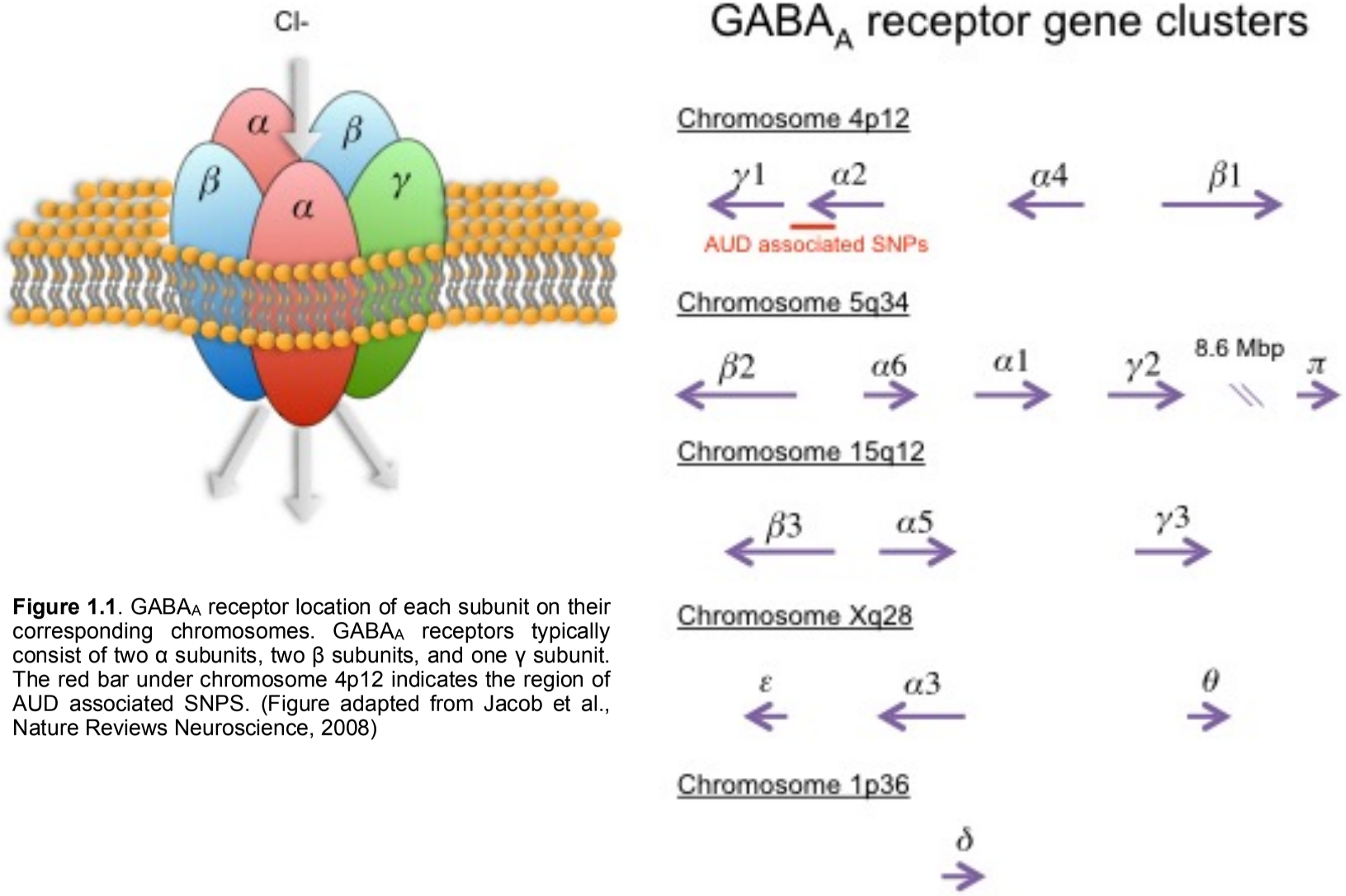


Figure 1.1. GABA$_A$ receptor location of each subunit on their corresponding chromosomes. GABA$_A$ receptors typically consist of two α subunits, two β subunits, and one γ subunit. The red bar under chromosome 4p12 indicates the region of AUD associated SNPS. (Figure adapted from Jacob et al., Nature Reviews Neuroscience, 2008)

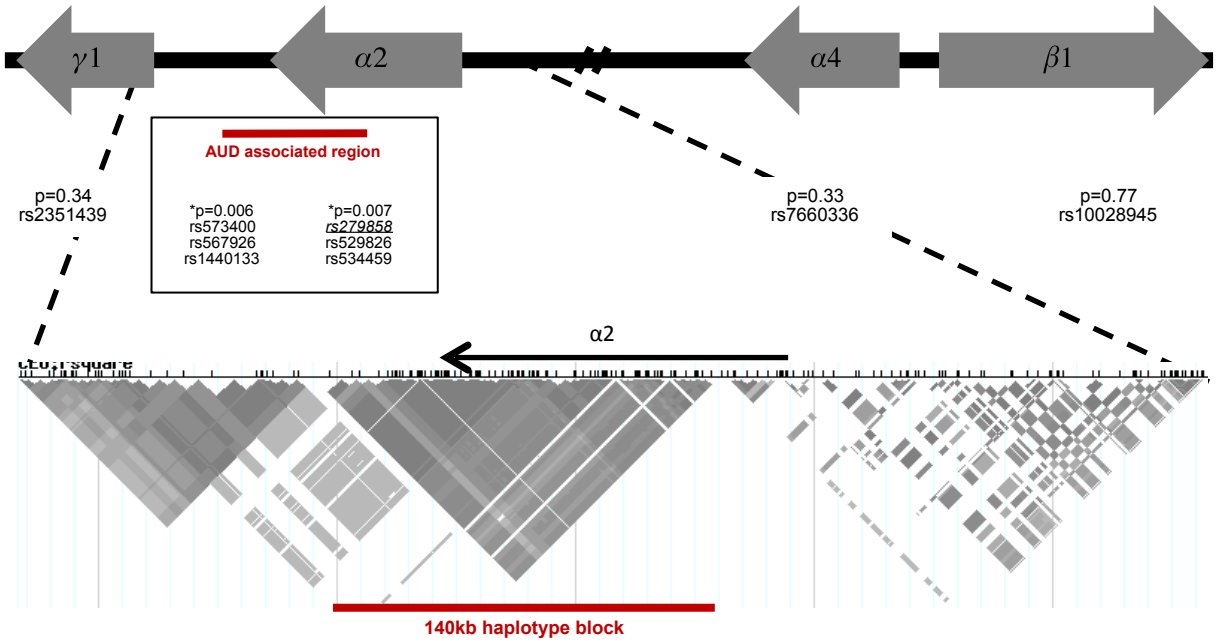


Figure 1.2. Diagram of the $GABA_A$ receptor subunit gene cluster on chromosome 4p12. The image shooting out of the chromosome is the 140kb haplotype block diagram generated by Hapmap. The SNPs with p-values are relative to the association of genotype to iPSC-derived neural cell culture chromosome 4 $GABA_A$ receptor subunit gene expression phenotype. P-values within the box have strong associations within this region and those outside of the box do not have strong associations.

synonymous tag-SNP for this region, rs279858, which is located in exon 5 of the *GABRA2* gene and involves a T to C variation; the C-allele being associated with AUD risk. It has been shown that this AUD risk C-allele has a frequency of 40% in controls and 46% in alcoholics (Covault, Gelernter et al. 2008). *GABRA2* is widespread in early development and becomes more focalized in adult mice showing the highest expression in the hippocampus and limbic regions (Pirker, Schwarzer et al. 2000). The *GABRA2* alcohol risk allele is also associated with intermediate neural phenotypes such as fast beta-frequency EEG activity (Edenberg, Dick et al. 2004), activation of the insular cortex during reward anticipation (Villafuerte, Trucco et al. 2014), changes in activation of the medial frontal cortex by alcohol cues (Kareken, Liang et al. 2010), and increased activation of the nucleus accumbens in reward anticipation (Heitzeg, Villafuerte et al. 2014). $GABA_A$ receptors containing the α2 subunit mediate the anxiolytic effects of benzodiazepines, which are used to lessen withdrawal symptoms in alcoholics. However, the more proximal molecular effects associated with the SNPs in this haplotype block have yet to be elucidated.

A New Model to Study AUDs

The study of $GABA_A$ receptor subunit genes in mice has been highly informative for the study of human diseases and disorders, including AUDs. The $GABA_A$ receptor subunit genes in mice located on chromosomes 5, 11, 7, X, and 4 are orthologous to the placement on human chromosomes 4p12, 5q34, 15q12, Xq28, and 1p36, respectively (Uusi-Oukari, Heikkila et al. 2000). However, while there are many similarities present, a very important component of AUD risk is SNPs that are specific to the human genome.

Mouse models have a great ability to show us what happens when full genes are knocked out, or knocked in, and what the consequences are, but human disease-associated genetics are more nuanced than this. So while rodent research has brought the field a long way in terms of researching basic biological genetics related to AUD risk, the next leg of research must be catered to the human condition.

One way to study disease associated genetic variants in human tissue is through iPSC technology. Since 2007 when Takahashi and Yamanaka et al. showed the reprogramming of human fibroblasts into induced pluripotent stem cells, the use of this technology has blossomed (Takahashi, Tanabe et al. 2007). Reprogramming is achieved by introducing four factors (Oct3/4, Sox2, Klf4, and c-Myc) into fibroblasts to reset cells back to a pluripotent state. One of the greatest difficulties when researching human neurological disorders and diseases is obtaining viable tissue with which to perform research. The options were either to mimic the illness in an animal model, or obtain post-mortem brain tissue. With animal models there is hope the findings are translatable to humans. However, the ability to translate findings from these approaches has limitations. It's been shown that less than 8% of cancer clinical trials have translated from mice to humans (Mak, Evaniew et al. 2014). One example of these trials that showed promise in mice, but failed in humans, is the TGN1412 trial. In this trial, all of the six healthy volunteers who received the TGN1412 monoclonal antibody, which was supposed to activate T-cells without toxic effects, had multi-organ failure and were admitted to the hospital (Suntharalingam, Perry et al. 2006). Additionally, all treatments for amyotrophic lateral sclerosis (ALS) in the last ten years that showed effective results in mouse models (primarily TDP43 and SOD1 mutants) have failed (or been minimally

effective) in human trials (Perrin 2014). In terms of post-mortem brain tissue, obtaining these samples from relevant subjects at early as well as late stages of disease is not always easy, and when the tissue is obtained the quality and integrity of the cells is often severely compromised. In addition, post-mortem brain tissue typically only gives a look at the outcome of the illness at later stages after years of variable treatments and possible co-morbidities, which is also subject to limitations. IPSC technology has allowed studies of a number of neurological illnesses at a cellular and molecular level never before achievable. Studies done with iPSCs and iPSC-derived neurons have provided insights and advancements in the research of many neurological illnesses including schizophrenia (Brennand, Simone et al. 2011), bipolar disorder (O'Shea and McInnis 2015), ALS (Cashman and Lazzerini Ospri 2013, Chen, Qian et al. 2014), and Prader-Willi and Angelman syndromes (Chamberlain, Chen et al. 2010, Cruvinel, Budinetz et al. 2014). The use of iPSCs to create live neurons from multiple subjects carrying defined gene variants, while maintaining the genetic makeup of the subjects, allows us to choose subjects with and without disease associated variants and study the possible differences in the neurons associated with these specific human genetic variants.

Research done in the lab using iPSC-derived forebrain neural cultures has provided us with some interesting questions regarding the use of iPSCs to study AUDs. A gene expression analysis using quantitative PCR (qPCR) was done to examine genes previously associated with AUDs (genes encoding $GABA_A$ receptor subunits and NMDA receptor subunits) in control subjects and alcoholic subjects with different genotypes of the rs279858 SNP in *GABRA2*; two T alleles (TT), one T allele and one C allele (TC), or

two C alleles (CC). When looking at the $GABA_A$ receptor genes, it was discovered that gene expression of a cluster of genes encoding $GABA_A$ subunits located on chromosome 4p12 (*GABRG1, GABRA2, GABRA4, GABRB1*) was very low in approximately half of the lines examined. Ten of 11 lines with the TT genotype expressed this gene cluster (both control and alcoholic subjects), while only four of 14 CC lines expressed the gene cluster, and five of 11 CT lines has the expresser phenotype. Cluster analysis separated the lines into two groups: lines which express the $GABA_A$ receptor subunit gene cluster on chromosome 4p12, or “expressers,” and lines which show reduced or minimal expression of the $GABA_A$ receptor subunit gene cluster, or “non-expressers” (Lieberman, Kranzler et al. 2015).

In addition to qPCR data, RNA-seq data was also available for a subset of samples providing genome wide gene expression data. RNAseq data was used to calculate correlations between $GABA_A$ receptor subunit genes on chromosome 4p12 and other $GABA_A$ receptor subunit genes located on other chromosomes. Genes on chromosome 4p12 were found to be highly correlated, whereas genes on other chromosomes showed little correlation to *GABRA2* (chromosome 4p12). Utilizing this RNA-seq data, we generated a graph showing the average values of the reads per kilobase of transcript per million reads mapped (RPKM) for $GABA_A$ subunits on chromosomes 1, 4, 5, 15, and X. In **Figure 1.3A**, there is a clear difference between the expresser lines and the non-expresser lines in the gene cluster on chromosome 4p12, as expected and the non-expresser lines have virtually no expression of these genes. **Figure 1.3B** shows that for $GABA_A$ subunit genes on other chromosomes there is little change in expression among expresser lines and non-expresser lines. While these

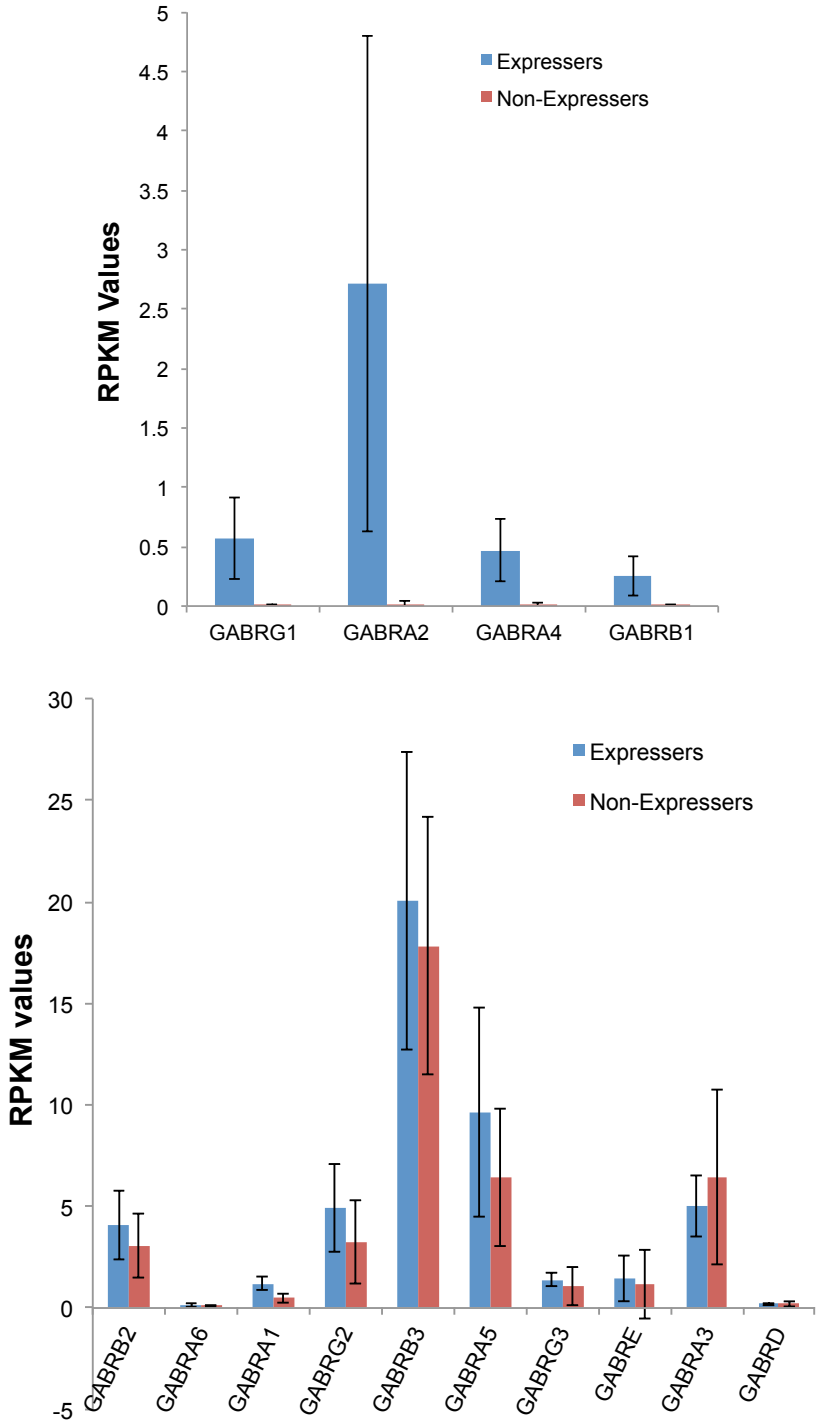


Figure 1.3. Graphical representation of RPKM values from RNA-seq data in Lieberman et al., 2015. **(A)** Expression of $GABA_A$ receptor subunit genes located on chromosome 4p12 in expressers and non-expressers. Expressers show significantly greater RPKM values than the non-expressers. **(B)** Expression of $GABA_A$ receptor subunit genes located on chromosomes 5, 15, and X. Expression does not differ between expressers and non-expressers.

chromosome 4p12 $GABA_A$ receptor subunit genes have low RPKM values in these cultures (low RPKM values can be unreliable (Mortazavi, Williams et al. 2008)), the qPCR data in Lieberman et al., 2015 verify the findings from the RNA-seq data and suggest that what we are seeing is a reproducible phenotype present in ~40% of lines. Additionally, Lieberman et al., 2015 looked at results from post-mortem tissue databases and saw no genotype related changes in *GABRA2* expression in any of the samples indicating this could potentially be a developmental phenotype.

Data from the RNA-seq studies in Lieberman et al., 2015, show that the genes flanking the chromosome 4p12 $GABA_A$ receptor subunit gene cluster are expressed similarly in both expresser and non-expresser lines. Since this phenotype is confined to the chromosome 4p12 $GABA_A$ receptor subunit gene cluster, we have reason to believe that there is a common regulatory element for this ~1.4mb gene cluster. There is evidence that this may be a common pattern for clustered $GABA_A$ receptor subunit genes. As mentioned above, the distribution of $GABA_A$ receptor subunit genes on the chromosomes of mice is orthologous to that of humans. In mice it has be demonstrated that long distance regulation of multiple $GABA_A$ receptor subunit genes on chromosome 11 (chromosome 5 in humans) exist. A study done by Uusi-Oukari et al., 2000, found that α6 -/- mice showed decreased mRNA expression of the β2 and α1 subunits suggesting a common regulatory element that acts over long distances (Uusi-Oukari, Heikkila et al. 2000). This finding in mice, together with the orthologous layout of the $GABA_A$ receptor subunit genes on chromosomes, led us to believe we are seeing a similar mechanism in the human $GABA_A$ receptor subunit gene cluster on chromosome 4p12.

The question remains, what is going on that has linked the genotype of a synonymous tag-SNP to expression levels of the $GABA_A$ receptor subunit genes clustered on chromosome 4p12 in iPSC-derived neurons? One potential modulator of this cluster could be a change in epigenetic modifications.

Epigenetic modifications and gene expression

Epigenetics involves mitotically heritable modifications beyond the genomic DNA sequence. In recent years, the field has been expanding rapidly. What was once focused on primarily X-linked phenotypes and imprinted genes is now opening into a world where expression of autosomal genes is found to be heavily influenced by developmental and environmental, non-genetic features via epigenetic regulation.

In order for DNA to be more compact yet efficiently accessed, it is wrapped around histone octamers made up of two H2A, two H2B, two H3, and two H4 histone subunits. Approximately 146bp of double stranded DNA is wrapped around a histone octamer creating what is called a nucleosome. Each nucleosome is connected by linker DNA bound to an H1 subunit. The histone subunits each have an N-terminal tail that can be modified by a number of post-translational modifications including methylation, acetylation, ubiquitylation, and phosphorylation (Cedar and Bergman 2009). The histone subunit affected, the location of the residue modified on the tail of the histone subunit, the type of modification, and the degree of modification can all alter the 3D chromatin structure and influence the availability of DNA for transcription. For example, the lysine at position 4 (K4) on the tail of the histone 3 subunit (H3) can be tri-methylated (H3K4me3) and this mark is seen in up to ~75% of gene promoters and is associated

with active gene expression (Gu and Lee 2013). Another common histone modification is the tri-methylation (me3) of lysine 27 (K27) on the histone 3 subunit (H3K27me3). This mark is associated with repression of transcription. Examples of other lysine modifications on the histone 3 subunit tail include H3K9me3, a repressive mark found in differentiated cells, and H3K36me3, a mark found in actively transcribed genes (Zhou, Goren et al. 2011).

Methylation of histone lysine residues is catalyzed by a number of different histone methyltransferases (HMTs) also known as “methyl writers” (Ruthenburg, Allis et al. 2007). H3K4, for example, is methylated by at least 14 known HMTs, all of which contain a SET (a protein domain characterized in *Drosophila:* **S**u(var)3-9, **E**nhancer of zeste, **T**rithorax (Cheng, Collins et al. 2005)) domain. They include: SET1A, SET1B, SET7/9, MLL1-5, SETMAR, ASH1L, PRDM9, and SMYD1-3 (Gu and Lee 2013). The SET domains are either related to yeast Set1 and *Drosophila* Trx as with MLL1-5, SET1A, and SET1B, or unrelated as with SET7/9, SETMAR, ASH1L, PRDM9, and SMYD1-3 (Ruthenburg, Allis et al. 2007, Gu and Lee 2013). The SET domain in these proteins catalyzes the transfer of the methyl group of S-adenosylmethionine to H3K4 (Gu and Lee 2013). In the case of H3K27 methylation, EZH2, which is part of the polycomb repressive complex 2 (PRC2), initiates methylation (Cedar and Bergman 2009).

Similar to methylation, demethylation of histone lysine residues is not only dependent on the position of the lysine, but is also sometimes methylation state specific. For example, Jumonji C (JmjC) domain proteins UTX and JMJD3 are specific demethylases for di-methylated H3K27 and tri-methylated H3K27 residues, with

preference for the latter (Swigut and Wysocka 2007). Demethylation of di- and tri-methylated H3K4 is initiated by KDM5A-D (Gu and Lee 2013).

Evaluation of histone modification profiles for specific loci is done using chromatin immunoprecipitation (ChIP) assays, which examines the interaction of proteins with DNA. For example, ChIP can test to see if a transcription factor interacts with a promoter of interest, or whether a particular histone modification is enriched in a given location on the genome. The ChIP process (depicted in **Figure 1.4**) involves the cross-linking of histone proteins to DNA with formaldehyde followed by controlled sonication. The chromatin is then split equally into separate reactions; one for input (unimmunoprecipitated DNA) and the remaining into different immunoprecipitations with antibodies to histone proteins of interest (i.e. H3K4me3 or H3K27me3). All samples are then reverse cross-linked and treated with RNase A and proteinase K, and then the DNA is purified. The purified DNA from the input sample and IPs are then analyzed via sequencing or quantitative PCR (qPCR).

Initial research done by Ernst and Kellis in 2010 found 51 annotated chromatin states in human T-cells using a Hidden Markov Model (Ernst and Kellis 2010). The following year, Ernst et al. (2011) looked at nine different chromatin marks in nine different cell types to create regulatory predictions and they defined 15 broad annotations of chromatin states. The frequency of each chromatin mark was analyzed and charted to show which annotations best fit each chromatin mark. For example, H3K4me3 was found to be associated with strong promoters with 99% frequency. CTCF, a CCCTC-binding protein, was annotated as an insulator with 92% frequency, and H3K27me3 was associated and annotated as an inactive/poised promoter with 72%

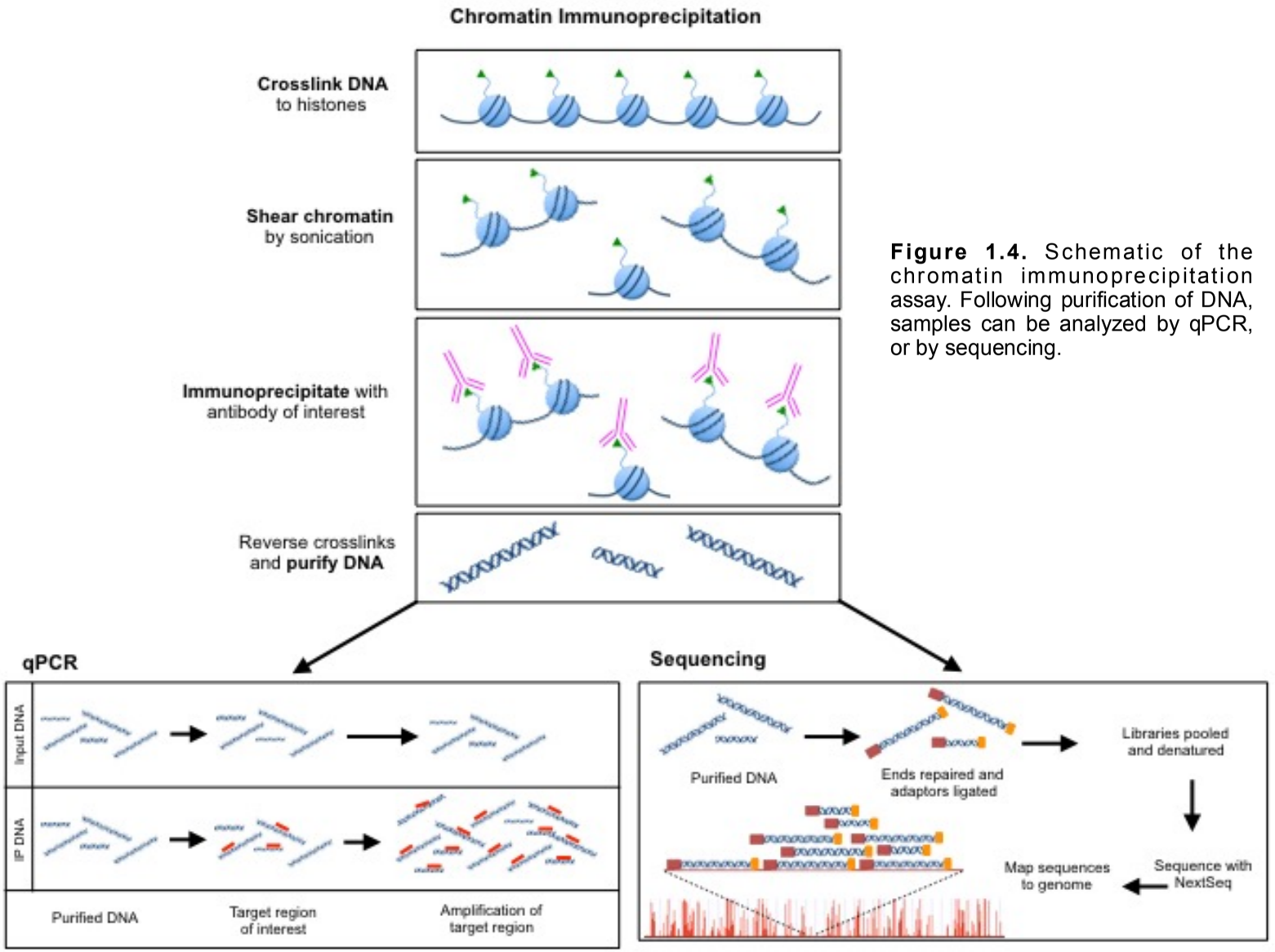


Figure 1.4. Schematic of the chromatin immunoprecipitation assay. Following purification of DNA, samples can be analyzed by qPCR, or by sequencing.

frequency (Ernst, Kheradpour et al. 2011). An example of the data from Ernst et al. 2011, is represented in the pink bar at the top of **Figure 1.5.** This shows that this region was found to be in a bivalent/poised state in embryonic stem (ES) cells. Additionally, other archived ChIP-sequencing data on the UCSC Genome Browser (http://genome.ucsc.edu/) also shows a bivalent/poised promoter for *GABRA2* in H1 and H9 ES cells and iPSCs (iPS6.9) (**Figure 1.5**). A poised promoter has both the active, H3K4me3, and repressive, H3K27me3, marks present. In the pluripotent state, the gene is inactive, but poised for expression once its fate has been determined. For example, differentiated neurons maintain the active mark, and the the enrichment of the repressive mark decreases in this cell type. The bottom two tracks on **Figure 1.5** show profiles for the active mark and repressive mark in post-mortem mid-frontal lobe brain tissue in which the bivalent state modifications have resolved (H3K4me3 is relatively more absent than the H3K27me3); consistent with *GABRA2* expression in the brain, the poised promoter was "switched" to an active promoter.

DNA methylation and CpG islands

Another feature of epigenetic regulation is DNA methylation. While most CpG sites throughout the genome are methylated, there are regions (typically ~1kb in length) in the genome with a higher than average GC content, known as CpG islands (Jones 2012), which are typically unmethylated and are found at the promoter of many genes (Moore, Le et al. 2013). While uncommon, DNA methylation at CpG islands does occur and is associated with transcriptional silencing and is important in gene regulation and development (Rose and Klose 2014). In addition, aberrant DNA methylation of cancer

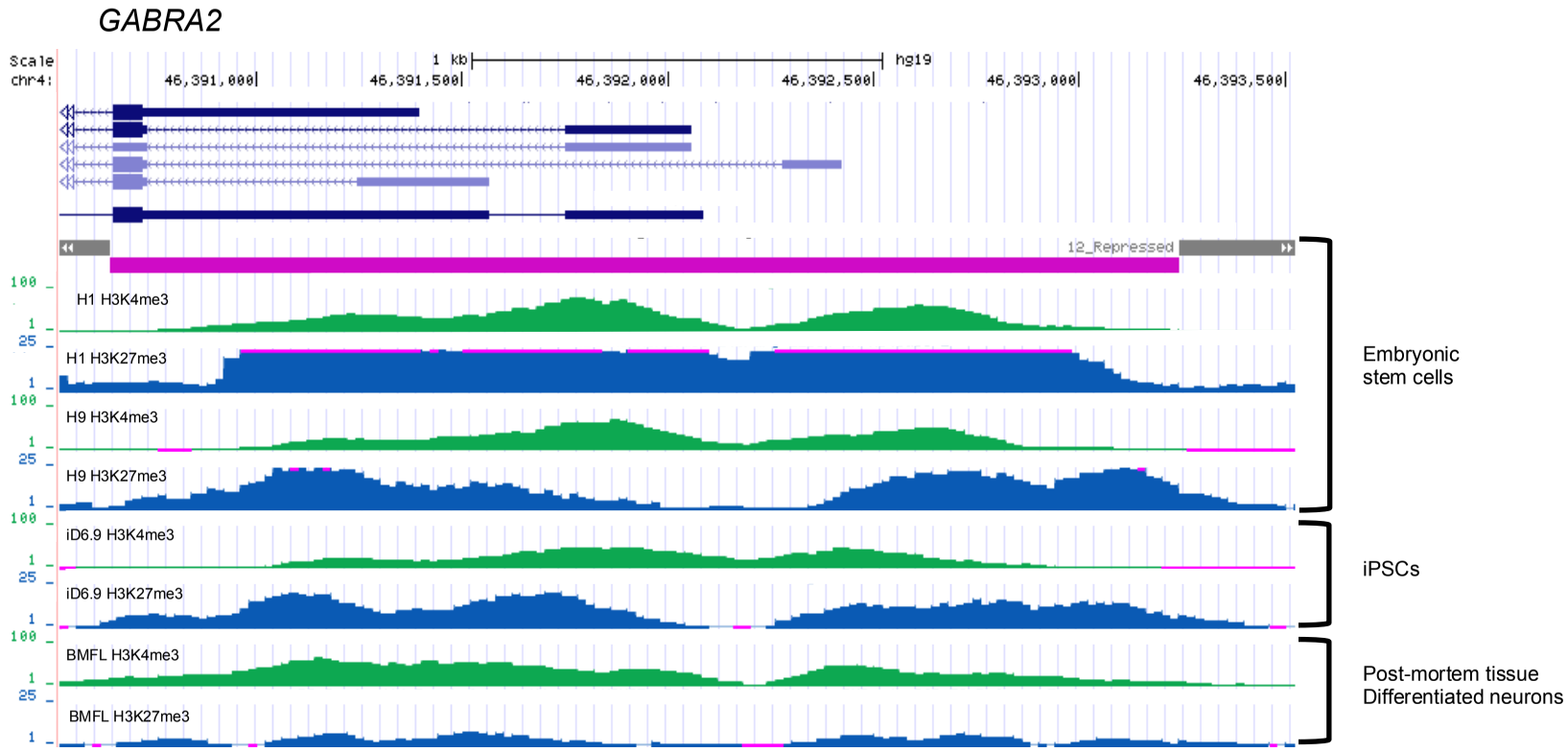


Figure 1.5. Raw ChIP-seq data tracks showing enrichment of H3K4me3 and H3K27me3 in ES cells, iPSCs, and post-mortem brain tissue at the *GABRA2* promoter. The pink bar at the top shows that this region is in a poised state in ES cells (from Ernst et al. 2011 and UCSC Genome Browser.

suppresser genes results in a variety of cancers (Cedar and Bergman 2009).

DNA methylation occurs when a methyl group is added to the cytosine ring at CpG dinucleotides. This process is mediated by a family of DNA methyltransferases (DNMTs) (Moore, Le et al. 2013). DNMT1 is necessary for the maintenance of DNA methylation. During DNA replication, DNMT1 methylates the daughter strands in reference to the parental methylation patterns (Esteller 2007). DNMT1 is crucial for development, and studies have shown that the knockout of DNMT1 in mice is embryonically lethal (Li, Bestor et al. 1992). *De novo* methylation is mediated by DNMT3A and DNMT3B. While knockout of DNMT3B in mice is embryonically lethal, DNMT3A knockout mice survive, but typically not more than 4 weeks, (Okano, Bell et al. 1999).

It is known that there is some opposition between DNA methylation and the presence of H3K4 methylated histones. DNMT3L, a catalytically inactive DNMT, binds nucleosomes at the H3 N-terminus and can initiate *de novo* DNA methylation with the aid of DNMT3A2 and DNMT3B. However, when H3 is methylated (mono-, di-, tri-) at lysine 4 (K4), DNMT3L can no longer bind and DNA methylation is inhibited (Ooi, Qiu et al. 2007).

DNA methylation is ‘read’ by a number of proteins belonging to three separate families. Methyl-CpG-binding domain (MBD) proteins and zinc-finger domain proteins generally bind methylated DNA and recruit repressor complexes leading to transcriptional repression. UHRF (ubiquitin-like containing PHD and RING finger domain) proteins are important for DNA methylation maintenance as they bind DNMT1,

which (as mentioned above) aids in DNA methylation maintenance on daughter strands during replication (Moore, Le et al. 2013).

Thesis Objectives

Hypothesis

Our hypothesis is that a yet-to-be-identified functional SNP represented by the tag-SNP rs279858 in the 140kb *GABRA2* haplotype block is influencing the 1.5mb chromosome 4p12 $GABA_A$ receptor subunit gene cluster through epigenetic mechanisms including changes in histone modifications and DNA methylation. Specifically, we hypothesize that the non-expresser neurons are retaining repressive (H3K27me3) epigenetic marks and/or are losing the activating (H3K4me3) mark during development. In addition, we hypothesize that the absence of the H3K4me3 modification in these lines allows for DNA methylation of the CpG island located in the promoters of the $GABA_A$ receptor subunit gene cluster on chromosome 4p12. Alternatively, DNA methylation is inhibiting the ability of histone methyltransferases to methylate H3K4 in this region.

Aims

1. Compare the histone modification profiles for an active mark, H3K4me3, and a repressive mark, H3K27me3, in the promoters of the $GABA_A$ receptor subunit genes in iPSC-derived neural cultures that express the chromosome 4p12 gene cluster (expresser lines) as well as those with reduced expression (non-expressers).

2. Characterize whether a bivalent “poised” promoter type histone modification profile is present in iPSCs and whether this differs by neural chromosome 4p12 $GABA_A$ receptor subunit gene cluster expression phenotype.

3. Evaluate and compare DNA methylation profiles at a site in the CpG island in the promoter of *GABRA2,* as well as neighboring genes in expresser and non-expresser iPSC lines.

Chapter 2.

Histone Modification Profiles in iPSCs and iPSC-derived cells

Background

The process of reprogramming fibroblasts into iPSCs changes epigenetic features, including histone modifications, chromatin structure, and DNA availability (Liang and Zhang 2013). To evaluate the potential for epigenetic regulation via changes in histone methylation on the chromosome 4p12 $GABA_A$ receptor subunit gene cluster, ChIP-seq and ChIP-qPCR were conducted. We used an antibody specific for the H3K4me3 modification, which is associated with the promoters of active genes, and an antibody specific for the H3K27me3 modification, which is associated with poised promoters in pluripotent cells and with facultative heterochromatin and repressed genes in differentiated cells (Ernst, Kheradpour et al. 2011). Using the UCSC Genome Broswer, we evaluated archived raw ChIP-seq data tracks for H3K4me3 and H3K27me3 in ES cells and iPSCs. These tracks, shown in **Figure 1.5**, were used to indicate the location of enrichment of our two marks of interest. In pluripotent stem cells, archived ChIP-seq data indicates the *GABRA2* promoter is bivalent, or "poised," showing enrichment of both H3K4me3 and H3K27me3 marks, readying the gene to either be expressed, or repressed during differentiation depending on cell type and location provided by chemical and developmental cues. In contrast, ChIP-seq tracks from post-mortem brain mid frontal lobe (BMFL) tissue show decreased H3K27me3 enrichment, and maintenance of the active H3K4me3 mark. Our hypothesis was that non-expresser lines retain the H3K27me3 repressive enrichment and lose enrichment of

H3K4me3, the active mark. To assess this, we used ChIP with H3K4me3 followed by genome wide sequencing to assess and global changes that may be affecting the cluster of $GABA_A$ receptor subunit genes on chromosome 4p12 and ChIP followed by quantitative PCR (qPCR) to target specific genes/promoters of interest.

Methods

Cell lines

Lines were the same as used in the 2015 Lieberman study (Lieberman, Kranzler et al. 2015). Briefly, skin punch biopsies were taken from 21 patients and were cultured as fibroblasts. Fibroblasts were reprogrammed into induced pluripotent stem cells (iPSCs) by the Stem Cell Core at the University of Connecticut Health Center.

IPSC differentiation into forebrain neurons

Neuronal differentiation was done using the method described in Zeng et al. (Zeng, Guo et al. 2010). Briefly, iPSCs were cultured on a feeder layer of irradiated mouse embryonic fibroblasts with human Embryonic Stem Cell Media (hESM) comprised of DMEM/F12, ß-ME, L-glutamine, Knockout SR, MEM NEAA, and bFGF to reduce differentiation of colonies. To passage the cells, colonies were treated with 1mg/ml dispase (in DMEM/F12) for five minutes to loosen attachment to the plate and feeder cells. Dispase was removed and colonies were washed one time with DMEM/F12. HESM was added and plates were scraped with a 5ml pipette to break up and release colonies. A small portion of the colony pieces was passaged to another plate of irradiated mouse embryonic fibroblasts (iMEFs), and for neuronal differentiation, the

remaining broken up colonies were transferred to a T-75 suspension flask containing ESM without bFGF. *Stage 1*: Medium was changed every day for six days by letting the cell clusters, also known as embryoid bodies (EBs), settle to the bottom before removing old medium. *Stage 2*: On day seven, EBs were spun down at 1,000 rpm for five minutes and transferred to a new T-75 flask with Neuronal Induction Media (NIM) containing DMEM/F12, heparin, MEM NEAA, and N2 supplement. EBs were cultured in NIM for three to four days changing media every other day. *Stage 3*: EBs were then plated onto tissue culture treated plates with NIM containing 10% FBS to aid in EB attachment. Twelve to 14 hours after plating, medium was changed to NIM without FBS, and then every other day for 10-14 days. *Stage 4*: Neuroepithelial cell colonies from plates were dislodged from the plates with a stream of medium from a 1ml pipette and the detached clusters were transferred to a T-75 suspension flask to culture as neurospheres. These neurospheres were cultured for seven days in NIM with B27. *Stage 5*: Neurospheres greater than 300μm were separated from the other clusters and dissociated by incubating cells with Accutase for five minutes at 37˚C, spun down, and then rinsed with Neurobasal medium. Neurospheres were plated on 12-well plates coated with 20ng/ml laminin. Neurons were cultured in Neuronal Differentiation Medium (NDM) containing BDNF, GDNF, IGF, laminin, L-glucose, B27, and N2. Medium was changed every three to four days for 12 weeks, during which time they become differentiated neurons expressing GABA and glutamate receptor genes. IPSC lines and iPSC-derived neural cultures from these lines have been previously characterized in our lab (see Lieberman et al., 2015).

Chromatin immunoprecipitation (ChIP)

For neurons, ChIP was done using the Millipore EZ-Magna ChIP (17-409) protocol. For iPSCs and neuroepithelial cells, a combination of multiple protocols was used including the Millipore EZ-Magna ChIP protocol, a high-throughput ChIP protocol (Blecher-Gonen, Barnett-Itzhaki et al. 2013), and the Cotney et al. protocol (Cotney and Noonan 2015). IPSCs from a 10cm petri dish, or neurons from 6-12 24-well wells, were fixed with a final concentrations of 1% formaldehyde (freshly made) to crosslink histones to DNA and 1X glycine was then added to quench the excess formaldehyde. Cells were washed twice with 1X PBS and then a third time with 1X PBS with Protease Inhibitor Cocktail II (Millipore). Cells were either processed immediately or frozen at -80˚C until further use. Cells were pelleted at 1,000 rpm for five minutes. Cell pellets were suspended in 1ml of cell lysis buffer containing 50mM Tris-HCl (pH 8.0), 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, and 0.25% Triton X-100. Samples were incubated on ice for 20 minutes. Cells were disrupted using a 2ml Dounc homogenizer with pestle B (Kontes) for 30 strokes. Lysates were spun down at 2,000*g* for 5 minutes at 4˚C and supernatant was discarded. Pellets were resuspended in 300µl nuclear lysis buffer containing 10mM Tris-HCl (pH 8.0), 1mM EDTA, 0.5mM EGTA, and 0.2% SDS and incubated on ice for 20 minutes. Chromatin was then sheared into 200-800bp fragments using controlled sonication with a QSonica sonicator for 30 minutes in 10 second bursts. To evaluate sonication efficiency, 5µl of sonicated material was treated with RNase A for 30 minutes at 37˚C, followed by a proteinase K treatment at 37˚C and then reversed crosslinks at 65˚C. The DNA was then run on an agarose gel. The DNA was a smear localized to 200-800bp. Sonicated material was spun down to remove

debris and kept at -80˚C until further analysis. Frozen pellets were thawed on ice, and sheared chromatin was then immunoprecipitated with ChIP-validated antibodies from Millipore against the active mark, histone 3 tri-methylated lysine 4 (H3K4me3; Millipore, 17-614), and against the repressive mark, histone 3 tri-methylated lysine 27 (H3K27me3; Millipore, 17-622). Antibodies were tested for binding specificity using the Millipore AbSurance Histone H3 Antibody Specificity Array (Millipore, 16-667). Antibodies were coupled to washed protein-G magnetic beads (Millipore, 16-662) in blocking binding buffer containing 1% BSA and 1% Tween-20 in PBS in low-bind 1.5ml tubes and incubated at room temperature for one hour. Beads were then separated from the buffer using a magnetic rack. The supernatant was discarded and the beads were resuspended in a portion of sonicated material diluted to 500µl with dilution buffer consisting of 0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris-HCl, 167mM NaCl, and protease inhibitor cocktail II (Millipore, 539132). Samples were incubated overnight at 4˚C on a rotator. The following day, beads were separated using the magnetic rack and the supernatant was discarded. The beads were resuspended in 180µl dilution buffer with protease inhibitors and transferred to a well in a pre-chilled 96-well PCR plate. Beads were separated using a magnetic rack for 96-well plates and the supernatant was discarded. Beads were washed with 180µl cold Low Salt Immune Complex Wash Buffer (Millipore, 20-154) twice, 180µl High Salt Immune Complex Wash Buffer (Millipore, 20-155) twice, LiCl Immune Complex Wash Buffer (Millipore, 20-156) twice, and TE buffer (Millipore, 20-157). Beads were washed by moving the plate from side to side on the magnetic rack. After the removal of the TE buffer, beads were resuspended in 50µl elution buffer consisting of 10 mM Tris-HCl (pH 8.0), 5 mM EDTA

(pH 8.0), 300 mM NaCl, and 0.5% SDS. Samples were treated with 1µg RNase A for 30 minutes at 37˚C and then proteinase K treated for 2 hours at 37˚C followed by a 65˚C incubation overnight. The following day, beads were separated using the magnetic rack and the supernatant was moved to a new plate. To purify the DNA, Agencourt AMPure XP beads (Beckman Coulter, A63881) were added at 2.3x the volume of supernatant and mixed by pipetting up and down 25 times. Samples were incubated at room temperature for 2 minutes and then the beads were allowed to be separated using the magnetic rack for 4 minutes. Supernatant was removed and beads were washed twice with 100µl 70% EtOH for 30 seconds while the plate remained on the rack. After all of the ethanol was removed the plate was removed from the rack. The beads were dried at room temperature for 4 minutes. Beads were then resuspended in 60µl 10mM Tris-HCl (pH 8.0) and mixed by pipetting up and down 25x. Samples were incubated at room temperature for 2 minutes. The plate was returned to the rack for 4 minutes to separate out the beads and the supernatant was transferred to another plate.

The concentration of ChIP DNA was measured using the Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, Q32851). ChIP DNA was then used to construct sequencing libraries (undiluted), or diluted to 50pg/µl for qPCR.

ChIP-seq

ChIP was done with both iPSCs and iPSC-derived neuroepithelial cells from both an expresser line, 563-9, and a non-expresser line, 559-5, and immunoprecipitated with H3K4me3. Resulting ChIP-DNA from the immunoprecipitations, as well as inputs for each cell type (563-9 iPSC, 563-9 neuroepithelial cells, 559-5 iPSCs, and 559-5

neuroepithelial cells), was then processed using the Illumina TruSeq ChIP Sample Preparation protocol. Briefly, End Repair was done on 50µl of 100-200pg/µl ChIP or input DNA to blunt the ends by removing 3' overhangs and filling in 5' overhangs. Next the 3' ends were adenylated to prevent fragments from ligating to one another in the next step where multiple indexing adaptors are ligated to the DNA fragments for hybridization. The ligation products were run on an agarose gel, the region of the gel containing DNA in the 200-350bp size-range was excised, and DNA was extracted using a gel extraction kit (Qiagen, 28704). DNA was examined using a TapeStation instrument (Agilent) to confirm DNA fragments were in the 250-300bp size-range, which is optimal for efficient cluster generation. The purified adapter modified DNA products were amplified using PCR. The ChIP DNA target size range was validated again using the TapeStation and the library DNA was quantified using the NEBNext Quant Kit for Illumina (New England Biolabs, E7630S). The DNA library samples were normalized and pooled for sequencing using the Illumina NextSeq 500 system with a High Output Kit (75 cycles; Illumina, FC-404-1005).

ChIP-seq data was analyzed first by SICKLE for adaptive trimming of the reads to remove poor quality ends. Reads were aligned to the genome (hg19) using STAR. HOMER was then used to create bedGraph files to import into the UCSC Genome Browser to view raw data pile-up tracks, find peak calls, annotate peaks, and create a meta-gene for each line. DiffBind was used to contrast peaks between samples with ChIP-seq data.

Quantitative real-time PCR for ChIP

To obtain multiple primers specific for the $GABA_A$ receptor subunit genes' promoter regions, the target genome sequence 1000 bases upstream of each gene (guided from archived ChIP-seq data from the UCSC genome browser) was entered into PCR Tiler (http://pcrtiler.alaingervais.org:8080/PCRTiler/), which calculates primer pairs with specified annealing temperatures, GC content, and product specificity. Primers were designed for the *GABRA2* (gene of interest; chromosome 4) promoter, *GABRB2* (chromosome 5) promoter as a comparative control, *GUF1* (a gene flanking the chromosome 4 cluster) as another comparative control, a region negative for both H3K4me3 and H3K27me3 enrichment near *GABRA2* (negative control), and *MYOD1* (a positive control for H3K27me3 enrichment). A positive control for H3K4me3 enrichment were the control primers for *GAPDH*, supplied with the ChIP antibody. For primer sequences, see **Table 2.1**. QPCR was done in duplicate using 20µl reactions with 1µl 50pg ChIP DNA, primers, and SYBR Green Master Mix (Thermo Fisher Scientific, 4309155). QPCR was conducted using an ABI 7500 PCR System. The protocol was 95˚C for 10 minutes, and 40 cycles of 95˚C for 15 seconds and 60˚C for 1 minute. To calculate fold over input, the average Ct values of duplicates were calculated and then entered into the equation: $2^{(\text{Input Ct average} - \text{IP Ct average})}$. Fold enrichments were compared between groups using a 2-tailed student's t-test.

Results

To assess possible differences in histone modification enrichment between expresser and non-expresser lines, we examined 12- to 15-week (post plating) neural

Primer	Target	Sequence (5'-3')	Tm (˚C)
ChpGA2_C Forward	GABRA2	CTC CCA AGT TTC CTA TCT CGT CAA	56.2
ChpGA2_C Reverse		AGT GTC TCT ATC GGG ACC AAC G	58.2
GAPDH Forward	GAPDH	TAC TAG CGG TTT TAC GGG CG	57.2
GAPDH Reverse		TCG AAC AGG AGG AGC AGA GAG CGA	63.3
MYOD1 Forward	MYOD1	CGC CAG GAT ATG GAG CTA CT	56.2
MYOD1 Reverse		CGG GTC GTC ATA GAA GTC GT	56.3
ChpGB2_B Forward	GABRB2	CTC TGG GTG TGC GAG TCC	58.1
ChpGB2_B Reverse		CGG AGC GGT CCC TAG AAG	57.3
GUF1 Forward	GUF1	GGC TAC CGA CAG GCT CTA	56.1
GUF1 Reverse		GCC CTA TAC ATG TCC AAG ATC G	55
Negative control Forward	Unenriched region in *GABRA2*	TGA GGC TAG AGG GTA TAA AGT G	53.3
Negative control Reverse		ACT GGT AAC CAA CTT CAA ACA A	52.7

Tabel 2.1. Primer sequences and melting temperatures for ChIP-qPCR.

cultures of previously identified lines from Lieberman et al. 2015. Using archived ChIP-seq data tracks from the Genome Browser, we targeted the sequence of the promoter region of *GABRA2* using PCR tiler to obtain primers that coincided with the H3K4me3 and H3K27me3 peaks. Histone modification enrichment was analyzed by ChIP-qPCR using antibodies to H3K4me3 and H3K27me3, and data was measured as a percent of the input DNA. We found that the active mark, H3K4me3, was robustly enriched at the *GABRA2* promoter in expresser neural cultures (n=7), but not in non-expresser lines (n=4) (**Figure 2.1A,** *p=0.02). The repressive mark, H3K27me3, showed no significant difference between expresser (n=6) and non-expresser (n=2) neurons (**Figure 2.1B**, p=0.7). *GAPDH*, a housekeeping gene, was used as a positive control for H3K4me3 enrichment and showed no significant difference between expresser (n=7) and non-expresser (n=4) lines (**Figure 2.1C**, p=0.1). The *GAPDH* promoter, which is a strong promoter, had no enrichment of H3K27me3 in either expressers (n=6) or non-expressers (n=2) as expected (**Figure 2.1D**, p=0.25). This suggested a link between expression status and H3K4me3 enrichment in neurons.

ChIP-seq to evaluate genome wide enrichment of H3K4me3

Messenger RNA (mRNA) expression timelines were previously done in the lab looking at *GABRA1* and *GABRG2,* located on chromosome 5, and *GABRA2* and *GABRG1*, located on chromosome 4p12. Expression data from two TT lines (expressers) and two CC lines (non-expressers) from iPSCs, neuroepithelial cells (NE), and neural cell cultures 2-, 6-, and 15-weeks (post-plating). mRNA expression data for genes located on chromosome 5, *GABRA1* and *GABRG2*, show no expression in

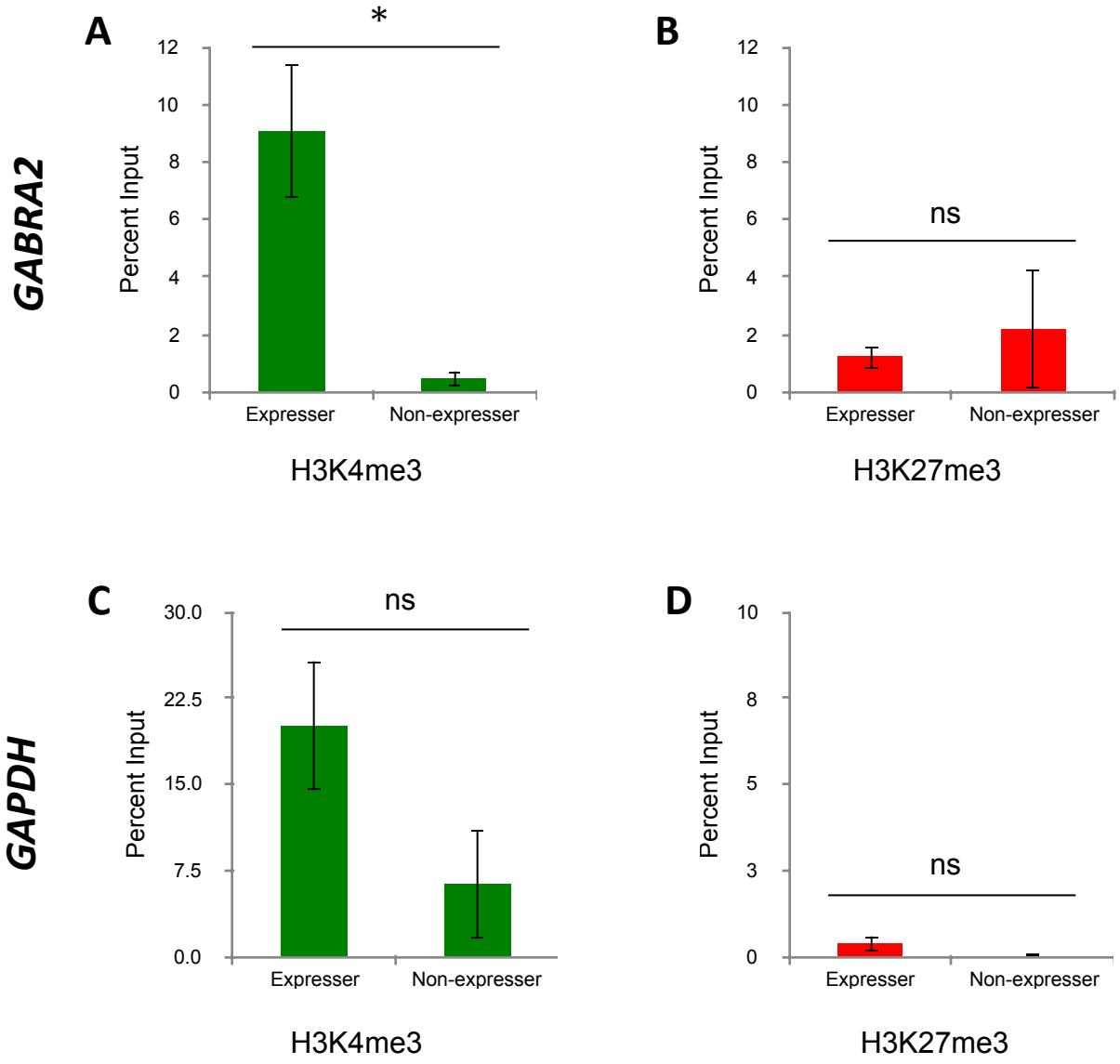
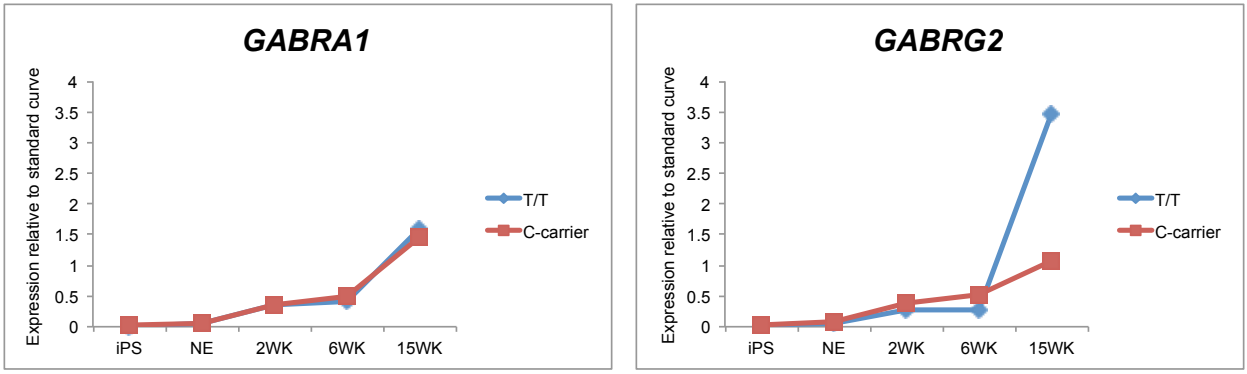


Figure 2.1. ChIP-qPCR data for expresser and non-expresser 12-15 week iPSC-derived neurons. (**A**) H3K4me3 enrichment at the *GABRA2* promoter (expressers n=7 and non-expressers n=4; *p=0.02). (**B**) H3K27me3 enrichment at the *GABRA2* promoter (expressers n=6 and non-expressers n=2; p=0.74). (**C**) H3K4me3 enrichment at the *GAPDH* promoter (expressers n=7 and non-expressers n=4; p=0.1). (**D**) H3K27me3 enrichment at the *GAPDH* promoter (expressers n=7 and non-expressers n=4; p=0.25).

iPSCs and NE, but begin to show low expression in 2-week neural cell cultures (**Figure 2.2A, B**) in both TT (blue) and CC (red) lines, and robust expression in 15-week neural cell cultures. However, expression of *GABRA2* and *GABRG1* shows differences between TT and CC lines starting in 2-week neural cell cultures (**Figure 2.2C, D**); expression in the TT lines progresses, but the CC lines show no expression for all time-points. As epigenetic changes precede gene expression, and because we saw differences in expression levels as early as in 2-week neural cell cultures, we predicted that there could be different histone modification profiles for expresser and non-expresser lines as early as at the neuroepithelial cell (NE) state. Knowing pluripotent cells (hESC lines H1 and H9, and iPSC line 6.9) have bivalent/poised promoters (see **Figure 1.5**), we set out to find when the switch from the poised promoter to a resolved active promoter occurs during differentiation.

First, to get a genome-wide perspective of H3K4me3 enrichment to see if there was reduction in H3K4me3 enrichment selectively in the chromosome 4p12 $GABA_A$ receptor subunit gene cluster in non-expresser lines, we performed ChIP followed by whole genome sequencing on iPSCs (one expresser and one non-expresser line) and neuroepithelial cells (one expresser and one non-expresser). A ChIP library was made using H3K4me3 IP DNA and input DNA and run on a NextSeq instrument. Analysis of the data was done using the program SICKLE first to trim low quality ends, STAR to align to the hg19 genome, and then Homer to create bedGraph files, obtain peak calls, annotate peaks, and create meta-gene profiles. Files were uploaded to the UCSC Genome Browser to visualize pile-up maps of the raw ChIP-seq data peaks and the peak calls. The distribution of peak calls is shown in **Figure 2.3** for each line. Each pie

A



B

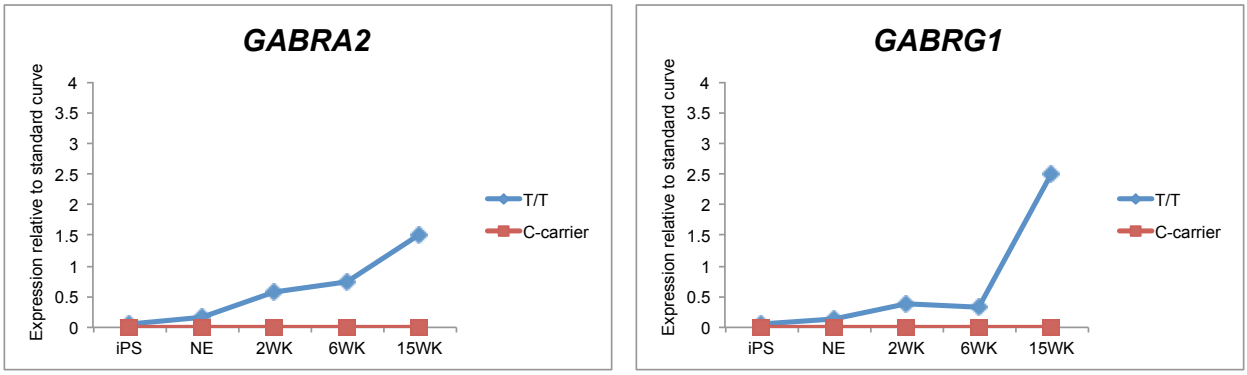


Figure 2.2. mRNA expression time course of $GABA_A$ receptor subunit genes in expresser (TT) (n=2) and non-expresser (C-carrier) (n=2) lines. (**A**) Expression of two genes, *GABRA1* and *GABRG2*, located on chromosome 5q34. Both non-expresser and expresser lines show expression in neurons. (**B**) Expression of two genes, *GABRA2* and *GABRG1*, located on chromosome 4p12. Expression begins in two week neurons in expresser lines, but non-expressers never express either gene up to 15 week neurons.

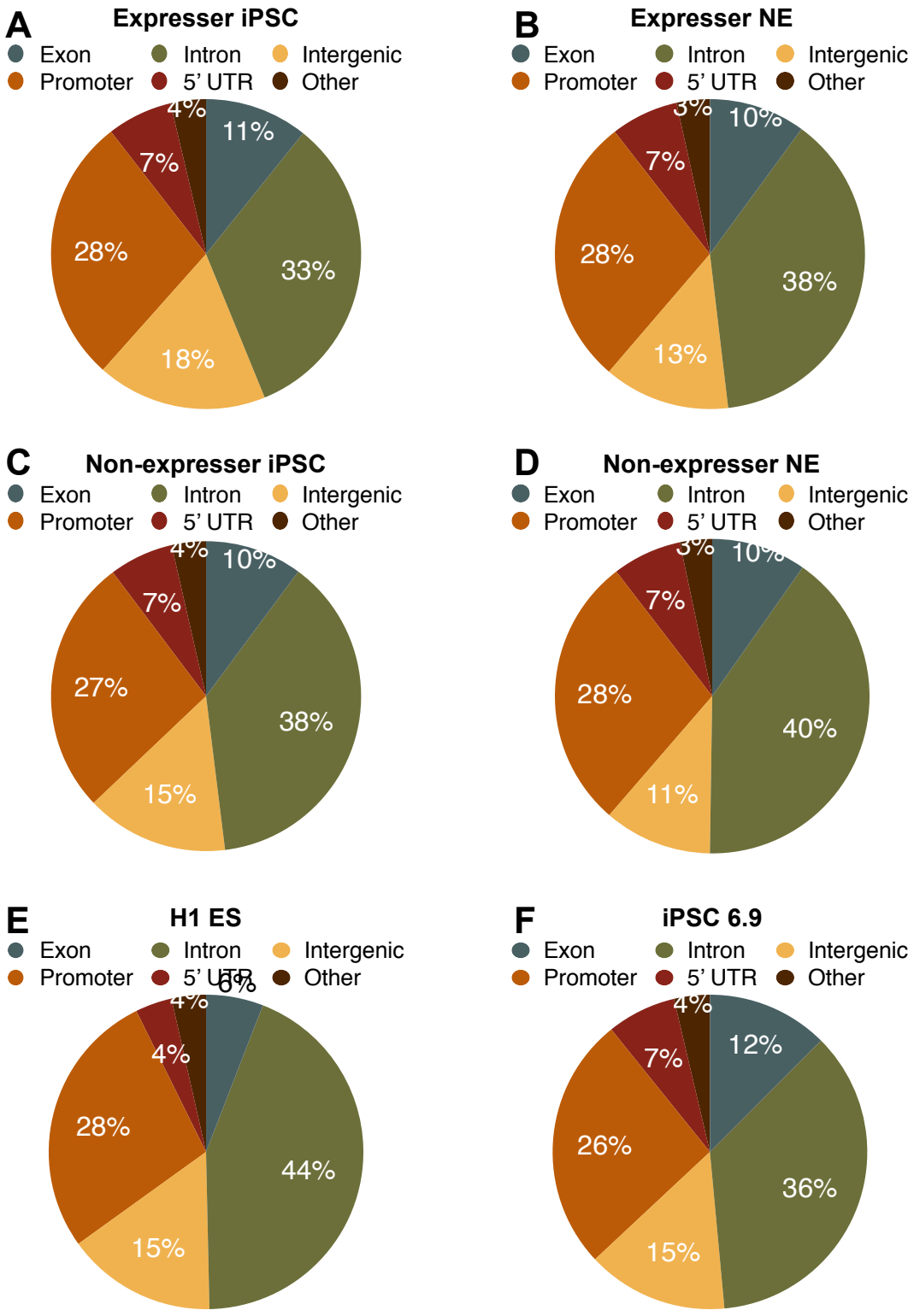


Figure 2.3. Distribution of H3K4me3 peaks from ChIP-seq. **(A-D)** Experimental lines. **(E,F)** Data from Genome Browser archive for comparison.

chart shows the percentage of peaks within exons, introns, intergenic regions, promoters, 5' UTRs, and an 'other' group for the remaining peaks. Archived data from the Genome Browser for H1 ES cells and iPSC 6.9 were analyzed to create pie charts to use as a comparison to our samples and show similar profiles indicating our ChIP-seq was successful. Additionally, the distribution of alignments were nearly identical between samples. The meta-gene profile for each line is shown in **Figure 2.4** showing similar distribution of ChIP-seq alignment along a meta-gene and a peak at the transcript start site. All genes are scaled to be 10,000bp (gene linear distance equivalent of 0% to 100%) with 5,000bp upstream and downstream of the gene (**Figure 2.4 A-D**). Again, archived data from the Genome Browser for H1 ES cells and iPSC 6.9 (**Figure 2.4 E, F**) were used for comparison and showed similar profiles to our samples, again validating our ChIP-seq protocol.

Pile-up maps of raw ChIP-seq data and the peak calls visualized in the UCSC genome browser show robust enrichment of H3K4me3 in a pluripotent-associated gene, *NANOG*, in both expresser and non-expresser iPSC lines, but as expected, not in the iPSC-derived neuroepithelial cells from either line (data not shown). In contrast, there was robust enrichment of H3K4me3 in a neuronal frontal cortex-associated gene, *FOXG1*, in the neuroepithelial cells, but not in the iPSCs (data not shown). These results further validate the ChIP-seq datasets.

At the *GABRA2* promoter, there is enrichment of H3K4me3 in expresser iPSCs and iPSC-derived neuroepithelial cells, but not in the non-expresser neuroepithelial cells and unexpectedly also in the non-expresser iPSC sample (**Figure 2.5**). To see if this difference between expresser and non-expresser lines was present in the other

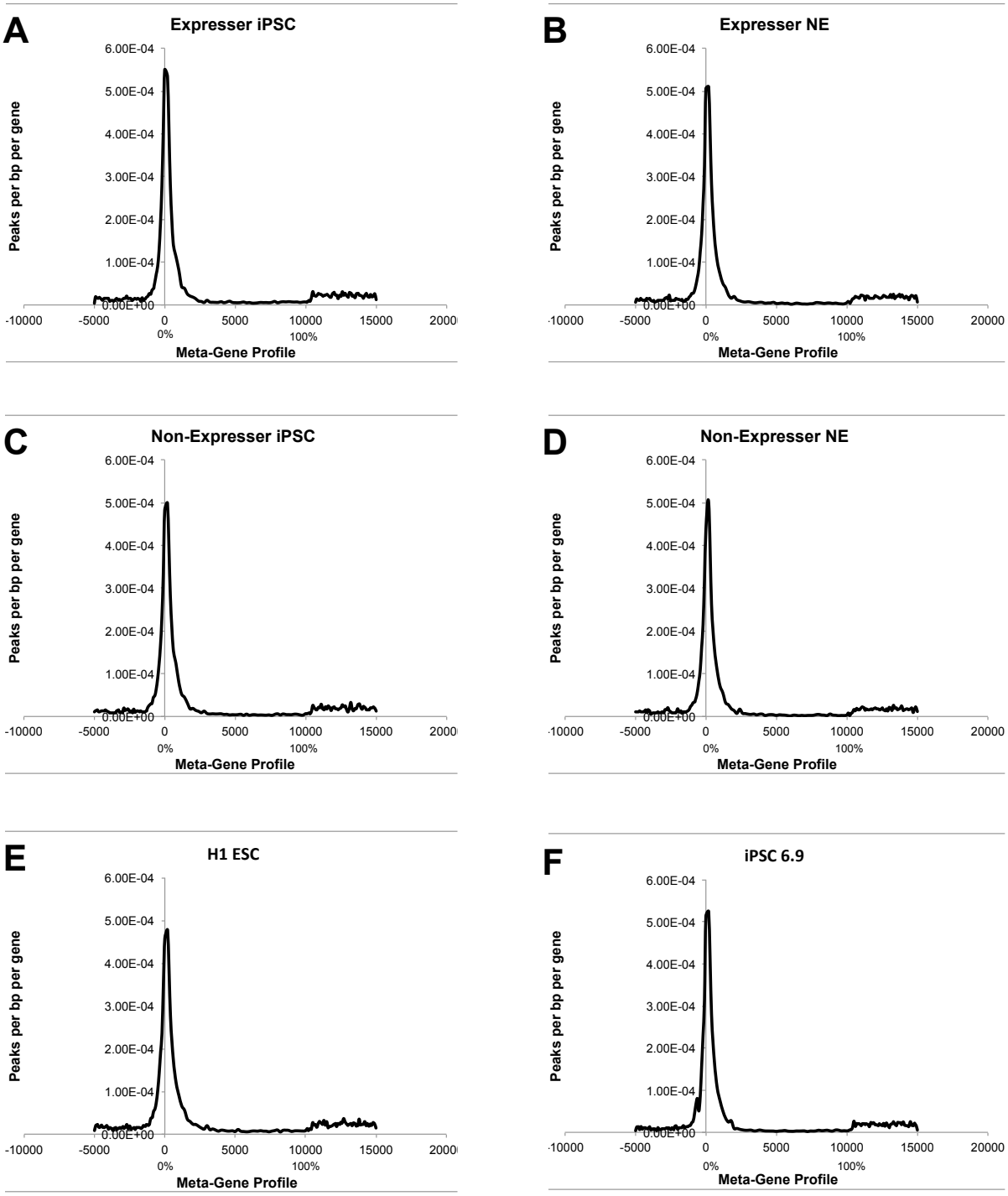


Figure 2.4. Meta-gene profiles for **(A)** expresser iPSC, **(B)** expresser neuroepithelial cells, **(C)** non-expresser iPSCs, and **(D)** non-expresser neuroepithelial cells. H1 **(E)** and iPSC 6.9 **(F)** were used for comparison. All genes are scaled to 10,000bp (equal to 100%) with 5,000bp upstream and downstream of the gene included.

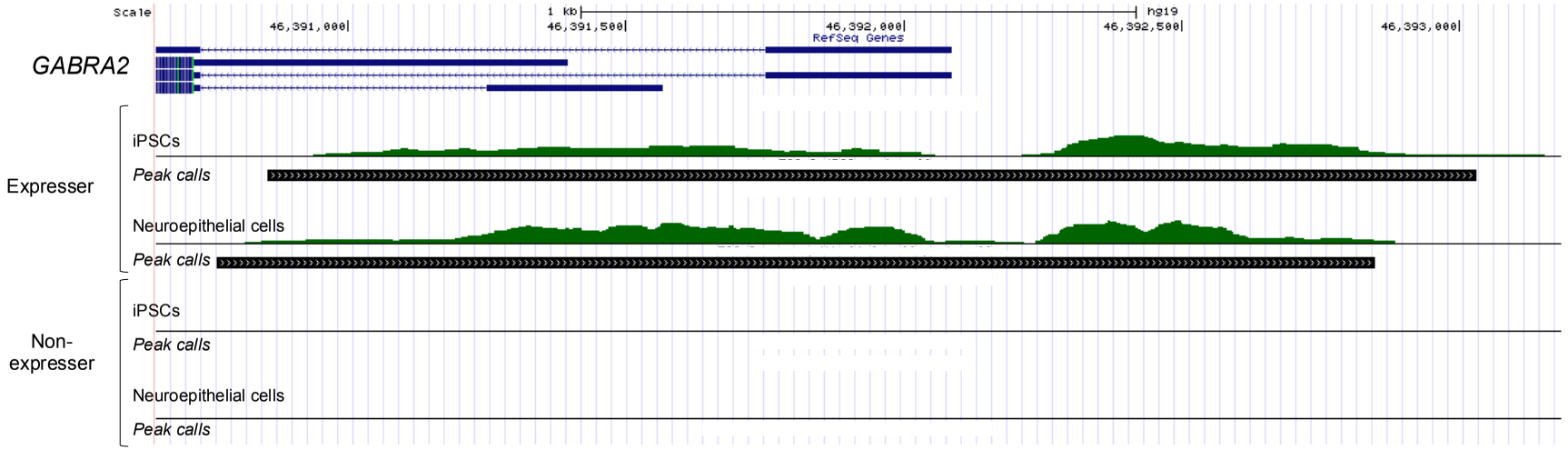


Figure 2.5. Raw ChIP-seq data tracks (green peaks) and peak calls (black bars) showing enrichment of H3K4me3 at the *GABRA2* promoter in an iPSC expresser, neuroepithelial cell expresser, iPSC non-expresser, and neuroepithelial cell non-expresser.

chromosome 4p12 $GABA_A$ receptor subunit genes, enrichment of H3K4me3 was examined in the promoters of *GABRG1*, *GABRA4*, and *GABRB1* (**Figure 2.6A-C**). Similar to *GABRA2*, enrichment of H3K4me3 in expresser iPSCs and iPSC-derived neuroepithelial cells was significantly higher at these promoters than enrichment in non-expresser iPSCs and iPSC-derived neuroepithelial cells.

We also assessed the flanking genes of the $GABA_A$ receptor subunit gene cluster on chromosome 4p12. *GUF1*, *GNPDA2*, and *COMMD8* all showed robust expression in all lines (**Figure 2.7A-C**).

When comparing peak calls between $GABA_A$ receptor subunit gene clusters on different chromosomes, there were no significant differences between the expresser or the non-expresser line in these clusters as there was with the chromosome 4p12 cluster (data not shown).

To contrast peaks in the ChIP-seq data, the application DiffBind was used. The number of H3K4me3 binding regions in the four samples was 21,032. When contrasting samples by expresser and non-expresser lines, there were 46 regions that were significantly different at a false discovery rate (FDR) <0.01 including the four $GABA_A$ receptor subunit genes clustered at chromosome 4p12 (see **Table 2.2** for full list). Contrasting iPSC and neuroepithelial cell samples, there were 5,960 significant differences found at a FDR <0.01 reflecting the large change in the enrichment distribution of this epigenetic mark associated with the differentiation process.

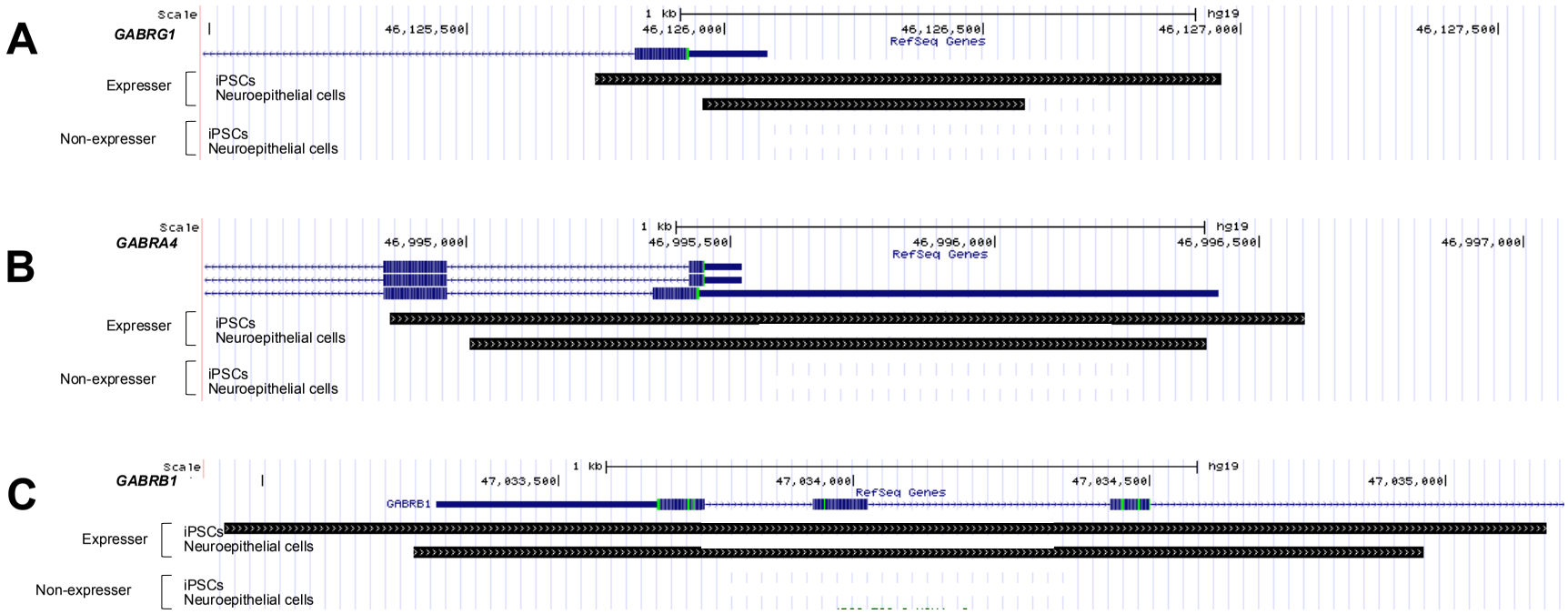


Figure 2.6. Peak calls (black bars) showing enrichment of H3K4me3 at promoters of the other genes within the chromosome 4p12 $GABA_A$ receptor subunit gene cluster. Graphs show enrichment at the (**A**) *GABRG1* promoter (**B**) *GABRA4* promoter, and (**C**) *GABRB1* promoter in an iPSC expresser, neuroepithelial cell expresser, iPSC non-expresser, and neuroepithelial cell non-expresser.

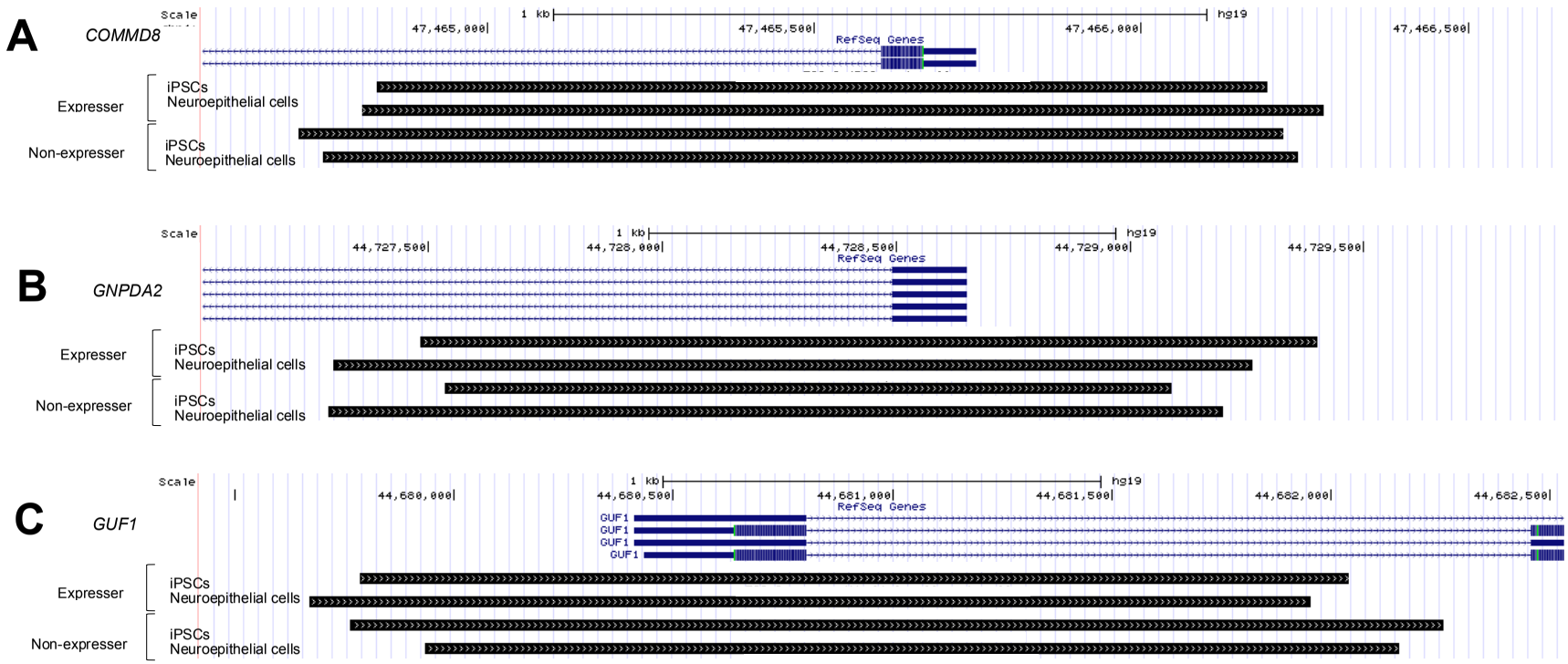


Figure 2.7. Peak calls (black bars) showing enrichment of H3K4me3 at promoters of genes flanking the chromosome 4p12 $GABA_A$ receptor subunit gene cluster. Graphs show enrichment at the (**A**) *COMMD8* promoter (**B**) *GNPDA2* promoter, and (**C**) *GUF1* promoter in expresser iPSCs, expresser neuroepithelial cells, non-expresser iPSCs, and non-expresser neuroepithelial cells.

Gene Name	Gene Description	Chromosome
ACOT1	acyl-CoA thioesterase 1	chr14
ANKRD20A5P	ankyrin repeat domain 20 family, member A5, pseudogene	chr18
BMS1P5	BMS1 pseudogene 5	chr10
CHCHD2	coiled-coil-helix-coiled-coil-helix domain containing 2	chr7
CNTNAP3	contactin associated protein-like 3	chr9
COL22A1	collagen, type XXII, alpha 1	chr8
CRYZ	crystallin, zeta (quinone reductase)	chr1
EIF3CL	eukaryotic translation initiation factor 3, subunit C-like	chr16
FAM86B3P	family with sequence similarity 86, member A pseudogene	chr8
GABRA2	gamma-aminobutyric acid (GABA) A receptor, alpha 2	chr4
GABRA4	gamma-aminobutyric acid (GABA) A receptor, alpha 4	chr4
GABRA5	gamma-aminobutyric acid (GABA) A receptor, alpha 5	chr15
GABRB1	gamma-aminobutyric acid (GABA) A receptor, beta 1	chr4
GABRG1	gamma-aminobutyric acid (GABA) A receptor, gamma 1	chr4
GRM7	glutamate receptor, metabotropic 7	chr3
HKR1	HKR1, GLI-Kruppel zinc finger family member	chr19
HLA-C	major histocompatibility complex, class I, C	chr6
HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	chr6
HLA-DRB1	major histocompatibility complex, class II, DR beta 1	chr6
LINC01060	long intergenic non-protein coding RNA 1060	chr4
LINC01194	long intergenic non-protein coding RNA 1194	chr5
LOC100287846	patched 1 pseudogene	chr8
LOC101927815	uncharacterized LOC101927815	chr8
LOC101927815	uncharacterized LOC101927815	chr8
LOC101927815	uncharacterized LOC101927815	chr8
LYNX1	Ly6/neurotoxin 1	chr8
NEUROD4	neuronal differentiation 4	chr12
NLRP2	NLR family, pyrin domain containing 2	chr19
NNAT	neuronatin	chr20
PARG	poly (ADP-ribose) glycohydrolase	chr10
PHACTR3	phosphatase and actin regulator 3	chr20
PHACTR3	phosphatase and actin regulator 3	chr20
PLS3	plastin 3	chrX
PNRC2	proline-rich nuclear receptor coactivator 2	chr1
SEPT7P9	septin 7 pseudogene 9	chr10
SLURP1	secreted LY6/PLAUR domain containing 1	chr8
SRSF10	serine/arginine-rich splicing factor 10	chr1
STAG3L2	stromal antigen 3-like 2 (pseudogene)	chr7
THEM6	thioesterase superfamily member 6	chr8
TRIM43	tripartite motif containing 43	chr2
TRPC7	transient receptor potential cation channel, subfamily C, member	chr5
TTTY3	testis-specific transcript, Y-linked 3 (non-protein coding)	chrY
TULP1	tubby like protein 1	chr6
ZNF132	zinc finger protein 132	chr19
ZNF558	zinc finger protein 558	chr19
ZSCAN23	zinc finger and SCAN domain containing 23	chr6

Table 2.2. DiffBind regions significantly different at a false discovery rate (FDR) <0.01 including the chromosome 4p12 $GABA_A$ receptor subunit genes (highlighted in orange).

Validating ChIP-seq with ChIP-PCR

ChIP-qPCR was first done in iPSCs to evaluate the presence or absence of a bivalent promoter state in our cell lines as the ChIP-seq data showed a lack of H3K4me3 enrichment in the non-expresser line. In these experiments, primers targeting the promoter of *GABRA2* were used, primers targeting the promoter of *GABRB2* were used to evaluate enrichment in a $GABA_A$ receptor subunit gene located on a different chromosome (chromosome 5), and primers targeting the promoter of *GUF1*, a flanking gene of the chromosome 4p12 $GABA_A$ receptor subunit gene cluster shown to have mRNA expression in both expresser and non-expresser lines. A negative control targeting a region within the *GABRA2* gene lacking enrichment of H3K4me3 and H3K27me3 (based on archived ChIP-seq data (see Figure 1.5)) was also examined. After qPCR with the *GABRA2* promoter region primers, we found that expresser iPSC lines showed robust enrichment of the active H3K4me3 mark and low, but present, enrichment of the repressive H3K27me3 mark (**Figure 2.8A, B**). However, the non-expresser iPSC lines, while showing some enrichment of the repressive H3K27me3 mark, showed significantly less enrichment of the active H3K4me3 mark compared to the expresser lines (**Figure 2.8A, B**) consistent with the ChIP-seq findings. This suggests that the promoter of the non-expresser cells was in fact never poised in iPSCs as there is a lack of H3K4me3 enrichment, and that the differences in mRNA expression fate between lines seen in neural cell cultures may be set not during neural differentiation, but in iPSCs.

To examine whether the differences seen between the iPSCs in the two expression groups were specific to the chromosome 4p12 $GABA_A$ receptor subunit

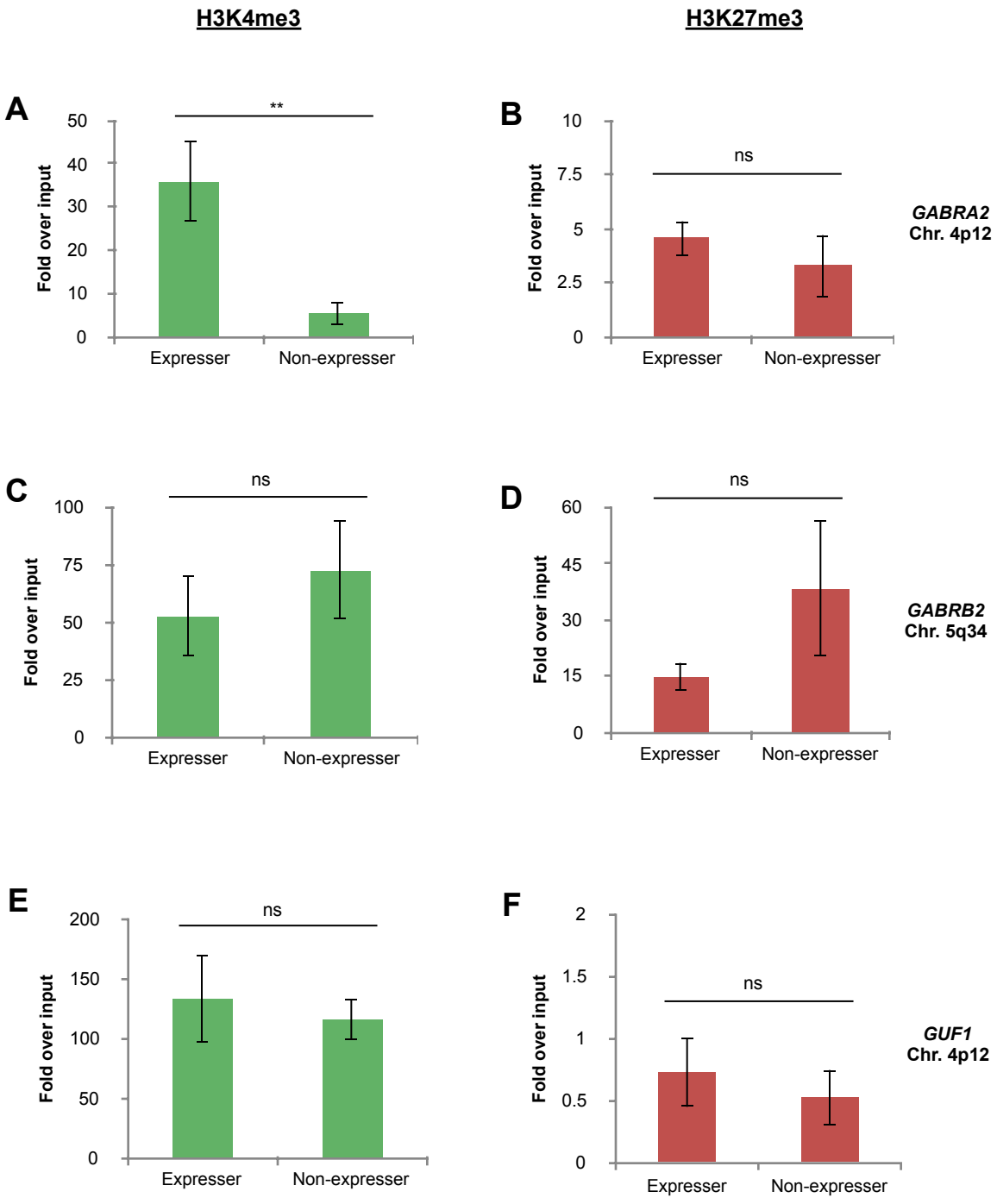


Figure 2.8. Enrichment of histone modifications in expresser and non-expresser iPSCs. **(A)** H3K4me3 enrichment at the *GABRA2* promoter (expresser n=7 and non-expresser n=4; **p=0.01). **(B)** H3K27me3 enrichment at the *GABRA2* promoter (expresser n=7 and non-expresser n=4; p=0.24). **(C)** H3K4me3 enrichment at the *GABRB2* promoter (expresser n=6 and non-expresser n=4; p=0.65). **(D)** H3K27me3 enrichment at the *GABRB2* promoter (expresser n=5 and non-expresser n=4; p=0.74). **(E)** H3K4me3 enrichment at the *GUF1* promoter (expresser n=7 and non-expresser n=3; p=0.83). **(F)** H3K27me3 enrichment at the *GUF1* promoter (expresser n=7 and non-expresser n=4; p=0.42).

gene cluster, we examined the promoter of *GABRB2*, a $GABA_A$ receptor subunit gene located on chromosome 5. Enrichment of H3K4me3 was robust in both expresser and non-expresser lines, with no significant difference between the two groups (**Figure 2.8C, D**). In addition, H3K27me3 enrichment was robust in both groups indicating that this gene promoter is in a poised promoter state in iPSCs with no significant difference between expressers and non-expresser lines (**Figure 2.8C, D**).

As mentioned in Chapter 1, Lieberman et al. (2015) discovered that the genes flanking the chromosome 4p12 $GABA_A$ receptor subunit gene cluster were expressed in neural cell cultures derived from both expresser and non-expresser lines indicating that transcriptional repression in this region is localized to the $GABA_A$ receptor subunit gene cluster in non-expressers. Because of this, we evaluated the histone modification enrichment at the promoter of *GUF1*, a flanking gene downstream of *GABRG1*. This showed a strong promoter with high H3K4me3 enrichment, but no H3K27me3 enrichment and with no difference seen between the expressers and non-expressers at this site (**Figure 2.8E, F**) consistent with the gene expression data (Lieberman, Kranzler et al. 2015). To measure background enrichment, a region within *GABRA2* that was unenriched (based on the Genome Browser data in Figure 1.5) was used as a negative control. As expected, both H3K4me3 and H3K27me3 enrichment was very low in both expresser and non-expresser lines (**Figure 2.9A, B**). *GAPDH* was used as a positive control for H3K4me3 enrichment showing robust enrichment in both expressers and non-expressers (**Figure 2.9C, D**), and *MYOD1* was used as a positive control for H3K27me3 enrichment showing robust enrichment in both expressers and non-expressers (**Figure 2.9E, F**).

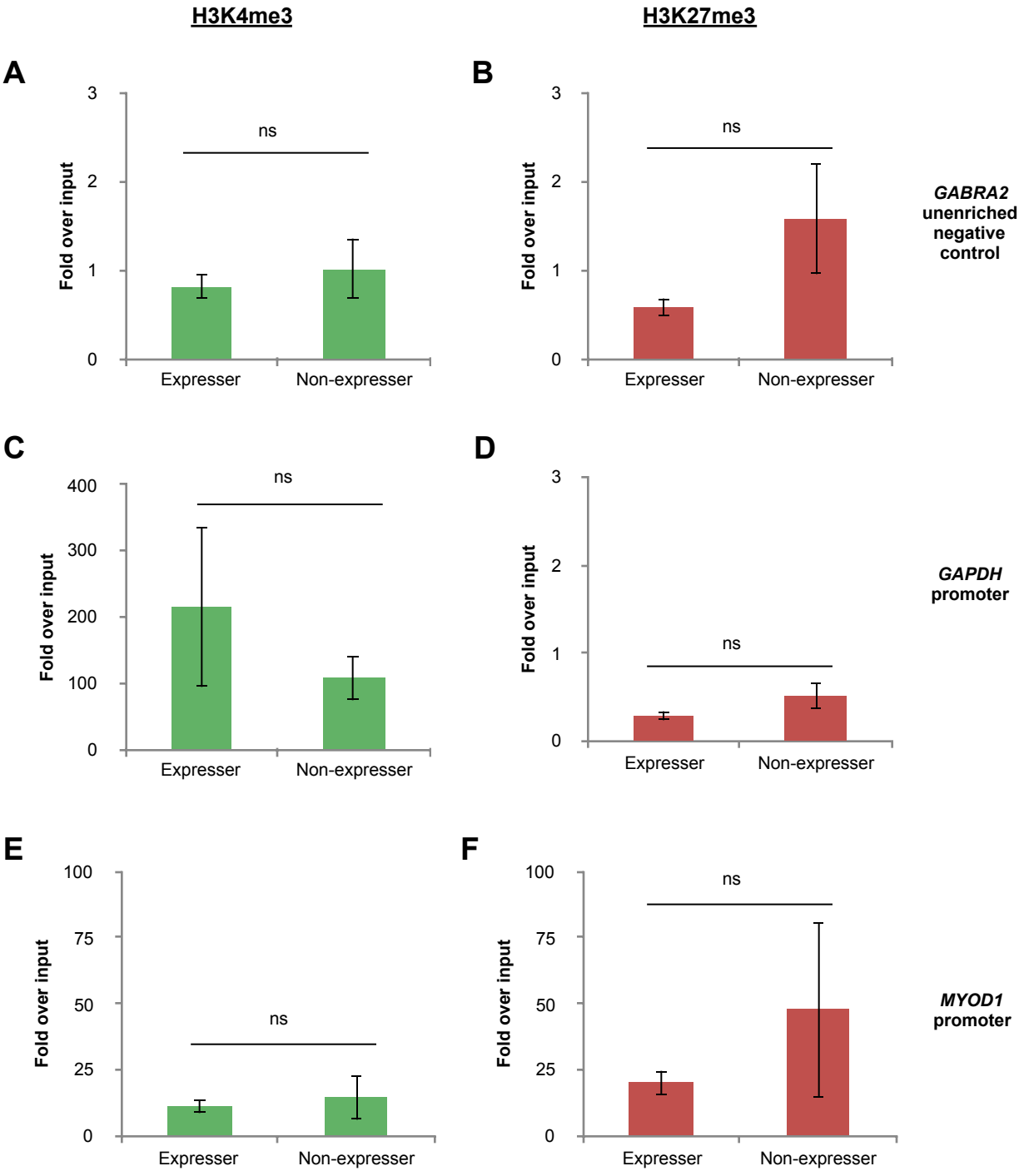


Figure 2.9. Enrichment of histone modifications in expresser and non-expresser iPSCs. **(A)** H3K4me3 enrichment at the *GABRA2* unenriched negative control (expresser n=7 and non-expresser n=4; p=0.68). **(B)** H3K27me3 enrichment at the *GABRA2* unenriched negative control (expresser n=7 and non-expresser n=4; p=0.3). **(C)** H3K4me3 enrichment at the *GAPDH* promoter (expresser n=6 and non-expresser n=4; p=0.42). **(D)** H3K27me3 enrichment at the *GAPDH* promoter (expresser n=6 and non-expresser n=4; p=0.25). **(E)** H3K4me3 enrichment at the *MYOD1* promoter (expresser n=6 and non-expresser n=4; p=0.71). **(F)** H3K27me3 enrichment at the *MYOD1* promoter (expresser n=5 and non-expresser n=4; p=0.52).

Discussion

Based on archived ChIP-seq data from the UCSC Genome Browser from H1 and H9 embryonic stem cell and iPSC 6.9 tracks (Figure 1.5) and the data from Ernst et al., 2011, we expected to see a basal poised/bivalent H3K4me3 and H3K27me3 epigenetic state at the *GABRA2* gene promoter in all iPSC lines followed by divergence of epigenetic marks reflecting gene expression status after beginning neural differentiation. However, we found that there is a difference in expresser and non-expresser lines already in iPSCs. While expresser lines showed enriched H3K4me3 and H3K27me3 marks, the non-expresser lines showed significantly less enrichment of the H3K4me3 mark, but showed similar levels of H3K27me3 enrichment as the expresser lines.

As there were significant differences of H3K4me3 enrichment between the expresser and non-expresser lines in the ChIP-seq experiments, it would be interesting to get a genome wide evaluation of H3K27me3 enrichment in expresser and non-expresser iPSCs. It was hard to tell in the qPCR experiments if there was much H3K27me3 enrichment in the non-expresser lines and having this additional ChIP-seq data may help give a comprehensive view of H3K4me3 versus H3K27me3 modifications in the chromosome 4p12 receptor subunit gene cluster region. Additionally, as mentioned in Chapter 1, the H3K27me3 mark is regulated by the polycomb repressive complex 2 (PRC2) and is normally associated with poised promoters and pluripotent cells. During differentiation if the gene is to be repressed, the H3K27me3 mark is replaced with H3K9me3, which is associated with gene repression in differentiated cells. It would be interesting to see if the non-expresser lines show

replacement of H3K27me3 with H3K9me3 in iPSCs or neuroepithelial cells in response to this yet-to-be identified functional variant.

The limited H3K4me3 enrichment in the *GABRA2* promoter in the non-expresser iPSC lines observed in our ChIP-seq and ChIP-qPCR datasets suggests that epigenetic reprogramming during iPSC generation is targeting the chromosome 4p12 receptor subunit gene cluster for inhibition in non-expresser lines. It is possible that this region in non-expressers acquires another repressive epigenetic mark that is impacting the ability of H3K4 methyltransferases to methylate K4. Another possibility, discussed in Chapter 3, is that there is differential DNA methylation at the CpG islands in the promoters of the chromosome 4p12 $GABA_A$ receptor subunit genes between expresser and non-expresser lines and this is inhibiting transcription activation.

Chapter 3.

DNA Methylation of $GABA_A$ Receptor Subunit Genes

Background

DNA methylation is a common epigenetic modification that can be altered when reprogramming fibroblasts to iPSCs. It is also known that there is a strong relationship between DNA methylation and histone modifications. In early development, CpG islands are typically unmethylated. This is perhaps due to the presence of methylation of the lysine 4 residue on the H3 tail (H3K4me). DNMT3L, a catalytically inactive binding partner of DNMT3A2 and DNMT3B, initiates the interaction between chromatin and the DNMT complex. It does this by recognizing the K4 residue on the H3 subunit tail. However, when H3K4 is methylated, it blocks the ability for DNMTL to bind inhibiting DNA methylation at this site. Thus, there is a negative correlation between methylation of H3K4 (associated with activation of genes) and DNA methylation of CpG islands (associated with gene repression). In addition, DNA methylation has been linked to H3K27me3, a mark associated with gene repression. It has been suggested that DNA methylation of CpG islands at promoters during differentiation is more likely if the promoter is marked with H3K27me3 in embryonic stem cells (Cedar and Bergman 2009, Rose and Klose 2014).

A study done in Bulgaria looking at genome-wide methylation profiles in schizophrenia found DNA hypermethylation of the CpG island in the *GABRA2* promoter in male schizophrenic subjects (Rukova, Staneva et al. 2014). In addition, a study looking at DNA methylation in aging, a progressive increase in DNA methylation of the

GABRA2 CpG island over the lifespan was observed in men (Siegmund, Connor et al. 2007). These studies suggest that differential DNA methylation of the *GABRA2* promoter CpG island could be a possible explanation for the two gene expression phenotypes we observe. Based on these studies and the ChIP findings in Chapter 2, we set out to evaluate a possible link between reprogramming fibroblasts into iPSCs and DNA methylation states of the *GABRA2* promoter CpG island, as well as the link to chromosome 4p12 $GABA_A$ receptor subunit gene cluster expression status. We hypothesize that the fibroblast cells will have DNA methylation present at the *GABRA2* promoter CpG island which are reset to an unmethylated state in $GABA_A$ receptor subunit gene cluster expresser lines but remain methylated in non-expresser lines.

Methods

Cell lines

Lines were the same as used in the 2015 Lieberman study (Lieberman, Kranzler et al. 2015). Briefly, skin punch biopsies were taken from 21 patients and were used to generate fibroblast cell lines. Fibroblasts were reprogrammed into induced pluripotent stem cells (iPSCs) by the Stem Cell Core at the University of Connecticut Health Center. This study utilized 24 expresser lines from 17 subjects and 15 non-expresser lines from 11 subjects to measure DNA methylation at the CpG island in *GABRA2*. For targeting *GABRA4* promoter CpG island DNA methylation, eight expresser lines lines from six subjects, and seven non-expresser lines from seven subjects were examined. For *GABRB1*, five expresser lines from four subjects, and five non-expresser lines from five subjects were examined. A gene flanking the $GABA_A$ receptor subunit gene cluster,

GNPDA2, whose RNA levels are not linked to the cluster expression pattern, was examined using two expresser lines from two subjects and three non-expresser lines from three subjects.

DNA methylation assay

Evaluation of DNA methylation was done using a combination of the method in Holemon et al. (Holemon, Korshunova et al. 2007) and the Qiagen Epitect Methyl-qPCR protocol. Briefly, 250ng of fibroblast, or iPSC DNA was split into four tubes: mock-enzyme (Mo), a CpG methylation sensitive (Ms) enzyme, HhaI (New England Biosystems), a CpG methylation dependent (Md) enzyme McrBC (New England Biosystems), and a double enzyme mix with both HhaI and McrBC (Msd). Samples were incubated overnight at 37˚C. Digested DNA was then analyzed by qPCR using primers for a site within the *GABRA2* promoter CpG island (**Figure 3.1**) using the RT^2 SYBR Green ROX qPCR Mastermix (Qiagen), a master mix designed for high GC content products. In addition, primer sets were designed for a site within the CpG islands in the promoters of *GABRA4*, *GABRB1*, and *GNPDA2* (**Table 3.1**). The qPCR amplification protocol was 95˚C for 10 minutes, 3 cycles of 99˚C for 30 seconds and 72˚C for 1 minute, and 40 cycles of 97˚C for 15 seconds and 72˚C for 1 minute. A dissociation curve was included at the end of the protocol to monitor melting curves for single melt peaks. Resulting Ct values for Mo, Ms, Md, and Msd were then entered into an Excel spreadsheet (www.sabiosciences.com/dna_methylation_data_analysis.php), that calculates digestion efficiency by giving an analytical window (W) and the percentage of DNA refractory (R) to restriction enzyme digestion, and also the

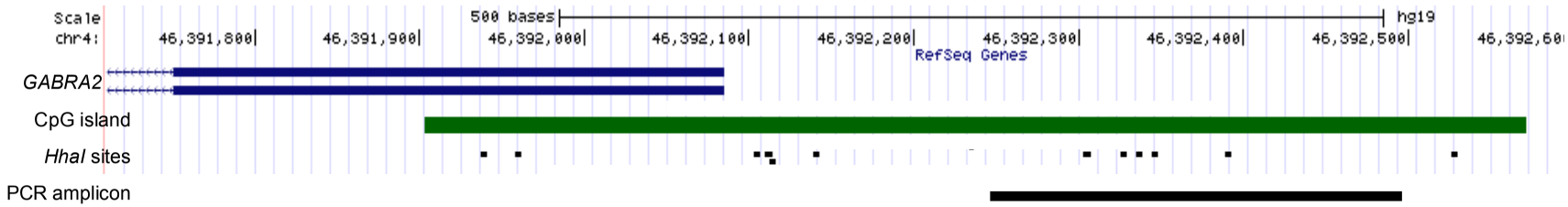


Figure 3.1. CpG island in the *GABRA2* promoter (green bar). The dots below represent *HhaI* (methylation sensitive) restriction enzyme sites within the CpG island. The black bar at the bottom marks the amplicon of interest used to assay methylation status in *GABRA2.*

Primer	Target	Sequence (5'-3')	Tm (˚C)
GA2CpG Forward v2	CpG island in the promoter of *GABRA2*	CTCGATCCCTCGATCCCTGGTAACT	61.2
GA2CpG Reverse v2		CGCCACACTACATTCCCGGCTTTA	61.3
GA4CpG Forward v3	CpG island in the promoter of *GABRA4*	CAGGGTGCGAGGAGAGGGC	63.1
GA4CpG Reverse v3		ACGAAAGGGTGTGGAGCGGT	61.7
GB1CpG Forward v2	CpG island in the promoter of *GABRB1*	TCTGCGTTTCATTGGCGGTCAC	60.5
GB1CpG Reverse v2		GGCAGGGCCTTGGTGCAT	61.7
GNPDA2CpG Forward	CpG island in the promoter of *GNPDA2*	GGCGTCCCAGTCTCAGCACAAA	62
GNPDA2CpG Reverse		CGGTTCCTTCCTGCGCCTTTATCT	61.2

Table 3.1. Primer sequences for DNA methylation studies.

percentage of unmethylated DNA and methylated DNA using the following equations, where F is the fraction of DNA evaluated:

DNA refractory to enzyme digestion: $$F_R = 2^{-\Delta CT(Msd-Mo)}$$

Unmethylated DNA: $$F_{UM} = \frac{2^{-CT(Md)}}{2^{-CT(Mo)} - 2^{-CT(Msd)}}$$

Hypermethylated DNA: $$F_{HM} = \frac{2^{-CT(Ms)}}{2^{-CT(Mo)} - 2^{-CT(Msd)}}$$

Intermediately methylated DNA: $$F_{IM} = 1 - F_{HM} - F_{UM}$$

Percentages were averaged among groups, standard error was calculated, and a Student's T-test was used to analyze significance between the percentage of total methylation between groups.

Results

De novo DNA methylation of the CpG island of the *GABRA2* promoter has been shown to occur in both schizophrenia and aging (Siegmund, Connor et al. 2007, Rukova, Staneva et al. 2014), possibly suggesting that DNA methylation at this locus can vary based on genetic variation and developmental processes. To evaluate the possible presence of DNA methylation at the *GABRA2* promoter, we used a DNA methylation assay which uses a methylation sensitive enzyme (HhaI), a methylation dependent enzyme (McrBC), and a double digest with both enzymes (see methods for further explanation). Digestion is followed by qPCR and Ct values are used to calculate

the percentage of unmethylated and methylated DNA. For example, if the target region is methylated, the Ms digest will result in little to no degradation of template resulting in qPCR Ct values equal to the mock digest, while the majority of the template will be cut by the Md enzyme resulting in an increase in Ct value.

We first examined fibroblast DNAs to evaluate CpG methylation status at the *GABRA2* promoter CpG island site. We had hypothesized that non-expresser fibroblasts possessed DNA methylation at the *GABRA2* promoter and that this was potentially maintained during the reprogramming process. Unexpectedly, we found that in dermal fibroblasts the *GABRA2* promoter CpG island was unmethylated in all lines tested, regardless of *GABRA2* tag-SNP genotype or expression phenotype (**Figure 3.2A**). After the reprogramming process, iPSC clones generated from the same fibroblasts became differentially methylated (**Figure 3.2B,** the two iPSC clones color-coded to parental fibroblast strain) indicating that reprogramming was inducing DNA methylation at this site, and that the level of methylation could differ between individual iPSC clones generated from the same fibroblast line. Further, there were discordant iPSC clones from the same parental fibroblasts. One such example is the CC line 703 (pink dots), which gave rise to one CpG hypermethylated clone and one unmethylated clone. Discordance among clones was also seen in chromosome 4p12 $GABA_A$ receptor subunit gene expression status in iPSC-derived neurons (Lieberman, Kranzler et al. 2015).

While the sample size in **Figure 3.2B** is small, there is a noticeable difference between TT lines, where only one of six clones shows greater than 80% CpG methylation at this site, and C-carrier lines where six out of eight have greater than 80%

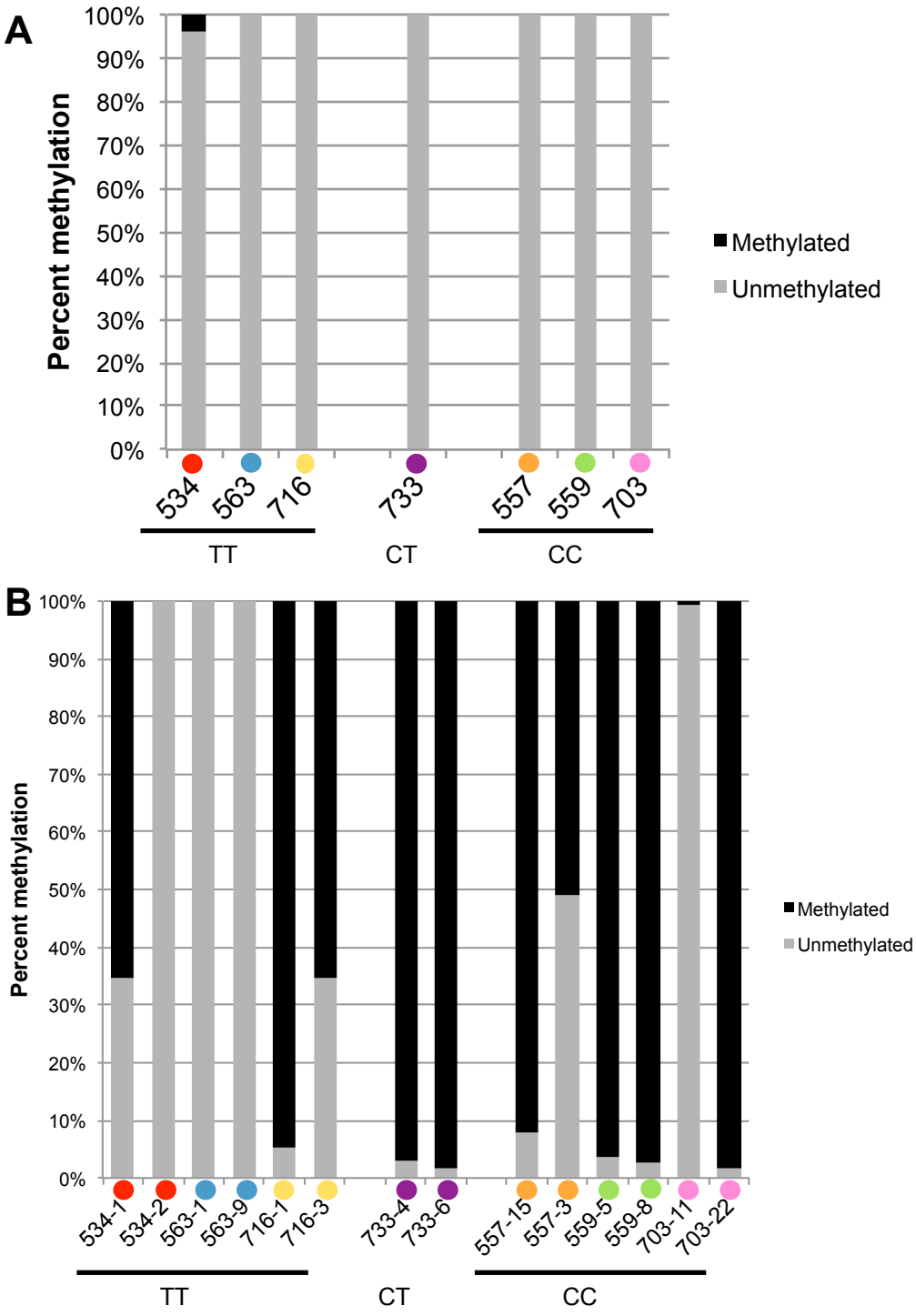


Figure 3.2. DNA methylation at the CpG island in the *GABRA2* promoter. **(A)** DNA methylation in fibroblasts in all three *GABRA2* rs279858 genotypes. Fibroblasts give rise to both expresser and non-expresser iPSC lines. **(B)** Two iPSC lines generated from each of the fibroblasts from **A.** Graph shows varying levels of DNA methylation at the *GABRA2* CpG island among clones from the same fibroblast line. Parental fibroblasts are color coded to their resulting iPSC clones.

methylation at this site. This suggests that there is potentially some relation of CpG methylation to genotype. To evaluate the potential difference between expresser and non-expresser lines, 24 expresser lines and 16 non-expresser lines were analyzed. From this, we found that *GABRA2* promoter CpG island methylation was greater in iPSC lines whose neural products were non-expressers, compared with expresser iPSC lines (**Figure 3.3**) (75% vs. 43%, respectively; *p=0.02).

To see if this difference in DNA methylation between expresser and non-expresser lines was specific to the *GABRA2* CpG island, or might be present for other $GABA_A$ receptor subunit genes on chromosome 4p12, we also examined sites within the CpG islands of the *GABRA4* and *GABRB1* promoters. The *GABRG1* promoter does not contain an annotated CpG island and therefore was not examined. Similar to *GABRA2*, non-expresser lines showed more DNA methylation at the CpG islands in both the *GABRA4* (**p=0.01) and *GABRB1* (***p=0.003) promoters (**Figure 3.4A, B**). Based on the histone modification data in Chapter 2 and the gene expression data published in Lieberman et al 2015, we predicted that this variation in CpG methylation would be localized to the $GABA_A$ cluster genes so we targeted a site in the CpG island in the promoter of a flanking gene, *GNPDA2*. We observed low levels of methylation for *GNPDA2* (**Figure 3.4C**); 0.04% +/- 0.01% expresser (n=2); 16.83% +/- 16.59% non-expresser (n=3) and no significant difference between expressers and non-expresser lines (p=0.4) at this site suggesting that this difference in DNA methylation is specific to the $GABA_A$ receptor subunit gene cluster on chromosome 4p12.

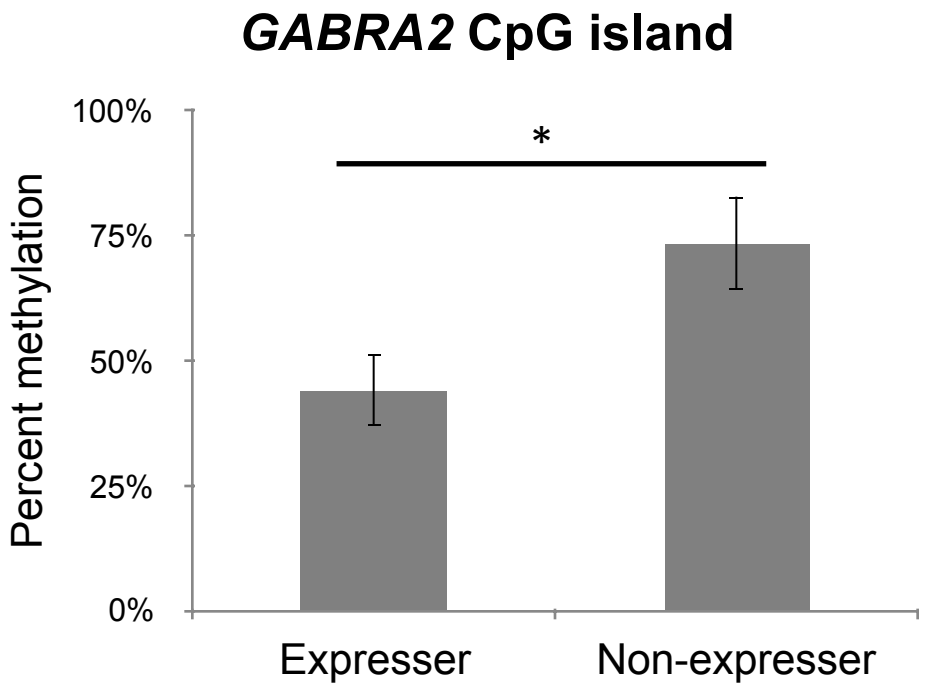


Figure 3.3. DNA methylation in iPSCs targeting a site in the CpG island in the promoter of GABRA2. Non-expresser lines had significantly more DNA methylation than expresser lines (expresser n=24 and non-expresser n=16; *p=0.02).

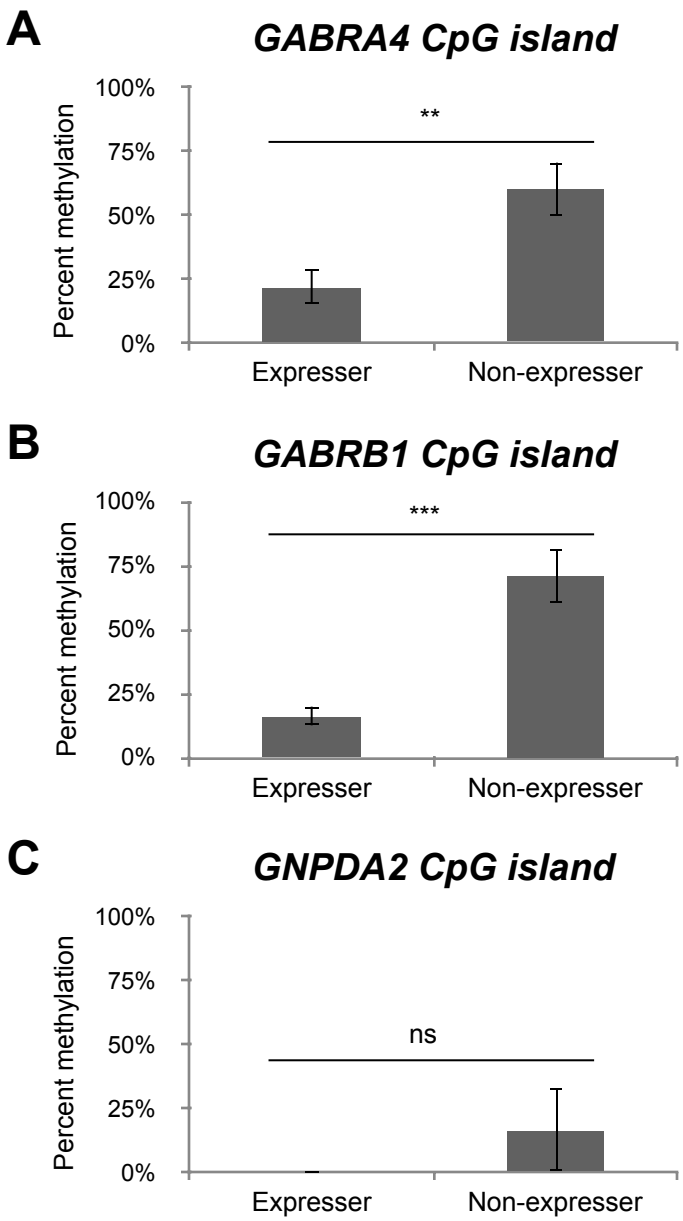


Figure 3.4. DNA methylation in expresser and non-expresser iPSCs. Percent DNA methylation at sites within the CpG island in the *GABRA4* promoter (expresser n=8 and non-expresser n=7; **p=0.01) **(A)**, the CpG island in the *GABRB1* promoter (expresser n=5 and non-expresser n=5; ***p=0.003) **(B)**, and the CpG island in the promoter of a flanking gene of the $GABA_A$ receptor subunit gene cluster, *GNPDA2* (expresser n=2 and non-expresser n=3; p=0.42) **(C).**

Discussion

It is known that the reprogramming process can result in aberrant side effects including changes in DNA methylation that are probably intrinsic to correct reprogramming (Lister, Pelizzola et al. 2011, Liang and Zhang 2013). However, for the lines evaluated in this thesis, this does not seem to be a random phenomenon. We found that DNA methylation at the *GABRA2* gene promoter in iPSCs is significantly related to the future chromosome 4p12 $GABA_A$ receptor subunit gene cluster expression status in neural cells. It is possible that the DNA methylation differences dissipate after sufficient passages; however, there were no noticeable changes in methylation of lines in early passages versus the same lines at later passages (data not shown). In addition, preliminary studies with neuroepithelial cells show maintenance of the DNA methylation status reflecting the iPSC line used (data not shown).

One thing we do not know, however, is whether the DNA methylation we are seeing in the lines is coming from the AUD risk C-allele, the T-allele, or both. Allelic discrimination assays done in Lieberman et al. 2015 looking at SNPs in *GABRG1*, *GABRA4*, and *GABRB1* suggest there is some allelic bias in gene expression in this region. This could mean there is also bias in DNA methylation, especially since DNA methylation is related to gene expression, or rather, repression.

A possible explanation of this increased DNA methylation at the *GABRA2* CpG island in non-expresser lines could be that the functional variant(s) associated with the tag-SNP rs279858 cause an increase in DNMT3A/B activity causing *de novo* DNA methylation. Because only one amplicon at each CpG island was evaluated, it is unknown if this is specific to the chromosome 4 $GABA_A$ receptor subunit gene cluster or

if there is a more global phenomenon in these lines. To test this, full genome sequencing after a methyl-ChIP pull-down could be used to examine this.

Another interesting point with these findings is that while we saw methylation differences between expressers and non-expressers, when the methylation data were grouped according to rs279858 genotype, there was a trend, but no significant differences (data not shown). This would suggest that the initial DNA methylation generated during reprogramming is random with respect to rs279858 tag-haplotype, but that the maintenance of the DNA methylation states may be biased toward the loss of methylation during neural differentiation in rs297858 TT lines and maintenance of *GABRA2* CpG methylation in CC lines. This would predict an association of methylation with genotype in neural cells, which had previously been associated with expression phenotype.

Chapter 4.

Discussion and Future Directions

Discussion

Chapter 1 presented an interesting phenotype seen when studying an AUD associated SNP in the *GABRA2* gene using iPSC-derived neural cell cultures. Lieberman et al., 2015, showed that lines derived from subjects carrying the rs279858 AUD risk C-allele showed decreased expression of a cluster of $GABA_A$ receptor subunit genes on chromosome 4p12; *GABRG1, GABRA2, GABRA4, and GABRB1* were termed “non-expresser” lines. They also found that the flanking genes of this cluster were expressed in both expresser and non-expresser lines. Because this phenotype was seen exclusively in this cluster of $GABA_A$ receptor subunit genes on chromosome 4p12, we hypothesized that they share a common epigenetic regulatory element.

In Chapter 2, we explored the idea that the cluster of $GABA_A$ receptor subunit genes on chromosome 4p12 was being regulated by changes in histone modifications within this region. We looked at two modifications, one associated with active chromatin, H3K4me3, and one associated with transcriptional repression, H3K27me3. Using ChIP followed by genome-wide sequencing and qPCR, we found significant differences between expresser and non-expresser lines not only in iPSC-derived neurons, but also, and perhaps unexpectedly, in iPSCs. While previous ChIP-seq data from the UCSC Genome Browser indicated a poised promoter in iPSCs and ES cells, we did not see expression of the H3K4me3 mark in non-expresser lines indicating not a poised gene, but a repressed one.

There is a known correlative relationship between the H3K27me3 mark and DNA methylation as they are both marks for repression, and a known inverse relationship between the H3K4me3 mark and CpG island DNA methylation (Cedar and Bergman 2009). In Chapter 3, we set out to elucidate possible differences in DNA methylation at CpG islands in the promoters of the $GABA_A$ receptor subunit genes on chromosome 4p12. We found that non-expresser lines showed significantly higher average DNA methylation compared to expresser lines, which were largely unmethylated. Only five of 24 lines were >80% methylated in expresser lines versus 11 of 16 in the non-expresser lines. The fact that we see not only a reduction in H3K4me3 enrichment in non-expresser lines, but also increased DNA methylation of the $GABA_A$ receptor subunit gene CpG islands, is consistent with the idea that histone methylation of H3K4 inhibits DNA methylation at CpG islands.

During differentiation, when promoters change from a poised state and begin to shift to either an active promoter or repressed, there can also be changes in the types of histone modifications present. The repressive H3K27me3 mark is common in pluripotent cells and is common in development, whereas the repressive H3K9me3 mark is associated with differentiated cells and is a more stable repressive mark. One possibility is that the H3K27me3 mark has been replaced by the H3K9me3 mark in the non-expresser lines so we are missing a key repressive mark, which will be looked at in future studies.

Future Directions

An alternative regulatory mechanism not explored in this research may involve the rs279858 SNP itself. This synonymous SNP changes a TTT-lysine into a TTC-lysine, with virtually no codon preference in humans. While previously thought of as "silent mutations," GWAS studies have consistently shown synonymous SNPs correlated to many illnesses (Plotkin and Kudla 2011). Though there is no real codon preference here, the reason for this expresser versus non-expresser phenotype could be linked to the change in codon itself. Changing the codon has been shown to adjust the mRNA structure, potentially leaving it less or more stable as a result, thus affecting protein generation, or protein structure and folding (Plotkin and Kudla 2011). However, based on our findings here, it would seem that this is not the case. It seems that there is in fact a functional variant yet to be discovered that is influencing the creation of histone modification profiles and DNA methylation in iPSCs.

While expresser and non-expresser status has been linked to rs279858 genotype as reported in Lieberman et al. 2015, there are outliers present. The trend is that TT homozygotes are expressers and that C-carriers are non-expressers. However, there are cases where a CT or CC subject can produce both an expresser line and a non-expresser line. In addition there are two TT subjects which produce non-expresser iPSC lines. This suggests that we are close to the functional variant, but that it is not directly linked to rs279858 genotype. However, using the tag-SNP rs279858 as a marker that we see is linked to epigenetic features like DNA methylation of CpG islands, we can use this SNP and this assay to screen additional lines in the future.

In addition, the findings in this study could be relevant on a greater scope. If we are seeing such a dramatic phenotype in iPSC lines with different genotypes at this one SNP, it is possible similar phenotypes exist in iPSCs from other illnesses. It is possible that other SNPs associated with other illnesses are also somehow indirectly influencing DNA methylation of CpG islands, or histone modifications as were associated with the rs279858 genotype.

Considering the fact that we see such a strong phenotype in iPSCs and in the early development of neurons, the effects that this phenotype might have in the developing brain are of interest. Perhaps this phenotype only occurs in regions linked to AUDs such as the frontal cortex and the reward pathway including the ventral tegmental area and the nucleus accumbens. This would lead to a change in the normal pattern of $GABA_A$ subunit utilization within these regions with different electrophysiological characteristics and subtle variation and result in the disruption of these neuronal systems.

Long distance chromatin looping

Another possible epigenetic feature is long distance chromatin looping. In addition to transcription factors binding promoters of genes, chromatin can also be looped bringing a regulatory element, such as an enhancer, into close proximity to a promoter affecting overall transcriptional output of the gene. A common protein that initiates long distance chromatin looping is the CCCTC-binding factor (CTCF), a 11-zinc finger protein. CTCF binds a specific sequence in the genome, CCCTC, and with the help of cohesin (DeMare, Leng et al. 2013, Merkenschlager and Odom 2013), CTCF

proteins can loop together bringing distant regulatory elements, such as enhancers, in close contact with promoters creating an active chromatin hub. This leads to higher levels of transcription (**Figure 4.1**; adapted from (Stadhouders, Thongjuea et al. 2012)). Based on ChIP-seq data (from the UCSC Genome Browser) done with an antibody to CTCF in iPSCs, there are multiple CTCF sites in and around the $GABA_A$ receptor subunit gene cluster on chromosome 4p12. One way to look at the 3D structure of these loops, and to see which sites interact with which, is to use Chromatin Conformation Capture (3C). This process involves crosslinking cells, digesting the DNA, ligating the ends together, and reverse crosslinking. The products are then detected by PCR, and the quantified products are mapped to the genome and interactions can be visualized based on a known anchor site to other distant sites (Naumova, Smith et al. 2012). A variation of this called Chromatin Interaction Analysis by Paired-End Tag sequencing, or ChIA-PET, uses this technique combined with immunoprecipitation. For example, studies done by Ruan et al., (2012) as part of the ENCODE project, used ChIA-PET in K562 cells to visualize long-distance chromatin looping between CTCF sites. The data from these studies indicate a possible loop encompassing the chromosome 4p12 $GABA_A$ receptor subunit gene cluster (**Figure 4.2**). It is possible that the functional SNP(s) linked to the tag-SNP rs279858 change an anchor point or distal point for long distance chromatin looping.

Targeted resequencing

The goal of this research is to aid in finding the functional genetic variant(s) within the *GABRA2* haplotype block associated with AUDs. In order to do so, our lab

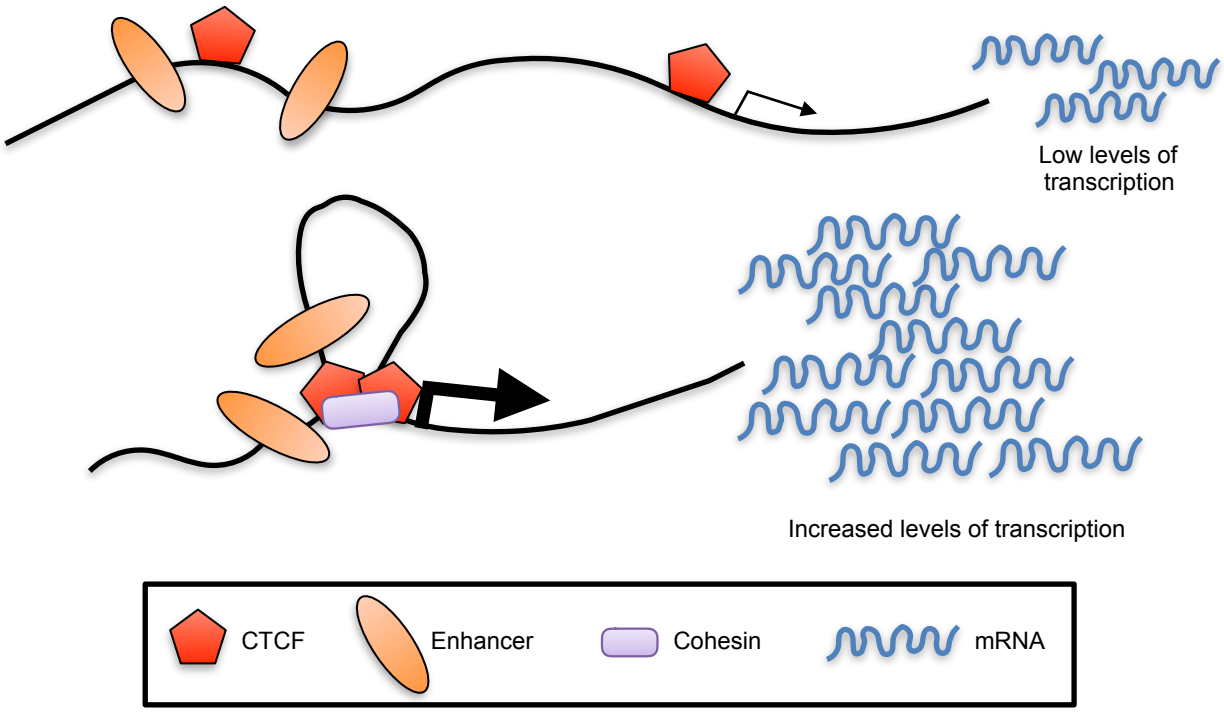


Figure 4.1. CTCF looping. CTCF binds a specific sequence in the genome and with the help of cohesin, CTCF proteins loop together bringing elements, such as enhancers, in close contact of promoters creating an active chromatin hub. This leads to higher levels of transcription (Figure adapted from Stadhouders, Thongjuea et al., 2012).

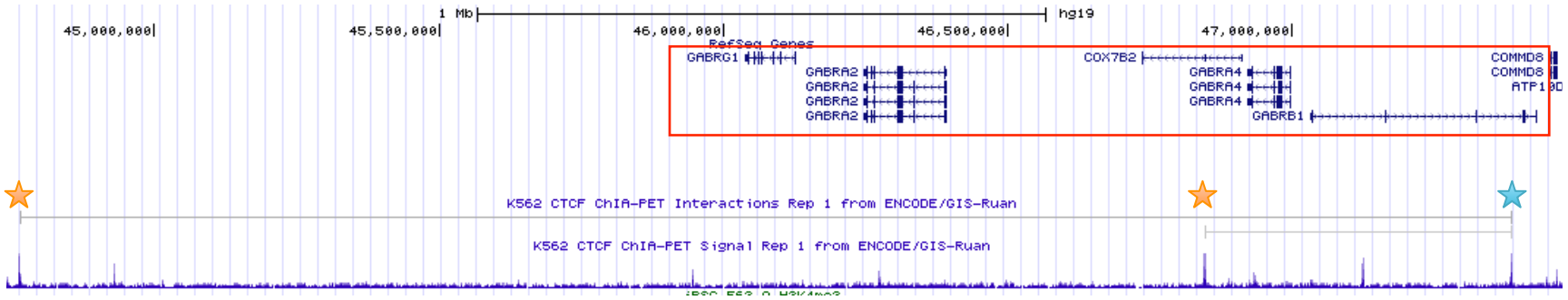


Figure 4.2. ChIA-PET from the UCSC Genome Browser looking at long distance chromatin looping initiated by CTCF in K562 cells. The red box encompasses the GABAA receptor subunit gene cluster on chromosome 4p12. The blue star represents a possible CTCF anchor site. The other two yellow stars show CTCF binding sites that the anchored CTCF may interact with. This suggests there is a short loop and a longer loop present in this region.

has done some targeted resequencing of the 140kb haplotype block using the Haloplex system. We looked at three TT and three CC expressers, and one TT and two CC non-expressers. While we did not find any SNPs associated with expression status stronger than the tag-SNP rs279858, we did find 224 SNPs in linkage disequilibrium with rs279858. In addition, we found a ~2kb deletion in the intergenic region between *GABRA2* and *GABRG1* in TT subjects, regardless of chromosome 4p12 $GABA_A$ receptor subunit gene cluster expression status (**Figure 4.3**). Data from the 1000 Genomes Project shows this deletion present in approximately 47% of Caucasian chromosomes. It is possible that CC subjects, which do not have this ~2kb deletion, have a sequence within this region prone to altering regional epigenetics, which ultimately influences the expression of the chromosome 4p12 $GABA_A$ receptor subunit genes. In fact, there is a *THE1B* long terminal repeat (LTR) sequence within this deletion, which has previously been shown to activate the proto-oncogene *CSF1R* in human lymphoma cells (Lamprecht, Walter et al. 2010). The activation of this LTR could potentially be recruiting DNMTs to this region of the genome, or be inhibiting methylation of H3K4 in non-expresser lines during neural differentiation.

Intriguingly, the frequency of this deletion is seen far less in African Americans, affecting only 22% of chromosomes and is not linked to rs279858, making these subjects an asset to finding the possible functional variant in this haplotype block. We have obtained ten African American fibroblast lines and genotyped them (**Table 4.1**). There is one line of particular interest as it is homozygous for TT at the rs279858 genotype. As mentioned above, Caucasian lines homozygous for TT also contain a ~2kb deletion in the intergenic region between *GABRA2* and *GABRG1*, this line,

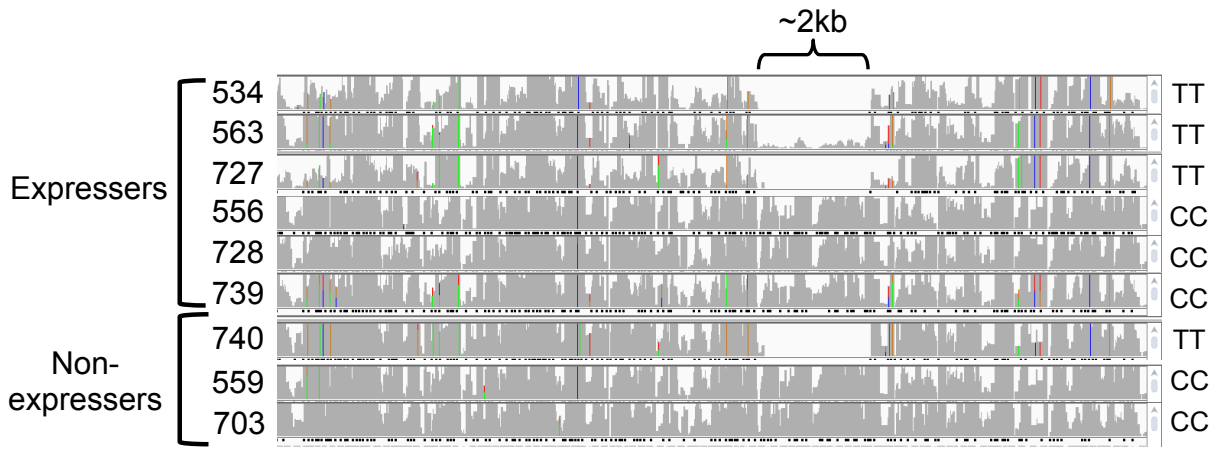


Figure 4.3. Targeted resequencing pile-up data of the region spanning the 140kb haplotype block (3' end of GABRA2 and the intergenic region between GABRA2 and GABRG1. TT lines show a 2kb deletion in this region, regardless of expression status.

Line	rs279858	2kb INDEL
CD2	CT	DEL/INS
CD3	TT	INS/INS
CD4	CT	INS/INS
CD5	CT	INS/INS
AG14298	CT	DEL/INS
GM03349C	TT	DEL/INS
AG09699	TT	DEL/DEL
AG09555	TT	DEL/DEL
AG09150B	TT	DEL/DEL
GM05757C	CT	DEL/INS

Table 4.1. African-American fibroblast lines with corresponding rs279858 genotype and the presence of a 2kb insertion/deletion in the intergenic region between *GABRA2* and *GABRG1*. The highlighted line is of particular interest.

however, does not have this deletion (highlighted line in **Table 4.1**). These lines will be reprogrammed into iPSCs by the UConn Stem Cell Core, and future studies will look at the histone modification profiles and DNA methylation pattern at CpG islands of these lines, and mRNA expression of $GABA_A$ receptor subunits genes will be measured in iPSC-derived neurons.

Targeted editing with CRISPR/Cas9

The clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR associated protein 9 (Cas9) (CRISPR/Cas9) system is based on an adaptive immune response against microphages in prokaryotes and is the newest genome editing tool first used in mammalian cells in 2013 (Hsu, Lander et al. 2014). A single guide RNA (sgRNA), composed of a CRISPR RNA (cRNA) fused to a trans-activating cRNA (tracrRNA) can be engineered to target a specific sequence in the genome where the Cas9 nuclease creates a double-stranded break allowing the genome to be edited at this site (Ceasar, Rajan et al. 2016). By utilizing this technology, it is possible to efficiently target and modify genetic mutations that cause disease. For example, when we discover candidates for the functional variant linked to our tag-SNP rs279858, we can use the CRISPR/Cas9 system to target and edit this variation in non-expresser lines to create isogenic versions of the previously deemed ‘non-expresser’ lines. By doing this, we should see changes in DNA methylation and histone modification enrichment at the promoters of the chromosome 4p12 $GABA_A$ receptor subunit gene cluster in iPSCs, and gene expression of the chromosome 4p12 $GABA_A$ receptor subunit gene cluster in neurons.

Alternatively, we could use the CRISPR/Cas9 system to find the functional variant. As mentioned above, our targeted resequencing of the 140kb haplotype block uncovered a 2kb deletion seen in the TT lines, regardless of expression status, that the CC lines did not possess. It was also mentioned that there is a *THE1B* sequence within this sequence that could potentially be creating the expresser and non-expresser lines. We could use CRISPR/Cas9 to target and this sequence in CC lines to see if DNA methylation decreases, histone modification enrichment at the promoters of the chromosome 4p12 $GABA_A$ receptor subunit gene cluster increases in iPSCs, and gene expression of the chromosome 4p12 $GABA_A$ receptor subunit gene cluster in neurons is restored.

Allelic imbalance

While once thought to be limited to X-linked and imprinted genes, recent studies have found that many autosomal genes also show monoallelic expression (Nag, Savova et al. 2013). In the recent publication from the lab, Lieberman et al., 2015, showed allelic imbalance in a subset of lines looking at iPSC-derived neural culture cDNA at SNPs in genes nearby rs279858 (Lieberman, Kranzler et al. 2015).

We found a SNP located in a H3K4me3 enriched region of the *GABRA4* promoter, rs2229940. In a preliminary experiment, we used a TaqMan genotype assay targeting this SNP with ChIP DNA from H3K4me3 and H3K27me3 immunoprecipitations to look at the presence of allelic imbalance of histone modification enrichment in this $GABA_A$ receptor subunit gene cluster. Though inconclusive, preliminary results suggest the presence of possible allelic bias suggesting there is random event that shows allelic

preference during reprogramming, that is possibly influenced by a functional variant in the 140kb haplotype block.

In addition to possible allelic bias of histone modifications, it is also unknown if the DNA methylation we are seeing is coming from only one allele, or a combination of the two. Using digested DNA from the DNA methylation assays, allelic imbalance assay can be done to assess allelic imbalance of DNA methylation in fibroblasts and iPSCs.

Comparing iPSCs to hESCs

We found a significant difference in DNA methylation at the CpG island in the *GABRA2* promoter depending on rs279858 genotype. An interesting experiment would be to look at DNA methylation at the CpG island in the *GABRA2* promoter as well as H3K4me3 and H3K27me3 enrichment in human embryonic stem cell (ESC) lines. Looking at hESC lines would shed light on whether what we are seeing in iPSCs is an actual developmental difference linked to rs279858 genotype, or an effect of reprogramming skin cell fibroblasts into iPSCs. Preliminary data in our lab has looked at DNA methylation at the *GABRA2* CpG island in five ESC lines. Among the lines were H1 (CC genotype), H9 (TT), CT2 (TT), CT3 (CT), and CT4 (TT). While the sample size was small, we saw no significant correlation to rs279858 genotype to DNA methylation, as most lines, regardless of genotype were unmethylated. This would suggest that what we are seeing in the iPSC lines is due to an event occurring during the reprogramming process. It is known that many aberrant effects can take place during reprogramming, such as random mutations, DNA methylation, and changes in histone modification profiles. However, even if what we are seeing is a reprogramming side effect, we have

shown that it is significantly related to rs279858 genotype indicating some link between genotype and epigenetic regulation. Further studies should be done to look at more hESC lines to see if this is a pluripotent stem cell phenomenon, or a reprogramming artifact cued by the functional variant linked to the rs279858 genotype.

Working model

The findings in this thesis have provided insights into the epigenetic factors affecting gene expression in iPSC-derived neural cultures from lines with different rs279858 genotypes. Based on these findings, we have established a working model (**Figure 4.4**). This model suggests that fibroblasts with different rs279858 genotypes have no epigenetic differences between lines. However, during the reprogramming process, placement of H3K4me3 marks and *de novo* DNA methylation at the CpG islands of the chromosome 4p12 $GABA_A$ receptor subunit gene cluster differs creating expresser lines and non-expresser lines. As mentioned in Chapter 3 discussion, there were no significant differences in the amount of DNA methylation between genotypes, but there were significant differences between expressers and non-expressers. This could indicate that the initial DNA methylation during reprogramming is random, but the maintenance of this methylation could differ between genotypes. The absence of H3K4me3 in non-expresser lines also provides an opportunity for the CpG islands in this region to be methylated. Then when iPSCs are differentiated into neurons, the DNA methylation resolves and is either present (in non-expresser lines) or absent (in expresser lines).

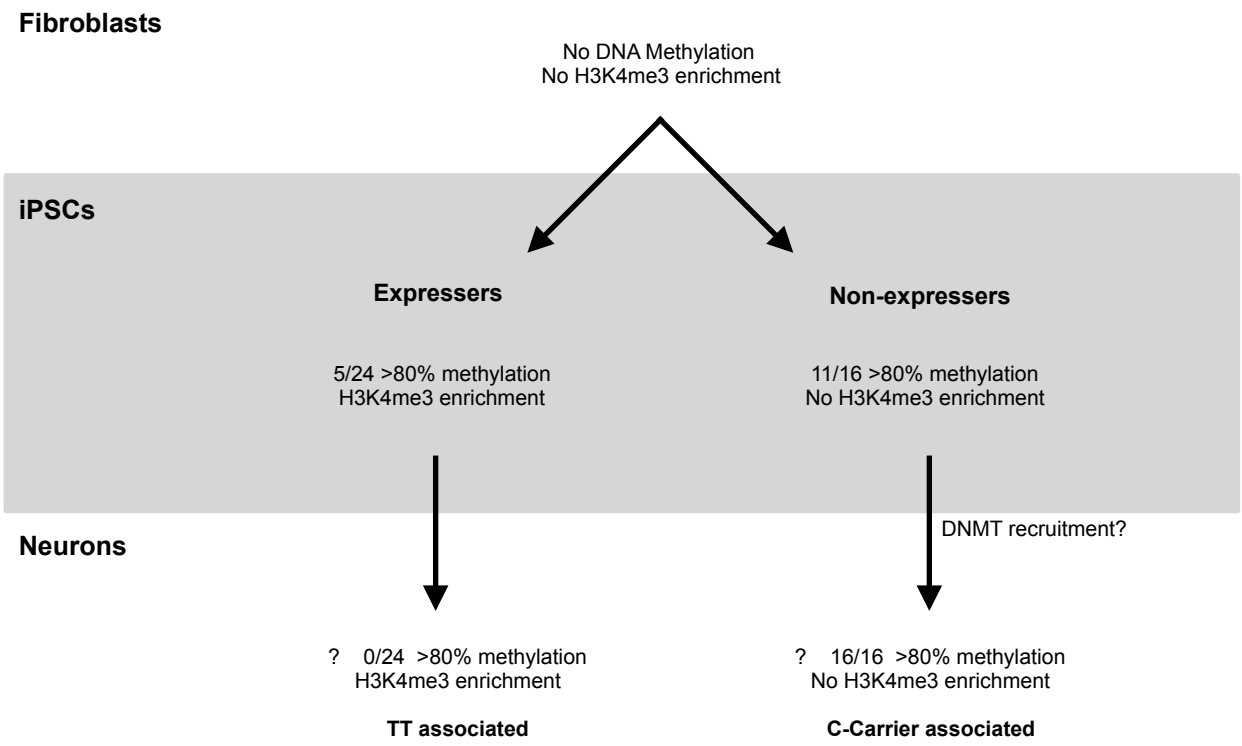


Figure 4.4. Working model.

Our revised hypothesis is that the yet-to-be-identified functional variant inhibits long distance chromatin looping, inhibits H3K4 methylation, and/or recruits DNMT3A and DNMT3B for *de novo* DNA methylation, resulting in repression of this region.

Conclusion

This research provides a small entrée into understanding the epigenetic changes seen in iPSCs after reprogramming and how they may regulate and affect gene expression in later development, especially as it relates to AUDs. Research models are never perfect or completely applicable, and the use of iPSCs in researching neurological illnesses certainly has its limitations. However, IPSCs and iPSC-derived neurons have provided a new unparalleled way to study neurological illnesses. They have given us access to live human neurons and allowed us to study the electrophysiological, morphological, and molecular phenotypes previously unknown. Now that we have the tools to study AUDs at a cellular and molecular level, we can focus on more specific elements of why certain SNPs increase the risk of developing an AUD.

References

- Aguayo, L. G. (1990). "Ethanol potentiates the GABAA-activated Cl- current in mouse hippocampal and cortical neurons." Eur J Pharmacol **187**(1): 127-130.
- Barnard, E. A., P. Skolnick, R. W. Olsen, H. Mohler, W. Sieghart, G. Biggio, C. Braestrup, A. N. Bateson and S. Z. Langer (1998). "International Union of Pharmacology. XV. Subtypes of gamma-aminobutyric acidA receptors: classification on the basis of subunit structure and receptor function." Pharmacol Rev **50**(2): 291-313.
- Blecher-Gonen, R., Z. Barnett-Itzhaki, D. Jaitin, D. Amann-Zalcenstein, D. Lara-Astiaso and I. Amit (2013). "High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states." Nat Protoc **8**(3): 539-554.
- Boffetta, P., M. Hashibe, C. La Vecchia, W. Zatonski and J. Rehm (2006). "The burden of cancer attributable to alcohol drinking." Int J Cancer **119**(4): 884-887.
- Brennand, K. J., A. Simone, J. Jou, C. Gelboin-Burkhart, N. Tran, S. Sangar, Y. Li, Y. Mu, G. Chen, D. Yu, S. McCarthy, J. Sebat and F. H. Gage (2011). "Modelling schizophrenia using human induced pluripotent stem cells." Nature **473**(7346): 221-225.
- Cashman, C. R. and L. Lazzerini Ospri (2013). "Induced pluripotent stem cells and motor neuron disease: toward an era of individualized medicine." J Neurosci **33**(20): 8587-8589.
- Ceasar, S. A., V. Rajan, S. V. Prykhozhij, J. N. Berman and S. Ignacimuthu (2016). "Insert, remove or replace: A highly advanced genome editing system using CRISPR/Cas9." Biochim Biophys Acta **1863**(9): 2333-2344.
- Cedar, H. and Y. Bergman (2009). "Linking DNA methylation and histone modification: patterns and paradigms." Nat Rev Genet **10**(5): 295-304.
- Chamberlain, S. J., P. F. Chen, K. Y. Ng, F. Bourgois-Rocha, F. Lemtiri-Chlieh, E. S. Levine and M. Lalande (2010). "Induced pluripotent stem cell models of the genomic imprinting disorders Angelman and Prader-Willi syndromes." Proc Natl Acad Sci U S A **107**(41): 17668-17673.
- Chen, H., K. Qian, Z. Du, J. Cao, A. Petersen, H. Liu, L. W. t. Blackbourn, C. L. Huang, A. Errigo, Y. Yin, J. Lu, M. Ayala and S. C. Zhang (2014). "Modeling ALS with iPSCs reveals that mutant SOD1 misregulates neurofilament balance in motor neurons." Cell Stem Cell **14**(6): 796-809.
- Cheng, X., R. E. Collins and X. Zhang (2005). "Structural and sequence motifs of protein (histone) methylation enzymes." Annu Rev Biophys Biomol Struct **34**: 267-294.
- Conigrave, K. M., B. F. Hu, C. A. Camargo, Jr., M. J. Stampfer, W. C. Willett and E. B. Rimm (2001). "A prospective study of drinking patterns in relation to risk of type 2 diabetes among men." Diabetes **50**(10): 2390-2395.
- Cotney, J. L. and J. P. Noonan (2015). "Chromatin immunoprecipitation with fixed animal tissues and preparation for high-throughput sequencing." Cold Spring Harb Protoc **2015**(2): 191-199.

- Covault, J., J. Gelernter, V. Hesselbrock, M. Nellissery and H. R. Kranzler (2004). "Allelic and haplotypic association of GABRA2 with alcohol dependence." Am J Med Genet B Neuropsychiatr Genet **129b**(1): 104-109.
- Covault, J., J. Gelernter, K. Jensen, R. Anton and H. R. Kranzler (2008). "Markers in the 5'-region of GABRG1 associate to alcohol dependence and are in linkage disequilibrium with markers in the adjacent GABRA2 gene." Neuropsychopharmacology **33**(4): 837-848.
- Cruvinel, E., T. Budinetz, N. Germain, S. Chamberlain, M. Lalande and K. Martins-Taylor (2014). "Reactivation of maternal SNORD116 cluster via SETDB1 knockdown in Prader-Willi syndrome iPSCs." Hum Mol Genet **23**(17): 4674-4685.
- DeMare, L. E., J. Leng, J. Cotney, S. K. Reilly, J. Yin, R. Sarro and J. P. Noonan (2013). "The genomic landscape of cohesin-associated chromatin interactions." Genome Res **23**(8): 1224-1234.
- Devor, E. J. and C. R. Cloninger (1989). "Genetics of alcoholism." Annu Rev Genet **23**: 19-36.
- Djousse, L., M. L. Biggs, K. J. Mukamal and D. S. Siscovick (2007). "Alcohol consumption and type 2 diabetes among older adults: the Cardiovascular Health Study." Obesity (Silver Spring) **15**(7): 1758-1765.
- Edenberg, H. J. (2007). "The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants." Alcohol Res Health **30**(1): 5-13.
- Edenberg, H. J., D. M. Dick, X. Xuei, H. Tian, L. Almasy, L. O. Bauer, R. R. Crowe, A. Goate, V. Hesselbrock, K. Jones, J. Kwon, T. K. Li, J. I. Nurnberger, Jr., S. J. O'Connor, T. Reich, J. Rice, M. A. Schuckit, B. Porjesz, T. Foroud and H. Begleiter (2004). "Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations." Am J Hum Genet **74**(4): 705-714.
- Enoch, M. A. (2008). "The role of GABA(A) receptors in the development of alcoholism." Pharmacol Biochem Behav **90**(1): 95-104.
- Enoch, M. A., C. A. Hodgkinson, Q. Yuan, B. Albaugh, M. Virkkunen and D. Goldman (2009). "GABRG1 and GABRA2 as independent predictors for alcoholism in two populations." Neuropsychopharmacology **34**(5): 1245-1254.
- Ernst, J. and M. Kellis (2010). "Discovery and characterization of chromatin states for systematic annotation of the human genome." Nat Biotechnol **28**(8): 817-825.
- Ernst, J., P. Kheradpour, T. S. Mikkelsen, N. Shoresh, L. D. Ward, C. B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis and B. E. Bernstein (2011). "Mapping and analysis of chromatin state dynamics in nine human cell types." Nature **473**(7345): 43-49.
- Esteller, M. (2007). "Epigenetic gene silencing in cancer: the DNA hypermethylome." Hum Mol Genet **16 Spec No 1**: R50-59.
- Fehr, C., T. Sander, A. Tadic, K. P. Lenzen, I. Anghelescu, C. Klawe, N. Dahmen, L. G. Schmidt and A. Szegedi (2006). "Confirmation of association of the GABRA2 gene with alcohol dependence by subtype-specific analysis." Psychiatr Genet **16**(1): 9-17.
- Fillman, S. G., C. E. Duncan, M. J. Webster, M. Elashoff and C. S. Weickert (2010). "Developmental co-regulation of the beta and gamma GABAA receptor subunits with

distinct alpha subunits in the human dorsolateral prefrontal cortex." Int J Dev Neurosci **28**(6): 513-519.
- Gelernter, J. and H. R. Kranzler (2009). "Genetics of alcohol dependence." Hum Genet **126**(1): 91-99.
- Goldman, D., G. Oroszi and F. Ducci (2005). "The genetics of addictions: uncovering the genes." Nat Rev Genet **6**(7): 521-532.
- Grant, B. F., R. B. Goldstein, T. D. Saha, S. P. Chou, J. Jung, H. Zhang, R. P. Pickering, W. J. Ruan, S. M. Smith, B. Huang and D. S. Hasin (2015). "Epidemiology of DSM-5 Alcohol Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions III." JAMA Psychiatry **72**(8): 757-766.
- Gu, B. and M. G. Lee (2013). "Histone H3 lysine 4 methyltransferases and demethylases in self-renewal and differentiation of stem cells." Cell Biosci **3**(1): 39.
- Hashimoto, T., Q. L. Nguyen, D. Rotaru, T. Keenan, D. Arion, M. Beneyto, G. Gonzalez-Burgos and D. A. Lewis (2009). "Protracted developmental trajectories of GABAA receptor alpha1 and alpha2 subunit expression in primate prefrontal cortex." Biol Psychiatry **65**(12): 1015-1023.
- Heitzeg, M. M., S. Villafuerte, B. J. Weiland, M. A. Enoch, M. Burmeister, J. K. Zubieta and R. A. Zucker (2014). "Effect of GABRA2 genotype on development of incentive-motivation circuitry in a sample enriched for alcoholism risk." Neuropsychopharmacology **39**(13): 3077-3086.
- Holemon, H., Y. Korshunova, J. M. Ordway, J. A. Bedell, R. W. Citek, N. Lakey, J. Leon, M. Finney, J. D. McPherson and J. A. Jeddeloh (2007). "MethylScreen: DNA methylation density monitoring using quantitative PCR." Biotechniques **43**(5): 683-693.
- Hsu, P. D., E. S. Lander and F. Zhang (2014). "Development and applications of CRISPR-Cas9 for genome engineering." Cell **157**(6): 1262-1278.
- Husain, K., R. A. Ansari and L. Ferder (2014). "Alcohol-induced hypertension: Mechanism and prevention." World J Cardiol **6**(5): 245-252.
- Jacob, T. C., S. J. Moss and R. Jurd (2008). "GABA(A) receptor trafficking and its role in the dynamic modulation of neuronal inhibition." Nat Rev Neurosci **9**(5): 331-343.
- Jones, P. A. (2012). "Functions of DNA methylation: islands, start sites, gene bodies and beyond." Nat Rev Genet **13**(7): 484-492.
- Kareken, D. A., T. Liang, L. Wetherill, M. Dzemidzic, V. Bragulat, C. Cox, T. Talavage, S. J. O'Connor and T. Foroud (2010). "A polymorphism in GABRA2 is associated with the medial frontal response to alcohol cues in an fMRI study." Alcohol Clin Exp Res **34**(12): 2169-2178.
- Lamprecht, B., K. Walter, S. Kreher, R. Kumar, M. Hummel, D. Lenze, K. Kochert, M. A. Bouhlel, J. Richter, E. Soler, R. Stadhouders, K. Johrens, K. D. Wurster, D. F. Callen, M. F. Harte, M. Giefing, R. Barlow, H. Stein, I. Anagnostopoulos, M. Janz, P. N. Cockerill, R. Siebert, B. Dorken, C. Bonifer and S. Mathas (2010). "Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma." Nat Med **16**(5): 571-579, 571p following 579.
- Lappalainen, J., E. Krupitsky, M. Remizov, S. Pchelina, A. Taraskina, E. Zvartau, L. K. Somberg, J. Covault, H. R. Kranzler, J. H. Krystal and J. Gelernter (2005). "Association between alcoholism and gamma-amino butyric acid alpha2 receptor subtype in a Russian population." Alcohol Clin Exp Res **29**(4): 493-498.

- Laurie, D. J., W. Wisden and P. H. Seeburg (1992). "The distribution of thirteen GABAA receptor subunit mRNAs in the rat brain. III. Embryonic and postnatal development." J Neurosci **12**(11): 4151-4172.
- Leitzmann, M. F., E. L. Giovannucci, M. J. Stampfer, D. Spiegelman, G. A. Colditz, W. C. Willett and E. B. Rimm (1999). "Prospective study of alcohol consumption patterns in relation to symptomatic gallstone disease in men." Alcohol Clin Exp Res **23**(5): 835-841.
- Li, E., T. H. Bestor and R. Jaenisch (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." Cell **69**(6): 915-926.
- Liang, G. and Y. Zhang (2013). "Genetic and epigenetic variations in iPSCs: potential causes and implications for application." Cell Stem Cell **13**(2): 149-159.
- Lieberman, R., H. R. Kranzler, P. Joshi, D. G. Shin and J. Covault (2015). "GABRA2 Alcohol Dependence Risk Allele is Associated with Reduced Expression of Chromosome 4p12 GABAA Subunit Genes in Human Neural Cultures." Alcohol Clin Exp Res **39**(9): 1654-1664.
- Lister, R., M. Pelizzola, Y. S. Kida, R. D. Hawkins, J. R. Nery, G. Hon, J. Antosiewicz-Bourget, R. O'Malley, R. Castanon, S. Klugman, M. Downes, R. Yu, R. Stewart, B. Ren, J. A. Thomson, R. M. Evans and J. R. Ecker (2011). "Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells." Nature **471**(7336): 68-73.
- Litten, R. Z., M. Egli, M. Heilig, C. Cui, J. B. Fertig, M. L. Ryan, D. E. Falk, H. Moss, R. Huebner and A. Noronha (2012). "Medications development to treat alcohol dependence: a vision for the next decade." Addict Biol **17**(3): 513-527.
- Maheswaran, R., J. S. Gill, P. Davies and D. G. Beevers (1991). "High blood pressure due to alcohol. A rapidly reversible effect." Hypertension **17**(6 Pt 1): 787-792.
- Mak, I. W., N. Evaniew and M. Ghert (2014). "Lost in translation: animal models and clinical trials in cancer treatment." Am J Transl Res **6**(2): 114-118.
- Merkenschlager, M. and D. T. Odom (2013). "CTCF and cohesin: linking gene regulatory elements with their targets." Cell **152**(6): 1285-1297.
- Moore, L. D., T. Le and G. Fan (2013). "DNA methylation and its basic function." Neuropsychopharmacology **38**(1): 23-38.
- Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer and B. Wold (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq." Nat Methods **5**(7): 621-628.
- Nag, A., V. Savova, H. L. Fung, A. Miron, G. C. Yuan, K. Zhang and A. A. Gimelbrant (2013). "Chromatin signature of widespread monoallelic expression." Elife **2**: e01256.
- Naumova, N., E. M. Smith, Y. Zhan and J. Dekker (2012). "Analysis of long-range chromatin interactions using Chromosome Conformation Capture." Methods **58**(3): 192-203.
- O'Keefe, J. H., K. A. Bybee and C. J. Lavie (2007). "Alcohol and cardiovascular health: the razor-sharp double-edged sword." J Am Coll Cardiol **50**(11): 1009-1014.
- O'Shea, K. S. and M. G. McInnis (2015). "Induced pluripotent stem cell (iPSC) models of bipolar disorder." Neuropsychopharmacology **40**(1): 248-249.
- O'Shea, R. S., S. Dasarathy and A. J. McCullough (2010). "Alcoholic liver disease." Am J Gastroenterol **105**(1): 14-32; quiz 33.

- Okano, M., D. W. Bell, D. A. Haber and E. Li (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-257.
- Olfson, E. and L. J. Bierut (2012). "Convergence of genome-wide association and candidate gene studies for alcoholism." Alcohol Clin Exp Res **36**(12): 2086-2094.
- Ooi, S. K., C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage, P. Tempst, S. P. Lin, C. D. Allis, X. Cheng and T. H. Bestor (2007). "DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA." Nature **448**(7154): 714-717.
- Perrin, S. (2014). "Preclinical research: Make mouse studies work." Nature **507**(7493): 423-425.
- Picton, A. J. and J. L. Fisher (2007). "Effect of the alpha subunit subtype on the macroscopic kinetic properties of recombinant GABA(A) receptors." Brain Res **1165**: 40-49.
- Pirker, S., C. Schwarzer, A. Wieselthaler, W. Sieghart and G. Sperk (2000). "GABA(A) receptors: immunocytochemical distribution of 13 subunits in the adult rat brain." Neuroscience **101**(4): 815-850.
- Plotkin, J. B. and G. Kudla (2011). "Synonymous but not the same: the causes and consequences of codon bias." Nat Rev Genet **12**(1): 32-42.
- Prescott, C. A. and K. S. Kendler (1999). "Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins." Am J Psychiatry **156**(1): 34-40.
- Rose, N. R. and R. J. Klose (2014). "Understanding the relationship between DNA methylation and histone lysine methylation." Biochim Biophys Acta **1839**(12): 1362-1372.
- Rukova, B., R. Staneva, S. Hadjidekova, G. Stamenov, t. Milanova and D. Toncheva (2014). "Genome-wide methylation profiling of schizophrenia." Balkan J Med Genet **17**(2): 15-23.
- Rukova, B., R. Staneva, S. Hadjidekova, G. Stamenov, V. Milanova and D. Toncheva (2014). "Whole genome methylation analyses of schizophrenia patients before and after treatment." Biotechnol Biotechnol Equip **28**(3): 518-524.
- Ruthenburg, A. J., C. D. Allis and J. Wysocka (2007). "Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark." Mol Cell **25**(1): 15-30.
- Siegmund, K. D., C. M. Connor, M. Campan, T. I. Long, D. J. Weisenberger, D. Biniszkiewicz, R. Jaenisch, P. W. Laird and S. Akbarian (2007). "DNA methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons." PLoS One **2**(9): e895.
- Stadhouders, R., S. Thongjuea, C. Andrieu-Soler, R. J. Palstra, J. C. Bryne, A. van den Heuvel, M. Stevens, E. de Boer, C. Kockx, A. van der Sloot, M. van den Hout, W. van Ijcken, D. Eick, B. Lenhard, F. Grosveld and E. Soler (2012). "Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development." EMBO J **31**(4): 986-999.
- Suntharalingam, G., M. R. Perry, S. Ward, S. J. Brett, A. Castello-Cortes, M. D. Brunner and N. Panoskaltsis (2006). "Cytokine storm in a phase 1 trial of the anti-CD28 monoclonal antibody TGN1412." N Engl J Med **355**(10): 1018-1028.

- Swigut, T. and J. Wysocka (2007). "H3K27 demethylases, at long last." Cell **131**(1): 29-32.
- Takahashi, K., K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda and S. Yamanaka (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." Cell **131**(5): 861-872.
- Uusi-Oukari, M., J. Heikkila, S. T. Sinkkonen, R. Makela, B. Hauer, G. E. Homanics, W. Sieghart, W. Wisden and E. R. Korpi (2000). "Long-range interactions in neuronal gene expression: evidence from gene targeting in the GABA(A) receptor beta2-alpha6-alpha1-gamma2 subunit gene cluster." Mol Cell Neurosci **16**(1): 34-41.
- Villafuerte, S., E. M. Trucco, M. M. Heitzeg, M. Burmeister and R. A. Zucker (2014). "Genetic variation in GABRA2 moderates peer influence on externalizing behavior in adolescents." Brain Behav **4**(6): 833-840.
- Zeng, H., M. Guo, K. Martins-Taylor, X. Wang, Z. Zhang, J. W. Park, S. Zhan, M. S. Kronenberg, A. Lichtler, H. X. Liu, F. P. Chen, L. Yue, X. J. Li and R. H. Xu (2010). "Specification of region-specific neurons including forebrain glutamatergic neurons from human induced pluripotent stem cells." PLoS One **5**(7): e11853.
- Zhou, V. W., A. Goren and B. E. Bernstein (2011). "Charting histone modifications and the functional organization of mammalian genomes." Nat Rev Genet **12**(1): 7-18.
-