Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Kristina Butze Mercer Date

Investigation of genetic variation and molecular mechanisms associated with risk for posttraumatic stress disorder

By

Kristina Butze Mercer
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Genetics and Molecular Biology

_____	_____
Dr. Kerry Ressler Advisor	Dr. Jennifer Mulle Co-Advisor
_____	_____
Dr. Karen Conneely Committee Member	Dr. M. Katharine Rudd Committee Member
_____	
Dr. Alicia Smith Committee Member	

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Investigation of genetic variation and molecular mechanisms associated with risk for posttraumatic stress disorder

By

Kristina Butze Mercer
M.P.H., Emory University, 1999
B.S., University of New Hampshire, 1993

Advisors: Kerry Ressler, M.D., Ph.D., Jennifer Mulle, M.H.S, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Graduate Division of Biological and Biomedical Science,

Genetics and Molecular Biology

2016

Abstract

Investigation of genetic variation and molecular mechanisms associated with risk for posttraumatic stress disorder

By Kristina Butze Mercer

Post-traumatic stress disorder (PTSD) is a debilitating psychiatric condition that affects 5-12% of the U.S. adult population. Symptoms of the disorder include re-experiencing the trauma through nightmares or flashbacks, avoiding places or things that trigger memory of the trauma, and an increased arousal response that can cause problems with concentration or sleep. Those who suffer from PTSD find day-to-day life difficult and are more likely to have comorbid psychiatric conditions, chronic health problems, abuse drugs and alcohol, and attempt suicide. Treatment options have limited efficacy and most individuals with PTSD remain symptomatic for years to a lifetime, leading to an economic burden estimated in the billions of dollars. To reduce the impact of this disorder on public health, we must understand the underlying mechanisms that lead to PTSD in some individuals but not others when exposed to a similarly traumatic event. Estimates of heritability range from 30-40%, indicating a large fraction of risk is attributable to genetics. This dissertation focuses on potential molecular mechanisms responsible for increased PTSD symptoms and the identification of novel genetic variants that associate with risk for PTSD. I provide evidence that a single nucleotide polymorphism (SNP) that we previously associated with PTSD (rs2267735) is located within an estrogen response element that binds to estradiol-activated estrogen receptor, increasing gene expression of *ADCYAP1R1* and activating a mechanism hypothesized to be involved in the normal stress response. Homozygosity for the C allele at this SNP alters estrogen-dependent regulation, resulting in stress response dysregulation that may lead to a preferential increase in PTSD symptoms among women. To uncover other types of genomic variants that have not yet been examined in risk for PTSD, I also explored the association of PTSD with copy number variants (CNVs), large and relatively rare duplications and deletions found throughout the genome. I present preliminary findings of CNV associations with PTSD, involving *CNTN5* and *IMMP2L*, two genes that have been implicated in neuronal synaptic plasticity and other psychiatric conditions, respectively. This work has expanded our knowledge of genetic factors and the mechanisms associated with risk and resilience for PTSD in the aftermath of severe trauma.

Investigation of genetic variation and molecular mechanisms associated with risk for posttraumatic stress disorder

By

Kristina Butze Mercer
M.P.H., Emory University, 1999
B.S., University of New Hampshire, 1993

Advisors: Kerry Ressler, M.D., Ph.D., Jennifer Mulle, M.H.S, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Graduate Division of Biological and Biomedical Science,

Genetics and Molecular Biology

2016

Acknowledgments

I want to thank each of my committee members for providing invaluable expertise and support. A special thanks to my co-advisor Jen Mulle for taking me in, providing help whenever needed, and teaching me how to be strong. I would like to acknowledge my advisor Kerry Ressler and former mentor Guy Benian for always believing in me and giving me the "push" I needed to pursue my degree. To my classmates, Kelly Shaw, Annie McPherson, Liz Kennedy, and Marko Bajiic- thank you for your amazing friendship. Most of all, I would like to acknowledge my family without whom this dissertation would not have been possible- the patience and understanding of my children, the unrelenting support of my husband, and the every morning reminder of "you can do this!" from my father. This dissertation is dedicated to my husband, Jon Mercer and my father, Gene Larry Butze.

Table of Contents

Table of Tables

Table of Supplementary Tables

Table of Figures

Table of Supplementary Figures

Table of Appendices

Chapter 1: The genetics of posttraumatic stress disorder (PTSD)

Clinical characteristics and epidemiology of PTSD

Posttraumatic stress disorder (PTSD) is a neuropsychiatric disorder that can result from exposure to a traumatic event involving threatened death, serious injury or sexual violation. Examples of traumatic events include a natural disaster, severe automobile accident, military combat, being diagnosed with a life-threatening illness, the sudden unexpected death of a loved one, physical or sexual assault, or robbery/mugging. Exposure can be experiencing the event directly, witnessing the event, having a family member or close friend experience the event, or being repeatedly exposed first-hand to the details of the event. According to the DSM-V, the diagnostic manual for classifying mental disorders, one of these types of events must occur in order to be diagnosed with PTSD [1].

The DSM-V classifies the symptoms of PTSD into four clusters: 1) re-experiencing (having spontaneous flashbacks or recurrent memories of the traumatic event); 2) avoidance (avoiding activities, places or people that may trigger a reminder of the event); 3) negative cognitions and mood (emotional numbing where sufferers become estranged from others, experience a lack of interest in activities, and have a distorted sense of self); and 4) arousal (heightened vigilance leading to irritability and concentration problems). A diagnosis of PTSD is given if symptoms from each of the four clusters persist for more than a month.

PTSD is a disorder than can affect anyone, whether child or adult, civilian or military combat veteran [2]. It is estimated that 2-7% of adolescents ages 13 to 17, living

in the U.S., have had PTSD during their lifetime [3]. A prevalence of 4-12% for lifetime PTSD among U.S adults has been reported [3–6]. Women are over-represented in the upper range of this estimate, with a prevalence of PTSD twice that of men [7]. Living environment can also greatly influence the prevalence of PTSD. For example, among those residing in areas where civilian trauma is common, lifetime PTSD has been reported in up to 46% of the population [8].

PTSD symptoms can last from several months to several years, with symptom duration four times longer in females than males [9]. A recent study (2014) assessing PTSD remission among 81,000 individuals across 42 studies reported an average of 44% of those diagnosed with PTSD became non-cases over a period of 40 months [10]. This means more than half of several thousand people diagnosed with PTSD were still suffering from the disorder after more than 3 years. A study by Kessler et al. found that—regardless of treatment—40% of individuals with PTSD remain symptomatic at 72 months and show negligible improvement, even after several years [5].

The impact of PTSD on public health

Not only can the symptoms of PTSD be debilitating for years to a lifetime: those with the disorder are also at higher risk for comorbid psychiatric disorders, tend to be more likely to abuse drugs and alcohol, exhibit suicidal ideation, and be predisposed to other physical disease. Family relationships also suffer as a result of PTSD and the degree of symptoms are positively correlated with likelihood of unemployment [11–13]. In a 2001 review, Breslau et al. reported that those with PTSD have functional impairment that results in reduced activity (~19%), the inability to work or attend school

(~14%) and less time spent with people (~16%) over a 30 day period when PTSD symptoms were at their worst [9]. The lost days of work and health care costs associated with PTSD result in an increased economic burden to both the individual and the community at large. Among U.S. servicemembers returning from deployment, economic burden was estimated (in 2008) at 4 to 6 billion dollars over two years from medical care costs and lost productivity [14]. Given the prevalence, disease burden and overall cost to patients, their family and society, PTSD is a public health concern.

PTSD is comorbid with several other psychiatric conditions including affective disorders (major depressive disorder, dysthymia, and mania) and anxiety disorders (generalized anxiety disorder, panic disorder, simple phobia, social phobia, and agoraphobia) [5]. The odds of having one of these other disorders is 2 to 10 times greater (depending on the disorder) for those with PTSD compared to those without PTSD. Most commonly associated with PTSD is major depressive disorder, with estimates across several studies revealing that nearly half of those with PTSD also suffer from this neuropsychiatric condition [5,9,15]. Men are 7 times more likely, and women 4 times more likely, to have major depression compared to individuals of the same gender without PTSD [5].

Substance abuse is also common among sufferers of PTSD; alcohol and/or drugs are used to self-medicate and have been reported to alleviate symptoms associated with PTSD [16,17]. Estimates for prevalence of alcohol abuse from 30-40% have been reported among individuals with PTSD [5,15,18]. In a review of the literature between 2007 and 2012, which included data from 42 published manuscripts, Debell et al. summarized comorbid alcohol misuse among those with PTSD in the U.S. and several

other countries [19]. For all studies examined, the relationship between alcohol use and PTSD was statistically significant with odds ratios ranging from 1.4 to 4.6. These findings, in agreement with several other studies, provide convincing data that PTSD associates with increased alcohol use across different populations, sample characteristics and study designs. The same relationship exists with increased drug use among those with PTSD. Approximately 30% of PTSD sufferers also abuse drugs, with 3-4 times increased likelihood of abuse compared to individuals without PTSD [5]. Estimates in excess of 50% have been reported among male veterans [15]. Variable incidences of substance abuse have been observed based on the population being studied. Gender, type of trauma, and severity of the trauma each contribute to the prevalence of alcohol use and drug abuse among those with PTSD [17].

Suicidality is also common among individuals with PTSD. One study performed by Marshall et al. examined suicidal ideation among a cohort of approximately 1,500 individuals across the U.S. [20]. Among those with one symptom of PTSD, thoughts of suicide were present in 13% of individuals. This number increased to 33% among those experiencing at least 4 symptoms of PTSD. In the same study, current suicidal ideation was measured among individuals with a diagnosis of PTSD while controlling for age, sex and race. This analysis revealed a statistically significant, 2.09-fold increase in suicidal ideation for individuals with PTSD compared to those with other psychiatric disorders. Breslau et al. also report the prevalence of suicidal thoughts and suicide attempts among a smaller cohort of individuals with PTSD (N=93) but observed similarly high rates of suicidality: 28% report ever wanting to die, 46% report having thought about suicide, and 17% had attempted suicide [9]. In another study, Wilcox et al. found that, among young

adults (mean age = 21), 10% of those with PTSD had attempted suicide compared to only 2.5% of those who were exposed to trauma but did not develop PTSD [21]. Using data collected from the National Comorbidity Survey, Kessler et al. determined that those with PTSD were 6 times more likely to attempt suicide and 5 times more likely to experience suicidal ideation [22].

Individuals with PTSD are at higher risk for several physical ailments and chronic disease, which contribute further to the functional impairment caused by the primary symptoms alone. According to one study, those with PTSD are 2 times more likely to have arthritis, 1.7 times more likely to have asthma, and 2 times more likely to have an ulcer (all statistically significant) [23]. Compared to the population rate (CDC statistic), statistically significant rates of diabetes (type I or II), stroke, myocardial infarction, cancer (of any type) and cirrhosis of the liver have been observed among veterans with PTSD [24] . Other studies have reported higher rates of cardiovascular, musculoskeletal, gastrointestinal, neurological and pseudoneurological (amnesia, paralysis, fainting, double vision) symptoms [25,26]. PTSD is comorbid with somatization disorder, which is defined as having persistent symptoms of at least 13 types within six symptom categories: 1) gastrointestinal, 2) pain, 3) cardiopulmonary, 4) pseudoneurological, 5) sexual (e.g. pain during intercourse, impotence), and 6) female reproductive (e.g. painful menstruation, irregular menstrual periods, excessive menstrual bleeding) [9,27]. Andreski et al. found a strong association between PTSD and symptom groups defined by this disorder [27]. Total symptoms experienced across all somatic symptom categories were approximately 2 times greater for both men and women with PTSD compared to a non-PTSD comparison group.

Current treatment options for PTSD

Though there are treatment options for PTSD, the disorder is difficult to treat and no cure exists. The limited efficacy of current treatment options is compounded by patient incompliance [28]. Individuals with PTSD become frustrated with treatments that offer little relief from their symptoms and, as previously mentioned, may turn to alcohol and drugs to help them cope. Others with PTSD may not even seek help [29], with many suffering in silence due to the stigma associated with mental illness. This is particularly true among trauma exposed military personnel, who would benefit from treatment of any type [30]. Public awareness of PTSD continues to increase and will hopefully result in greater treatment-seeking behaviors and treatment adherence.

New treatment options for PTSD continue to be investigated. The more we learn about the disorder the better equipped we are to develop novel and effective treatment options. Cognitive-behavioral and exposure therapies as well as eye movement desensitization and reprocessing are commonly practiced psychotherapies used to treat PTSD symptoms. A meta-analysis of studies investigating the efficacy of these methods reveals a 56% recovery rate of PTSD [31]. Pharmacotherapy is another line of treatment for PTSD with serotonin re-uptake inhibitors (SSRIs) used most prominently, and the only FDA approved medications for an indication of PTSD. In a meta-analysis of 21 studies, the efficacy of SSRIs was compared to placebo in reducing symptoms of PTSD. Three drugs in particular, fluoxetine, paroxetine and venlafaxine, are useful in the treatment of PTSD but have limited efficacy [32]. Several other treatments for PTSD

exist [33,34], including virtual reality [35], group therapy [36], meditation [37], and acupuncture [38]. According to one study, patients reported preference for cognitive therapy or cognitive therapy with exposure therapy. Virtual reality, computer-based therapy and eye movement desensitization were among the less-favored types of therapy, despite favorable outcomes associated with EMDR, in particular [31,39].

Early intervention after trauma has been explored as an option for the prevention of PTSD [40]. Psychological debriefing, which involves helping individuals cope with their trauma immediately after the event, has been used with no apparent success in PTSD prevention [41] . Conversely, pharmacological treatments that include the administration of high-dose glucocorticoid (hydrocorticosterone or dexamethasone), morphine, and ketamine have been used within hours of the trauma with promising success. Glucocorticoid activation appears to only work in the short-term but it may have a positive impact on the hyper-arousal and fear responses associated with PTSD [42,43]. Only preliminary data exist for the efficacy of using morphine or ketamine directly after trauma but, so far, both have resulted in lower rates of PTSD development [44,45].

In some cases PTSD will resolve; for the majority it remains a chronic psychiatric condition. Although there are several treatment options for PTSD to date, none are 100% effective. It is likely that prevention of disease development may be the best “cure” for PTSD. Preventing trauma, which is often an inevitable consequence of life, is not an option. Therefore, identifying the underlying differences between those who develop PTSD and those who do not could aid in the early identification of those at risk. A better understanding of the biology of PTSD could then help determine the best methods for treating at risk individuals either before or immediately after trauma exposure.

Known risk factors for PTSD

PTSD is a complex trait that is the result of both environmental and genetic risk factors. While experiencing a traumatic event is necessary for the development of PTSD, other environmental exposures or conditions can increase the likelihood of developing the disorder in the aftermath of such trauma. For example, adults who were neglected or physically or sexually abused as children are at increased risk for PTSD [46–49]. This early life stress has been implicated in neuroendocrine alterations during development [50]. Cumulative trauma is also a risk factor for PTSD. Those who experience multiple traumatic events, particularly assaultive violence, are more likely to develop PTSD than those experiencing one trauma [51]. The type of trauma experienced can also greatly affect PTSD outcome. Violent assault and the sudden unexpected death of a loved one account for 39.5 and 31.1 percent of PTSD cases, respectively [7]

Twin studies have provided heritability estimates for PTSD, which reveal a role for genetics in PTSD susceptibility. Using twin pairs that were in active duty during the Vietnam War, True et al. assessed the correlation of PTSD symptoms between siblings while controlling for degree of combat exposure. He and his colleagues reported that genetic factors account for approximately 30% of the variance observed in PTSD symptoms [52]. Stein et al. observed a similar heritability estimate of 38% in their study of PTSD among trauma-exposed, monozygotic and dizygotic twins from the general population of British Columbia, Canada [53]. The advent of genome-wide genotyping

platforms have facilitated better estimates of heritability for complex traits [54]. Using the DSM-V criteria for PTSD and genotype data among pedigrees, adjusted heritability estimates of 60% have more recently been reported [55].

History and challenges in determining genetic causes of PTSD

With the knowledge that PTSD is heritable and some variability in PTSD is due to genetic variation, several research studies have focused on identifying genes involved in risk for this disorder. This research began with the investigation of what are known as candidate genes. These are genes presumed to play a role in pathways–such as fear learning and stress response—that could lead to the development of PTSD when dysregulated. Candidate genes for PTSD risk have included those involved in the regulation of stress, particularly via the hypothalamic-pituitary-adrenal (HPA) axis, and genes whose function is related to neurodevelopment or neurotransmitter activity. Genes that associate with other neuropsychiatric conditions have also been explored as candidates.

Until recently, single nucleotide polymorphisms (SNPs) have been the main type of genetic variation investigated, either in candidate gene or genome-wide association studies (GWAS). SNPs have been a primary target in the search for disease heritability because they are easy to genotype and represent the most common type of sequence variation in the genome [56]. The first challenge in conducting these genetic studies was identifying the polymorphic nucleotides. The International HapMap project, which began in 2002, identified 3 million SNPs through the genome sequencing of individuals of African (Yoruba people of Ibadan, Nigeria), Asian (Japanese from Tokyo and Chinese

from Beijing) and European ancestry (US residents with northern and western European ancestry) [57]. Later phases of the HapMap and 1000 genomes projects have now collectively provided SNP genotypes for 26 populations [58]. All the SNPs found through these efforts have been deposited in a database known as dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) and have been used widely by the research community to investigate disease associations [59].

While most polymorphisms are innocuous, some may result in the disruption of gene expression, gene regulation and/or protein function. Without *a priori* knowledge of which nucleotide differences may contribute to gene dysregulation, most research scientists use a wide-net approach to SNP selection, choosing several SNPs within a gene to be tested for association in genetic analyses. This leads to a second challenge, which is choosing the most informative SNPs to genotype among the several hundred that may exist within the candidate gene. Using a method known as tag SNP selection, population measures of linkage disequilibrium between SNPs can be utilized to impute SNP genotypes for un-typed loci with a pre-defined level of confidence [60,61]. This method provides an affordable and practical approach to test genetic associations. However, intergenic regions of unknown significance and genes not yet identified as having a role in certain diseases are inevitably missed.

In the past decade, the release of commercially available and increasingly affordable genotyping arrays have made it possible to genotype SNPs across the entire genome, eliminating the need to carefully select which SNPs to genotype and allowing for a hypothesis-free approach to identify novel genes involved in disease mechanisms. The first GWAS, which uncovered a SNP associated with age-related macular

degeneration, was published in 2005 [62]. Since then, 2,370 GWAS studies have been published and 18,542 SNPs associated with disease have been discovered [63]. While GWAS have been successful in the identification of meaningful SNP associations for diseases such as coronary artery disease, Crohn's disease, rheumatoid arthritis, and type 1 diabetes, fewer confident findings have been reported for psychiatric disorders, PTSD in particular [64,65].

There are several challenges associated with identifying genetic risk for PTSD by GWAS [66]. To reduce false positive associations, population stratification must be considered in regression analyses to appropriately account for allele frequency differences that are due to ancestry rather than disease outcome [67,68]. The impact of environmental factors on PTSD susceptibility necessitates the inclusion of gene by environment (GxE) interactions in genetic analyses of risk [69–71]. Large sample size is also a significant contributor to the detection of genome-wide significance, as evidenced by the success of large GWAS performed by consortia for complex diseases such as schizophrenia [72,73]. Studies of genetic risk that utilize small sample sizes do not have the statistical power to detect disease associations with small effect sizes even if they exist, especially when considering models of disease associations such as GxE interactions. With meta-analysis assessment using genome-wide genotype data for several hundred thousand samples from numerous studies, future genetic inquiry into the genetics of PTSD will be better powered to detect true associations. Recently, a PTSD consortium was assembled to combine data from nearly 200,000 subjects in an attempt to increase sample size and aid in the discovery of meaningful genetic associations for PTSD [74].

Summary of current PTSD-associated polymorphisms

Both candidate gene analyses and GWAS have resulted in the discovery of statistically significant associations between PTSD and SNP variants within 36 genes or intergenic regions. Table 1-1 provides a summary of the PTSD associations identified to date. For most of the associations listed, either a GxE analysis was performed or environmental or population-specific factors were accounted for in the analysis. These variables are included in the Table as "interaction variables" or "environmental context." Note that associations for the majority of SNPs have not been replicated. Among the few that have replicated, studies have also been published that refute the association. These variants are denoted by "CR". While some associations that fail to replicate may truly be false positives, there are several factors, as discussed above that can complicate reproducibility.

To date, association with PTSD has been replicated for only a few genes, generating hypotheses of a relationship between dysregulation of gene function and development of PTSD. Experimental examination of the effects of such variants could offer key insights into the biology of PTSD. Loci within two genes, *COMT* and *FKBP5*, have been replicated in more than one study with no conflicting reports. Individuals with the "AA" genotype at rs4860 in *COMT* at are at increased risk for PTSD. The "A" variant (a.k.a. Val158Met) is a non-synonymous polymorphism that results in the substitution of a valine for a methionine at amino acid 158 of Catechol-O-Methyltransferase, the COMT

protein. COMT is an enzyme involved in the catalysis of neurotransmitters. Genetic variants within the gene are thought to alter enzyme activity, resulting in abnormalities of catecholamine neurotransmission and increased risk for neuropsychiatric conditions [75]. The *FKBP5* gene encodes a protein involved in glucocorticoid receptor sensitivity and regulation of the stress hormone pathway via the HPA axis [76]. The association between PTSD symptoms and SNPs within *FKPB5* were identified in a GxE association test accounting for degree of child abuse [77]. The associated SNPs were found to result in prolonged activation of the stress hormone axis and a dysregulated response to stress [76]. Thus, it has been hypothesized that carriers of the *FKBP5* risk genotypes who have also been abused as children acquire altered sensitization of the stress-response pathway during development, which puts them at increased risk for PTSD later in life when exposed to further trauma. The phenotypes resulting from genetic variation within *FKPB5* have been well studied and associate with outcomes such as cognitive function, hippocampal structure, gray matter volumes and white matter integrity [78–81]. Our understanding of FKBP5 continues to improve; this growing knowledge may provide options for prevention or treatment of PTSD and other stress related disorders in the future [82].

For genetic variants in *SLC6A3* and *ADCYAP1R1*, several studies supported an association with PTSD with one study to the contrary. The single study that failed to replicate the *SLC6A3* variable number tandem repeat (VNTR) (9 repeats) association with PTSD had fewer than 200 study participants [83]. A SNP in *ADCYAP1R1*, rs2267735, was replicated in three studies, but one study reported no association. Although the conflicting study by Chang et al. was sufficiently powered to detect an

association with sample sizes of 2500 and 6000 for two independent studies, the authors did not utilize a GxE approach to account for trauma load in their analyses [84]. Both of these genes remain promising prospects in PTSD risk but further studies are necessary to determine the true significance of these associations. In this dissertation (Chapter 2) I test a model wherein one PTSD-associated SNP, rs2267735, results in an altered stress response through estradiol regulation of *ADCYAP1R1* (PAC1).

While numerous studies have examined the relationship between *SLC6A4* (sodium-dependent serotonin transporter) and PTSD and reported positive associations, one large meta-analysis consisting of 1,874 cases and 7,785 controls from 13 studies found no relationship with PTSD and either the biallelic (long or short) or triallelic (extra-long, long, or short) length polymorphisms of 5-HTTLPR (serotonin-transporter-linked polymorphic region) [85]. A second meta-analysis of 12 studies also failed to identify a main effect between PTSD and the 5-HTTLPR alleles but did find a statistically significant association with PTSD and the homozygous short genotype (S/S) in a GxE with high-trauma exposure [86].

The first GWAS for PTSD, published in 2013, reported an association for the variant rs8042149 in the gene *RORA* [87]. A second SNP in the gene, rs893290, has since been reported as increasing risk for PTSD [88]. *RORA*, encodes RAR-related orphan receptor A, a nuclear receptor protein involved in maturation of photoreceptors in the retina and circadian rhythm [89]. It remains unclear how *RORA* may play a role in the development of PTSD.

The intergenic variant, rs406001, located on chromosome 7 at band p12, was originally identified in a GWAS by Xie et al. and replicated in GxE analysis with

childhood trauma [90,91]. This variant has no known function but is close to the gene *COBL,* which may have a role in neuronal development and function. Five other PTSD-associated variants have been identified, and more are likely to follow through the efforts of the Psychiatric Genomics Consortium (PGC) Posttraumatic Stress Disorder workgroup.

Using copy number variation to gain novel insight into the genetic architecture of PTSD

Given the sufficient depth of SNP coverage across the genome and the thousands of samples assessed, it is likely that some of the missing genetic variability for PTSD may be found through the characterization of rare copy number variants. Compared to SNPs, these variants are more likely to be penetrant, with greater effect size [92]. CNVs are genomic gains and losses that range in size from 1 kilobase to several megabases [93]. Given the relatively large size of some CNVs and the number of genes possibly affected, the potential for negative health impact can be significant for this variant type [94]. One way CNVs can result in disease is through heterozygous duplication or deletion of genes that are sensitive to dosage effects and critical for normal cellular processes. To date, CNVs have been associated with multiple diseases including type 1 and type 2 diabetes, Crohn's disease, rheumatoid arthritis, asthma, and Parkinson's [64,95]. Over the past five years, a large body of research has also begun to define the role of CNVs in genetic susceptibility to neuropsychiatric conditions such as autism spectrum disorders and schizophrenia [96–107]. Recent research has also shown that schizophrenia-related CNVs might also increase risk for major depressive and panic disorders [108,109]. These

findings underline the importance of examining the prevalence of CNVs in other neuropsychiatric cohorts to more fully characterize genetic contributors to disease [110]. Based on the success of CNV discovery and association with other neuropsychiatric disorders, research teams have just begun to test for association between CNVs and PTSD, but no findings have yet been published.

CNVs can be detected using existing GWAS data from SNP arrays [111–116]. Signal intensities at each of the probes can be used to infer copy number of an allele in addition to genotype. Higher or lower signal intensities across multiple probes are indicative of a deviation from a copy number leading to 3 (duplication) or 1 (deletion) copies, respectively. The Log R ratio (LRR) is the log base 2 of the ratio of observed intensity over expected intensity. When the observed and expected intensities are equivalent, the LRR value for a normal diploid copy number (2) is zero. LRR values of less than or greater than zero are indicative of a copy number loss or gain, respectively (Figure 1-1 A and Figure 1-2 A). The B allele frequency (BAF) is a measure of the frequency of one of two alleles, arbitrarily assigned as A and B. For a locus with normal copy number of 2 alleles the genotypes can be AA, AB or BB giving BAF values of 0.0, 0.5, and 1.0, respectively. Deviation from these expected frequencies (e.g. 0.33 from genotype AAB and 0.66 from genotype ABB) across multiple loci provides evidence of a change in copy number (Figure 1-1 B and Figure 1-2 B). Because it is impossible to interrogate potential copy number variants for thousands of samples across all chromosomes by visual inspection, CNV calling algorithms are employed to identify CNVs. The current gold standard for CNV calling is PennCNV [116,117]. However several other programs can also be used such as QuantiSNP [118], R-Gada [115], and

iPattern (unpublished). Once CNVs have been identified, PLINK and R can be utilized to test for associations [119].

Using SNP genotyping data from the GTP and Army STARRS [120] cohorts, in this dissertation I provide preliminary findings of rare CNVs that associates with PTSD. Also, during the investigation of CNVs among the GTP cohort we identified an increased burden of undiagnosed genomic disorder CNVs as an incidental finding (Chapter 4).

Conclusion

Over the past decade, we have learned a lot from the successes and failures of PTSD genetic association studies. Armed with the knowledge that we must address complex factors such as gene by environment interactions, population stratification, and genomic variation beyond just SNPs, our ability to identify genetic risk loci has and will continue to improve. Collaborative efforts to increase sample size for genetic analyses will undoubtedly be a significant contributor to future success. As the list of genomic variation associated with PTSD risk grows, elucidation of the mechanisms responsible will enable us to better treat—and hopefully even prevent—this debilitating disorder and mitigate its huge impact on public health.

Table 1-1. Statistically significant associations between PTSD and genetic variants

Gene/ Genomic Region	SNP/Variant type [a]	Interaction variable or environmental context	Reference
Candidate Gene -findings			
ADCYAP1R1	rs2267735*,CR	Childhood maltreatment, trauma load	[121–124]
ADRB2	rs2400707	Childhood trauma	[125]
ALOX12	rs1042357, rs10852889	Thickness of Pre-frontal cortex	[126]
ANK3	rs9804190, rs28932171, rs11599164, rs17208576		[127]
APOE2	rs7412	Combat	[128]
BDNF	rs6265*,CR	Earthquake	[129–131]
CCK	rs1799923	Combat	[132]
CHRNA5	rs12898919 , rs16969968	Age, trauma exposure, smoking	[133,134]
CNR1	rs1049353 + rs806368		[135]
COMT	rs4860*, rs4633	Race, trauma load	[55,134,136–139]
CRHR1	rs12944712, rs12938031, rs4792887	Hurricane , pediatric injury	[140,141]
CRP	rs1130864	Civilian trauma	[142]
SLC6A3	VNTR*,CR , rs27072	Hurricane, preschoolers, combat	[143,143–147]
DRD2	rs6277, rs2075652, rs7131056	Combat	[148,149]
DRD3	rs2134655, rs201252087, rs4646996, rs9868039	Combat	[150]
DTNBP1	rs9370822		[151]
FKBP5	rs9296158*, rs3800373*, rs1360780*, rs9470080*	Race, Child abuse	[77,152,153]
GABRA2	rs279836, rs279826, rs279871	Childhood trauma	[154]
GLUT1	rs710218	Civilian trauma	[155]

Gene/ Genomic Region	SNP/Variant type[a]	Interaction variable or environmental context	Reference
NOS1AP	rs386231	Combat	[156]
OXTR	rs53576	Social and economic + stress	[157]
PRKCA	rs4790904	Trauma-exposed veterans	[137,158]
RGS2	rs4606	High hurricane exposure and low social support	[159]
WWC1	rs10038727, rs4576167	Survivors of Rwandan genocide and conflict with the Lord's Resistance Army (Uganda)	[160]
SLC18A2	Haplotype including rs363276		[161]
SLC1A1	rs10739062	Combat-exposed veterans	[162]
SLC6A4	5-HTTLPR alleles*,CR, rs25531	Childhood adversity	[85,86,163,164]
SRD5A2	rs523349	Gender	[165]
TPH2	rs11178997	Earthquake	[166]
GWAS -findings			
RORA	rs893290, rs8042149	Childhood physical abuse	[87,88]
PRTFDC1	rs6482463	Child and adult trauma	[167]
Intergenic -7p12	rs406001*	Race	[90,91]
DICER1	rs10144436	Comorbid depression	[168]
Intergenic - 2q32.1	rs10170218		[169]
Intergenic - 4p15	rs717947	Gender	[170]
TLL1	rs6812849		[171]

[a] Variants starting with "rs" represent SNPs based on dbSNP nomenclature , * PTSD association was replicated in more than one study, CR Conflicting results published for the same variant

Figure 1-1. Graphing of LRR and BAF for a deletion. The shaded region represents a deletion of chromosome 15. A.) A deletion is represented by Log R ratio values of less than zero. B) A deletion is represented by B allele frequencies of 0 or 1 and are missing frequencies of 0.5 representative of a heterozygous genotype.

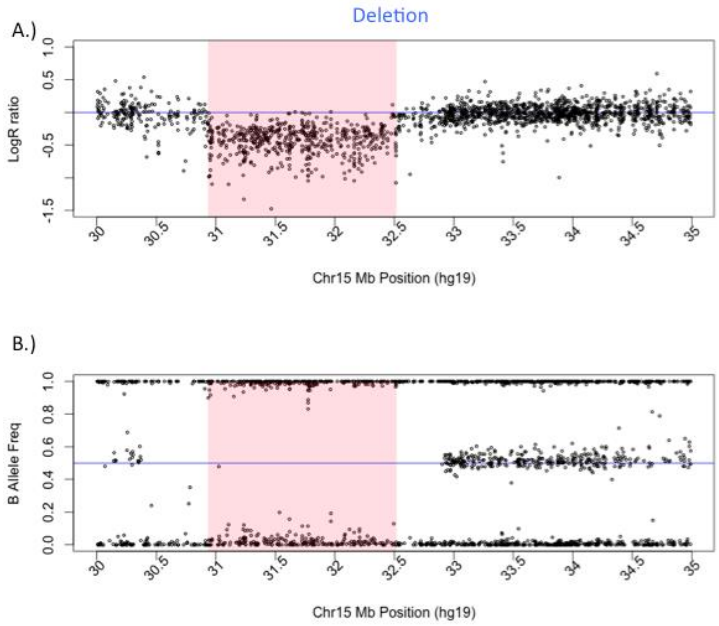
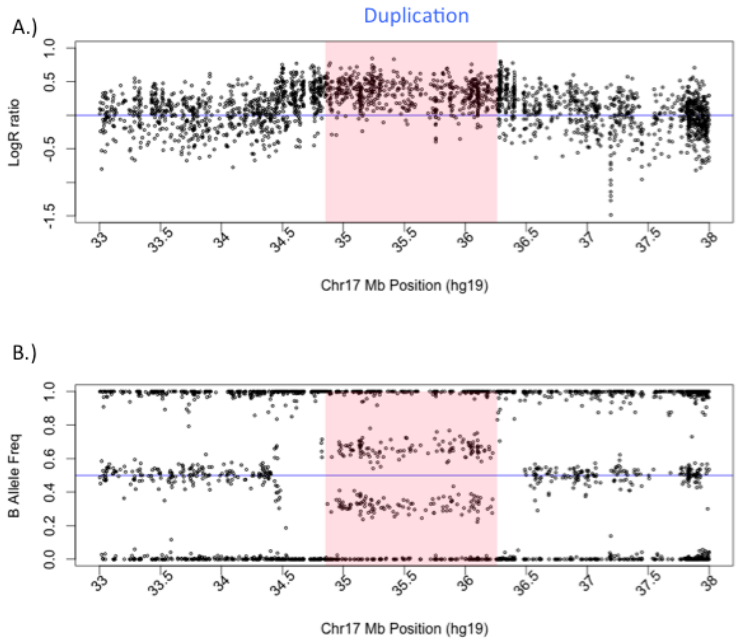


Figure 1-2. Graphing of LRR and BAF for a duplication. The shaded region represents a duplication of chromosome 17. A.) A duplication is represented by Log R ratio values of greater than zero. B) A duplication is represented by B allele frequencies of 0, 0. 33, 0.66, or 1.



References

1. American Psychiatric Association. (2013) Diagnostic and statistical manual of mental disorders, (5th ed.). Washington, DC.

2. Yule W: Posttraumatic stress disorder in the general population and in children. J Clin Psychiatry 2001;62 Suppl 17:23–28.

3. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen H-U: Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012 Sep;21:169–184.

4. Resnick HS, Kilpatrick DG, Dansky BS, Saunders BE, Best CL: Prevalence of civilian trauma and posttraumatic stress disorder in a representative national sample of women. J Consult Clin Psychol 1993 Dec;61:984–991.

5. Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB: Posttraumatic stress disorder in the National Comorbidity Survey. Arch Gen Psychiatry 1995 Dec;52:1048–1060.

6. Breslau N, Davis GC, Andreski P, Peterson E: Traumatic events and posttraumatic stress disorder in an urban population of young adults. Arch Gen Psychiatry 1991 Mar;48:216–222.

7. Breslau N, Kessler RC, Chilcoat HD, Schultz LR, Davis GC, Andreski P: Trauma and posttraumatic stress disorder in the community: The 1996 detroit area survey of trauma. Arch Gen Psychiatry 1998 Jul 1;55:626–632.

8. Gillespie CF, Bradley B, Mercer K, Smith AK, Conneely K, Gapen M, et al.: Trauma exposure and stress-related disorders in inner city primary care patients. Gen Hosp Psychiatry 2009 Dec;31:505–514.

9. Breslau N: Outcomes of Posttraumatic Stress Disorder. J Clin Psychiatry 2001 Jan 3;62:1,478–59.

10. Morina N, Wicherts JM, Lobbrecht J, Priebe S: Remission from post-traumatic stress disorder in adults: A systematic review and meta-analysis of long term outcome studies. Clin Psychol Rev 2014 Apr;34:249–255.

11. Smith MW, Schnurr PP, Rosenheck RA: Employment outcomes and PTSD symptom severity. Ment Health Serv Res 2005 Jun;7:89–101.

12. Sayers SL, Farrow VA, Ross J, Oslin DW: Family problems among recently returned military veterans referred for a mental health evaluation. J Clin Psychiatry 2009 Feb;70:163–170.

13. Riggs DS, Byrne CA, Weathers FW, Litz BT: The quality of the intimate relationships of male Vietnam veterans: problems associated with posttraumatic stress disorder. J Trauma Stress 1998 Jan;11:87–101.

14. Tanielian T, Jaycox LH, Adamson DM, Burnam MA, Burns RM, Caldarone LB, et al.: Invisible Wounds of War [Internet] 2008 [cited 2016 Jan 12];Available from: http://www.rand.org/pubs/monographs/MG720.html

15. Zatzick DF, Marmar CR, Weiss DS, Browner WS, Metzler TJ, Golding JM, et al.: Posttraumatic stress disorder and functioning and quality of life outcomes in a nationally representative sample of male Vietnam veterans. Am J Psychiatry 1997 Dec;154:1690–1695.

16. Kosten TR, Krystal J: Biological mechanisms in posttraumatic stress disorder. Relevance for substance abuse. Recent Dev Alcohol Off Publ Am Med Soc Alcohol Res Soc Alcohol Natl Counc Alcohol 1988;6:49–68.

17. Bremner JD, Southwick SM, Darnell A, Charney DS: Chronic PTSD in Vietnam combat veterans: Course of illness and substance abuse. Am J Psychiatry 1996 Mar;153:369–75.

18. Mills KL, Teesson M, Ross J, Peters L: Trauma, PTSD, and substance use disorders: findings from the Australian National Survey of Mental Health and Well-Being. Am J Psychiatry 2006 Apr;163:652–658.

19. Debell F, Fear NT, Head M, Batt-Rawden S, Greenberg N, Wessely S, et al.: A systematic review of the comorbidity between PTSD and alcohol misuse. Soc Psychiatry Psychiatr Epidemiol 2014 Sep;49:1401–1425.

20. Marshall RD, Olfson M, Hellman F, Blanco C, Guardino M, Struening EL: Comorbidity, impairment, and suicidality in subthreshold PTSD. Am J Psychiatry 2001 Sep;158:1467–1473.

21. Wilcox HC, Storr CL, Breslau N: Posttraumatic stress disorder and suicide attempts in a community sample of urban american young adults. Arch Gen Psychiatry 2009 Mar;66:305–311.

22. Kessler RC, Borges G, Walters EE: PRevalence of and risk factors for lifetime suicide attempts in the national comorbidity survey. Arch Gen Psychiatry 1999 Jul 1;56:617–626.

23. Greenberg PE, Sisitsky T, Kessler RC, Finkelstein SN, Berndt ER, Davidson JR, et al.: The economic burden of anxiety disorders in the 1990s. J Clin Psychiatry 1999 Jul;60:427–435.

24. Buckley TC, Mozley SL, Bedard MA, Dewulf A-C, Greif J: Preventive health behaviors, health-risk behaviors, physical morbidity, and health-related role

functioning impairment in veterans with post-traumatic stress disorder. Mil Med 2004 Jul;169:536–540.

25. Escobar JI, Canino G, Rubio-Stipec M, Bravo M: Somatic symptoms after a natural disaster: a prospective study. Am J Psychiatry 1992 Jul;149:965–967.

26. McFarlane AC, Atchison M, Rafalowicz E, Papay P: Physical symptoms in post-traumatic stress disorder. J Psychosom Res 1994 Oct;38:715–726.

27. Andreski P, Chilcoat H, Breslau N: Post-traumatic stress disorder and somatization symptoms: a prospective study. Psychiatry Res 1998 Jun 15;79:131–138.

28. Schottenbauer MA, Glass CR, Arnkoff DB, Tendick V, Gray SH: Nonresponse and dropout rates in outcome studies on PTSD: review and methodological considerations. Psychiatry 2008;71:134–168.

29. Sayer NA, Friedemann-Sanchez G, Spoont M, Murdoch M, Parker LE, Chiros C, et al.: A Qualitative Study of Determinants of PTSD Treatment Initiation in Veterans. Psychiatry 2009 Sep 1;72:238–255.

30. Pietrzak RH, Johnson DC, Goldstein MB, Malley JC, Southwick SM: Perceived stigma and barriers to mental health care utilization among OEF-OIF veterans. Psychiatr Serv Wash DC 2009 Aug;60:1118–1122.

31. Bradley R, Greene J, Russ E, Dutra L, Westen D: A multidimensional meta-analysis of psychotherapy for PTSD. Am J Psychiatry 2005 Feb;162:214–227.

32. Hoskins M, Pearce J, Bethell A, Dankova L, Barbui C, Tol WA, et al.: Pharmacotherapy for post-traumatic stress disorder: systematic review and meta-analysis. Br J Psychiatry 2015 Feb 1;206:93–100.

33. Foa EB, Tolin DF: Comparison of the PTSD Symptom Scale-Interview Version and the Clinician-Administered PTSD scale. J Trauma Stress 2000 Apr;13:181–191.

34. Cukor J, Spitalnick J, Difede J, Rizzo A, Rothbaum BO: Emerging treatments for PTSD. Clin Psychol Rev 2009 Dec;29:715–726.

35. Botella C, Serrano B, Baños RM, Garcia-Palacios A: Virtual reality exposure-based therapy for the treatment of post-traumatic stress disorder: a review of its efficacy, the adequacy of the treatment protocol, and its acceptability. Neuropsychiatr Dis Treat 2015;11:2533–2545.

36. Lubin H, Loris M, Burt J, Johnson DR: Efficacy of Psychoeducational Group Therapy in Reducing Symptoms of Posttraumatic Stress Disorder Among Multiply Traumatized Women. Am J Psychiatry 1998 Sep 1;155:1172–1177.

37. Heffner KL, Crean HF, Kemp JE: Meditation Programs for Veterans With Posttraumatic Stress Disorder: Aggregate Findings From a Multi-Site Evaluation. Psychol Trauma Theory Res Pract Policy 2016 Jan 11; DOI: 10.1037/tra0000106

38. Hollifield M, Sinclair-Lian N, Warner TD, Hammerschlag R: Acupuncture for Posttraumatic Stress Disorder: A Randomized Controlled Pilot Trial. J Nerv Ment Dis 2007 Jun;195:504–513.

39. Tarrier N, Liversidge T, Gregg L: The acceptability and preference for the psychological treatment of PTSD. Behav Res Ther 2006 Nov;44:1643–1656.

40. Kearns MC, Ressler KJ, Zatzick D, Rothbaum BO: Early Interventions for Ptsd: A Review. Depress Anxiety 2012 Oct 1;29:833–842.

41. Rose S, Bisson J, Churchill R, Wessely S: Psychological debriefing for preventing post traumatic stress disorder (PTSD). Cochrane Database Syst Rev 2001;CD000560.

42. Jovanovic T, Phifer JE, Sicking K, Weiss T, Norrholm SD, Bradley B, et al.: Cortisol suppression by dexamethasone reduces exaggerated fear responses in posttraumatic stress disorder. Psychoneuroendocrinology 2011 Nov;36:1540–1552.

43. Surís A, North C, Adinoff B, Powell CM, Greene R: Effects of exogenous glucocorticoid on combat-related PTSD symptoms. Ann Clin Psychiatry Off J Am Acad Clin Psychiatr 2010 Nov;22:274–279.

44. Bryant RA, Creamer M, O'Donnell M, Silove D, McFarlane AC: A study of the protective function of acute morphine administration on subsequent posttraumatic stress disorder. Biol Psychiatry 2009 Mar 1;65:438–440.

45. McGhee LL, Maani CV, Garza TH, Gaylord KM, Black IH: The correlation between ketamine and posttraumatic stress disorder in burned service members. J Trauma 2008 Feb;64:S195–198; Discussion S197–198.

46. Rodriguez N, Ryan SW, Rowan AB, Foy DW: Posttraumatic stress disorder in a clinical sample of adult survivors of childhood sexual abuse. Child Abuse Negl 1996 Oct;20:943–952.

47. Twaite JA, Rodriguez-Srednicki O: Childhood sexual and physical abuse and adult vulnerability to PTSD: the mediating effects of attachment and dissociation. J Child Sex Abuse 2004;13:17–38.

48. Widom CS: Posttraumatic stress disorder in abused and neglected children grown up. Am J Psychiatry 1999 Aug;156:1223–1229.

49. Rowan AB, Foy DW, Rodriguez N, Ryan S: Posttraumatic stress disorder in a clinical sample of adults sexually abused as children. Child Abuse Negl 1994 Jan;18:51–61.

50. Yehuda R, Flory JD, Pratchett LC, Buxbaum J, Ising M, Holsboer F: Putative biological mechanisms for the association between early life adversity and the subsequent development of PTSD. Psychopharmacology (Berl) 2010 Aug 13;212:405–417.

51. Breslau N, Chilcoat HD, Kessler RC, Davis GC: Previous exposure to trauma and PTSD effects of subsequent trauma: results from the Detroit Area Survey of Trauma. Am J Psychiatry 1999 Jun;156:902–907.

52. True WR, Rice J, Eisen SA, Heath AC, Goldberg J, Lyons MJ, et al.: A twin study of genetic and environmental contributions to liability for posttraumatic stress symptoms. Arch Gen Psychiatry 1993;50:257–264.

53. Stein MB, Jang KL, Taylor S, Vernon PA, Livesley WJ: Genetic and environmental influences on trauma exposure and posttraumatic stress disorder symptoms: A twin study. Am J Psychiatry 2002;159:1675–1681.

54. Zaitlen N, Kraft P: Heritability in the genome-wide association era. Hum Genet 2012 Oct;131:1655–1664.

55. Goenjian AK, Noble EP, Steinberg AM, Walling DP, Stepanyan ST, Dandekar S, et al.: Association of COMT and TPH-2 genes with DSM-5 based PTSD symptoms. J Affect Disord 2014 Oct 22;172C:472–478.

56. Collins FS, Brooks LD, Chakravarti A: A DNA Polymorphism Discovery Resource for Research on Human Genetic Variation. Genome Res 1998 Dec 1;8:1229–1231.

57. International HapMap Consortium: A haplotype map of the human genome. Nature 2005 Oct 27;437:1299–1320.

58. The 1000 Genomes Project Consortium: A global reference for human genetic variation. Nature 2015 Oct 1;526:68–74.

59. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al.: dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001 Jan 1;29:308–311.

60. Stram DO: Tag SNP selection for association studies. Genet Epidemiol 2004 Dec;27:365–374.

61. Spencer CCA, Su Z, Donnelly P, Marchini J: Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip. PLoS Genet 2009 May 15;5:e1000477.

62. Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, et al.: Complement factor H polymorphism in age-related macular degeneration. Science 2005 Apr 15;308:385–389.

63. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al.: The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 2014 Jan 1;42:D1001–D1006.

64. Wellcome Trust Case Control Consortium, Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, et al.: Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 2010 Apr 1;464:713–720.

65. Craddock N, O'Donovan MC, Owen MJ: Genome-wide association studies in psychiatry: lessons from early studies of non-psychiatric and psychiatric phenotypes. Mol Psychiatry 2008 May 27;13:649–653.

66. Cornelis MC, Nugent NR, Amstadter AB, Koenen KC: Genetics of post-traumatic stress disorder: review and recommendations for genome-wide association studies. Curr Psychiatry Rep 2010 Aug;12:313–326.

67. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006 Aug;38:904–909.

68. Cardon LR, Bell JI: Association study designs for complex diseases. Nat Rev Genet 2001 Feb;2:91–99.

69. Almli LM, Duncan R, Feng H, Ghosh D, Binder EB, Bradley B, et al.: Correcting systematic inflation in genetic association tests that consider interaction effects: application to a genome-wide association study of posttraumatic stress disorder. JAMA Psychiatry 2014 Dec 1;71:1392–1399.

70. Broadaway KA, Duncan R, Conneely KN, Almli LM, Bradley B, Ressler KJ, et al.: Kernel Approach for Modeling Interaction Effects in Genetic Association Studies of Complex Quantitative Traits. Genet Epidemiol 2015 Jul;39:366–375.

71. Voisey J, Young RM, Lawford BR, Morris CP: Progress towards understanding the genetics of posttraumatic stress disorder. J Anxiety Disord 2014 Dec;28:873–883.

72. Consortium TSPG-WAS (GWAS): Genome-wide association study identifies five new schizophrenia loci. Nat Genet 2011 Oct;43:969–976.

73. Sullivan PF: The Psychiatric GWAS Consortium: Big Science Comes to Psychiatry. Neuron 2010 Oct 21;68:182–186.

74. Logue MW, Amstadter AB, Baker DG, Duncan L, Koenen KC, Liberzon I, et al.: The Psychiatric Genomics Consortium Posttraumatic Stress Disorder Workgroup: Posttraumatic Stress Disorder Enters the Age of Large-Scale Genomic Collaboration. Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol 2015 Sep;40:2287–2297.

75. Lachman HM, Papolos DF, Saito T, Yu YM, Szumlanski CL, Weinshilboum RM: Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. Pharmacogenetics 1996 Jun;6:243–250.

76. Binder EB: The role of FKBP5, a co-chaperone of the glucocorticoid receptor in the pathogenesis and therapy of affective and anxiety disorders. Psychoneuroendocrinology 2009 Dec;34 Suppl 1:S186–195.

77. Binder EB, Bradley RG, Liu W, Epstein MP, Deveau TC, Mercer KB, et al.: Association of FKBP5 polymorphisms and childhood abuse with risk of posttraumatic stress disorder symptoms in adults. JAMA J Am Med Assoc 2008 Mar 19;299:1291–1305.

78. Fujii T, Ota M, Hori H, Hattori K, Teraishi T, Sasayama D, et al.: Association between the common functional FKBP5 variant (rs1360780) and brain structure in a non-clinical population. J Psychiatr Res 2014 Nov;58:96–101.

79. Fujii T, Ota M, Hori H, Hattori K, Teraishi T, Matsuo J, et al.: The common functional FKBP5 variant rs1360780 is associated with altered cognitive function in aged individuals. Sci Rep 2014;4:6696.

80. Fani N, King TZ, Reiser E, Binder EB, Jovanovic T, Bradley B, et al.: FKBP5 genotype and structural integrity of the posterior cingulum. Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol 2014 Apr;39:1206–1213.

81. Fani N, Gutman D, Tone EB, Almli L, Mercer KB, Davis J, et al.: FKBP5 and attention bias for threat: associations with hippocampal function and shape. JAMA Psychiatry 2013 Apr;70:392–400.

82. Zannas AS, Wiechmann T, Gassen NC, Binder EB: Gene-Stress-Epigenetic Regulation of FKBP5: Clinical and Translational Implications. Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol 2016 Jan;41:261–274.

83. Bailey JN, Goenjian AK, Noble EP, Walling DP, Ritchie T, Goenjian HA: PTSD and dopaminergic genes, DRD2 and DAT, in multigenerational families exposed to the Spitak earthquake. Psychiatry Res 2010 Aug 15;178:507–510.

84. Chang S-C, Xie P, Anton RF, De Vivo I, Farrer LA, Kranzler HR, et al.: No association between ADCYAP1R1 and post-traumatic stress disorder in two independent samples. Mol Psychiatry 2012 Mar;17:239–241.

85. Navarro-Mateu F, Escámez T, Koenen KC, Alonso J, Sánchez-Meca J: Meta-analyses of the 5-HTTLPR polymorphisms and post-traumatic stress disorder. PloS One 2013;8:e66227.

86. Gressier F, Calati R, Balestri M, Marsano A, Alberti S, Antypa N, et al.: The 5-HTTLPR polymorphism and posttraumatic stress disorder: a meta-analysis. J Trauma Stress 2013 Dec;26:645–653.

87. Logue MW, Baldwin C, Guffanti G, Melista E, Wolf EJ, Reardon AF, et al.: A genome-wide association study of post-traumatic stress disorder identifies the retinoid-related orphan receptor alpha (RORA) gene as a significant risk locus. Mol Psychiatry 2013 Aug;18:937–942.

88. Lowe SR, Meyers JL, Galea S, Aiello AE, Uddin M, Wildman DE, et al.: RORA and posttraumatic stress trajectories: main effects and interactions with childhood physical abuse history. Brain Behav 2015 Apr;5:e00323.

89. Jetten AM: Retinoid-related orphan receptors (RORs): critical roles in development, immunity, circadian rhythm, and cellular metabolism. Nucl Recept Signal 2009;7:e003.

90. Xie P, Kranzler HR, Yang C, Zhao H, Farrer LA, Gelernter J: Genome-wide association study identifies new susceptibility loci for posttraumatic stress disorder. Biol Psychiatry 2013 Nov 1;74:656–663.

91. Almli LM, Srivastava A, Fani N, Kerley K, Mercer KB, Feng H, et al.: Correspondence. Biol Psychiatry 2014 Aug 15;76:e3–e4.

92. Beckmann JS, Estivill X, Antonarakis SE: Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet 2007 Aug;8:639–646.

93. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al.: Global variation in copy number in the human genome. Nature 2006 Nov 23;444:444–454.

94. Stankiewicz P, Lupski JR: Structural variation in the human genome and its role in disease. Annu Rev Med 2010;61:437–455.

95. Almal SH, Padh H: Implications of gene copy-number variation in health and diseases. J Hum Genet 2012 Jan;57:6–13.

96. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, et al.: Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature 2009 May 28;459:569–573.

97. Bucan M, Abrahams BS, Wang K, Glessner JT, Herman EI, Sonnenblick LI, et al.: Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. PLoS Genet 2009 Jun;5:e1000536.

98. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al.: Functional impact of global rare copy number variation in autism spectrum disorders. Nature 2010 Jul 15;466:368–372.

99. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, et al.: Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 2011 Jun 9;70:863–885.

100. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al.: Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 2008 Feb;82:477–488.

101. Moreno-De-Luca D, Mulle JG, Kaminsky EB, Sanders SJ, Myers SM, Adam MP, et al.: Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. Am J Hum Genet 2010 Nov 12;87:618–630.

102. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, et al.: Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet 2009 Nov;41:1223–1227.

103. Ingason A, Rujescu D, Cichon S, Sigurdsson E, Sigmundsson T, Pietiläinen OPH, et al.: Copy number variations of chromosome 16p13.1 region associated with schizophrenia. Mol Psychiatry 2011 Jan;16:17–25.

104. Mulle JG, Dodd AF, McGrath JA, Wolyniec PS, Mitchell AA, Shetty AC, et al.: Microdeletions of 3q29 confer high risk for schizophrenia. Am J Hum Genet 2010 Aug 13;87:229–236.

105. Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J, et al.: Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am J Psychiatry 2011 Mar;168:302–316.

106. Vacic V, McCarthy S, Malhotra D, Murray F, Chou H-H, Peoples A, et al.: Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. Nature 2011 Mar 24;471:499–503.

107. Kotlar AV, Mercer KB, Zwick ME, Mulle JG: New discoveries in schizophrenia genetics reveal neurobiological pathways: A review of recent findings. Eur J Med Genet 2015 Dec;58:704–714.

108. Degenhardt F, Priebe L, Herms S, Mattheisen M, Mühleisen TW, Meier S, et al.: Association between copy number variants in 16p11.2 and major depressive disorder in a German case-control sample. Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet 2012 Apr;159B:263–273.

109. Kawamura Y, Otowa T, Koike A, Sugaya N, Yoshida E, Yasuda S, et al.: A genome-wide CNV association study on panic disorder in a Japanese population. J Hum Genet 2011 Dec;56:852–856.

110. Malhotra D, Sebat J: CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell 2012 Mar 16;148:1223–1241.

111. Wang K, Bucan M: Copy Number Variation Detection via High-Density SNP Genotyping. Cold Spring Harb Protoc 2008 Jun 1;2008:pdb.top46.

112. LaFramboise T: Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Res 2009 Jul 1;37:4181–4193.

113. Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, et al.: High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. Genome Res 2006 Sep;16:1136–1148.

114. Winchester L, Yau C, Ragoussis J: Comparing CNV detection methods for SNP arrays. Brief Funct Genomic Proteomic 2009 Sep 1;8:353–366.

115. Pique-Regi R, Cáceres A, González JR: R-Gada: a fast and flexible pipeline for copy number analysis in association studies. BMC Bioinformatics 2010 Jul 16;11:380.

116. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al.: PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 2007 Nov 1;17:1665–1674.

117. Nutsua ME, Fischer A, Nebel A, Hofmann S, Schreiber S, Krawczak M, et al.: Family-Based Benchmarking of Copy Number Variation Detection Software. PloS One 2015;10:e0133465.

118. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, et al.: QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res 2007;35:2013–2025.

119. Lin C-F, Naj AC, Wang L-S: Analyzing Copy Number Variation using SNP Array Data: Protocols for Calling CNV and Association Tests. Curr Protoc Hum Genet Editor Board Jonathan Haines Al 2013 Oct 18;79:Unit–1.27.

120. Kessler RC, Colpe LJ, Fullerton CS, Gebler N, Naifeh JA, Nock MK, et al.: Design of the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). Int J Methods Psychiatr Res 2013 Dec;22:267–275.

121. Ressler KJ, Mercer KB, Bradley B, Jovanovic T, Mahan A, Kerley K, et al.: Post-traumatic stress disorder is associated with PACAP and the PAC1 receptor. Nature 2011 Feb 24;470:492–497.

122. Uddin M, Chang S-C, Zhang C, Ressler K, Mercer KB, Keyes KM, et al.: Adcyap1r1 genotype, posttraumatic stress disorder, and depression among women exposed to childhood maltreatment. Depress Anxiety 2013 Mar;30:251–258.

123. Almli LM, Mercer KB, Kerley K, Feng H, Bradley B, Conneely KN, et al.: ADCYAP1R1 genotype associates with post-traumatic stress symptoms in highly traumatized African-American females. Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet 2013 Apr;162B:262–272.

124. Wang L, Cao C, Wang R, Qing Y, Jianxin Zhang, Zhang XY: PAC1 receptor (ADCYAP1R1) genotype is associated with PTSD's emotional numbing symptoms in Chinese earthquake survivors. J Affect Disord 2013 Aug 15;150:156–159.

125. Liberzon I, King AP, Ressler KJ, Almli LM, Zhang P, Ma ST, et al.: Interaction of the ADRB2 Gene Polymorphism With Childhood Trauma in Predicting Adult Symptoms of Posttraumatic Stress Disorder. JAMA Psychiatry 2014 Aug 27; DOI: 10.1001/jamapsychiatry.2014.999

126. Miller MW, Wolf EJ, Sadeh N, Logue M, Spielberg JM, Hayes JP, et al.: A novel locus in the oxidative stress-related gene ALOX12 moderates the association between PTSD and thickness of the prefrontal cortex. Psychoneuroendocrinology 2015 Dec;62:359–365.

127. Logue MW, Solovieff N, Leussis MP, Wolf EJ, Melista E, Baldwin C, et al.: The ankyrin-3 gene is associated with posttraumatic stress disorder and externalizing comorbidity. Psychoneuroendocrinology 2013 Oct;38:2249–2257.

128. Kim TY, Chung HG, Shin H-S, Kim SJ, Choi JH, Chung MY, et al.: Apolipoprotein E gene polymorphism, alcohol use, and their interactions in combat-related posttraumatic stress disorder. Depress Anxiety 2013 Dec;30:1194–1201.

129. Zhang L, Benedek DM, Fullerton CS, Forsten RD, Naifeh JA, Li XX, et al.: PTSD risk is associated with BDNF Val66Met and BDNF overexpression. Mol Psychiatry 2014 Jan;19:8–10.

130. Li RH, Fan M, Hu MS, Ran MS, Fang DZ: Reduced severity of posttraumatic stress disorder associated with Val allele of Val66Met polymorphism at brain-derived neurotrophic factor gene among Chinese adolescents after Wenchuan earthquake. Psychophysiology 2016 Jan 11; DOI: 10.1111/psyp.12603

131. Wang T: Does BDNF Val66Met Polymorphism Confer Risk for Posttraumatic Stress Disorder? Neuropsychobiology 2015;71:149–153.

132. Badour CL, Hirsch RL, Zhang J, Mandel H, Hamner M, Wang Z: Exploring the association between a cholecystokinin promoter polymorphism (rs1799923) and posttraumatic stress disorder in combat veterans. J Anxiety Disord 2015 Sep 28;36:78–83.

133. Kimbrel NA, Garrett ME, Dennis MF, Liu Y, Patanam I, Workgroup VM-AM, et al.: Effect of genetic variation in the nicotinic receptor genes on risk for posttraumatic stress disorder. Psychiatry Res 2015 Sep 30;229:326–331.

134. Boscarino JA, Erlich PM, Hoffman SN, Zhang X: Higher FKBP5, COMT, CHRNA5, and CRHR1 allele burdens are associated with PTSD and interact with trauma exposure: implications for neuropsychiatric research and treatment. Neuropsychiatr Dis Treat 2012;8:131–139.

135. Lu AT, Ogdie MN, Järvelin M-R, Moilanen IK, Loo SK, McCracken JT, et al.: Association of the cannabinoid receptor gene (CNR1) with ADHD and post-traumatic stress disorder. Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet 2008 Dec 5;147B:1488–1494.

136. Valente NLM, Vallada H, Cordeiro Q, Bressan RA, Andreoli SB, Mari JJ, et al.: Catechol-O-methyltransferase (COMT) val158met polymorphism as a risk factor for PTSD after urban violence. J Mol Neurosci MN 2011 Mar;43:516–523.

137. Kolassa I-T, Kolassa S, Ertl V, Papassotiropoulos A, De Quervain DJ-F: The risk of posttraumatic stress disorder after trauma depends on traumatic load and the catechol-o-methyltransferase Val(158)Met polymorphism. Biol Psychiatry 2010 Feb 15;67:304–308.

138. Clark R, DeYoung CG, Sponheim SR, Bender TL, Polusny MA, Erbes CR, et al.: Predicting post-traumatic stress disorder in veterans: interaction of traumatic load with COMT gene variation. J Psychiatr Res 2013 Dec;47:1849–1856.

139. Humphreys KL, Scheeringa MS, Drury SS: Race moderates the association of Catechol-O-methyltransferase genotype and posttraumatic stress disorder in preschool children. J Child Adolesc Psychopharmacol 2014 Oct;24:454–457.

140. Amstadter AB, Nugent NR, Yang B-Z, Miller A, Siburian R, Moorjani P, et al.: Corticotrophin-releasing hormone type 1 receptor gene (CRHR1) variants predict posttraumatic stress disorder onset and course in pediatric injury patients. Dis Markers 2011;30:89–99.

141. White S, Acierno R, Ruggiero KJ, Koenen KC, Kilpatrick DG, Galea S, et al.: Association of CRHR1 variants and posttraumatic stress symptoms in hurricane exposed adults. J Anxiety Disord 2013 Oct;27:678–683.

142. Michopoulos V, Rothbaum AO, Jovanovic T, Almli LM, Bradley B, Rothbaum BO, et al.: Association of CRP genetic variation and CRP level with elevated PTSD symptoms and physiological responses in a civilian population with high levels of trauma. Am J Psychiatry 2015 Apr;172:353–362.

143. Drury SS, Brett ZH, Henry C, Scheeringa M: The association of a novel haplotype in the dopamine transporter with preschool age posttraumatic stress disorder. J Child Adolesc Psychopharmacol 2013 May;23:236–243.

144. Segman RH, Cooper-Kazaz R, Macciardi F, Goltser T, Halfon Y, Dobroborski T, et al.: Association between the dopamine transporter gene and posttraumatic stress disorder. Mol Psychiatry 2002;7:903–907.

145. Chang S-C, Koenen KC, Galea S, Aiello AE, Soliven R, Wildman DE, et al.: Molecular variation at the SLC6A3 locus predicts lifetime risk of PTSD in the Detroit Neighborhood Health Study. PloS One 2012;7:e39184.

146. Bailey JN, Goenjian AK, Noble EP, Walling DP, Ritchie T, Goenjian HA: PTSD and dopaminergic genes, DRD2 and DAT, in multigenerational families exposed to the Spitak earthquake. Psychiatry Res 2010 Aug 15;178:507–510.

147. Valente NLM, Vallada H, Cordeiro Q, Miguita K, Bressan RA, Andreoli SB, et al.: Candidate-gene approach in posttraumatic stress disorder after urban violence: association analysis of the genes encoding serotonin transporter, dopamine transporter, and BDNF. J Mol Neurosci MN 2011 May;44:59–67.

148. Duan Z, He M, Zhang J, Chen K, Li B, Wang J: Assessment of functional tag single nucleotide polymorphisms within the DRD2 gene as risk factors for post-traumatic stress disorder in the Han Chinese population. J Affect Disord 2015 Dec 1;188:210–217.

149. Voisey J, Swagell CD, Hughes IP, Morris CP, van Daal A, Noble EP, et al.: The DRD2 gene 957C>T polymorphism is associated with posttraumatic stress disorder in war veterans. Depress Anxiety 2009;26:28–33.

150. Wolf EJ, Mitchell KS, Logue MW, Baldwin CT, Reardon AF, Aiello A, et al.: The dopamine D3 receptor gene and posttraumatic stress disorder. J Trauma Stress 2014 Aug;27:379–387.

151. Voisey J, Swagell CD, Hughes IP, Connor JP, Lawford BR, Young RM, et al.: A polymorphism in the dysbindin gene (DTNBP1) associated with multiple psychiatric disorders including schizophrenia. Behav Brain Funct BBF 2010;6:41.

152. Xie P, Kranzler HR, Poling J, Stein MB, Anton RF, Farrer LA, et al.: Interaction of FKBP5 with childhood adversity on risk for post-traumatic stress disorder. Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol 2010 Jul;35:1684–1692.

153. Sarapas C, Cai G, Bierer LM, Golier JA, Galea S, Ising M, et al.: Genetic markers for PTSD risk and resilience among survivors of the World Trade Center attacks. Dis Markers 2011;30:101–110.

154. Nelson EC, Agrawal A, Pergadia ML, Lynskey MT, Todorov AA, Wang JC, et al.: Association of childhood trauma exposure and GABRA2 polymorphisms with risk of posttraumatic stress disorder in adults. Mol Psychiatry 2009 Mar;14:234–235.

155. Neigh G, Jovanovic T, Smith A, Almli L, Gillespie C, Kilaru V, et al.: GLUT1 polymorphism decreases incidence of depression and ptsd after trauma. Psychoneuroendocrinology 2015 Nov;61:59.

156. Lawford BR, Morris CP, Swagell CD, Hughes IP, Young RM, Voisey J: NOS1AP is associated with increased severity of PTSD and depression in untreated combat veterans. J Affect Disord 2013 May;147:87–93.

157. Lucas-Thompson RG, Holman EA: Environmental stress, oxytocin receptor gene (OXTR) polymorphism, and mental health following collective stress. Horm Behav 2013 Apr;63:615–624.

158. Liu Y, Rimmler J, Dennis MF, Ashley-Koch AE, Hauser MA, Mid-Atlantic Mental Illness Research Education and Clinical Center Workgroup, et al.: Association of Variant rs4790904 in Protein Kinase C Alpha with Posttraumatic Stress Disorder in a U.S. Caucasian and African-American Veteran Sample. J Depress Anxiety 2013 Jun 1;2:S4001.

159. Amstadter AB, Koenen KC, Ruggiero KJ, Acierno R, Galea S, Kilpatrick DG, et al.: Variant in RGS2 moderates posttraumatic stress symptoms following potentially traumatic event exposure. J Anxiety Disord 2009 Apr;23:369–373.

160. Wilker S, Kolassa S, Vogler C, Lingenfelder B, Elbert T, Papassotiropoulos A, et al.: The role of memory-related gene WWC1 (KIBRA) in lifetime posttraumatic stress disorder: evidence from two independent samples from African conflict regions. Biol Psychiatry 2013 Nov 1;74:664–671.

161. Solovieff N, Roberts AL, Ratanatharathorn A, Haloosim M, De Vivo I, King AP, et al.: Genetic Association Analysis of 300 Genes Identifies a Risk Haplotype in SLC18A2 for Posttraumatic Stress Disorder in two Independent Samples. Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol 2014 Feb 14; DOI: 10.1038/npp.2014.34

162. Zhang J, Sheerin C, Mandel H, Banducci AN, Myrick H, Acierno R, et al.: Variation in SLC1A1 is related to combat-related posttraumatic stress disorder. J Anxiety Disord 2014 Dec;28:902–907.

163. Walsh K, Uddin M, Soliven R, Wildman DE, Bradley B: Associations between the SS variant of 5-HTTLPR and PTSD among adults with histories of childhood emotional abuse: results from two African American independent samples. J Affect Disord 2014 Jun;161:91–96.

164. Liu Y, Garrett ME, Dennis MF, Green KT, VA Mid-Atlantic MIRECC Registry Workgroup, Ashley-Koch AE, et al.: An examination of the association between 5-HTTLPR, combat exposure, and PTSD diagnosis among U.S. veterans. PloS One 2015;10:e0119998.

165. Gillespie CF, Almli LM, Smith AK, Bradley B, Kerley K, Crain DF, et al.: Sex dependent influence of a functional polymorphism in steroid 5-α-reductase type 2 (SRD5A2) on post-traumatic stress symptoms. Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet 2013 Apr;162B:283–292.

166. Goenjian AK, Bailey JN, Walling DP, Steinberg AM, Schmidt D, Dandekar U, et al.: Association of TPH1, TPH2, and 5HTTLPR with PTSD and depressive symptoms. J Affect Disord 2012 Nov;140:244–252.

167. Nievergelt CM, Maihofer AX, Mustapic M, Yurgil KA, Schork NJ, Miller MW, et al.: Genomic predictors of combat stress vulnerability and resilience in U.S. Marines: A genome-wide association study across multiple ancestries implicates PRTFDC1 as a potential PTSD gene. Psychoneuroendocrinology 2015 Jan;51:459–471.

168. Wingo AP, Almli LM, Stevens JJ, Klengel T, Uddin M, Li Y, et al.: DICER1 and microRNA regulation in post-traumatic stress disorder with comorbid depression. Nat Commun 2015;6:10106.

169. Guffanti G, Galea S, Yan L, Roberts AL, Solovieff N, Aiello AE, et al.: Genome-wide association study implicates a novel RNA gene, the lincRNA AC068718.1, as a risk factor for post-traumatic stress disorder in women. Psychoneuroendocrinology 2013 Dec;38:3029–3038.

170. Almli LM, Stevens JS, Smith AK, Kilaru V, Meng Q, Flory J, et al.: A genome-wide identified risk variant for PTSD is a methylation quantitative trait locus and confers decreased cortical activation to fearful faces. Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet 2015 Jul;168B:327–336.

171. Xie P, Kranzler HR, Yang C, Zhao H, Farrer LA, Gelernter J: Genome-wide Association Study Identifies New Susceptibility Loci for Posttraumatic Stress Disorder. Biol Psychiatry 2013 Nov 1;74:656–663.

Chapter 2: Functional evaluation of a PTSD-associated genetic variant: estradiol regulation and *ADCYAP1R1*

This chapter has been submitted to Translational Psychiatry and is currently under review for publication

Introduction

Posttraumatic stress disorder (PTSD) is a psychiatric disorder that can affect children and adults as well as civilian and non-civilian populations. Although anyone is susceptible to developing PTSD when exposed to a traumatic event, a greater prevalence has been reported among women compared to men. In fact, within the US, the percent of females that suffer from PTSD during their lifetime is nearly twice that of men [1,2]. This observation has led to renewed interest in the contribution of hormones, particularly estrogen, in psychiatric outcomes among women. Several research studies support an association between low levels of estrogen and increased fear and anxiety related behaviors [3–7]. In line with the implications of these research findings, genetic variants that interfere with estrogen regulation of the stress pathway may play a role in neuropsychiatric disease.

One example of estrogen regulation on cellular processes involves sequence specific enhancers known as estrogen response elements (EREs) [8]. The ligand receptor complex formed by binding of estradiol (E2) to estrogen receptors (ER) alpha or beta can translocate to the nucleus where it can then interact with EREs to activate gene transcription [9–11]. Sequence alterations in the relatively conserved ERE can interfere with the binding efficiency of E2-activated ER and negatively affect its enhancer activity

[12]. As such, genomic polymorphisms or other mutations within an ERE can result in dysregulation of normal responses triggered by estradiol. We previously identified a single nucleotide polymorphism (SNP), rs2267735, that is associated with decreased expression of *ADCYAP1R1* cortex mRNA in females only [13,14]. Interestingly, this genomic variant is located within a putative ERE in an intron of *ADCYAP1R1*, the gene encoding pituitary adenylate cyclase-activating polypeptide type 1 receptor (PAC1).

The PAC1 receptor, when bound to its ligand PACAP (pituitary adenylate cyclase-activating polypeptide), contributes to the stress response system through the hypothalamic-pituitary-adrenal (HPA) axis by activating production of cortisol [15–17]. Impairment of the HPA axis and related signaling pathways has been attributed to several psychiatric conditions including PTSD [18–24]. Consequently, irregularities in the mRNA expression of neuropeptide hormones and receptors involved in related stress pathways could contribute to neuropsychiatric disorders. Consistent with this paradigm, we have shown that carriers of the rs2267735 risk genotype CC, compared to those with the CG or GG genotype, have not only a statistically significant decrease in expression levels of *ADCYAP1R1* but also higher levels of PTSD symptoms [13]. Taken together, these findings have led us to hypothesize that, via an ERE, estradiol and the estrogen receptor may play a significant role in the regulation of PAC1 and regulation of stress response.

In the current manuscript, we sought to determine whether or not estradiol-activated ERα is sufficient to induce transcription of *ADCYAP1R1;* if the putative *ADYCAP1R1* ERE binds ERα differentially dependent on genotype at rs2267735; if serum estradiol has an effect on expression of *ADCYAP1R1* in humans; and if expression

levels of *ADCYAP1R1* correlate with PTSD symptoms. The results presented in this chapter provide insight into a mechanism that may partly explain the biological effects of estradiol on PTSD outcome and other stress- and trauma-related disorders.

Materials and Methods

Expression analysis of *Adcyap1r1* in mouse brain

8-week old week female mice were ovariectomized after which a 2mm pellet containing either sesame oil or 17β-estradiol (E2) dissolved in sesame oil at 1μg/μl (cat # E8875; Sigma Aldrich, St. Louis, MO) was placed subcutaneously into 16 mice per treatment group (32 total). Ten days later, half of these mice (N=16; 8 from each treatment group) were fear conditioned with exposure to a tone paired with foot shock for 5 times using the following conditions: 6 kHz for 30 seconds followed by 0.6 mA foot shock for 1 second and a 2 minute interval before the next set of tone-shocks. The remaining 16 mice were kept in their home cages in the vivarium. Within 2 hours of exposure to fear conditioning, the brains of each mouse were removed and immediately frozen on dry ice. 1mM tissue punches were collected from the bed nucleus of the stria terminalis (BNST), a brain region previously found to associate with estradiol induced expression of Adcyap1r1 [13]. mRNA was extracted using the RNeasy Mini kit (cat 74104; Qiagen, Inc., Valencia, CA) and converted to cDNA using the RT^2 First Strand Kit (cat # 330401, Qiagen, Inc.). For gene expression analysis we performed quantitative PCR (qPCR) using TaqMan Gene Expression assays for *Adcyap1r1* (Mm01326453_m1; cat # 4351372, ThermoFisher Scientific, Waltham, MA) and *Gapdh,* as an endogenous control (cat # 4352932, ThermoFisher Scientific). The $2^{-\Delta\Delta Ct}$ method was used to compare

fold-change in expression between the condition and treatment groups [25]. 2-tailed t tests were used to determine statistical significance of fold differences.

Competitive ELISA to determine ERE binding

To test the binding capabilities of ERα to the putative estrogen response elements (ERE) containing either the C or G allele of rs2267735, we performed a competitive binding ELISA using the TransAM ER kit (cat # 41396; Active Motif, Carlsbad, CA). To each microwell coated with canonical ERE sequence oligonucleotide (oligo), we added estradiol treated MCF-7 nuclear extract (5μg) (cat # 36016, Active Motif) and 10-100 pmoles of one of four double-stranded oligo DNAs (2 competing/experimental sequences plus a negative and positive control). The sequences of the oligos used for this experiment are provided in Supplementary Table 2-1. Each experimental assay was performed in triplicate. The controls were performed in duplicate. Individual oligos and nuclear extract were added to each microwell and incubated for one hour. Next, a primary antibody (ERα; 1:1000) then secondary antibody (HRP-conjugated IgG; 1:1000) were added for one hour each. After several rinses, a colorimetric reaction to horseradish peroxidase (HRP) was initiated and absorbance was measured at a wavelength of 450nm to determine binding efficiencies. Antibodies and other reagents were provided in the TransAM ER kit.

Expression analysis of *ADCYAP1R1* in transfected HEK293T cells

HEK293T cells were transfected with a plasmid containing either green fluorescent protein (GFP) or the full-length human estrogen receptor (pCMV-hERα) [26,27] and treated with E2 or ethanol only. mRNA was extracted and gene expression of

ADCYAP1R1 was measured with commercially available assays. The $2^{-\Delta\Delta Ct}$ method was used to compare fold-change in expression between each condition and treatment group [25]. See Supplemental Materials and Methods for more experimental details.

Cross-linking chromatin immunoprecipitation (X-ChIP)

HEK293T cells were transfected with pCMV-hERα and treated with E2. Following treatment, the cells were exposed briefly to formaldehyde to cross-link DNA/protein complexes, and then lysed and sonicated. A monoclonal antibody to ERα was used to isolate DNA bound by the receptor. After reversing the crosslinking, the ERα bound DNA was precipitated using standard phenol/chloroform extraction followed by ethanol precipitation. The ERE of interest within *ADCYAP1R1* was measured using region specific primers and quantitative PCR. The qPCR measurement for a transcriptionally inactive region of the genome was used as a negative control. See Supplemental Materials and Methods for more experimental details.

Collection of phenotype and genotype data from study participants

Patients waiting for an appointment with their primary care of OB/GYN doctor at the Grady Memorial Hospital in Atlanta, GA were recruited to participate in a research study aimed at identifying genetic factors that contribute to PTSD. Consenting participants were asked to provide a saliva sample for DNA collection and genotyping. From some participants, whole blood was also drawn in order to collect a serum sample from which estradiol was measured (by RIA, cat # KE2D1, Siemens Healthcare Diagnostics, Malvern, PA). Study participants completed the PTSD symptom scale (PSS)

to measure current PTSD symptoms [28]. Genotyping of rs2267735 was performed using either a TaqMan SNP genotyping assay (C_15872945_10, cat # 4351379, ThermoFisher Scientific) or the Sequenom MassArray iPlex system (Sequenom Inc., San Diego, CA) with custom designed primers.

Detection of *ADCYAP1R1* mRNA in human whole blood by quantitative PCR

Total mRNA was extracted from whole blood samples collected in Tempus Blood RNA tubes (cat # 432792; ThermoFisher Scientific) using the PerfectPure RNA 96 Cell Kit (cat #2900296; 5 PRIME, Inc., Gaithersburg, MD). The expression of *ADCYAP1R1* mRNA in blood was measured in duplicate by reverse transcriptase reactions (rtPCR) with target-specific primers, followed by qPCR with internal DNA Detection Switch (iDDS) probes and antiprobes [29,30]. rtPCR reactions were performed with 2 μL of total RNA and GoScript Reverse Transcriptase (cat # A5003, Promega, Madison, WI), according to the manufacturer's protocol. qPCR reactions were performed in 20μL volumes containing 10 μL of 2X HotStart-IT Probe qPCR Master Mix (cat # 75766, Affymetrix, Inc., Santa Clara, CA); additional $MgCl_2$ and dNTPs (final concentrations of 5 and 0.1mM, respectively); 3 μL of cDNA; internal primers, and an iDDS probe and antiprobe (GeneTAG Technology, Atlanta, GA). *ADCYAP1R1* expression levels were normalized to *GAPDH* using the $2^{-\Delta\Delta Ct}$ method [25]. The sequences and final concentrations for the primer, probe, and antiprobe concentrations are provided in Supplementary Table 2-1.

For this analysis, we selected 105 samples consisting of rs2267735 genotyped females with mRNA (RNA integrity number (RIN) > 5.0; concentration > 100 ng/μl), a

measure of estradiol, a PTSD symptoms score, and moderate trauma exposure. Samples with a gene expression threshold (C_t) value greater than 50 for duplicate reactions were excluded from further analysis resulting in a final sample size of 95 (range of C_t values: 30.84-45.88).

Results

Expression of *ADCYAP1R1* in the brain of mice exposed to estradiol treatment and fear conditioning

The PAC1 receptor, encoded by *Adcyap1r1*, and its ligand PACAP play a significant role in stress regulation within the bed nucleus of the stria terminalis (BNST). Within the BNST of rats, increased transcript levels of both PAC1 and PACAP are induced by chronic stress, which results in the release of corticosterone and anxiety related behavior [31–33]. Expression of *Adcyap1r1* in the BNST is also activated by estradiol. We have shown previously in ovariectomized rats that treatment with E2 results in higher levels of *Adcyap1r1* expression in the BNST when compared to vehicle-only (oil) treated controls [13]. We replicated the *Adcyap1r1* expression response to E2 in the BNST of ovariectomized mice, revealing a slightly smaller but statistically significant 1.4-fold increase over that measured for vehicle-only exposed mice ($p = 0.0005$) (Figure 2-1). We also show that fear conditioning (fc) results in a 4-fold increase in expression when compared to non-fear conditioned home-caged (hc) animals ($p=7.1x10^{-6}$) (Figure 2-1). Interestingly, when paired, the combination of fear conditioning plus E2 treatment (Estradiol (fc)) results in expression of *Adcyap1r1* that is

greater than that exhibited by either condition separately; p=.03 compared to fear conditioning only (vehicle (fc)) and p= 8.9×10^{-7} compared to estradiol treatment only (estradiol (hc)) (Figure 2-1). These data reveal that there is an additive effect of estradiol and fear conditioning on the expression levels of *Adcyap1r1* in the BNST.

ERα dependent activation of *ADCYAP1R1* expression

Estradiol regulates gene expression by binding to estrogen receptors alpha (ERα) or beta (ERα), activating a conformational change that allows the estradiol receptor complex to bind to chromatin and induce transcription. ERα has been implicated in mood regulation and is also highly expressed in the BNST of the human brain [34,35]. Therefore, we hypothesized that the estradiol effect on expression of *ADCYAP1R1* is likely occurring via interaction with ERα. To address this hypothesis, we sought to determine if introducing ERα into a cell line that expresses a detectable level of *ADCYAP1R1* but does not endogenously express ERα would result in enhanced expression of *ADCYAP1R1* in the presence of estradiol. Full-length hERα was transiently transfected into HEK293T (human embryonic kidney) cells and treated with estradiol for several hours. The cells were also transfected with a GFP plasmid to control for the effects of transfection on changes in *ADCYAP1R1* expression. Expression was also examined in GFP transfected cells treated with estradiol to control for effects of estradiol on expression that do not involve ligand activation of ERα. Among HEK293 cells treated with estradiol, there is a 3.9-fold increase in *ADCYAP1R1* expression for cells transfected with hERα compared to those transfected with GFP ($p=5.8 \times 10^{-6}$) (Figure 2-2). These data reveal that ERα is sufficient to induce expression of *ADCYAP1R1*. Cells

that were treated with estradiol and transfected with GFP show a minor and non-significant increase in fold change expression (1.0 versus 1.2) compared to vehicle-transfected cells (Figure 2-2). These data further support the role of estrogen activated ERα specifically, in estradiol-induced expression of *ADCYAP1R1*.

in vivo **binding of ERα to an *ADCYAP1R1* intronic ERE**

E2/ERα activation of transcription occurs by binding to estrogen response elements (EREs), which are specific DNA sequences located throughout the genome [9]. Within a particular intron of *ADCYAP1R1* there are several predicted EREs (purple boxes, Supplementary Figure 2-1). One of these EREs (underlined, Supplementary Figure 2-1) contains a SNP, rs2267735 (in red, Supplementary Figure 2-1), which correlates with allele-specific differential expression of *ADCYAP1R1* in the cortex of human brain [13]. To test whether or not this specific ERE is capable of binding ligand-activated ERα *in vivo*, consistent with that of a functional ERE, we performed cross-linking chromatin immunoprecipitation (ChIP). HEK293T cells, which do not endogenously express ERα, were transiently transfected with full-length human ERα-encoding cDNA. After the cells were treated with estradiol, then formaldehyde cross-linked to preserve DNA/protein complexes, the chromatin was incubated with a monoclonal antibody to ERα to "pull-down" ERα-bound DNA. DNA that bound ERα was detected and measured using quantitative PCR (qPCR). Primers (in green, Supplementary Figure 2-1) were specifically designed to amplify a 55 bp region (chr7: 31,095,865-31,095,920) containing the predicted ERE sequence of interest. Relative to the percent input (non-immunoprecipitated DNA) observed for a transcriptionally

inactive region of the genome (negative control). We found that binding of ERα to this particular *ADCYAP1R1* ERE is approximately 9-fold greater (.50% versus 0.06%) compared to a control DNA region (p=0.002) (Figure 2-3 A). These data suggest that the human ERα is indeed able to bind to the predicted ERE containing a PTSD-associated SNP within *ADCYAP1R1*.

Differential binding of ERα based on the SNP variant present within the ERE

With evidence of ERα binding to the ERE of interest, we then wanted to test whether there may be differential allele-specific binding efficiencies between the C versus G allele of rs2267735. Since the ERE is a regulatory sequence which activates transcription when bound to ligand-activated ERα, the efficiency of binding is correlated with enhancer activity [36]. Thus, we hypothesized that the C allele, for which we observed lower *ADCYAP1R1* expression among homozygous carriers in our previous work, [13] would have lower binding affinity compared to the G allele. Double stranded, oligonucleotides (oligos) containing the genomic sequences of the ERE (in parenthesis, Supplementary Figure 2-1) with either the "C" or "G" allele were used in a competitive ELISA to compare binding of ERα to that of the canonical ERE sequence. We used 20pmol of oligonucleotide for four sequences (non-canonical, C allele, G allele, and canonical; see Supplementary Table 2-1) to measure the ability of each oligo to "out-compete" canonical ERE for ERα binding. The data for each oligo (non-canonical, C allele, and G allele) were transformed to represent a measure of fluorescence relative to the positive control (canonical ERE) (Figure 2-3 B). As expected, the non-canonical oligo has the least affinity to ERα with a binding efficiency of 50% compared to that of the

positive control. The second lowest binding was observed for the C allele oligo (68%). The *ADCYAP1R1* intronic ERE that contains the rs2267735 C allele shows a statistically significant reduction in binding to ERα (p= .0005) in comparison to the canonical ERE (Figure 2-3 B). Compared to binding of ERα to the G allele, which binds nearly as well as the canonical sequence, the C allele also binds with less efficiency (statistically significant; p= .03). The directions of these results are consistent with the hypothesis that the C allele would bind least well of the two experimental conditions. The dosage effect of binding was also examined for a range of oligo concentrations (Supplementary Figure 2-2). Interestingly, while all other oligo binding measures remained relatively constant, the C allele oligo eventually reaches near normal binding at 100pmol of oligo. This leads us to speculate that, regardless of having a C allele present at this particular *ADCYAP1R1* ERE, some binding may still occur in a dose-dependent fashion.

ADCYAP1R1 expression relative to rs2267735 genotype and PTSD symptoms

The C allele of rs2267735 results in reduced binding of E2/ERα to the ERE with a binding efficiency that appears to be conditional. This reduced binding is presumably associated with lower levels of *ADCYAP1R1* transcript previously observed among study participants with the CC genotype [13]. Although we have shown an estradiol induced increase in expression of *ADCYAP1R1* in mice, we have not yet observed this phenomenon in humans nor have we tested the effect of variable estradiol concentrations on expression levels. In addition to better understanding the effects of serum estradiol on *ADCYAP1R1* expression, we were also interested in testing the effect of estradiol on expression among those with the CC genotype at rs2267735. Assuming lower expression

among women with the CC genotype, we wanted to determine if having higher levels of estradiol could compensate, leading to increased levels of *ADCYAP1R1* expression.

Serum estradiol and transcript levels of *ADCYAP1R1* were measured among 95 genotyped females from whom serum (for measuring estradiol) and mRNA were collected from whole blood draws performed on the same day. In human whole blood, higher levels of serum estradiol correlate with increased *ADCAYP1R1* mRNA (Table 2-1 A). Among the women with high estradiol (top of median split; range = 33.84-528.35 pg/ml), 61.7% have high expression. Correspondingly, 60.4% of women with low estradiol (5.70-32.74 pg/ml) have low expression. This association is statistically significant (p= .031). Next, we tested the relationship between estradiol, genotype and *ADCYAP1R1* expression. Women with levels of serum estradiol in the lowest quartile, have a higher, averaged *ADCYAP1R1* ΔCt value (corresponding to lower expression) if they have the CC risk genotype (ΔCt=22.02) compared to women with either the GG or GC genotype (ΔCt=20.86) (2-tailed t-test; p=.049) (Table 2-1 B). Among women with estradiol measures in the top quartile, we observe a more even distribution of *ADCYAP1R1* expression between the CC and GG or GC genotype groups (average ΔCt=21.23 and 21.39, respectively; data not shown). We examined the expression of *ADCYAP1R1* (by median split) among women with the CC genotype and either high or low estradiol measures and found that expression increases in the higher estradiol group (p=.034) (Table 2-1 C). This same analysis was not statistically significant for those with the GG or GC genotype (data not shown). These latter results suggest that increased estradiol may act in a compensatory way to increase levels of *ADCYAP1R1* observed among woman with the CC genotype at rs2267725.

We have shown an increase of *ADCYAP1R1* expression in response to acute stress in mice. However, it remained unclear if expression changes would also occur in humans, particularly in those with PTSD. In the current study we found a negative relationship between blood mRNA *ADCYAP1R1* expression and PTSD symptoms among our study participants. The data reveal that those with low *ADCYAP1R1* expression have a statistically significant increase in current PTSD symptoms measured with the PSS ($p = .026$) (Table 2-1 D). Although there was a moderate increase for current PTSD symptoms among those in the lower range of serum estradiol (median split) compared to those in the upper range (mean PSS of 11.05 vs. 9.23) these findings were not statistically significant ($p>0.05$; data not shown).

Discussion

We previously reported a genetic association between a SNP (rs2267735) within the gene *ADCYAP1R1* (PAC1) and PTSD in females. Because the rs2267735 SNP lies within a putative ERE, we hypothesized a role for estrogen in regulating *ADCYAP1R1* expression. We now report evidence of a functional role for rs2267735 in the dysregulation of ERα/ERE transcriptional activation of *ADCYAP1R1* and provide further insight and rationale for the sex-specific effects of this polymorphism on PTSD risk.

Based on our current data, we have developed a model for the putative cellular mechanisms through which the C allele of rs2267735 results in differential expression of *ADCYAP1R1* and increased susceptibility to PTSD (Figure 2-4). Previously reported in rats, here we confirm that induced stress in mice also results in an increased expression of

ADCYAP1R1 mRNA in the BNST (#1 in model, Figures 2-4 A & 2-4 B). We have now shown that both stress and estradiol additively increase expression of *ADCYAP1R1* (#2, Figures 2-4 A & 2-4 B). Estradiol regulates expression of *ADCYAP1R1* through ligand activation of ERα and binding to an ERE within the gene (#3, Figure 2-4 A). When the risk allele ("C") is present within the ERE sequence, binding of E2/ERα is compromised (#4, Figure 2-4 B). As such, reduced binding results, which inhibits the activation of *ADCYAP1R1* transcription (#5, Figure 2-4 B) and results in a dysregulated response to stress or trauma. Finally, as a result of these altered mechanisms of regulation, lower expression of *ADCYAP1R1* is associated with PTSD (#6, Figure 2-4 B). It is possible that decreased *ADCYAP1R1* leads to impaired feedback to the stress axis, consistent with an inability to mount a proper compensatory stress response.

The findings we present here are consistent with and help clarify the results of our original PAC1 association with PTSD finding [13]. We previously found that increased methylation at a CpG within the promoter of *ADCYAP1R1* correlated with PTSD symptoms ($r=0.354$; $p<.0005$) [13]. Typically, promoter methylation results in reduced transcription of a gene. Although this is not conclusive evidence of an inverse relationship between *ADCYAP1R1* expression and PTSD symptoms, it was the first piece of evidence to suggest this direction of association. We have also previously shown that females with the PTSD risk genotype at rs2267735 ("CC"), have a statistically significant reduction in cortex *ADCYAP1R1* mRNA compared to females with the CG or GG genotype and males with the CC genotype [13]. Deductively, these data further insinuate that the degree of *ADCYAP1R1* expression and PTSD symptoms are inversely correlated. In support of this reasoning, results presented in this manuscript reveal that either those

with high PTSD symptoms have lower expression levels of *ADCYAP1R1* or that low expression levels increases risk for having greater symptoms of PTSD.

The signaling pathway associated with *ADCYAP1R1* (PAC1) and stress regulation requires binding of the PACAP ligand to the PAC1 G protein-coupled receptor. Thus, the association between PACAP and PTSD must also be considered when trying to understand the relationship between PAC1 and PTSD. An increase in PACAP is regularly observed as a biological response to stress [37]. We have previously shown that high levels of serum PACAP38 correlate with increased symptoms of PTSD in our study participants ($r=0.497$, $p\leq .005$) [13]. However, despite increased levels of PACAP, the rate-limiting factor for signaling activation is the number of receptors available to bind. Thus, for PTSD, in which expression of *ADCYAP1R1* and presumably levels of PAC1 are reduced, PACAP would be insufficient to compensate for the outcome associated with decreased receptor levels. One explanation would be that decreased PAC1 may lead to an inability to down regulate PACAP release in a compensatory fashion in response to stress.

There are several limitations to the current study. In the course of these experiments we found that *ADCYAP1R1* is not detectably expressed in all cell types. Lymphoblast cell lines with the CC and GG genotype of rs2267735 were originally obtained in order to measure the association between genotype and expression as a factor of variable E2 treatment, and to assess allele-specific binding efficiency of ERα to ERE *in vivo* with increasing exposure of E2. Unfortunately, we were unable to measure *ADCYAP1R1* expression in lymphoblast cell lines. Additionally, it would have been preferable to measure ERα binding in a neuronal cell line that likely contains

transcription factors that contribute to activation of *ADCYAP1R1* expression by facilitating E2/ERα binding to the ERE. The neuronal cell line we chose to utilize, SH-SY5Y, was difficult to grow and we therefore discontinued working with these cells. However, we were able to detect low but convincing levels of ERa binding to the *ADCYAP1R1* ERE in HEK293T cells, but more accurate measures of binding may have been obtained using cells with greater similarity to those present in brain tissue where *ADCYAP1R1* is highly expressed. The relationship between PTSD and *ADCYAP1R1* expression remains speculative given that it was tested using whole blood and should be explored further by examining expression in post-mortem brain tissue from individuals with and without PTSD. The amount of receptor protein produced in human brain tissue, particularly in the BNST, should also be measured and compared among individuals with differing genotypes at rs2267735. Although we have measured the levels of *ADCYAP1R1* mRNA in our experimental models, we cannot make any conclusions about the amount of receptor protein being produced. As mentioned above, the amount of receptor is likely key to the functionality of stress-related PACAP/PAC1 signaling.

The effect of estradiol on stress and anxiety behaviors has been well established in the scientific literature. Recent findings show differences in brain activity in response to psychosocial stress depending on whether women are in the low or high estrogen phase of their menstrual cycle [7]. Research from Glover et al. shows that women in the low estrogen phase of their cycle have impaired fear inhibition [3]. This was supported by another study by Wegerer et al. which showed that women with lower levels of estradiol have stronger intrusive memories - one of the symptoms of PTSD [38]. If low levels of estradiol increase risk for negative psychiatric outcomes, it is possible that administration

of estradiol during a time of trauma or anxiety might provide protective effects against the development of stress related disorders. Data supporting this hypothesis comes from Ferree et al., where the effects of being on oral contraceptives or receiving emergency contraception immediately after being sexual assaulted resulted in fewer PTSD type symptoms [39]. These data favor the hypothesis that susceptibility to developing PTSD may be related to levels of estradiol at the time of trauma.

As genetic variants associated with neuropsychiatric disorders continue to be identified, functional analyses into the consequences of such variants will be necessary to provide insight into the molecular mechanism of disease. Uncovering these mechanisms can provide invaluable clues that may result in prevention or more effective treatment options. In this manuscript, we provide one of the first analyses into the molecular effects of a PTSD genetic risk variant. In doing so we also identify a mechanism of estrogen regulation that may partially explain increased risk of PTSD among females. Our study illuminates the sex-dependent relationship between *ADCYAP1R1* polymorphisms, trauma, and estrogen on risk for PTSD. It also provides additional support for estradiol as a potential therapeutic treatment in the prevention of PTSD, particularly for those with the *ADCYAP1R1* risk genotype.

Supplementary Materials and Methods

Expression analysis of *ADCYAP1R1* in transfected HEK293T cells

HEK293T cells were plated at approximately 5 X 10^5 cells in 2, 6-well culture plates with phenol red free DMEM (cat # 12-917F; Lonza BioWhittaker, Walkersville, MD) supplemented with L-glutamine (4mM; cat # 25030-081; ThermoFisher Scientific) and charcoal-stripped FBS (10%; cat # S11695H; Atlanta Biologicals, Inc., Flowery Branch, GA). 25 hours later, the cells were transfected with either 2μg of a plasmid vector containing green fluorescent protein (GFP) or 3μg of a plasmid containing the full-length human estrogen receptor, pCMV-hERα (kindly provided by Dr. Ann Nardulli) in 2mls of Opti-MEM (cat # 31985-070; ThermoFisher Scientific) plus 10μl of Lipofectamine 2000 (cat # 11668; ThermoFisher Scientific). The cells were transfected for 6.5 hours after which the cell culture media was replaced with hormone free DMEM (as above) plus penicillin/streptomycin (1%; cat # 11668, ThermoFisher Scientific). After 21 hours, the cells from each transfection were combined and re-plated in hormone free media across 12 wells (24 total). After 24 more hours, 6-wells of cells for each transfection were grown in media with either 10nM 17β-Estradiol (E2) (cat # E2758; Sigma-Aldrich, St. Louis, MO) or EtOH only for 16.5 hours. mRNA was extracted from the cells using the RNeasy mini kit (cat # 74104; Qiagen, Inc., Valencia, CA). mRNA was converted to cDNA using the RT^2 First Strand Kit (cat # 330401, Qiagen, Inc.). Gene expression of *ADCYAP1R1* was assayed using TaqMan Gene Expression Assays: Hs01027974_m1 (*ADCYAP1R1*) and Hs02758991_g1 (*GAPDH*, endogenous control) with TaqMan Gene Expression Master Mix (cat # 4369016) (ThermoFisher Scientific). The $2^{-\Delta\Delta Ct}$ method was used to compare fold-change in expression between each

condition and treatment group. Successful transfection of hERα, which is not endogenously expressed in HEK293T cells, was confirmed using a TaqMan expression assay (*ESR1*, Hs00174860_m1, ThermoFisher Scientific). Additionally we observed 70-90% transfection efficiency for GFP by assessing florescence under a compound microscope.

Cross-linking chromatin immunoprecipitation (X-ChIP)

HEK293T cells were transfected with 3μg of pCMV-hERα (as described above) and treated with 10nM E2 for 2 hours. Prior to treatment with E2, cells were grown in hormone-free media: DMEM with phenol red (cat # 12-741F, Lonza BioWhittaker, Walkersville, MD) supplemented with additional L-glutamine (1%, cat # 25030-081; ThermoFisher Scientific) and charcoal stripped FBS (10%, cat # 12676011; ThermoFisher Scientific) for 41.5 hours. Following treatment with E2, the cells from each of 6-wells were gently pelleted and rinsed with 1 x phosphate buffered saline (PBS). After removing the PBS, 940ul of 1% formaldehyde in 1 x PBS (made from 37%; cat # F8775, Sigma Aldrich, St. Louis, MO) was added to the cells and placed at 37^0 C for 7 minutes. Glycine was then added at a final concentration of 125mM and incubated for 5 minutes. From this point forward all samples were kept on ice or at 4^0 Celsius (C). Cells in lysis buffer were sonicated using the Bioruptor (Diagenode Inc., Denville, NJ) for 30 cycles of 30 seconds on/ 30 seconds off. 10% of the total sonicated product was placed at -80^0 C to be used later as non-immunoprecipitated "input" DNA. For each of 6 replicates, 20μg of sheared, DNA-protein complex (in Chip dilution buffer) was immunoprecipitated using 2μg of a mouse monoclonal antibody to ERα (D-12) X (cat #

sc-8005 X; Santa Cruz, Dallas, TX). DNA-protein complexes bound to the antibody were incubated for 4.5 hours with a 50/50 mixture of magnetic beads covalently coupled with recombinant Protein A and Protein G (Dynabeads Protein A, cat # 10002D; Dynabeads Protein G, cat # 10004D, ThermoFisher Scientific) blocked with bovine serum albumin (BSA) and yeast RNA (cat # AM7118, ThermoFisher Scientific). Bound beads were rinsed with several buffers in the following order: low salt buffer, high salt buffer, lithium chloride buffer and Tris-EDTA (TE) buffer. The contents of each buffer are provided below. Remaining, bound DNA/protein complexes were eluted from the beads by adding elution buffer and heating the sample at 55^0 C for 15 minutes. To reverse the crosslinking we added RNase A (50μg, cat # 10109169001, Roche, Indianapolis, IN) for 2 hours at 37^0 C and then Proteinase K (approx. 35μg, cat # 03115828001, Roche, Indianapolis, IN) overnight (12+ hours) at 65^0 C. The DNA was purified using a standard phenol chloroform extraction and ethanol precipitation with glycogen (40μg; cat # 77534, Affymetrix, Santa Clara, CA). Quantitative PCR was performed using SYBR green master mix (cat # 330523, Qiagen Inc., Valencia, CA). Primers used for this experiment are listed in Supplementary Table 2-1. Percent input was calculated as described in the following document: http://www.thermofisher.com/us/en/home/life-science/epigenetics-noncoding-rna-research/chromatin-remodeling/chromatin-immunoprecipitation-chip/chip-analysis.html

ChIP solutions:

SDS Lysis Buffer (10mls)
50mM Tris HCl -- from 1M stock = 500ul
10mM EDTA -- from 0.5M stock = 200ul
1% SDS --- from 20% stock = 500ul
1mM PMSF – from 100mM stock isopropranol = 100ul
Bring up to volume in RNase-free water

+ 1 Roche Proteinase Inhibitor (PI) Complete tablet

ChIP Dilution Buffer (10mls)
16.7mM Tris-HCl -- from 1M stock = 167ul
1.2mM EDTA – from 0.5M stock = 24ul
167mM NaCl – from 5M stock =334ul
1.1% TritonX – from 100% = 110ul
0.01% SDS – from 20% stock = 5ul
Bring up to volume in RNase-free water
+1 Roche PI Complete tablet

High-Salt Buffer (10mls)
20mM Tris-HCl – from 1M stock = 200ul
2mM EDTA – from 0.5M stock = 40ul
1% Triton X = 100ul
0.1% SDS – from 20% stock = 50ul
500mM NaCl – from 5M stock = 1ml
Bring up to volume in RNase-free water
+1 Roche PI Complete tablet

Low-Salt Buffer (10mls)
20mM Tris-HCl – from 1M stock =200ul
2mM EDTA – from 0.5M stock =40ul
1% TritonX = 100ul
0.1% SDS – from 20% stock = 50ul
150mM NaCl – from 5M stock =300ul
Bring up to volume in RNase-free water
+1 Roche PI Complete tablet

LiCl Buffer (10mls)
10mM Tris HCl –from 1M stock = 100ul
1mM EDTA – from 0.5M stock = 20ul
1% Sodium Deoxycholate –from 10% stock = 1ml
250mM LiCl – from 1M stock =2.5ml
1% NP40 – from 10% stock = 1ml
Bring up to volume in RNase free water
+1 Roche PI Complete tablet

TE buffer (10mls)
10mM Tris HCl – from 1M stock = 100ul
1mM EDTA –from 0.5M stock = 20ul
Bring up to volume in RNase free water
+1 Roche PI Complete tablet

Elution Buffer (10mls)
50mM Tris HCl – from 1M stock = 500ul
10mM EDTA – from 0.5M stock =200ul
1% SDS – from 20% stock =500ul
Bring up to volume in RNase free water
(no PI—keep at room temp)

Table 2-1. The relationship between *ADCYAP1R1* expression and rs2267735 genotype, estradiol, and PTSD symptoms in whole blood mRNA from study participants

A.	*All genotypes:*	N	High *ADCYAP1R1* Expression	Low *ADCYAP1R1* Expression
	Low Serum Estradiol	48	39.6% (N=19)	60.4% (N=29)
	High Serum Estradiol	47	61.7% (N=29)	38.3% (N=18)
		Pearson Chi-square, p=.031		

B.	rs2267735 genotype:	N	Mean *ADCYAP1R1* dCT	Std. Deviation
	GG or GC	14	20.86	1.61
	CC	10	22.02	0.86
		2 tailed t-test, p=.049		

C.	*CC genotypes only:*	N	High *ADCYAP1R1* Expression	Low *ADCYAP1R1* Expression
	Low Serum Estradiol	23	30.4% (N=7)	69.6% (N=16)
	High Serum Estradiol	19	63.2% (N=12)	36.8% (N=7)
		Pearson Chi-square, p=.034		

D.		N	Mean PTSD symptoms	Std. Deviation
	High *ADCYAP1R1* expression	48	7.97	9.22
	Low ADCYAP1R1 expression	47	12.38	9.77
		2 tailed t-test, p=.026		

Figure 2-1. Ovariectomized female mice were given a subcutaneous pellet containing either estradiol or vehicle (sesame oil). The mice where then kept in their home cage in the vivarium (hc) or exposed to fear conditioning (fc) using a tone paired with foot shock. N=7 for each treatment group. Fold change in normalized expression was measured relative to normalized expression in vehicle plus home caged mice. The data shown are the average fold change ± standard deviation (SD) in expression of *Adcyap1r1* in the BNST per treatment group.

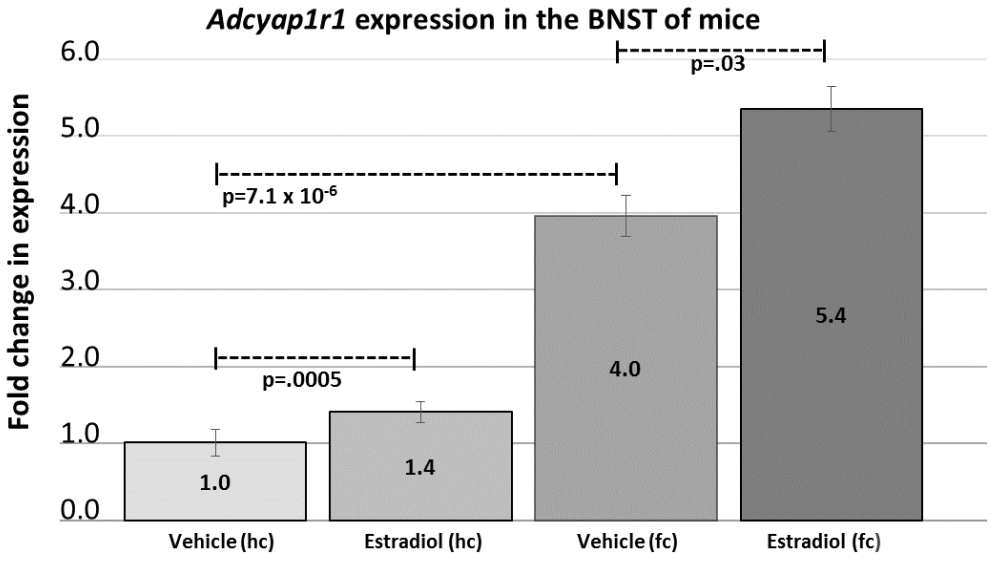


Figure 2-2. HEK293 cells were transiently transfected with full-length human estrogen receptor alpha (hERα) or GFP and treated with either estradiol (E2) or vehicle (ethanol) only. N=6 for each treatment group. Fold change in normalized expression is measured relative to normalized expression in cells transfected with GFP and treated with vehicle only. The data shown are the average fold change ± SD in expression of *ADCYAP1R1* per treatment group.

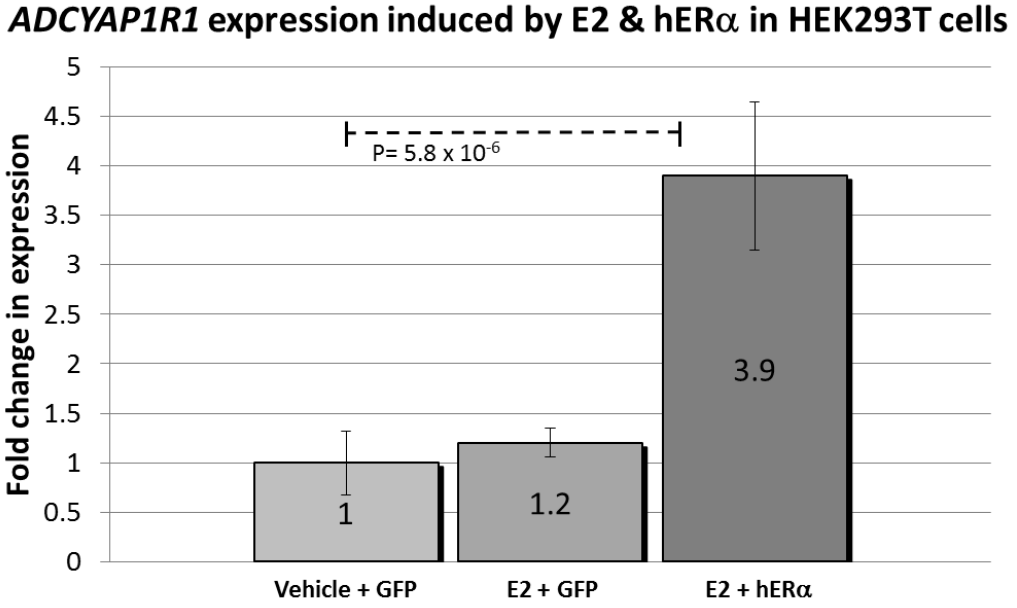


Figure 2-3. A.) Using ChIP followed by qPCR, we measured binding of ERα to two regions of the genome: an ERE, which contains rs2267735 in an intron of *ADCYAP1R1* (ERE region), and a transcriptionally inactive region on chromosome 4 (negative control region). N=6 for each group. The qPCR measures obtained from the immunoprecipitated chromatin were divided by the measure obtained from the non-immunoprecipitated input sample (the amount of chromatin used in the ChIP experiment) using the same primers. The data represent the average percent of input ± SD for the two regions. **B.**) A competitive ELISA was used to measure the binding of ERa to double stranded DNA sequences (oligos) relative to the canonical ERE binding sequence. The fluorescent measures obtained for the competing oligos were transformed by dividing these values by the fluorescent measure obtained for the canonical oligo (positive control). The data represent the averaged inverse of these values ± SD for each experimental oligo (C allele and G alelle) and the non-canonical (negative control).

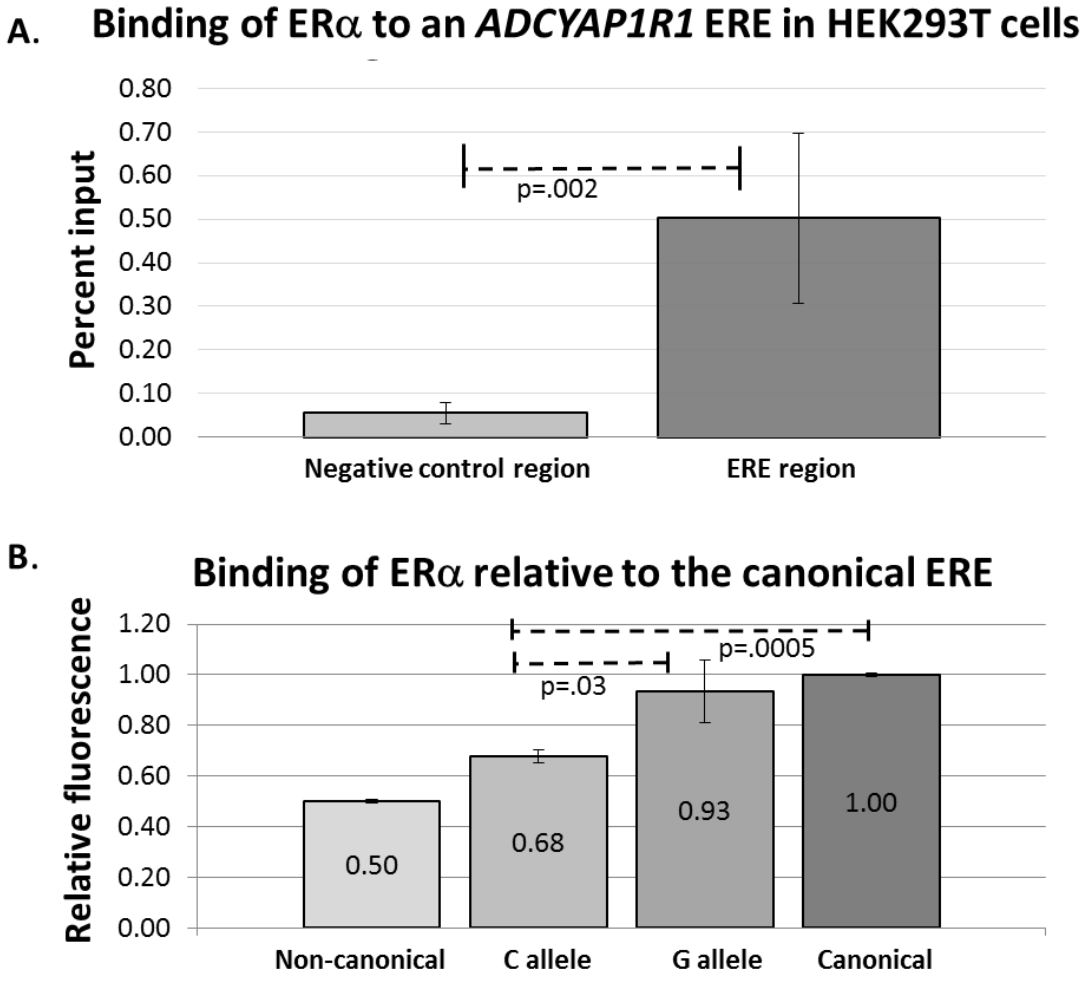
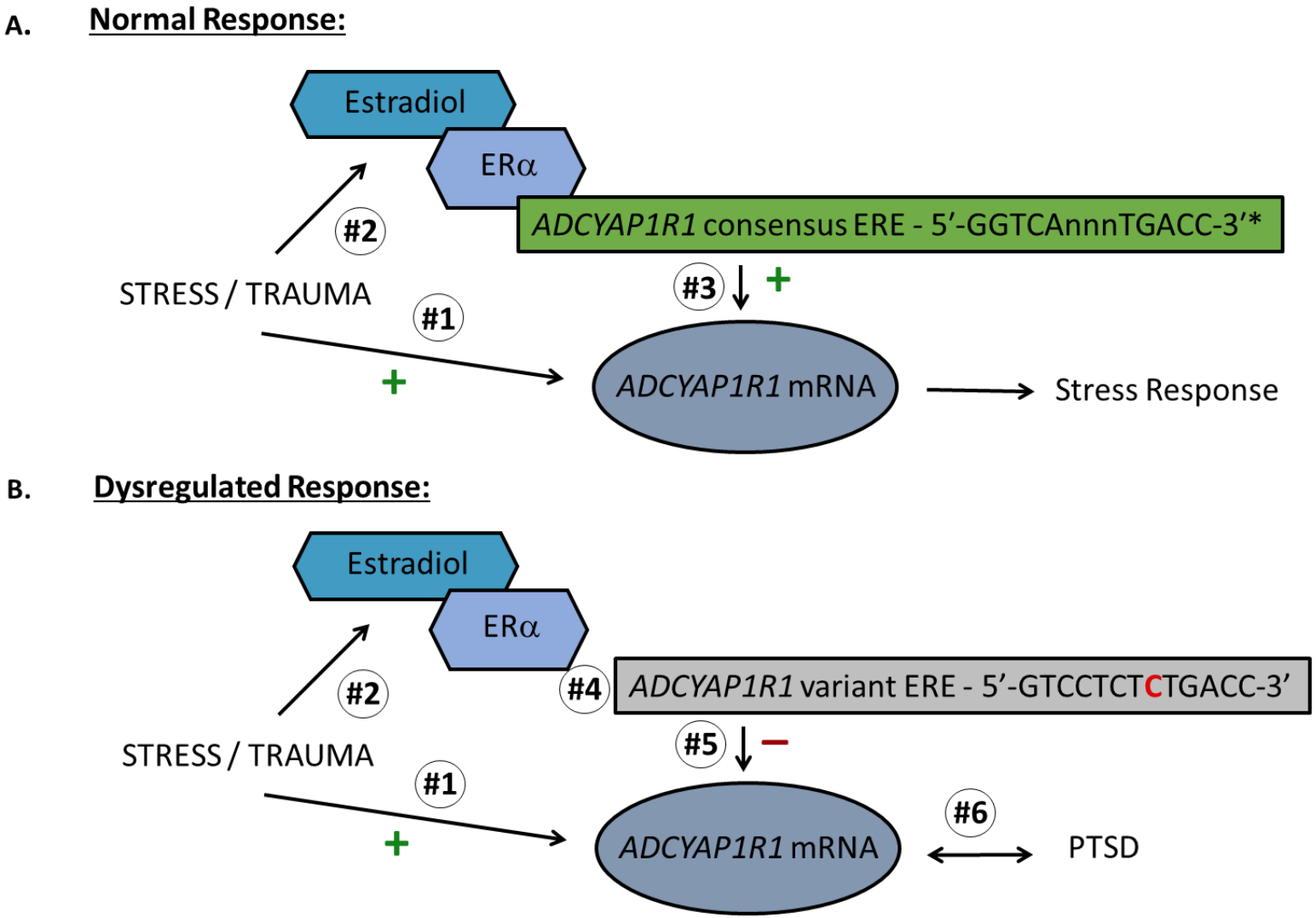


Figure 2-4. A.) A schematic of the hypothesized model for the role of stress, estradiol, ERα, and ERE on increased expression of *ADCYAP1R1*, which is observed in a normal stress response. **A.)** A schematic of the hypothesized model for the role of stress, estradiol, ERα, and ERE on reduced expression of *ADCYAP1R1*, which is observed in a dysregulated stress response.



Supplementary Table 2-1. Unless stated otherwise, all primers and double stranded oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA)

Double-stranded oligonucleotides for the competitive ELISA

1) Primer for "C" allele: 5' CTGAGCCACTCGTCCTCT**C**TGACCTGATGCTCCTG 3' (35nt)
2) Primer for "G" allele: 5' CTGAGCCACTCGTCCTCT**G**TGACCTGATGCTCCTG 3' (35nt)

3) Primer for wild type /canonical sequence (positive control):
5' GTCCAAAGTCAGGTCACAGTGACCTGATCAAAGTT 3' (35nt)

4) Primer for mutant sequence/non-canonical sequence (negative control):
5' GTCCAAAGTCACCGCACAGTGAAATGATCAAAGTT 3'(35nt)

Primers for chromatin immunoprecipitation

1) Forward primer= rs2267735_ChIP_FWD : 5' TGGCACTGAGCCACTCG 3'
2) Reverse primer = rs2267735_ChIP_REV : 5' CACTCATCCTACCAGGAGCATC 3'

3) Negative control primer pair = Human Negative Control Primer Set 2 (cat# 71002; Active Motif, Carlsbad, CA): proprietary sequence, relative location chr4: 188868000-188868500

Sequences and final concentrations of primers, probes and antiprobes used for rtPCR and qPCR of whole blood mRNA from study participant samples*

ADCYAP1R1

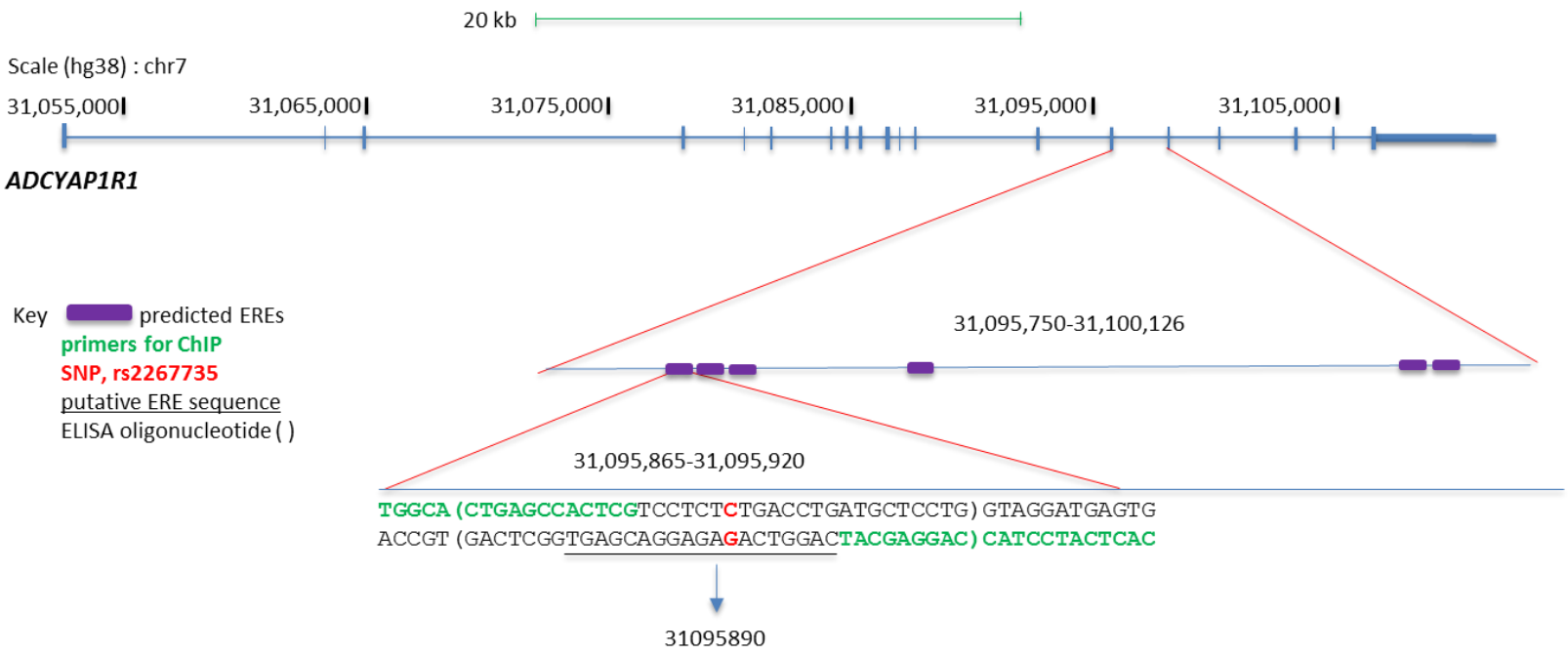
1) rtPCR primer: 5' GCTCTTGCTCAGGATGGAG 3'
2) Forward primer: 5' CTTTGTGGTGGCTGTTCTCT 3'; final concentration = 250 nM
3) Reverse primer: 5' AAGTCCACAGCGAAGTAACG 3'; final concentration = 250 nM
4) iDDS probe: 5' HEX-AAGCGAAAATGGCGAAGCTG-Phos 3'; final concentration = 200 nM
5) iDDS antiprobe: 5' CAGCTT**T**GCCATTTTCGCTT-BHQ-1 3'; final concentration = 400 nM

GAPDH

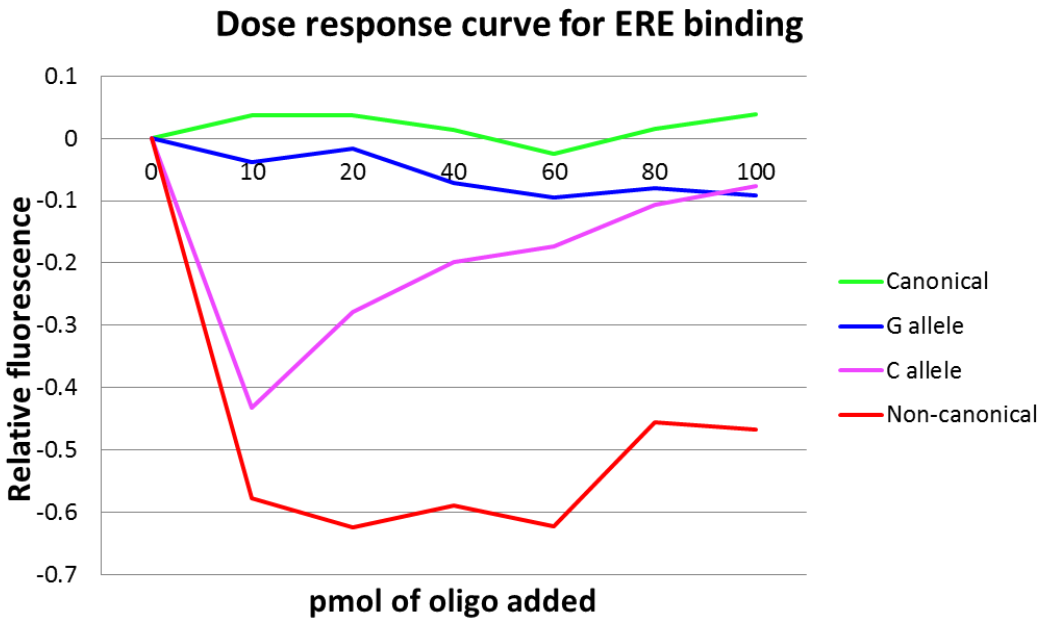
1) rtPCR primer: 5' ACAAGCTTCCCGTTCTCAG 3'
2) Forward primer: 5' TCAACGGATTTGGTCGTATT 3'; final concentration = 200 nM
3) Reverse primer: 5' ACAAGCTTCCCGTTCTCAG 3'; final concentration = 200 nM
4) iDDS probe: 5' FAM-TTGCCATCAATGACCCCTTC-Phos 3'; final concentration = 200 nM
5) iDDS antiprobe: 5' GAAGGGGT**A**ATTGATGGCAA-BHQ-1 3'; final concentration = 400 nM

The nucleotides in red indicate intentional mismatches in the antiprobes.
BHQ-1, black hole quencher 1; Phos, 3' phosphate group to block extension
*Primers and probes from GeneTAG Technology, Atlanta, GA

Supplementary Figure 2-1. An intronic region of *ADCYAP1R1* contains several predicted estrogen response elements (purple boxes), one of which (underlined) contains a SNP, rs2267735 (in red; chr7: 31095890) that correlates with *ADCYAP1R1* gene expression in the cortex of the brain (previously published by Ressler et al., 2011).



Supplementary Figure 2-2. To produce the dose response curve from competitive ELISA data, binding efficiencies were measured using increasing amounts of oligo added (by pmol). The non-canonical oligo shows the least amount of binding. Binding of the canonical and G allele show similar binding patterns. For the C allele, binding is low but reaches close to wild-type levels at 100pmol.



References

1. Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB: Posttraumatic stress disorder in the National Comorbidity Survey. Arch Gen Psychiatry 1995 Dec;52:1048–1060.

2. Breslau N, Kessler RC, Chilcoat HD, Schultz LR, Davis GC, Andreski P: Trauma and posttraumatic stress disorder in the community: The 1996 detroit area survey of trauma. Arch Gen Psychiatry 1998 Jul 1;55:626–632.

3. Glover EM, Mercer KB, Norrholm SD, Davis M, Duncan E, Bradley B, et al.: Inhibition of fear is differentially associated with cycling estrogen levels in women. J Psychiatry Neurosci JPN 2013 Sep;38:341–348.

4. Glover EM, Jovanovic T, Norrholm SD: Estrogen and Extinction of Fear Memories:Implications for Posttraumatic Stress Disorder Treatment. Biol Psychiatry 2015 Feb 16; DOI: 10.1016/j.biopsych.2015.02.007

5. Glover EM, Jovanovic T, Mercer KB, Kerley K, Bradley B, Ressler KJ, et al.: Estrogen levels are associated with extinction deficits in women with posttraumatic stress disorder. Biol Psychiatry 2012 Jul 1;72:19–24.

6. Shvil E, Sullivan GM, Schafer S, Markowitz JC, Campeas M, Wager TD, et al.: Sex differences in extinction recall in posttraumatic stress disorder: a pilot fMRI study. Neurobiol Learn Mem 2014 Sep;113:101–108.

7. Albert K, Pruessner J, Newhouse P: Estradiol levels modulate brain activity and negative responses to psychosocial stress across the menstrual cycle. Psychoneuroendocrinology 2015 Sep;59:14–24.

8. Cowley SM, Hoare S, Mosselman S, Parker MG: Estrogen Receptors α and β Form Heterodimers on DNA. J Biol Chem 1997 Aug 8;272:19858–19862.

9. Klein-Hitpass L, Tsai SY, Greene GL, Clark JH, Tsai MJ, O'Malley BW: Specific binding of estrogen receptor to the estrogen response element. Mol Cell Biol 1989 Jan 1;9:43–49.

10. Heldring N, Pike A, Andersson S, Matthews J, Cheng G, Hartman J, et al.: Estrogen Receptors: How Do They Signal and What Are Their Targets. Physiol Rev 2007 Jul 1;87:905–931.

11. J Gorski, Gannon F: Current Models of Steroid Hormone Action: A Critique. Annu Rev Physiol 1976;38:425–450.

12. Klinge CM: Estrogen receptor interaction with estrogen response elements. Nucleic Acids Res 2001 Jul 15;29:2905–2919.

13. Ressler KJ, Mercer KB, Bradley B, Jovanovic T, Mahan A, Kerley K, et al.: Post-traumatic stress disorder is associated with PACAP and the PAC1 receptor. Nature 2011 Feb 24;470:492–497.

14. Almli LM, Mercer KB, Kerley K, Feng H, Bradley B, Conneely KN, et al.: ADCYAP1R1 genotype associates with post-traumatic stress symptoms in highly traumatized African-American females. Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet 2013 Apr;162B:262–272.

15. Stroth N, Liu Y, Aguilera G, Eiden LE: Pituitary adenylate cyclase-activating polypeptide controls stimulus-transcription coupling in the hypothalamic-pituitary-adrenal axis to mediate sustained hormone secretion during stress. J Neuroendocrinol 2011 Oct;23:944–955.

16. Tsukiyama N, Saida Y, Kakuda M, Shintani N, Hayata A, Morita Y, et al.: PACAP centrally mediates emotional stress-induced corticosterone responses in mice. Stress Amst Neth 2011 Jul;14:368–375.

17. Stroth N, Eiden LE: Stress hormone synthesis in mouse hypothalamus and adrenal gland triggered by restraint is dependent on PACAP signaling. Neuroscience 2010 Feb 17;165:1025.

18. Rothbaum BO, Kearns MC, Reiser E, Davis JS, Kerley KA, Rothbaum AO, et al.: Early intervention following trauma may mitigate genetic risk for PTSD in civilians: a pilot prospective emergency department study. J Clin Psychiatry 2014 Dec;75:1380–1387.

19. White S, Acierno R, Ruggiero KJ, Koenen KC, Kilpatrick DG, Galea S, et al.: Association of CRHR1 variants and posttraumatic stress symptoms in hurricane exposed adults. J Anxiety Disord 2013 Oct;27:678–683.

20. Boscarino JA, Erlich PM, Hoffman SN, Zhang X: Higher FKBP5, COMT, CHRNA5, and CRHR1 allele burdens are associated with PTSD and interact with trauma exposure: implications for neuropsychiatric research and treatment. Neuropsychiatr Dis Treat 2012;8:131–139.

21. Amstadter AB, Nugent NR, Yang B-Z, Miller A, Siburian R, Moorjani P, et al.: Corticotrophin-releasing hormone type 1 receptor gene (CRHR1) variants predict posttraumatic stress disorder onset and course in pediatric injury patients. Dis Markers 2011;30:89–99.

22. Binder EB, Bradley RG, Liu W, Epstein MP, Deveau TC, Mercer KB, et al.: Association of FKBP5 polymorphisms and childhood abuse with risk of posttraumatic stress disorder symptoms in adults. JAMA J Am Med Assoc 2008 Mar 19;299:1291–1305.

23. Mehta D, Binder EB: Gene × environment vulnerability factors for PTSD: The HPA-axis. Neuropharmacology 2012 Feb;62:654–662.

24. Yehuda R, Cai G, Golier JA, Sarapas C, Galea S, Ising M, et al.: Gene expression patterns associated with posttraumatic stress disorder following exposure to the World Trade Center attacks. Biol Psychiatry 2009 Oct 1;66:708–711.

25. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods San Diego Calif 2001 Dec;25:402–408.

26. Reese JC, Katzenellenbogen BS: Differential DNA-binding abilities of estrogen receptor occupied with two classes of antiestrogens: studies using human estrogen receptor overexpressed in mammalian cells. Nucleic Acids Res 1991 Dec 11;19:6595–6602.

27. Petz LN, Nardulli AM: Sp1 binding sites and an estrogen response element half-site are involved in regulation of the human progesterone receptor A promoter. Mol Endocrinol Baltim Md 2000 Jul;14:972–985.

28. Foa EB, Tolin DF: Comparison of the PTSD Symptom Scale-Interview Version and the Clinician-Administered PTSD scale. J Trauma Stress 2000 Apr;13:181–191.

29. Murray JL, Hu P, Shafer DA: Seven novel probe systems for real-time PCR provide absolute single-base discrimination, higher signaling, and generic components. J Mol Diagn JMD 2014 Nov;16:627–638.

30. Storts DR: Alternative probe-based detection systems in quantitative PCR. J Mol Diagn JMD 2014 Nov;16:612–614.

31. Lezak KR, Roman CW, Braas KM, Schutz KC, Falls WA, Schulkin J, et al.: Regulation of bed nucleus of the stria terminalis PACAP expression by stress and corticosterone. J Mol Neurosci MN 2014 Nov;54:477–484.

32. Hammack SE, Roman CW, Lezak KR, Kocho-Shellenberg M, Grimmig B, Falls WA, et al.: Roles for pituitary adenylate cyclase-activating peptide (PACAP) expression and signaling in the bed nucleus of the stria terminalis (BNST) in mediating the behavioral consequences of chronic stress. J Mol Neurosci MN 2010 Nov;42:327–340.

33. Lezak KR, Roelke E, Harris OM, Choi I, Edwards S, Gick N, et al.: Pituitary adenylate cyclase-activating polypeptide (PACAP) in the bed nucleus of the stria terminalis (BNST) increases corticosterone in male and female rats. Psychoneuroendocrinology 2014 Jul;45:11–20.

34. Österlund MK, Keller E, Hurd YL: The human forebrain has discrete estrogen receptor α messenger RNA expression: high levels in the amygdaloid complex. Neuroscience 1999 Dec;95:333–342.

35. Östlund H, Keller E, Hurd YL: Estrogen Receptor Gene Expression in Relation to Neuropsychiatric Disorders. Ann N Y Acad Sci 2003 Dec 1;1007:54–63.

36. Kumar V, Green S, Stack G, Berry M, Jin J-R, Chambon P: Functional domains of the human estrogen receptor. Cell 1987 Dec 24;51:941–951.

37. Hammack SE, May V: Pituitary Adenylate Cyclase Activating Polypeptide in Stress-Related Disorders: Data Convergence from Animal and Human Studies. Biol Psychiatry 2014 Dec 9; DOI: 10.1016/j.biopsych.2014.12.003

38. Wegerer M, Kerschbaum H, Blechert J, Wilhelm FH: Low levels of estradiol are associated with elevated conditioned responding during fear extinction and with intrusive memories in daily life. Neurobiol Learn Mem 2014 Dec;116:145–154.

39. Ferree NK, Wheeler M, Cahill L: The influence of emergency contraception on post-traumatic stress symptoms following sexual assault. J Forensic Nurs 2012 Sep;8:122–130.

Chapter 3: Beyond SNPs: the discovery of CNVs associated with PTSD

Introduction

Posttraumatic stress disorder (PTSD) is a complex disease influenced by multiple genetic and environmental factors that each contribute to some fraction of the variance in PTSD phenotype following trauma [1–3]. Because there is no definitive cure for PTSD and the environmental exposures that lead to the disorder are often unavoidable, there is a great deal of interest in finding genetic risk factors that can help us identify better options for treatment or prevention. To date, geneticists who study PTSD have largely focused on changes at the individual nucleotide level by genotyping single nucleotide polymorphism (SNPs) throughout the genome. While this research has led to insightful discovery, these genetic variants confer only a marginal increase in risk for developing PTSD [4]. It is clear that the identification of genetic variation responsible for PTSD risk remains largely incomplete. Investigation of other genetic variant types, such as copy number variation (CNV) is a promising, second approach to uncovering genetic risk factors for PTSD. CNVs, which are large ($\geq$1 kb) genomic deletions or duplications, may be more likely to disrupt normal gene activity than individual SNPs, conferring higher penetrance and greater genetic effect on disease. In fact, compared to SNPs, there is considerable evidence for a more pronounced effect on disease outcome for CNVs associated with genetic susceptibility to schizophrenia [5].

As part of the Grady Trauma Project (GTP), our research team has collected extensive data, including trauma exposure, psychiatric symptoms, and genotypes on

4,607 African Americans (AAs), 30% of whom currently suffer from PTSD. Genome wide genotype data already generated from this study for a genome-wide association study (GWAS) can be further utilized to identify CNVs in this population. In collaboration with the Army STARRS project, we have also obtained GWAS data from 10,097 self-identified white, male veterans with or without combat exposure and a PTSD prevalence of approximately 20%. Data already available for large PTSD cohorts can be leveraged to investigate the association between CNVs and PTSD – an area of genetic inquiry that to date remains in its infancy in PTSD genetics research.

Identification of genetic variants that result in a heritable predisposition to PTSD may reveal genes involved in PTSD neurobiology, define molecular pathways involved in disease progression and offer potential options for prevention or cure. Our preliminary analysis reveals a large, 750 kb duplication at 11q22.1 that associates with PTSD in the GTP cohort ($p = 0.052$; permuted to correct for multiple testing). Within the boundaries of this copy number gain is the contactin-5 gene, *CNTN5*, which plays a role in neuropsychiatric and brain-related disorders as well as synaptic plasticity and neurodevelopment [6–9]. In an association test using the Army STARRS cohort, we have also identified a large, variably sized deletion (89 kb – 1.5 Mb) of chromosome 7 at band q31.1 that associates with PTSD. This CNV falls within a region identified as an autism spectrum locus [10] and contains the *IMMP2L* gene, which has been implicated in dyslexia [11], ADHD [12], and Tourette syndrome [13]. Further analysis and replication needs to be performed in order to confirm these associations. However, as discussed in this chapter, the relevant, biological function of the genes within the associated CNVs is encouraging.

Materials and Methods

Grady Trauma Project (GTP) cohort

GTP study participants were recruited from waiting rooms of either a primary care or OB/GYN clinic at Grady Memorial Hospital in Atlanta, GA. Patients waiting to be seen by a physician or individuals accompanying the patient were asked if they would like to participate in a study to identify genetic factors that increase a person's susceptibility to PTSD. Willing participants provided a saliva sample for DNA collection and completed a series of questionnaires to obtain demographics, and measures of trauma and PTSD. The majority of the study participants are self-identified African American (93%) and female (78%). The demographics of this study population have been described in detail previously be Gillespie et al. [14].

DNA from the GTP study participants was collected from saliva using an Oragene DNA Collection Kit (DNA Genotek, Inc., Ontario, Canada) and extracted using the Agencourt DNAdvance Nucleic Acid Isolation Kit (Beckman Coulter, Inc., Brea, CA). Whole blood was collected in a Vacutainer collection tube containing EDTA (Becton, Dickinson and Co., Franklin Lakes, NJ) from approximately 20% of the study population who agreed to participate further in the study. DNA from the whole blood samples was extracted using the E.Z.N.A. Mag-Bind Blood DNA Kit (Omega Bio-Tek, Inc., Norcross, GA).

To measure current PTSD symptoms among the study participants, we used the PTSD Symptom Scale (PSS) [15,16]. The PSS is a 17-item, self-report questionnaire that

is given to each study participant to measure the weekly frequency (0-5+ times per week) that a participant experiences symptoms of PTSD such as: re-experiencing the trauma; avoiding certain activities; and/or hyperarousal. The frequency scores on the PSS are totaled and used to obtain a continuous measure that positively correlates with severity of PTSD-like symptoms over the prior two weeks (alpha coefficient =0.93) [14]. A diagnosis of PTSD is determined based on positive responses to specific questions on the PSS that address the required symptom criteria (DSM-IV) for PTSD. PTSD diagnosis was used as categorical measure for current PTSD in the GTP by CNV analyses.

Study participants were also asked to self-report exposure to trauma using the Traumatic Events Inventory (TEI) [17]. The TEI is a 14-item instrument used to assess the type and frequency of traumatic events an individual has either witnessed or experienced during his or her lifetime. For analyses performed in this study, we utilized a measure of total types of trauma experienced or witnessed, derived from the TEI, to identify individuals who report having no trauma (variable measure = 0). Given that trauma exposure is necessary for the development of PTSD, it is possible that those who are not diagnosed with PTSD and do not report trauma, could in fact develop PTSD if exposed to a qualifying traumatic event. Thus, without taking trauma into account, these individuals are inappropriately categorized as controls. Therefore, in our assessment of risk for PTSD, these study participants are excluded from case control analyses. We also exclude individuals that do not have a measure for TEI (missing data) as we cannot be certain whether or not they experienced any trauma.

Genotyping and sample pruning of GTP cohort

To identify genetic risk for PTSD, we have collected genotypes from 4,607 unrelated African American GTP study participants across one million, single nucleotide polymorphisms (SNPs) throughout the genome. Participant samples that met each of two criteria were processed on the Illumina OmniQuad 1M BeadChip (Illumina Inc., San Diego, CA): 1) had completed measures for phenotypes of interest including PTSD, and 2) had a quantity of DNA ($\geq$400ng) that was sufficient to perform the genotyping experiments. There have been two phases of our genotype association analyses; one that included the first approximately 4,500 samples genotyped and the second that included roughly an additional 2,000 samples. In the first phase of experimentation, a total of 4,342 samples, excluding technical replicates and controls, were successfully genotyped. As a measure of quality control, we removed samples with a call rate of less than 99% (N=384), gender mismatches (N=13), and discordant replicates (N=1) [18]. To avoid false positive genetic associations due the inclusion of related individuals, we also excluded 1st and 2nd degree relatives (proportion identity by descent $\geq$ 0.125) (N=761). We also controlled for the effects of race in our CNV analysis by conducting principal component analysis to infer axes of ancestry and identify a sample with minimized ancestral differences (3 standard deviations from the medians of the 1st and 2nd principal components) [19]. In this step, an additional 197 samples were removed. Of the remaining 2,986 unrelated African Americans (AA), 131 were missing phenotypes for PTSD and 407 had no trauma history (or were missing data for this variable). The final sample size used for CNV calling was 2,448.

Identification of copy number variants (CNVs) in GTP cohort

Microarray intensity data generated from the Illumina OmniQuad 1M BeadChip was uploaded into Genome Studio (software; Illumina Inc., San Diego, CA) to generate Log R ratio (LRR) and B-allele frequency (BAF) values for each of the one million SNP loci. These values were then used to identify CNVs in our study population [20,21] by processing the data through three different CNV calling algorithms: PennCNV [22] (see Appendix 3-A for details and scripts), R-Gada (parameters: T=8, α=0.8, probe minimum 10) [23], and cnvPartition v3.2.0 (parameters: default except for GC Wave Adjust=True, Minimum Probe Count=10). Calls made by all algorithms were compared for each sample. To minimize false positives, a CNV call was considered valid if the following conditions held true: 1) a loss or gain consisted of a minimum of 10 consecutive probed loci, 2) the CNV was greater than 10 kb in size, and 3) a call was made by at least two of the 3 calling programs. As a final measure of quality control, 52 samples were removed due to an excess of CNV calls (> mean + 3*SD) resulting in a sample size of 2396 that were analyzed for CNV by PTSD association and burden.

Burden analysis and association tests

Association and burden analysis of PTSD risk due to CNVs were performed using PLINK, an open-source C/C++ genome wide association study (GWAS) tool set that allows for computationally efficient analysis of large datasets (http://pngu.mgh.harvard.edu/purcell/plink/) [24]. Deletion and duplication CNVs were analyzed separately based on size (>50 kb, >100 kb, >500 kb, and >1 Mb) and frequency

(all CNVs, CNV in 1% of samples, CNVs in 0.5% of samples). CNVs called in centromere regions were removed prior to analysis (see Appendix 3-B for genomic regions excluded). To determine the burden of CNVs in cases versus controls a burden analysis was performed in PLINK using a 1-sided test with 10,000 permutations (http://pngu.mgh.harvard.edu/~purcell/plink/cnv.shtml#burden). Tests were run to identify significant differences for the following statistics: 1) total number of CNV deletions or duplications greater than or equal to a defined size divided by the number of cases or controls (RATE), 2) number of cases or controls that have at least one deletion or duplication greater or equal in size to that defined for the CNV (PROP), 3) total length spanned (in kb) for CNV deletions or duplications in cases or controls for CNVs of defined size and type (KBTOT), and 4) average size of CNV per person in cases or controls among CNVs of defined size and type (KBAVG). To test for CNV by PTSD association, we also used PLINK to perform a 1 or 2-sided test along with 10,000 permutations to account for multiple testing (http://pngu.mgh.harvard.edu/~purcell/plink/cnv.shtml#assoc) (Appendix 3-C). A multiple test corrected p-value of ≤ 0.05 was considered statistically significant.

Characterization of an 11q22.1 CNV associated with PTSD

Visual Confirmation of CNVs

To confirm the CNV gain by visual inspection, we graphed the Log R ratio and B allele frequency values along the duplicated region of 11q22.1 for each of the CNV carriers using R statistical software.

Quantitative PCR

Quantitative real-time PCR (qPCR) was used to determine copy number in the region of the 750 kb duplication (hg19; chr11:99889749-100624538, Figure 3-1) for five study participant carriers. This was done using primers to three regions within the CNV duplication breakpoints (left, middle and right). qPCR using the three primer pairs was performed using 20ng of DNA, primers, and the Type-it CNV SYBR Green PCR +qC Kit (cat #206672; Qiagen Inc., Carol Stream, IL) according to the kit protocol. The sequences of the primers used for each of the three regions in the CNV are provided below. The primer pairs were designed to amplify the following regions on chromosome 11: left primers (hg19; chr11:99889768-99889857), middle primers (hg19; chr11:100224996-100225101), right primers (hg19; chr11:100624426-100624508). Primers provided with the kit were used as normal copy controls (proprietary sequence). All reactions were performed in duplicate using the ViiA 7 real time PCR instrument (Thermo Fisher Scientific). To determine fold-change in copy number we used the $2^{-\Delta\Delta Ct}$ method [25]. The Ct value obtained from qPCR using the control primers was subtracted from the Ct derived from the CNV region primer pairs. The ΔΔCt was calculated by subtracting the ΔCt values of a non-CNV carrier study participant from that of a CNV-carrier sample. Fold change in copy number was then determined using the formula: $2^{\wedge-\Delta\Delta Ct}$.

To determine PCR efficiency of each of the three CNV region primer pairs, we performed additional reactions in triplicate using serial dilutions of DNA with the

following amounts: 0.625 ng, 1.25 ng, 2.5 ng, and 5 ng. The value of the slope derived by graphing the log of DNA input by Ct value was used to calculate qPCR efficiency using the calculations available at: (https://www.thermofisher.com/us/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/qpcr-efficiency-calculator.html).

Primers:

Left FWD: 5'-TCTGGAAACAAGTCCTTTATCA-3'
Left REV: 5'-GAGACACACTAGTAAGGCAATAA-3'
Middle FWD: 5'-AGACTTCTGGCTTGAAGAACC-3'
Middle REV: 5'-GATCCAAGGTTGATTGGTCTGA-3'
Right FWD: 5'-CTTCTTGCTGCTGTTCCTACTT-3'
Right REV: 5'-AGAGGGAGGCCTTAATCCTATC-3'

A second method was also employed to determine copy number within the CNV region. qPCR was performed using a Taqman copy number assay designed to determine copy number at chr11:100176904 (Hs05250804_cn; Thermo Fisher Scientific). For normalization, a Taqman assay to *RPLPO* was used. All 5 carrier samples were assayed in addition to 2 non-carrier study participant samples and control DNA derived from a Coriell cell line. The reactions were performed in triplicate using 2ul of DNA (20ng), 1ul each of the target and reference 20x assays, 10ul of 2 x Taqman Gene Expression Master Mix (cat # 4369016; Thermo Fisher Scientific), and 6ul of PCR grade dH_2O. The reactions were performed using the ViiA 7 real time instrument and the following PCR conditions: 95°C for 10 minutes and then 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. To determine fold-change in copy number we used the $2^{-\Delta\Delta Ct}$ method [25]. The

Ct values from the *RPLPO* assay were used to calculate ΔCt and the ΔCt value for a normal copy, Coriell control DNA was used to determine ΔΔCt.

Breakpoint Mapping

We used a method known as Sequenom allelotyping to define the duplication breakpoints [26]. Using genome-wide SNP data for each of the 5 subjects with the chromosome 11 CNV gain, we identified several SNPs within the duplication breakpoint region that were heterozygous. These SNPs were genotyped using the Sequenom MassArray (MALDI-TOF) iPlex platform (Sequenom, San Diego, CA) [27]. The mass differential of each allele was determined using measurements obtained through mass spectrometry. A mass difference of >0.2 for polymorphic sites indicated the presence of duplication. DNA from both CNV carrier and non-carrier controls was used in this analysis. We excluded data for SNPs that revealed the presence of duplication for both carrier and non-carrier samples.

Mapping of breakpoint junctions

We hypothesized that the copy number event of interest was the result of non-allelic recombination, resulting in a tandem duplication. To test this hypothesis and determine the orientation of the duplication (tail to head or tail to tail) we used breakpoint junction mapping as described by Arlt et al. [28]. 20ng of genomic DNA from each of the CNV carriers and a non-carrier control was used in a 50-μl PCR with the following, final

concentrations of PCR reagents: 1x reaction buffer, 200 uM dNTPs, 0.5 μM each of forward and reverse primer, 2 mM $MgCl_2$, and 2.5 units of AmpliTaq Gold 360 DNA polymerase (cat # 4398823;Thermo Fisher Scientific). We used the reaction conditions recommended by the manufacturer for the DNA polymerase with an annealing temperature of 59^{o} C and a 5-minute extension. Several reactions were run using 2 sets of primers (A and B; primer sequences below) in the following three combinations: P1 + P1, P1 + P2, and P2 + P2. Given the orientation of the primers (see Figure 3-2), PCR using P1 + P2 primers would result in an amplicon only in the presence of a tail to head, tandem duplication. An amplicon resulting from a reaction including only a P2 primer would indicate a tail-to-tail tandem duplication. None of the primer combinations would result in a PCR product in the absence of a duplication or when using genomic DNA from a non-CNV carrier. Reactions using these conditions were included as negative controls to rule out false positive amplicons resulting from non-specific priming. Gel electrophoresis was used to visualize the presence or absence of a PCR product for each reaction. Because the breakpoints of the CNV are only estimates of the start and end and could potentially extend further 5' and/or 3' between consecutive probes, we designed a second set of primers in these regions (P1_B and P2_B; see Figure 3-3) to increase the likelihood of getting an amplicon in the event that the A primers were too distant to obtain a product by standard PCR.

Primers:

P1_A: 5'-AAAGACAGTCCCAGAGTGGG-3' (chr11: 99,889,815 – 99,889,834)
P1_B: 5'-ACATGGAACCTGTTCTGCACC-3' (chr11: 99,887,627-99,887,647)
P2_A: 5'-TCTGTAGGCTGGATTGCCAG-3' (chr11: 100,624,506-100,624,525)
P2_B: 5'-AGGCTCAACCCTGTTTGACAG-3' (chr11: 100,628,551-100,628,571)

Fusion Protein PCR

Given the two scenarios of a tandem repeat type, tail-to-head or tail-to-tail (Figure 3-2), the tail-to-head event could result in a fusion product between *ARHGAP42* and *CNTN5* and the translation of a truncated CNTN5 protein. The putative 5' and 3' breakpoints of the duplication fall within introns of the two genes. Also, the CNV is predicted to include the promoter region and at least the first exon of *ARHGAP42* (hg19; chr11:100,558,410-100,558,564). Thus, provided that normal splicing occurs and exons are maintained, an in-frame protein containing the first exon of *ARGHAP42* and the last several exons of *CNTN5* is a feasible outcome. Out of the 5 samples with the CNV duplication, we had mRNA for only one participant (PTSD_9646). Whole blood was collected in Tempus Blood RNA tubes (cat # 4342792; Thermo Fisher Scientific) and mRNA was extracted using the 5 PRIME PerfectPure RNA Purification kit. (cat# 2302110, Thermo Fisher Scientific) mRNA was converted to cDNA using the High Capacity RNA-to-cDNA Kit according to kit protocol using 600ng of RNA (RIN = 7.9) (cat # 4387406; Thermo Fisher Scientific). Using the ARHGAP_F primer and each of the 2 CNTN5_R primers (sequences below), we performed standard PCR using AmpliTaq Gold 360 Master Mix (cat # 4398876, Thermo Fisher Scientific), 100ng of cDNA, and 0.5 μM each of forward and reverse primer (final concentrations). We used reaction conditions consistent with those recommended for the DNA polymerase (annealing temperature of 59^{o} C and a 1-minute extension).

Primers:

5'-TTGCAGTGTCACGAGATTGAGC-3' (ARHGAP_F)
5'-CAGTAGCCTTACATTCCCATCG-3' (CNTN5_1_R)
5'-TTCAGTCGAAAAGTGGGAGCTG-3' (CNTN5_2_R)

Analysis of gene expression in brain and blood

2μg of human brain total RNA (cat# AM7962, Thermo Fisher Scientific) was reverse transcribed using the High Capacity RNA-to-cDNA Kit according to kit protocol (cat # 4387406; Thermo Fisher Scientific). 60ng of the resulting cDNA was used in each of 4 reactions to test the expression of *ARHGAP42* (1 Taqman assay: Hs00611831) and *CNTN5* (3 Taqman assays: Hs00544267, Hs00544274, and Hs00205041) in brain tissue. For each assay, the reaction was performed in quadruplicate and included a non-DNA template negative control. A Taqman assay for the housekeeping gene, *RPLPO (*Hs99999902), was used to normalize the data and calculate the ΔCt. The reactions were performed using Taqman Gene expression Master Mix (cat# 4369016, Thermo Fisher Scientific) and the ViiA 7 real time instrument (Thermo Fisher Scientific).

A similar experiment was performed using mRNA extracted from whole blood. For this experiment we used 3 *CNTN5* assays (as mentioned above) and *GAPDH* (Hs02758991) to normalize for the cDNA input. Based on data from a gene expression microarray, which revealed no detectable level of expression, we predicted that *CNTN5* does not typically express in whole blood. The goal of this experiment was to determine if we could detect *CNTN5* in whole blood mRNA as a result of the duplication. The expression experiments were performed on whole blood mRNA for one CNV carrier (PTSD_9646) and 8 non-CNV carriers with similar RNA quality (RIN) and concentrations. mRNA was extracted from whole blood stored in Tempus Blood Collection tubes (as described above). qPCR experiments

were performed using 45 cycles in order to capture any signal using the *CNTN5* assays.

SNP association testing using GWAS data

Using PLINK we analyzed the association of PTSD (using PSS; the continuous measure of PTSD symptoms) and SNPs (on the Illumina OmniQuad 1M BeadChip) present in both *CNTN5* and *ARGHAP42* gene regions (including 50 kb upstream of the transcription start). In the analyses we controlled for ancestry using he first two principal components, trauma history (using a continuous measure obtained from the traumatic experiences inventory (TEI)) [29], and gender. Only SNP associations that met Bonferonni correction for multiple testing were considered statistically significant.

Replication using additional GTP participant samples

Additional GTP samples were genotyped on the Illumina OmniQuad array after the initial finding of the CNV duplication. Excluding the samples analyzed in the discovery cohort (first phase), data for an additional 1,944 unrelated African Americans became available. CNVs in these samples were identified using PennCNV only. CNV calls were assessed for the presence of the 750 kb duplication among the new samples without any filtering.

Replication using other PTSD cohorts

In an attempt to replicate the association of the 11q22.1 duplication and PTSD in additional, large cohorts, we used LRR and BAF data (generated from the OmniExpressExome-8 v1 BeadChip) from two studies: the Marine Resiliency Study (MRS) (N=2585) [30] and the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS) (N=15953) [31]. The array, used by both studies, had sufficient coverage to detect the CNV with 250 probes spanning the region of duplication. PennCNV was used to make CNV calls just as described for GTP.

Army STARRS cohort

The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS) began in 2011 in an effort to better understand the increase in suicide rates among soldiers [31]. Phenotype and whole-genome genotype data have since been collected on nearly 20,000 study participants to date. From our collaborators at the University of Michigan (Murray Stein et al.), we received PTSD and trauma phenotypes as well as GWAS data for a total of 15,895 samples from the Army STARRS dataset. This dataset included a mix of population types including: 1) Non Hispanic White, 2) Non Hispanic Black, 3) Non Hispanic Asian, 4) Hispanic, 5) American Indian, 6) Hispanic White, 7) Hispanic Black, 8) Hispanic Asian 9) Hispanic Other, and 10) Other. Because it had the largest sample size, we decided to focus the primary analysis on Non-Hispanic Whites (N = 10,097).

Raw intensity values from the OmniExpressExome-8 v1 BeadChip were used to generate LRR and BAF values. We then used PennCNV to make CNV calls and PLINK to determine associations with PTSD (as described above). Stringent quality control (QC)

was performed on the PennCNV calls using a cutoff of Q3 + 3*IQR for the LogR Ratio SD, B allele frequency drift, waviness factor, and total number of CNV called. CNVs included in the analysis were also restricted to those called by greater than 10 probes. After this QC, the sample size decreased to 9,166. Next, we retained only individuals that experienced a criterion A trauma (as indicated in the phenotype files as "trauma_exposed_critA" = for the NSS samples and "nondeploy_traum_exposed_critA" = 1 or "deploy_trauma_exposed_critA"=1 for the PPDS samples). After removing non-CritA exposed individuals, the samples size was reduced to 7,190. For CNV by PTSD analysis, we used the dichotomous variable for PTSD (dsm_pts= yes for PPDS and dsm_pts =1 for NSS). Centromere, telomere, T-cell receptor and immunoglobulin regions were removed prior to further analyses (see Appendix 3-B for genomic regions excluded).

Using PLINK (as described earlier) we performed both burden analyses and association tests to identify statistically significant relationships between CNV deletion or duplications and PTSD. We focused our analysis specifically on deletions and duplication $\geq$50 kb (Appendix 3-D, 3-E, and 3-F).

Results

We ran several burden analyses to determine if deletions or duplications were more prevalent among GTP study participants with PTSD than those without PTSD. We did not detect a statistically significant burden for either deletions or duplications, greater or equal to 50 kb, 100 kb, 500 kb, or 1 Mb in size. We also ran statistical tests to identify individual CNVs that associate with PTSD, assessing deletions and duplications of

varying size and frequency separately. Among CNV duplications greater than 500 kb present in 0.5% of the study population, we found an approximately 750-kb duplication of chromosome 11 (11q22.1) that approaches significance in an association analysis with PTSD diagnosis (1-sided test; p= 0.052). This CNV was called in all three calling algorithms for 5 study participants with PTSD (Table 3-1) and no participants without PTSD. For each of the CNVs, we graphed the LRR and BAF values for the region to confirm the CNV by visual inspection (Figure 3-4 A through 3-4 E). LRR values greater than zero and BAF values of 0, 0.33, 0.66, and 1 are indicative of a duplicated region of the genome. The same region was graphed for a normal copy individual to serve as a non-duplication reference (Figure 3-4 F). The duplication coordinates determined graphically (shaded region) provide rough estimates of the CNV breakpoint (hg19; chr11: 99,885,000 – 100,630,000), which are consistent with the breakpoints determined by the calling algorithms (Table 3-1).

To validate the duplication, we performed quantitative PCR (qPCR) and determined the fold difference in copy number for three regions across the CNV among the five carrier samples relative to two non-carrier samples from the GTP cohort (Table 3-2 A and Table 3-2 B). The results are highly variable with fold change in copy number for the same sample differing based on which non-carrier DNA was used in the calculation of copy number. With the exception of PTSD_10456 that had the lowest fold change values, the data are consistent with the presence of a duplication but certainly not conclusive. The discrepancies in the data are due in large part to differences in PCR efficiency, which were observed between samples (likely due to DNA quality). This is not surprising given that the DNA for these samples was derived from saliva, which

typically results in DNA that is often contaminated with DNA from bacteria and food. Of all the samples only PTSD_7722 had PCR efficiencies >92% for each of the three primer pairs.

Using a second qPCR method that employed the use of a commercially available copy number assay, the data are much more convincing. In this experiment, unlike the one previous, we used qPCR data derived from better quality DNA (from a cell line) to calculate ΔCt. The non-carrier GTP samples DNAs were retained in the experiment as controls only. As expected, the fold-change in copy number determined for the non-CNV carriers samples were closer to 1.0 than the fold-change values determined for the CNV carrier samples. Also, for 4 of the 5 CNV carriers, the values are close to 1.5, which is indicative of a duplication (3copies / 2copies = 1.5; Table 3-2 C). Although sample PTSD_10456 had a fold-change of only 1.3, we still believe, based on the graphing the LRR and BAF, that this sample is a CNV carrier. These results are consistent with the presence of a duplication for a single region within the CNV, but provide no evidence regarding the size or extent of the duplicated region.

In an attempt to better define the breakpoints of the CNV we used Sequenom allelotyping. Due to significant variability in the data, we were only able to confirm the duplication for a 368 kb region within the predicted breakpoints. Using the data derived from this method, there were 4 SNPs (rs7126331, rs6590574, rs12174, and rs996858) that gave definitive results for all 5 CNV carrier samples, revealing the presence of a duplication for the following hg19 coordinates on chromosome 11: 99936389 – 100304613. Data derived from the use of other SNPs were either inconclusive or revealed a duplication for both carrier and non-carrier negative control samples; data

derived from the latter were excluded. Due to the limitations of this experiment, rs7126331 (hg19; chr11: 99936389) was the furthest 5' we were able to map the CNV breakpoint. For the 3' breakpoint, the majority of samples (at least 4 out of 5; excluding sample PTSD_11075) had convincing data consistent with the duplication extending as far as nucleotide 100546036 on chromosome 11.

Despite attempts to confirm the CNV breakpoints by molecular characterization, we remained confident that the CNV extended, both 5' and 3', to within a few kb of the breakpoints defined by the CNV calling algorithms. Using these breakpoints we attempted to map the breakpoint junctions to confirm our hypothesis that the duplication was in tandem in either a head to tail or tail to head orientation. Unfortunately, we were unable to amplify a PCR product that wasn't also present in non-carrier, negative controls. Our results are inconclusive and we therefore cannot conclude whether or not the copy number event resulted in a tandem duplication. Similarly, attempts to amplify a PCR product to show the existence of a fusion product between *ARHGAP42* and *CNTN5* also failed.

We were able to confirm that *CNTN5* is expressed in whole brain RNA. The following ΔCt values, relative to *RPLPO (*Hs99999902), were obtained for each of the three *CNTN5* assays: Hs00544267 = 7.8 $\pm$ 0.07, Hs00544274 = 8.0 $\pm$ 0.17, and Hs00205041 = 7.8 $\pm$ 0.13. Although we did not observe any detectable expression of *ARHGAP42* in whole brain RNA, we only used a single assay in our experiments. It remains possible that an alternatively spliced form(s) of *ARHGAP42*, which excludes one or both of the exons amplified in the reaction, does express in brain. Additional assays targeting other regions of the gene will need to be performed before excluding

ARHGAP42 as a gene expressed in the brain. Interestingly, using the gene expression assays to *CNTN5*, we observed measureable expression (Ct < 40) in whole blood mRNA derived from only the CNV carrier sample (PTSD_9646). None of 8 non-CNV carrier samples had detectable expression (Ct < 45). Additionally, for the duplication carrier sample, expression was only detected using 2 of the 3 assays (Ct = 39.3 for Hs00205041; Ct = 38.2 for Hs00544274) which are contained within the predicted duplication region (Figure 3-5). Though it appears that *CNTN5* is not normally expressed in whole blood, the duplication may result in increased *CNTN5* expression that for some unknown reason can be detected in this tissue type. This conclusion is supported by the fact that one of the assays (Hs00544267), which primes to a region outside of the duplication, did not result in detectable gene expression of *CNTN5* (Figure 3-5). Unfortunately with a sample size of only 1 subject, this finding while intriguing, is not definitive, and requires further investigation.

We used three strategies to support our original finding of an association between the duplication and PTSD. First, we tested the association between PTSD symptoms and SNPs within *CNTN5* and *ARGHAP42* (including regions 50 kb upstream of both genes). For *CNTN5* we tested the association of 500 SNPs. The lowest p-value was p=0.005 for the SNP, rs11223269. This was not statistically significant after multiple test correction. Only 91 SNPs were tested for association between PTSD and *ARHGAP42*. The association between PTSD and rs2513143 resulted in the lowest p-value of p=0.004. This was also non-significant. Secondly, we looked for the duplication within an additional 1,944 GTP samples as genotype data became available during the second phase of the GWAS. Based on a prevalence of approximately 1 in 500 (5 in 2,396), we

would expect to find at least 3 to 4 new CNV carriers. Unfortunately, we were unable to find any similar CNV duplications in this second dataset. The third approach involved the assessment of chromosome 11 CNV duplications among study participant from two large PTSD cohorts with GWAS data. No CNVs in the general region were found in the MRS dataset and only 4 were identified in the Army STARRS dataset (Figure 3-6). For the CNV carriers in the STARRS cohort, one had PTSD, two did not have PTSD, and the fourth was missing a phenotype for PTSD (not shown) (Figure 3-6). For the one affected CNV carrier, the CNV duplication was small and did not overlap with the CNV region identified in the GTP participants. Without CNVs of similar size and breakpoints identified in the new GTP samples, the MRS study or the STARRS study, we are unable to perform a true, replication analysis.

A burden analysis of deletions greater than 50 kb in the STARRS cohort revealed a statistically significant increase in the number of CNV deletions among non-Hispanic white participants with PTSD when compared to the number present in non-PTSD controls ($p=0.005$, Table 3-3). The number of individuals with and without PTSD that had at least one CNV $\geq$50 kb was not statistically significant ($p = 0.428$). Concerned that the deletion burden was only due to a large number of deletions among a small number of individuals, we examined the frequency distribution of CNVs $\geq$50 kb per person in PTSD cases and controls (Table 3-3). In neither cases nor controls were there greater than one person with more than 12 total CNVs 50 kb or larger. Thus it does not appear that the statistically significant finding is driven by the prevalence of CNVs in a few PTSD cases. The analysis also revealed a statistically significant differences in the total genomic coverage of deletions $\geq$50 kb. Compared to controls (total kb = 262), the combined size

of genomic deletion among cases (total kb = 289) was greater (p=0.004, Table 3-3). A similar analysis was performed for deletions ≥500 kb, revealing that both the rate and proportion of CNV deletions of this size are greater among PTSD cases versus controls (p = 0.013 and p = 0.22, respectively) (Table 3-4). The frequency distribution for deletions of this size is revealing: 3.8% of cases and only 2.8% of controls carry deletions ≥500 kb (Table 3-4).

We also conducted a PTSD by CNV association test using the STARRS cohort. Using a 2-sided test and 50,000 permutations (without restricting the analysis based on the frequency of CNVs) for CNVs ≥50 kb, we identified a CNV deletion on chromosome 7 (7q31.1) that is associated with PTSD at a level of statistical significance (multiple test corrected p-value = 0.025) (Table 3-5). 8 individuals with PTSD and only 1 without PTSD have a deletion ≥50 kb in this region (Figure 3-7). The deletions have variable, predicted breakpoints and the statistical significance applies only to a 20.4 kb region (noted by green lines; Figure 3-7) on chromosome 7 that contains 3 probes (at positions 110879251, 110879776 and 110879777) and represents the region of CNV overlap between all 9 carriers. With the exception of one large CNV identified among only one of the study participants, the remaining 8 CNVs are contained within the *IMMP2L* gene. The majority of the CNVs do not uncover exons within the gene nor does the 20.4 kb region of CNV overlap. However, we cannot rule out the possibility that these gene regions are regulatory.

Discussion

Due to rare prevalence and often deleterious nature, the main goal of identifying CNVs that associate with disease is to determine the genes or regulatory regions that are affected by the genetic variant and to then identify previously unknown components of biological processes involved in disease progression. We identified CNVs and analyzed the association between CNVs and PTSD in two very different study cohorts. The first analysis was performed in a primarily female, African American population exposed to civilian trauma (GTP study). The second analysis included trauma exposed male veterans that self-identified as non-Hispanic Caucasians (STARRS study). In both analyses we identified genes that are potentially involved in mechanisms associated with the development of PTSD.

In the GTP cohort we identified an approximately 750 kb duplication that associates with PTSD diagnosis. There are two genes partially covered by this particular CNV (*CNTN5* and *ARHGAP42*; Figure 3-1). Given that this duplication exists only in trauma exposed African American study participants diagnosed with PTSD, we predict that genes directly or indirectly dysregulated as a result of this variant are involved in the pathophysiology of PTSD. Since we were unable to determine the location and orientation of the duplication in experiments to molecularly characterize the CNV, we cannot predict what type of gene alterations may result from the duplication. However, we hypothesize that the aberrant expression of *CNTN5* is most likely associated with PTSD outcome. This prediction is based on the role of *CNTN5* in neuropsychiatric and brain-related disorders such as schizophrenia, autism, ADHD, anorexia nervosa and Alzheimer's [9,32–36]. *CNTN5* is involved in synaptic plasticity, which may be

important in the regulation of memory consolidation [8]. Additionally, *CNTN5* localizes to the amygdala in humans, a region of the brain that is involved in fear learning [6].

We were also able to confirm that *CNTN5* is expressed in human brain tissue. This finding adds further support to our hypothesis that *CNTN5* is likely the gene involved in in PTSD outcome as a result of the duplication, particularly given that we were unable to detect expression of *ARHGAP42* in the same brain tissue sample. Analysis of *CNTN5* expression in whole blood mRNA, which revealed detectable expression in only a carrier of the CNV duplication, provided additional support.

Experimental evidence for a role of *CNTN5* in PTSD would improve our understanding of the neurobiological mechanisms of PTSD and offer possibilities for prevention directly after trauma exposure when fear memory formation begins [37]. Unfortunately, attempts to identify a CNV with similar size and breakpoints in predominantly Caucasian study cohorts have failed and, thus, we have not yet been able to perform a true replication analysis of our association. It is possible that this particular CNV is more prevalent among African Americans and that an analysis of CNVs among large cohorts of African Americans will be necessary in order to replicate our finding. We had hoped that we could use the additional GTP samples from phase 2 of genotyping as a replication cohort. Unfortunately, we did not detect the CNV in the approximately 2,000 new, African American study participants. The reason for this remains unclear.

Several similarly sized CNVs have been detected in various other cohorts, suggesting that this CNV is not simply a novel event particular to our study population (Table 3-6). The Database of Genomic Variants (DGV), which catalogs CNVs among more than 22,000 healthy cohorts, reports only three comparably sized CNVs in this

region on chromosome 11 (http://dgv.tcag.ca/dgv/app/home) (Table 3-6). Thus, we believe that this 11q22.1 duplication may be more prevalent in a non-healthy sample cohort and could also result in a deleterious health outcome. In fact, Cooper et al. report finding 6 individuals (out of 15,767 with developmental delay/intellectual disability) with a duplication in *CNTN5* (see Supplementary Info., Cooper et al.) [38]. Individuals with these duplications were found to have neurological defects (N=2), abnormal craniofacial features (N=1), autism spectrum disorder (N=2), and epilepsy (N=1). Only 2 individuals among 8,329 healthy adult controls had a CNV duplication in this gene. In three other studies, CNV gains at 11q22.1 were also identified among individuals with autism or developmental delay (Table 3-6) [39–41].

Although we were able to provide relatively convincing data that the duplication exists, we were unable to determine breakpoints. Furthermore, we were unable to show any evidence of a tandem duplication or a fusion event between *ARHGAP42* and *CNTN5*. The inability to produce results from these experiments is likely due to the fact that the breakpoints are poorly mapped and performing PCR experiments without a more accurate idea of where to design the primers could result in an amplicon that is too large to amplify. Sequencing of at least one of the variants, presuming that that the duplications are similar, is necessary to discern both breakpoints as well as the type of duplication (e.g. tandem or translocation).

In the STARRS cohort, we identified a deletion that associate with PTSD and results in reduced copy number for the 3' gene region of *IMMP2L.* This finding is particularly intriguing given that both SNPs and deletions in this gene have been associated with neurodevelopmental disorders [11–13,42]. The CNV also falls within a

region of the genome implicated in autism susceptibility [10]. There are two drawbacks to this finding. Firstly, deletions in this region of the genome are relatively common with greater than 100 reported in DGV among healthy individuals. Secondly, there is only a small overlap between the CNV deletions identified among the carrier participants with PTSD. If the breakpoints for any of the CNV carriers are incorrect and the deletion for that individual falls outside of the overlap region, the association may lose statistical significance. Defining the breakpoints of each of the CNV deletions will be necessary to confirm that there is a deletion overlap for all 8 of the carriers with PTSD. Also, replication of this finding in a well-powered study of participants with similar ancestry will be necessary. Given how common this CNV is among presumed healthy controls, gene by environment interaction may be an important factor that accounts for the association of this CNV with PTSD.

The most reportable result, of all those presented, is the burden analysis that revealed a statistically significant increase in prevalence of deletions $\geq$50 kb among PTSD cases versus unaffected controls in the STARRS cohort. CNVs >500 kb were also tested by burden analysis and found to be more prevalent among cases. These results warrant further investigation, including a network analysis of the genes included in the deletion regions. An analysis of the genomic regions that are deleted by CNVs of this size among cases may provide further insight into genes or gene pathways involved in susceptibility to PTSD.

Table 3-1. The 750 kb CNV duplication on chromosome 11 was identified in 5 GTP study participants by each of three CNV calling programs. The start and end coordinates predicted by each algorithm are provided.

Sample ID	Chromosome	Start	End	CNV calling algorithm
PTSD_9327	11	99889749	100624538	GADA
PTSD_9327	11	99889749	100624538	PART
PTSD_9327	11	99889749	100624538	PCNV
PTSD_5543	11	99889749	100632773	GADA
PTSD_5543	11	99889749	100624096	PART
PTSD_5543	11	99889749	100615623	PCNV
PTSD_11075	11	99885930	100639190	GADA
PTSD_11075	11	99885930	100624538	PART
PTSD_11075	11	99889749	100615623	PCNV
PTSD_10456	11	99889749	100624538	GADA
PTSD_10456	11	99889749	100624538	PART
PTSD_10456	11	99889749	100624538	PCNV
PTSD_9646	11	99885930	100624538	GADA
PTSD_9646	11	99885930	100624538	PART
PTSD_9646	11	99932666	100624538	PCNV

Table 3-2. Fold change difference in copy number for 5 GTP study participants with the 750 kb duplication relative to GTP, non-CNV carriers: **A)** PTSD_7529 and **B)** PTSD_7722

A)

CNV Carrier Sample ID	CNV Region	Fold-change Copy number
PTSD_5543	left	1.9
PTSD_5543	middle	2
PTSD_5543	right	2.2
PTSD_9327	left	1.6
PTSD_9327	middle	1.8
PTSD_9327	right	1.8
PTSD_9646	left	1.8
PTSD_9646	middle	2
PTSD_9646	right	2.2
PTSD_10456	left	1.2
PTSD_10456	middle	1.3
PTSD_10456	right	1.4
PTSD_11075	left	1.6
PTSD_11075	middle	1.9
PTSD_11075	right	2

B)

CNV Carrier Sample ID	CNV Region	Fold-change Copy number
PTSD_5543	left	1.3
PTSD_5543	middle	1.1
PTSD_5543	right	2.2
PTSD_9327	left	1.4
PTSD_9327	middle	1.3
PTSD_9327	right	2.1
PTSD_9646	left	1.5
PTSD_9646	middle	1.5
PTSD_9646	right	2.6
PTSD_10456	left	1
PTSD_10456	middle	0.9
PTSD_10456	right	1.7
PTSD_11075	left	1.4
PTSD_11075	middle	1.4
PTSD_11075	right	2.3

Table 3-2C. Fold change difference in copy number relative to a control DNA (Coriell) for 2 GTP study participants without the 750 kb duplication and 5 GTP study participants with the duplication

Sample Name	CNV Carrier Status	Fold Change in Copy Number
PTSD_7529	non-carrier	1.14
PTSD_7722	non-carrier	1.22
PTSD_5543	carrier	1.45
PTSD_9327	carrier	1.44
PTSD_9646	carrier	1.48
PTSD_10456	carrier	1.32
PTSD_11075	carrier	1.40

Table 3-3. **A.)** The results of a burden analysis between CNV deletions ≥ 50 kb and PTSD cases in the STARRS cohort. **B)** The frequency distribution of number of CNV deletions ≥ 50 kb by control or case status

A.)

	Cases	Controls	p-value
Total CNVs	3094	9313	
RATE	1.803	1.701	0.005
PROP	0.8147	0.8122	0.428
KBTOT	288.8	262.4	0.004
KBAVG	128.1	123.7	0.135

B.)

Controls

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	1028	18.8	18.8	18.8
	1	1706	31.2	31.2	49.9
	2	1456	26.6	26.6	76.5
	3	758	13.8	13.8	90.4
	4	356	6.5	6.5	96.9
	5	75	1.4	1.4	98.3
	6	71	1.3	1.3	99.6
	7	6	.1	.1	99.7
	8	14	.3	.3	99.9
	10	3	.1	.1	100.0
	12	1	.0	.0	100.0
	Total	5474	100.0	100.0	

Cases

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	318	18.5	18.5	18.5
	1	502	29.3	29.3	47.8
	2	462	26.9	26.9	74.7
	3	234	13.6	13.6	88.3
	4	118	6.9	6.9	95.2
	5	37	2.2	2.2	97.4
	6	31	1.8	1.8	99.2
	7	1	.1	.1	99.2
	8	8	.5	.5	99.7
	10	4	.2	.2	99.9
	12	1	.1	.1	100.0
	Total	1716	100.0	100.0	

Table 3-4. **A.)** The results of a burden analysis between CNV deletions ≥ 500 kb and PTSD cases in the STARRS cohort. **B)** The frequency distribution of number of CNV deletions ≥ 500 kb by control or case status.

A.)

	Cases	Controls	p-value
Total CNVs	74	166	
RATE	0.04312	0.03033	0.013
PROP	0.03788	0.02795	0.022
KBTOT	1100	1040	0.319
KBAVG	958.8	947	0.447

B.)

Controls

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	5321	97.2	97.2	97.2
	1	142	2.6	2.6	99.8
	2	10	.2	.2	100.0
	4	1	.0	.0	100.0
	Total	5474	100.0	100.0	

Cases

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	1651	96.2	96.2	96.2
	1	56	3.3	3.3	99.5
	2	9	.5	.5	100.0
	Total	1716	100.0	100.0	

Table 3-5. *IMMP2L* CNV deletions associated with PTSD in the STARRS cohort

Sample Name	Chromosome	CNV Start	CNV End	CNV Size (kb)
R0171968	7	110338460	111849265	1511
R0044644	7	110811863	110899634	88
R0552882	7	110811863	110990633	179
R0671958	7	110811863	110915339	103
R0694413	7	110811863	111166165	354
R0517011	7	110840236	110932630	92
R0248628	7	110851553	111166165	315
R0520031	7	110879251	111265262	386
R0690037	7	110879251	111059886	181

Table 3-6. 11q22.1 CNV duplications and corresponding phenotypes from various studies

hg19 coordinates	CNV Type	Phenotype*	Study	Variant name
chr11:100163957-100681211	gain	No pheno. (DGV)	Shaikh [237]	nsv520995
chr11:99893282-100585889	gain	No pheno. (DGV)	Cooper [232]	nsv556122
chr11:99888390-100629136	gain	No pheno (DGV)	Coe [238]	dgv1265n100
chr11:99913394 -100627255	gain	Dev Delay	Kaminsky [233]	nsv529454
chr11:100196082-100713156	gain	Dev Delay	Kaminsky [233]	nsv530643
chr11:99889749-100624538	gain	PTSD	GTP	NA
chr11:99889749-100624538	gain	PTSD	GTP	NA
chr11:99889749-100624538	gain	PTSD	GTP	NA
chr11:99889749-100624538	gain	PTSD	GTP	NA
chr11:99889749-100624538	gain	PTSD	GTP	NA
chr11:99714232 100125599	gain	PTSD unknown	STARRS	NA
chr11:100192261-100863246	gain	No PTSD	STARRS	NA
chr11:100269350-101785248	gain	No PTSD	STARRS	NA
chr11:99715682-99841735	gain	PTSD	STARRS	NA
chr11:98907603-98971472	gain	Autism	Pinto [234]	NA
chr11:100211194-100784284	gain	Autism	Krumm [235]	NA
chr11:99931943-100558563	gain	Autism	Krumm [235]	NA

*DGV= variant reported in the Database of Genomic Variants

Figure 3-1. An approximately 750 kb duplication (indicated by the green bar) on chromosome 11 associates with PTSD. The copy number gain includes the 3' region of the gene *CNTN*5, which encodes for the contactin-5 protein known as NB2. Note that a non-coding transcript of *CNTN5* (in green text) is contained within the duplication in its entirety. The duplication also covers the 5' region and first exon of *ARHGAP42*, which encodes for a Rho GTPase activating protein.

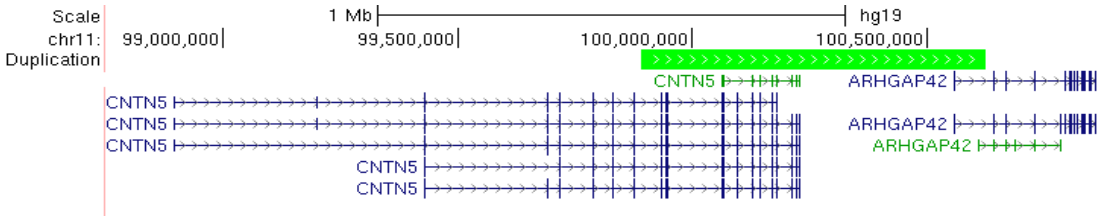


Figure 3-2. Breakpoint junction mapping to determine the orientation of a potential tandem repeat, resulting from a CNV duplication event involving *CNTN5* and *ARHGAP42*

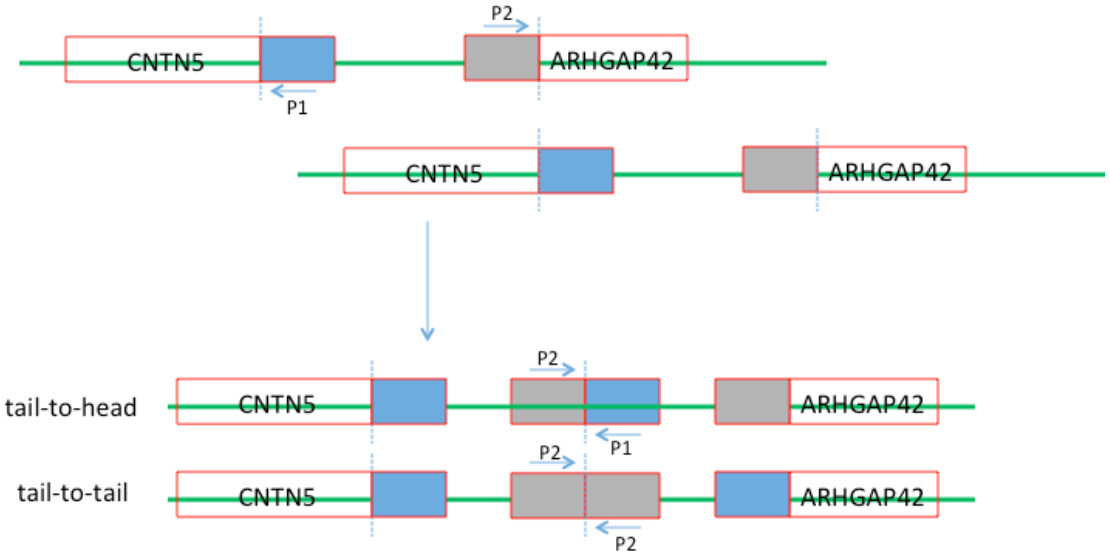


Figure 3-3. Locations of P1 and P2 primers relative to putative breakpoints on chromosome 11: 99889749 and 100624538

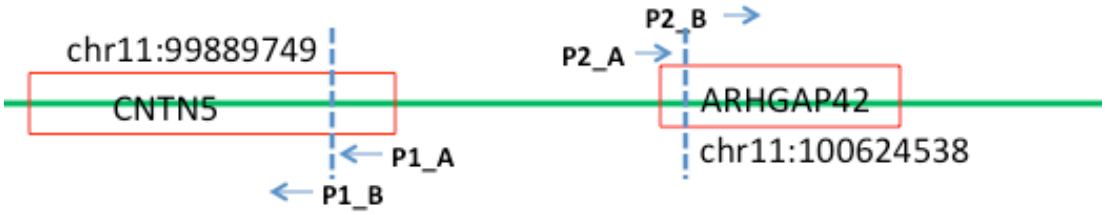
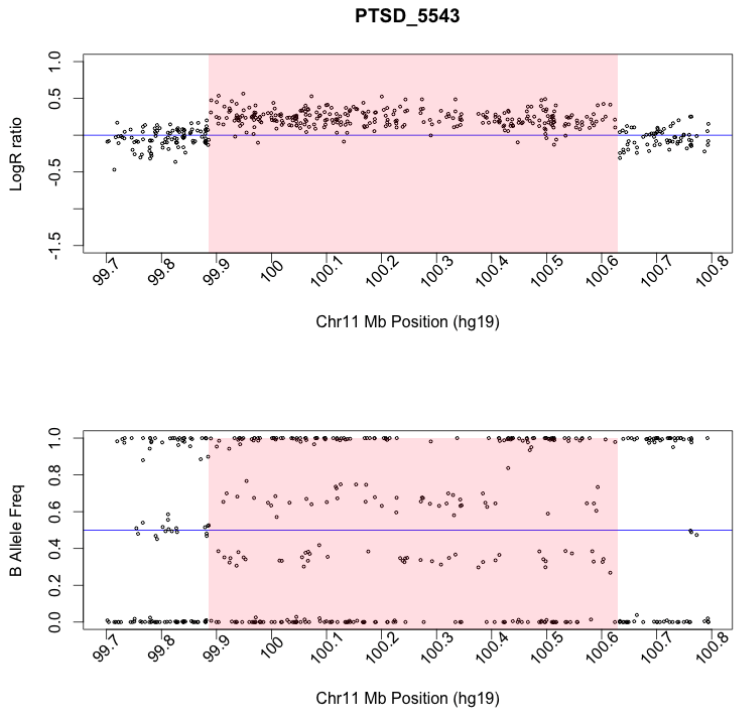


Figure 3-4. A 750 kb copy number duplication on chromosome 11 at 11q22.1, identified in 5, GTP study participants **A) – E)**, is shown visually by graphing of the Log R ratio and B allele frequency values. The region of the duplication is highlighted in pink (chr11: 99885000 – 100630000).

A.)



B.)

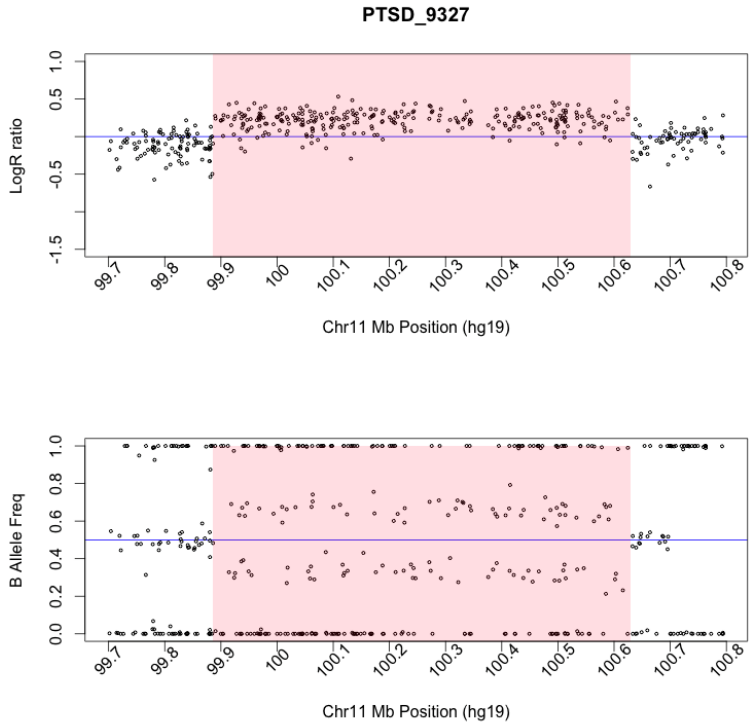
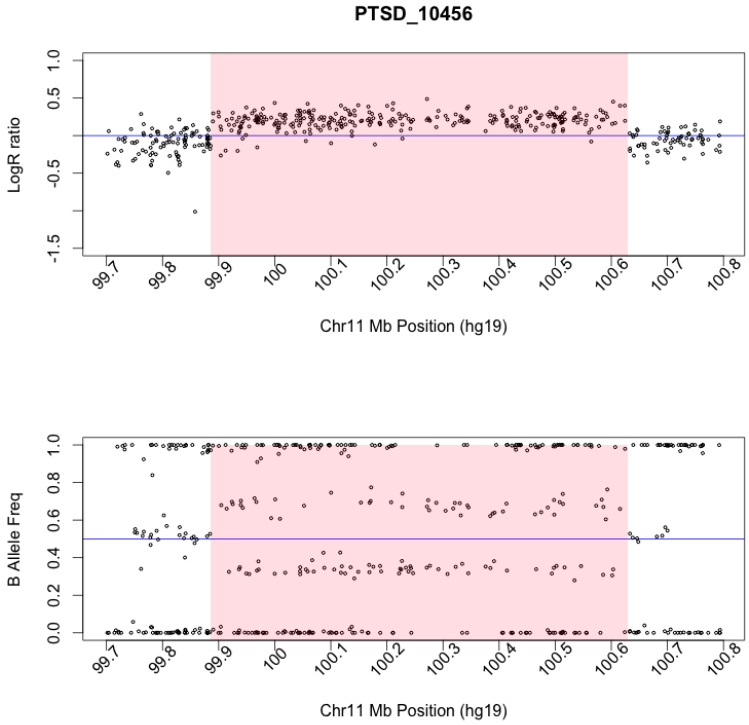
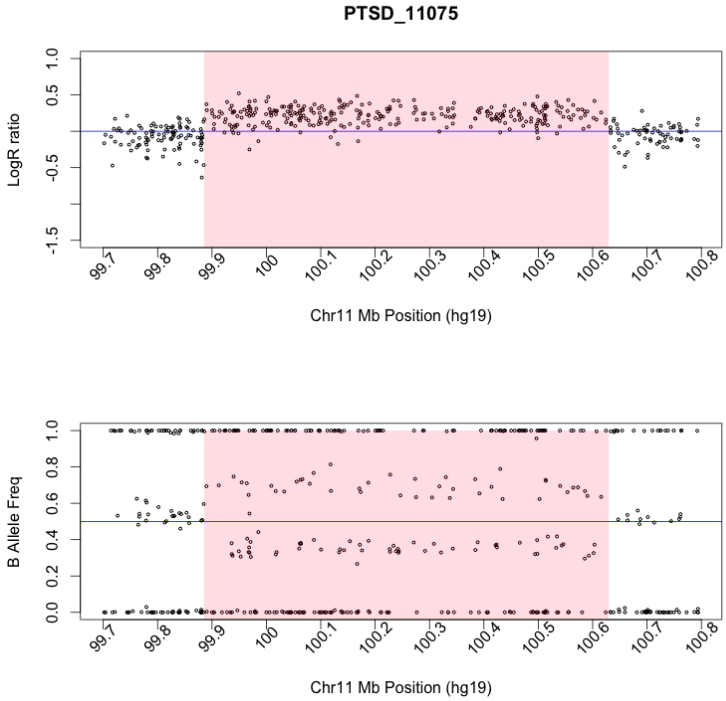


Figure 3-4. continued…

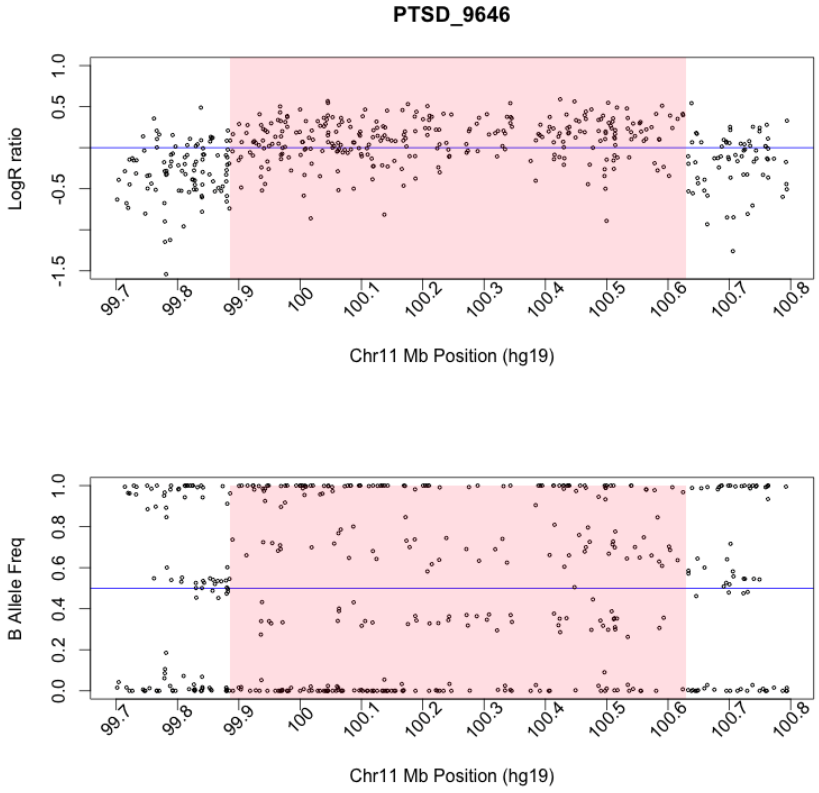
C.)



D.)



E.)


PTSD_9646
LogR ratio
1.0
0.5
-0.5
-1.5
99.7
99.8
99.9
100
100.1
100.2
100.3
100.4
100.5
100.6
100.7
100.8
Chr11 Mb Position (hg19)
B Allele Freq
1.0
0.8
0.6
0.4
0.2
0.0
Chr11 Mb Position (hg19)

F.)

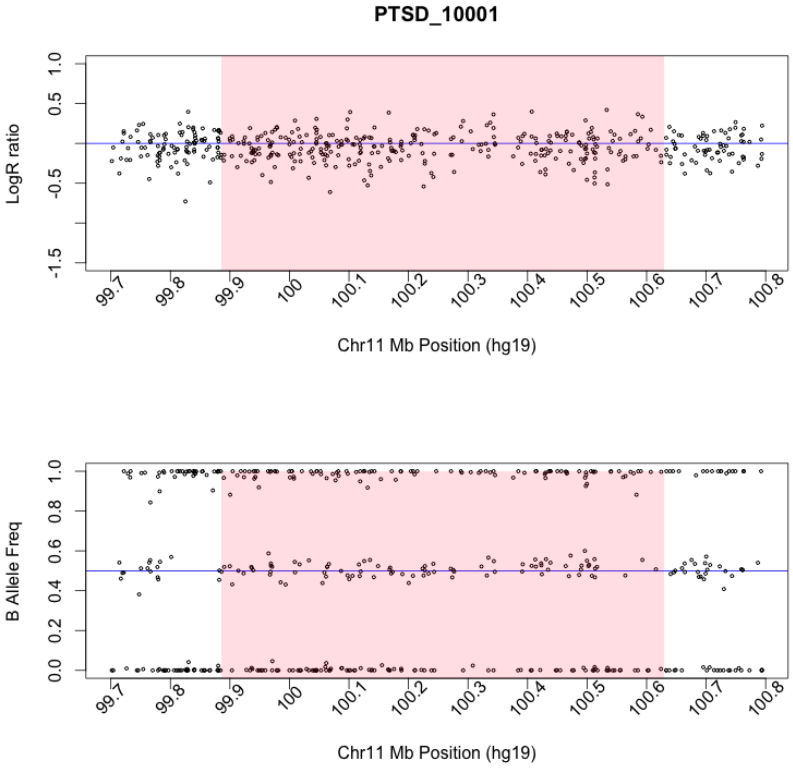

PTSD_10001
LogR ratio
1.0
0.5
-0.5
-1.5
99.7
99.8
99.9
100
100.1
100.2
100.3
100.4
100.5
100.6
100.7
100.8
Chr11 Mb Position (hg19)
B Allele Freq
1.0
0.8
0.6
0.4
0.2
0.0
Chr11 Mb Position (hg19)

Figure 3-5. 3 Taqman assays were used to detect gene expression of *CNTN5* in whole blood mRNA from one CNV carrier and eight non-CNV carriers. Gene expression was only detected in the carrier mRNA samples and only for the two assays (Hs00544274 and Hs00205041) that prime to regions within the duplication. The region shown in red, amplified by assay Hs00544267 did not result in detectable expression in any of the mRNA samples.

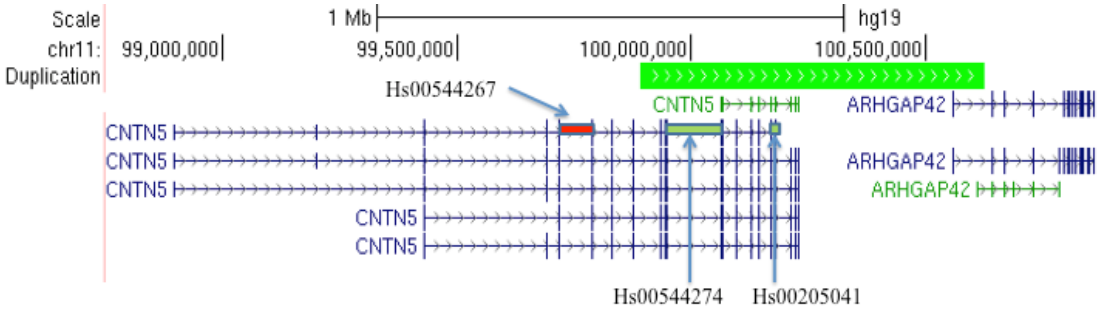


Figure 3-6. Chromosome 11 CNV gains in GTP and STARRS cohorts by PTSD diagnosis. The CNVs in blue were found in study participants with PTSD and those in red were found among study participants without PTSD. The studies in which each of these CNVs were identified are noted in the left hand

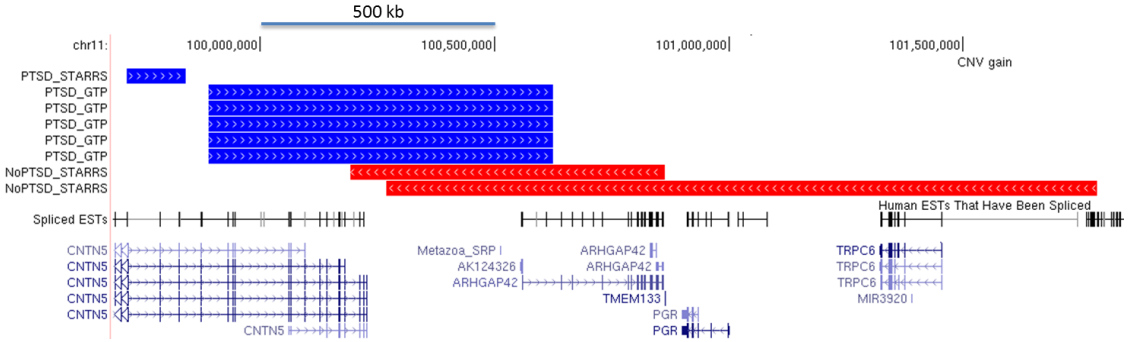
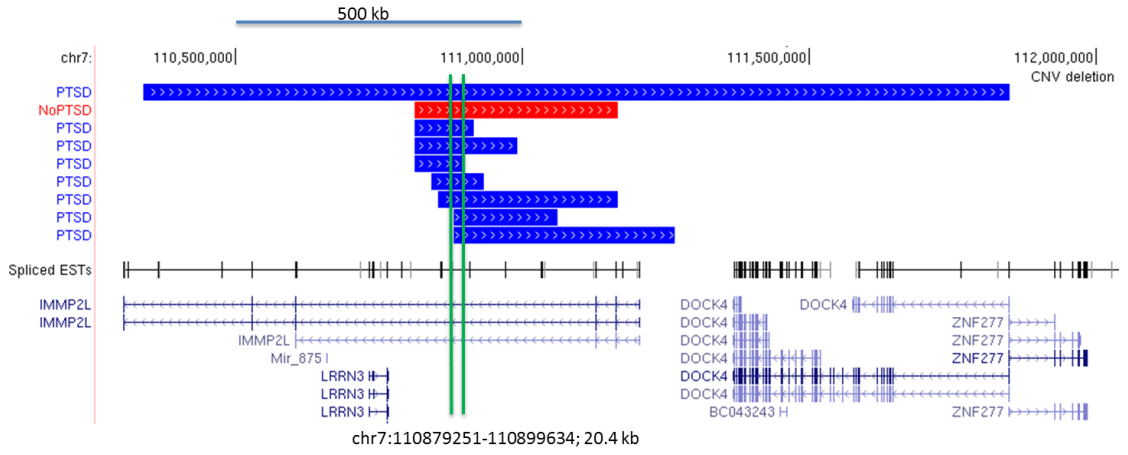


Figure 3-7. A CNV deletion at 7q31.1 which contains the gene *IMMP2L* is associated with PTSD in the STARRS cohort



Appendix 3-A. Steps employed using PennCNV for detection of CNVs in study cohorts

Running PennCNV:
http://penncnv.openbioinformatics.org/en/latest/user-guide/install/

Step 1: Export data from Genome Studio in the following format:

Name – Chr- Position-Genotype-Log R Ratio-B Allele Freq in this order. Use column chooser to select these columns and exclude all others.

Step 2: To generate working files for PennCNV this large, exported file will need to be split into individual files per sample. To do this use the Perl script, kcolumn.pl which is part of the penncnv package.

./kcolumn.pl GS_export_file.txt split 3 –heading 3 –tab –out sample –name –-start_split 1 –-end_split 1000

NOTE 1: If kcolumn.pl doesn't run it may be because you need to change file permissions to execute. This has to be done every time the script gets moved to a new folder

NOTE 2: kcolumn.pl can only be run on 1000 files at a time. Using the –-start and –-end commands will allow you to run only 1000 at a time, choosing which samples in the export file are processed each time. Also, the word before name (sample in this example) will be added to each file name before the SID number. This can be used to insert study name or other information to distinguish the samples.

Step 3: A pfb file specifically for your sample data set must be generated. To do this use the Perl script, compile_pfb.pl which is part of the penncnv package

./compile_pfb.pl -listfile signal_file_list (1000 at a time) -output out.pfb

1) If there are > 1000 samples you will need to generate more than 1 pfb file. Generate a text file with 1000 or fewer sample names (from step 2).
2) Next, average the pfb values for all the pfb files created using R script :

```
afile=read.delim("Study_A.pfb",sep="\t",as.is=TRUE,header=TRUE)
bfile=read.delim("Study_B.pfb",sep="\t",as.is=TRUE,header=TRUE)
cfile=read.delim("Study_C.pfb",sep="\t",as.is=TRUE,header=TRUE)
test=cbind.data.frame(afile$PFB,bfile$PFB,cfile$PFB)
test$average_PFB=apply(test,1,mean)
allfile=cbind.data.frame(afile$Name,afile$Chr,afile$Position,test$average_PFB)
names(allfile) <- c("Name", "Chr", "Position", "PFB")
write.table(allfile,file="study.pfb",sep="\t",row.names=FALSE,col.names=TRUE,
quote=FALSE)
```

Step 4: Generate a gc correction file using the pfb file from step 3. The gc5Base file can be obtained from the UCSC genome browser.

./cal_gc_snp.pl gc5Base.sorted "$study".pfb > "$study".gc_content

Step 5: Use detect_cnv.pl to make CNV calls.

1) Just as was done when generating the pfb files, a text file of sample files names needs to be created and fed into the script using –listfile. It is important to make sure the sample list is formatted correctly. Use the vi command to determine the text delimiter. To remove returns denoted by ^M use the following command to remove.
   sed –e "s/^M/\n/g" filename >> newfilename

/detect_cnv.pl –test –hmm lib/hhall.hmm –pfb study.pfb –listfile listfile.txt –conf –gcmodel lib/study.gc_content –log study.log –out study.rawcnv

Step 6: Filter out CNVs with fewer than 10 probes and less than 10 kb

./filter_cnv.pl 2986_gcWave.raw.cnv -qclogfile 2986_gcWave.log –qclrrsd 1.0 –qcbafdrift 0.1 –qcwf 1.0 –qcnumcnv 2000 –numsnp 10 –length 10k –qcpassout 2986_10_10_GCwave.qcpass –qcsumout 2986_10_10_GCwave.qcsum –out 2986_10_10_GCwave.raw.cnv

Appendix 3-B. CNVs within the regions below were excluded from association analyses

Centromeric Regions

chr1:121500000-128900000
chr2:90500000-96800000
chr3:87900000-93900000
chr4:48200000-52700000
chr5:46100000-50700000
chr6:58700000-63300000
chr7:58000000-61700000
chr8:43100000-48100000
chr9:47300000-50700000
chr10:38000000-42300000
chr11:51600000-55700000
chr12:33300000-38200000
chr13:16300000-19500000
chr14:16100000-19100000
chr15:15800000-20700000
chr16:34600000-38600000
chr17:22200000-25800000
chr18:15400000-19000000
chr19:24400000-28600000
chr20:25600000-29400000
chr21:10900000-14300000
chr22:12200000-17900000

Immunoglobulin and T-cell Receptor Regions

chr2:89156873-89630175
chr2:89890561-90471176
chr7:38397534-38398683
chr7:142139277-142139511
chr7:142470098-142470118
chr7:142008779-142008870
chr7:142180514-142180600
chr7:142231575-142231799
chr7:142239536-142239659
chr7:142495138-142495186
chr14:22293662-23021097
chr14:105994255-107283085
chr15:22099106-22797761
chr16:32914930-33208857
chr22:22385571-23265082

Telomeric Regions

chr1:1-100000
chr2:1-100000

Telomeric Regions cont…

chr3:1-100000
chr4:1-100000
chr5:1-100000
chr6:1-100000
chr7:1-100000
chr8:1-100000
chr9:1-100000
chr10:1-100000
chr11:1-100000
chr12:1-100000
chr13:1-100000
chr14:1-100000
chr15:1-100000
chr16:1-100000
chr17:1-100000
chr18:1-100000
chr19:1-100000
chr20:1-100000
chr21:1-100000
chr22:1-100000
chr21:48029895-48129895
chr22:51204566-51304566
chr19:59028983-59128983
chr20:62925520-63025520
chr18:77977248-78077248
chr16:90254753-90354753
chr15:102431392-102531392
chr14:107249540-107349540
chr13:115069878-115169878
chr12:133751895-133851895
chr11:134906516-135006516
chr10:135434747-135534747
chr9:141113431-141213431
chr8:146264022-146364022
chr7:159038663-159138663
chr6:171015067-171115067
chr5:180815260-180915260
chr4:191054276-191154276
chr3:197922430-198022430
chr2:243099373-243199373
chr1:249150621-249250621

Appendix 3-C. PLINK analysis log of CNV duplications (≥500 kb) by PTSD association in GTP cohort

```
@----------------------------------------------------------@
|      PLINK!      |    v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|      http://pngu.mgh.harvard.edu/purcell/plink/          |
@----------------------------------------------------------@
Options in effect:
          --cfile combrun3
          --remove SIDs_no_adult_trauma_or_missing.txt
          --cnv-exclude centromer_position_from_UCSC_hg19.txt
          --cnv-dup
          --cnv-kb 500
          --cnv-freq-exclude-above 10
          --cnv-test-1sided
          --mperm 10000

Reading marker information from [ combrun3.cnv.map ]
40242 (of 40242) markers to be included from [ combrun3.cnv.map ]
Reading individual information from [ combrun3.fam ]
Reading pedigree information from [ combrun3.fam ]
2934 individuals read from [ combrun3.fam ]
2803 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
898 cases, 1905 controls and 131 missing
843 males, 2091 females, and 0 of unspecified sex
131 individuals removed because of missing phenotypes
Reading individuals to remove [ SIDs_no_adult_trauma_or_missing.txt ] ... 407 read
407 individuals removed with --remove option
Reading CNV intersection list from [ centromer_position_from_UCSC_hg19.txt ]
Read 24 ranges to exclude from CNV list
Reading segment list (CNVs) from [ combrun3.cnv ]
Filtering segments based on frequencies
Will remove 4330 CNVs based on frequency (after other filters)
83715 mapped to a person, of which 4870 passed filters
4593 kept after excluding specific regions
263 of 104637 mapped as valid segments
 CopyN Case/Control
    3     94 / 169
Writing per-individual summary to [ plink.cnv.indiv ]
Writing positional summary to [ plink.cnv.summary ]
812 affected individuals out of 2396 in total
Set to permute within 1 cluster(s)
Writing permuted results for segment test to [ plink.cnv.summary.mperm ]
```

Appendix 3-D. PLINK analysis log of CNV deletions (≥50 kb) by PTSD association in STARRS cohort

```
@----------------------------------------------------------@
|      PLINK!      |    v1.07     |  10/Aug/2009    |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|      http://pngu.mgh.harvard.edu/purcell/plink/       |
@----------------------------------------------------------@

Options in effect:
        --cfile STARRS_filtered_all_IIDs
        --cnv-del
        --cnv-kb 50
        --cnv-test-2sided
        --mperm 50000
        --out del_50kb
        --noweb

Reading marker information from [ STARRS_filtered_all_IIDs.cnv.map ]
38019 (of 38019) markers to be included from [ STARRS_filtered_all_IIDs.cnv.map ]
Reading individual information from [ STARRS_filtered_all_IIDs.fam ]
Reading pedigree information from [ STARRS_filtered_all_IIDs.fam ]
7190 individuals read from [ STARRS_filtered_all_IIDs.fam ]
7190 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
1716 cases, 5474 controls and 0 missing
6548 males, 642 females, and 0 of unspecified sex

Reading segment list (CNVs) from [ STARRS_filtered_all_IIDs.cnv ]
57858 mapped to a person, of which 12407 passed filters
12407 of 57858 mapped as valid segments
 CopyN Case/Control
   0    276 / 835
   1   2818 / 8478

Writing per-individual summary to [ del_50kb.cnv.indiv ]
Writing positional summary to [ del_50kb.cnv.summary ]
1716 affected individuals out of 7190 in total
Set to permute within 1 cluster(s)
Writing permuted results for segment test to [ del_50kb.cnv.summary.mperm ]
```

Appendix 3-E. PLINK analysis log of burden of CNV deletions (≥ 50kb) in STARRS cohort

```
@----------------------------------------------------------@
|      PLINK!      |    v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|      http://pngu.mgh.harvard.edu/purcell/plink/       |
@----------------------------------------------------------@

Options in effect:
        --cfile STARRS_filtered_all_IIDs
        --cnv-del
        --cnv-kb 50
        --cnv-indiv-perm
        --mperm 10000
        --out del_50kb_burden
        --noweb

Reading marker information from [ STARRS_filtered_all_IIDs.cnv.map ]
38019 (of 38019) markers to be included from [ STARRS_filtered_all_IIDs.cnv.map ]
Reading individual information from [ STARRS_filtered_all_IIDs.fam ]
Reading pedigree information from [ STARRS_filtered_all_IIDs.fam ]
7190 individuals read from [ STARRS_filtered_all_IIDs.fam ]
7190 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
1716 cases, 5474 controls and 0 missing
6548 males, 642 females, and 0 of unspecified sex

Reading segment list (CNVs) from [ STARRS_filtered_all_IIDs.cnv ]
57858 mapped to a person, of which 12407 passed filters
12407 of 57858 mapped as valid segments
 CopyN Case/Control
    0    276 / 835
    1   2818 / 8478

Writing per-individual summary to [ del_50kb_burden.cnv.indiv ]
Writing positional summary to [ del_50kb_burden.cnv.summary ]
1716 affected individuals out of 7190 in total
Set to permute within 1 cluster(s)
Writing group summary statistics to [ del_50kb_burden.cnv.grp.summary ]
Writing permuted results for segment test to [
del_50kb_burden.cnv.summary.mperm ]
```

Appendix 3-F. PLINK analysis log of burden of CNV deletions (≥ 500kb) in STARRS cohort

```
@----------------------------------------------------------@
|      PLINK!      |    v1.07     |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|      http://pngu.mgh.harvard.edu/purcell/plink/          |
@----------------------------------------------------------@

Options in effect:
        --cfile STARRS_filtered_all_IIDs
        --cnv-del
        --cnv-kb 500
        --cnv-indiv-perm
        --mperm 10000
        --out del_500kb_burden
        --noweb

Reading marker information from [ STARRS_filtered_all_IIDs.cnv.map ]
38019 (of 38019) markers to be included from [ STARRS_filtered_all_IIDs.cnv.map ]
Reading individual information from [ STARRS_filtered_all_IIDs.fam ]
Reading pedigree information from [ STARRS_filtered_all_IIDs.fam ]
7190 individuals read from [ STARRS_filtered_all_IIDs.fam ]
7190 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
1716 cases, 5474 controls and 0 missing
6548 males, 642 females, and 0 of unspecified sex

Reading segment list (CNVs) from [ STARRS_filtered_all_IIDs.cnv ]
57858 mapped to a person, of which 240 passed filters
240 of 57858 mapped as valid segments
 CopyN Case/Control
   1    74 / 166

Writing per-individual summary to [ del_500kb_burden.cnv.indiv ]
Writing positional summary to [ del_500kb_burden.cnv.summary ]
1716 affected individuals out of 7190 in total
Set to permute within 1 cluster(s)
Writing group summary statistics to [ del_500kb_burden.cnv.grp.summary ]
Writing permuted results for segment test to [
del_500kb_burden.cnv.summary.mperm ]
```

References

1. True WR, Rice J, Eisen SA, Heath AC, Goldberg J, Lyons MJ, et al.: A twin study of genetic and environmental contributions to liability for posttraumatic stress symptoms. Arch Gen Psychiatry 1993;50:257–264.

2. Xian H, Chantarujikapong SI, Scherrer JF, Eisen SA, Lyons MJ, Goldberg J, et al.: Genetic and environmental influences on posttraumatic stress disorder, alcohol and drug dependence in twin pairs. Drug Alcohol Depend 2000;61:95–102.

3. Stein MB, Jang KL, Taylor S, Vernon PA, Livesley WJ: Genetic and environmental influences on trauma exposure and posttraumatic stress disorder symptoms: A twin study. Am J Psychiatry 2002;159:1675–1681.

4. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al.: Finding the missing heritability of complex diseases. Nature 2009 Oct 8;461:747–753.

5. Kotlar AV, Mercer KB, Zwick ME, Mulle JG: New discoveries in schizophrenia genetics reveal neurobiological pathways: A review of recent findings. Eur J Med Genet 2015 Dec;58:704–714.

6. Kamei Y, Takeda Y, Teramoto K, Tsutsumi O, Taketani Y, Watanabe K: Human NB-2 of the contactin subgroup molecules: chromosomal localization of the gene (CNTN5) and distinct expression pattern from other subgroup members. Genomics 2000 Oct 1;69:113–119.

7. Ogawa J, Lee S, Itoh K, Nagata S, Machida T, Takeda Y, et al.: Neural recognition molecule NB-2 of the contactin/F3 subgroup in rat: Specificity in neurite outgrowth-promoting activity and restricted expression in the brain regions. J Neurosci Res 2001 Jul 15;65:100–110.

8. Shimoda Y, Koseki F, Itoh M, Toyoshima M, Watanabe K: A cis-complex of NB-2/contactin-5 with amyloid precursor-like protein 1 is localized on the presynaptic membrane. Neurosci Lett 2012 Feb 29;510:148–153.

9. Biffi A, Anderson CD, Desikan RS, et al: GEnetic variation and neuroimaging measures in alzheimer disease. Arch Neurol 2010 Jun 1;67:677–685.

10. Maestrini E, Pagnamenta AT, Lamb JA, Bacchelli E, Sykes NH, Sousa I, et al.: High-density SNP association study and copy number variation analysis of the AUTS1 and AUTS5 loci implicate the IMMP2L-DOCK4 gene region in autism susceptibility. Mol Psychiatry 2010 Sep;15:954–968.

11. Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH, et al.: Relative burden of large CNVs on a range of neurodevelopmental phenotypes. PLoS Genet 2011 Nov;7:e1002334.

12. Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT, et al.: Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. Mol Psychiatry 2010 Jun;15:637–646.

13. Petek E, Windpassinger C, Vincent JB, Cheung J, Boright AP, Scherer SW, et al.: Disruption of a novel gene (IMMP2L) by a breakpoint in 7q31 associated with Tourette syndrome. Am J Hum Genet 2001 Apr;68:848–858.

14. Gillespie CF, Bradley B, Mercer K, Smith AK, Conneely K, Gapen M, et al.: Trauma exposure and stress-related disorders in inner city primary care patients. Gen Hosp Psychiatry 2009 Dec;31:505–514.

15. Foa EB, Riggs DS, Dancu CV, Rothbaum BO: Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. J Trauma Stress 1993 Oct 1;6:459–473.

16. Foa EB, Tolin DF: Comparison of the PTSD Symptom Scale-Interview Version and the Clinician-Administered PTSD scale. J Trauma Stress 2000 Apr;13:181–191.

17. Schwartz AC, Bradley RL, Sexton M, Sherry A, Ressler KJ: Posttraumatic stress disorder among African Americans in an inner city mental health clinic. Psychiatr Serv Wash DC 2005 Feb;56:212–215.

18. Almli LM, Duncan R, Feng H, Ghosh D, Binder EB, Bradley B, et al.: Correcting Systematic Inflation in Genetic Association Tests That Consider Interaction Effects: Application to a Genome-wide Association Study of Posttraumatic Stress Disorder. JAMA Psychiatry 2014 Oct 29; DOI: 10.1001/jamapsychiatry.2014.1339

19. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006 Aug;38:904–909.

20. Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, et al.: High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. Genome Res 2006 Sep;16:1136–1148.

21. Wang K, Bucan M: Copy Number Variation Detection via High-Density SNP Genotyping. Cold Spring Harb Protoc 2008 Jun 1;2008:pdb.top46.

22. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al.: PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 2007 Nov 1;17:1665–1674.

23. Pique-Regi R, Cáceres A, González JR: R-Gada: a fast and flexible pipeline for copy number analysis in association studies. BMC Bioinformatics 2010 Jul 16;11:380.

24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al.: PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007 Sep;81:559–575.

25. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods San Diego Calif 2001 Dec;25:402–408.

26. Parker B, Alexander R, Wu X, Feely S, Shy M, Schnetz-Boutaud N, et al.: Detection of copy number variation by SNP-allelotyping. J Neurogenet 2015 Mar;29:4–7.

27. Gabriel S, Ziaugra L, Tabbaa D: SNP genotyping using the Sequenom MassARRAY iPLEX platform. Curr Protoc Hum Genet Editor Board Jonathan Haines Al 2009 Jan;Chapter 2:Unit 2.12.

28. Arlt MF, Mulle JG, Schaibley VM, Ragland RL, Durkin SG, Warren ST, et al.: Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. Am J Hum Genet 2009 Mar;84:339–350.

29. Sprang G: The traumatic experiences inventory (TEI): A test of psychometric properties. J Psychopathol Behav Assess 1997 Sep;19:257–271.

30. Nievergelt CM, Maihofer AX, Mustapic M, Yurgil KA, Schork NJ, Miller MW, et al.: Genomic predictors of combat stress vulnerability and resilience in U.S. Marines: A genome-wide association study across multiple ancestries implicates PRTFDC1 as a potential PTSD gene. Psychoneuroendocrinology 2015 Jan;51:459–471.

31. STARRS: Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS): A Partnership Between NIMH and the U.S. Army [Internet]. [cited 2016 Feb 18].Available from: http://www.nimh.nih.gov/health/topics/suicide-prevention/suicide-prevention-studies/army-study-to-assess-risk-and-resilience-in-servicemembers-army-starrs-a-partnership-between-nimh-and-the-us-army.shtml

32. Daalen E van, Kemner C, Verbeek NE, Zwaag B van der, Dijkhuizen T, Rump P, et al.: Social responsiveness scale-aided analysis of the clinical impact of copy number variations in autism. neurogenetics 2011 Nov 1;12:315–323.

33. Lionel AC, Crosbie J, Barbosa N, Goodale T, Thiruvahindrapuram B, Rickaby J, et al.: Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. Sci Transl Med 2011 Aug 10;3:95ra75.

34. Maziade M, Raymond V, Cliche D, Fournier JP, Caron C, Garneau Y, et al.: Linkage results on 11Q21-22 in Eastern Quebec pedigrees densely affected by schizophrenia. Am J Med Genet 1995 Dec 18;60:522–528.

35. Nakabayashi K, Komaki G, Tajima A, Ando T, Ishikawa M, Nomoto J, et al.: Identification of novel candidate loci for anorexia nervosa at 1q41 and 11q22 in Japanese by a genome-wide association analysis with microsatellite markers. J Hum Genet 2009 Sep;54:531–537.

36. St Clair D, Blackwood D, Muir W, Carothers A, Walker M, Spowart G, et al.: Association within a family of a balanced autosomal translocation with major mental illness. Lancet 1990 Jul 7;336:13–16.

37. Parsons RG, Ressler KJ: Implications of memory modulation for post-traumatic stress and fear disorders. Nat Neurosci 2013 Feb;16:146–153.

38. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al.: A copy number variation morbidity map of developmental delay. Nat Genet 2011 Sep;43:838–846.

39. Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, et al.: An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med Off J Am Coll Med Genet 2011 Sep;13:777–784.

40. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al.: Functional impact of global rare copy number variation in autism spectrum disorders. Nature 2010 Jul 15;466:368–372.

41. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al.: Excess of rare, inherited truncating mutations in autism. Nat Genet 2015 Jun;47:582–588.

42. Goes FS, McGrath J, Avramopoulos D, Wolyniec P, Pirooznia M, Ruczinski I, et al.: Genome-wide association study of schizophrenia in Ashkenazi Jews. Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet 2015 Dec;168:649–659.

43. Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K, et al.: High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome Res 2009 Sep;19:1682–1690.

44. Coe BP, Girirajan S, Eichler EE: The genetic variability and commonality of neurodevelopmental disease. Am J Med Genet C Semin Med Genet 2012 May 15;160C:118–129.

Chapter 4: Enrichment of clinically significant copy number variants in a medically underserved population

INTRODUCTION

Human copy number variation (CNV) is a DNA dosage change that can take the form of a deletion (0 or 1 copies) or duplication (3 copies) and range in size from 1 kilobase (kb) to several megabases (Mb) [1]. Large deletions and duplications can be pathogenic and are typically rare in the general population [2]. Recurrent pathogenic CNVs mediated by recurrent segmental duplications are classified as "genomic disorders" (GD). GD are frequently associated with congenital anomalies, failure to thrive, developmental delay, intellectual disability, autism, and epilepsy, although many GD exhibit variable penetrance and expressivity [3–5]. These same CNVs can be associated with later-onset psychiatric syndromes including schizophrenia and bipolar disorder [6,7]. In a number of research studies, CNVs have been associated with these neurodevelopmental phenotypes by comparing clinically ascertained cohorts ("cases") with selected control populations. These controls are often not sampled in an epidemiological way and are typically higher functioning than the general population. Studies of unselected individuals reveal a more nuanced picture. For example, a recent study involving approximately 16,000 participants from Estonia, the United Kingdom, Italy and Minnesota (USA) found that CNVs ($\geq$250kb) associated with intellectual disability were undiagnosed but present among adults in this sample [8]. Adults with

these CNVs had lower educational attainment than those without, consistent with phenotypic consequences for carriers.

The Grady Trauma Project (GTP) was launched in Atlanta, GA in 2005 to determine the relative contribution of genetic risk factors to psychiatric disorders, particularly posttraumatic stress disorder (PTSD), in an underserved and impoverished urban population [9–13]. Ascertainment is ongoing, with current enrollment at nearly 10,000 study participants. Over 90% of individuals in the GTP report witnessing or experiencing a traumatic event, and 57% live below the poverty line, adding to the burden of life stressors. As such, this population is at high risk for PTSD. In 2010, our research team initiated a genome-wide association study (GWAS) to identify genetic variation (single nucleotide polymorphism (SNPs) and CNVs) that associate with psychiatric disorders in the GTP study population.

During our investigation of CNVs present among 4,100 unrelated study participants, we identified an unexpected number of large genomic deletions and duplications, and an apparent enrichment of GD-related CNVs. To formally assess this burden, we selected recurrent CNVs identified as pathogenic among three studies. In 2011, Cooper et al. reported 33 CNV regions (deletions and duplications) that were found significantly enriched among children with intellectual disability and various congenital defects (N=15,767) compared to unaffected adults controls (N=8,329) [3]. Using a similar case-control analysis approach to determine pathogenicity of recurrent deletions and duplication CNVs, Kaminsky et al. identified 21 CNV regions that were differentially represented between cases (N=15,749) and healthy controls (N=10,188) [14]. With the exception of a few non-recurrent CNVs reported only by Cooper et al., all the CNVs

reported in both the Cooper and Kaminsky papers were also identified as deleterious variants in the most recent and well powered analysis (N=31,516 cases; N=13,696 controls) performed by Moreno-De-Luca et al. [15].

Each of the GD-related CNV intervals reported by Moreno-De-Luca was evaluated in the current study. We included additional intervals recently recognized as pathogenic, for a total of 43 GD CNVs. This manuscript describes the prevalence of these CNVs in the GTP sample and in an equivalently sized and racially similar comparison sample (N=3,883). Differences in educational attainment and relationship status between pathogenic CNV carrier and non-carriers within our study cohort are also reported. We discuss the public health implications of our findings with an emphasis on the importance of access to genetic testing for early detection of the potential health risks associated with these CNVs.

MATERIALS AND METHODS

Selection of genomic disorder regions

Several studies have classified CNVs as pathogenic based on a statistically significant difference in prevalence among individuals with and without developmental delay, intellectual disability, congenital anomalies, epilepsy, autism, schizophrenia and other health disorders [3,14,15]. A comparison of the CNVs identified in each of the studies is provided in Supplementary Table 4-1. The most comprehensive list is provided by Moreno-De-Luca et al. [15]. In the current study we investigated the prevalence of all the CNVs reported by Moreno-De-Luca. After consultation with a clinical cytogeneticist

(MKR), we included additional regions that have recently been identified as pathogenic, including chromosome 2q11.2-2q13 and 2q13 [16]. We also included the 15q11.2 (BP1-BP2) deletion, which associates with several neurodevelopmental disorders [17,18] and the 10q23 duplication that can result in delayed speech and motor development [19]. Non-recurrent CNVs: 1p36, 2q37, 4p14, 9q34, and 17p13 have a range of breakpoints and sizes and were not included in our analysis. We investigated a total of 24 intervals, which include 43 pathogenic CNVs (26 deletions and 17 duplications) (Supplementary Table 4-6).

GTP sample

The GTP study cohort has been recruited over a 10-year span (from 2006 to 2015) from waiting rooms of an urban public hospital primarily serving African American individuals of low socioeconomic status (SES) in Atlanta, GA (study approved by Emory University IRB # 00002114). The majority of study participants were patients waiting to be seen by a health care provider; others were visitors accompanying the patient to their appointment. The study participants were not assessed for general physical health in our study, but were mostly outpatients visiting with a primary care doctor or attending an obstetrics/gynecology (OB/GYN) appointment. To our knowledge, subjects were not receiving direct treatment for any known genetic abnormality, and the clinics visited do not routinely perform any genetic testing. Approximately 20% of recruited participants volunteered to be screened further. These individuals provided blood samples and were interviewed by study personnel about their trauma exposure and symptoms of psychiatric

illness. No participant ever self-reported prior diagnosis with a genomic disorder, though this was not specifically asked by the interviewers.

Study participants were asked to complete several data collection instruments (questionnaires) related to demographics, trauma exposure, and mental health history. Survey measures are as previously described [20]. Participants who were incapable of completing questionnaires were excluded from the study. Most participating individuals in the GTP are low-income African American females. Many have been exposed to recurrent trauma across their life span, including high rates of childhood maltreatment. Individuals in this population also report disproportionately high rates of mental health disorders including depression and PTSD. A detailed description of the demographics for this study population (N= 9,553) is provided in Supplementary Table 4-2.

DNA sample collection and microarray processing

From our GTP study participants, we collected a whole blood sample (in a Vacutainer collection tube containing EDTA; Becton, Dickinson and Company, Franklin Lakes, NJ) and/or a saliva sample (in an Oragene DNA Collection Kit; DNA Genotek, Inc., Ontario, Canada). DNA was extracted from blood using the E.Z.N.A. Mag-Bind Blood DNA Kit (Omega Bio-Tek, Inc., Norcross, GA) or from saliva using the Agencourt DNAdvance Nucleic Acid Isolation Kit (Beckman Coulter, Inc., Brea, CA).

From 2010 to 2015, existing study participants who completed the modified PTSD symptom scale (mPSS) and from whom we were able to collect a sufficient quantity of DNA (> 400ng) were genotyped using the Illumina HumanOmni1-Quad

BeadChip (Illumina, Inc. San Diego, CA). Independent DNA samples for each study participant were processed using a single chamber on the microarray.

GTP sample quality control (QC)

6,858 participant DNA samples were processed on arrays and our standard QC pipeline was applied, as follows: samples with a SNP call rate < 98% were excluded (447 removed). 68 samples with discordant gender (between participant reported gender and that determined by array) were removed. Relatedness was determined by pairwise IBD (identity by descent) estimations in PLINK [21]; one member of a pair was removed when pairwise proportion IBD (pi_hat) was > 0.125 (identical individuals, or first or second degree relatives). 1,477 individuals were removed. This high number is largely attributable to individuals enrolling in the study multiple times as well as the enrollment of closely related individuals.

To control for the effects of geographical ancestry, we conducted principal component analysis, on pruned autosomal data, to identify a sample with minimized ancestral differences (3 standard deviations from the medians of the first and second principal components) [22,23]. As a result of this step, an additional 373 samples were removed. After this round of sample quality control, our GTP cohort consisted of 4,493 unique and unrelated African American study participants with high quality genotype data.

Comparison cohorts

In order to avoid issues of population stratification and obtain a sample that was matched to the GTP sample for population of origin, we thoroughly examined dbGaP (http://www.ncbi.nlm.nih.gov/gap) and identified three eligible studies that met two key criteria: 1) The sample consisted of a large number of self-identified African Americans, and 2) There were raw genome-wide data derived from an Illumina SNP array with sufficient density that we were well powered (equally powered as in the GTP sample) to detect any of the defined set of GD CNVs. The three studies and the number (N) of study participants self-identified as African American are as follows: 1) The Health and Retirement Study (HRS) (N=1,665; dbGaP study accession: phs000428.v1.p1), 2) A Multiethnic Genome-wide Scan of Prostate Cancer (N=2,751; dbGaP study accession: phs000306.v4.p1), and 3) The Multiethnic Cohort (MEC) Breast Cancer Genetics study (N=627; dbGaP accession: phs000517.v3.p). These studies used either the Illumina Human1M-Duo or Illumina HumanOmni2.5 SNP arrays. In total, the three studies combined provided data for 5,043 samples.

Comparison sample quality control

Using a QC pipeline identical to what was used in the GTP samples, we removed samples from the dbGaP studies that had a genotype call rate of < 98%. We also excluded singletons among related pairs with proportion IBD of > 0.125 and samples with excessive heterozygosity due to potential sample contamination. For the Health and Retirement Study we removed 74 samples for low call rate, 25 samples for relatedness and 2 samples for high heterozygosity (101 in total). A total of 179 samples were

removed from the Prostate Cancer study: 129 for low call rate and 50 for relatedness. Finally, among samples from the Breast Cancer study 70 were removed for call rates < 98%, 7 were removed due to relatedness, and 1 was removed for high heterozygosity (78 in total).

CNV detection and quality control of CNV calls

To ensure uniformity in CNV calling for all cohorts, our research team performed CNV calling and QC equivalently across all studies. For autosomal chromosomes, microarray intensity data at each probe were processed using the Genome Studio Software, (Illumina, Inc.) along with the manufacturer supplied SNP clustering file, to generate values used to infer DNA copy number (Log R Ratio (LRR) and B Allele Frequency (BAF)) [24]. LRR and BAF values were exported directly from Genome Studio, and used by the CNV calling algorithm PennCNV for identifying CNVs [25]. CNV calls were restricted to those spanning at least ten probes and 10 kb as similar criteria have been demonstrated to yield accurately called CNVs [26]. Then, to further improve the specificity of the CNVs identified, we removed samples with a Log R ratio standard deviation, B allele frequency drift, waviness factor, or number of total CNVs greater than the third quartile plus 3 times the interquartile range (Q3 + 3*IQR) for any of the four measures. In this step we removed an additional 371 samples from GTP, 117 from HRS, 355 from PC, and 33 from BC.

Each GD CNV identified by our bioinformatics pipeline was graphed in R using the LRR and BAF values. The deleted or duplicated regions associated with pathogenicity are highlighted in pink (Supplementary Figure 4-2). Using the graphed

data, a clinical cytogeneticist performed a visual assessment of the CNVs. A CNV was considered valid if the mapped breakpoints were consistent with a GD CNV. To ensure that there was enough probe coverage on the arrays used for the comparison cohorts to detect the GD CNVs identified in the GTP cohort we quantitated the number of probes for each of the CNV regions. One terminal CNV interval, 22q13.3, contained 27 probes (well over the 10 probe threshold required by our pipeline to make a CNV call); all other regions had coverage of > 150 probes (Supplementary Table 4-3).

qPCR validation of CNVs

Although we had high confidence in the validity of our pipeline based on visual inspection of the called CNVs intervals (Supplementary Figure 4-2), we sought to experimentally validate both deletions and duplications in four individuals at the 1q21.1 interval (Supplementary Figure 4-3). The PerfeCT SYBR Green SuperMix (cat. # 95055-100, Quanta Biosciences, Inc., Gaithersburg, MD) was used as the source of Taq polymerase, Sybr green, and dNTPs in each qPCR reaction. Experimental primer pairs were designed for three regions across the CNV. Control primers, assaying a region of normal copy number, were obtained from the Type-it CNV Sybr Green PCR kit (cat. # 206672, Qiagen, Inc., Valencia, CA) and were used to normalize input DNA. To determine a reference two-copy ΔCt value we performed qPCR on a normal copy DNA at the region of interest. Additionally, a positive control DNA (also two-copy number) was used to show reliability of our method and calculations. Validity of the qPCR results was measured by determining the slope of ΔCt values for a spread of input DNA concentrations (2.5– 20ng). A reaction was considered valid if the slope (absolute value)

of ΔCt by log DNA input was <0.1. Since not all primer pairs worked equally well, we chose to use one of the three primer pairs based on consistency to yield valid PCR results. Primer sequence (F): TGTTGAACTGTATCAAGTTGGTATG, (R): AGAATCTGTGCATATAGTGAGTACA. Four out of four (100%) CNV calls, including two deletions and two duplications, were validated for a CNV region spanning approximately 1Mb.

Ancestry analysis

After merging the samples from the GTP and 3 dbGAP studies (Prostate, Breast Cancer and Health Retirement) that survived the removal cut-offs for call rate, relatedness and CNV data quality, we then examined the cohort to confirm that samples were of similar ancestry. SNPs that had a call rate < 95% and a minor allele frequency (MAF) < .05 were removed. We used PLINK to prune the autosomal data in windows of 50 base pairs, removing one SNP from each pair of SNPs with r^2>0.05 to obtain a set of roughly independent markers [21]. 60,359 SNPs were retained. Principal-component analysis (PCA) was performed using R software (https://stat.ethz.ch/R-manual/R-devel/library/stats/html/prcomp.html) to infer axes of ancestry and remove outlier subjects [22]. Based on PCA, we retained African-American subjects who fell within three standard deviations of the medians of the first and second principal components (PCs) in our sample (samples inside the green lines; Supplementary Figure 4-1). The numbers of samples removed for each study were: GTP (22), HRS (56), PC (186), and BC (55). Using the final 7,983 samples (N=4,100 for GTP; N= 3,883 for HRS, BC, and

PC) we obtained principal components for each sample to control for ancestry in our statistical analysis of CNV burden (described below).

Statistical Analysis

Using R (https://www.r-project.org/), we tested for enrichment of GD CNVs in the GTP samples compared to the combined dbGaP samples using logistic regression adjusted for residual ancestry using principal components. To determine which principal components to include in our analysis, we tested for significance of eigenvalues using Tracy–Widom statistics [27]. Based on the results of this test, we included only the first principal component as a covariate in our primary logistic regression analysis. Results of an analysis using the first 10 principal components is provided in Supplementary Table 4-4. Pearson Chi-square tests were used to test for differences in various demographic measures between those in the total study cohort (N=4,100) with (N=72) or without (N=4,028) a GD CNV. We also show these measures for the larger GTP cohort (N= 9,553) and the GTP study cohort analyzed in the current study (N=4,100) (Supplementary Table 4-5). These latter analyses were performed using the SPSS statistical software package (IBM Corporation).

RESULTS

Twenty-four GD-related CNV regions were assessed in the current study (Table 4-1). GD CNVs within some of these regions were subdivided based on variable breakpoints, resulting from several clusters of segmental duplications in the region. These were carefully classified and reported separately based on the critical regions and

description in the literature. As a result, we assessed the presence of 26 deletions and 17 duplications in the GTP and dbGAP cohorts (Supplementary Table 4-6). For some CNV regions, the deletion is pathogenic but the reciprocal duplication is not enriched in cohorts of affected individuals and considered likely benign. Therefore, the duplication was not considered a GD CNV and was not analyzed in our study. Each of the pathogenic CNVs reported in this manuscript were examined by a clinical cytogeneticist and deemed consistent with a CNV that would be reported in a clinical setting. Additionally, each of the CNVs were graphed for visual confirmation (Supplementary Figure 4-2) and one CNV was validated using qPCR (Supplementary Figure 4-3). A complete list of the CNVs and the de-identified individuals in which they were identified is provided in Supplementary Table 4-7.

We identified 72 CNVs (49 deletions and 23 duplications) within our GTP sample and 30 CNVs (23 deletions and 7 duplications) in the three combined dbGaP samples (Table 4-2). We tested the significance of this enrichment using a logistic regression model, where we adjusted for residual ancestry difference using the first principal components as a covariate. We found that individuals within the GTP cohort are 2.22 times more likely (95% CI: 1.45-3.49; $p=3.2 \times 10^{-4}$) to have a GD CNV than are individuals within the combined cohort of the dbGaP studies (Table 4-3). When compared independently, deletions and duplications were each statistically enriched in the GTP study participants with odds ratios of 1.99 (95% CI: 1.21-3.37; $p= 8.0 \times 10^{-3}$) and 2.94 (95% CI: 1.31-7.46; p=0.01), respectively (Table 4-3). These latter results suggest that the increased burden of pathogenic CNVs is not driven solely by the number of deletions or duplications only. To confirm that pairs in the GTP cohort with the same

CNV deletion or duplication were unrelated, we estimated proportions of pairwise IBD specifically among the GD CNV carriers. All pairs of individuals with a shared CNV had an IBD proportion of < 0.035 (pi_hat), revealing that they are indeed not close relatives (Supplementary Table 4-8).

We also confirmed that the enriched prevalence of CNVs in the GTP cohort was not simply the result of shared ancestry. Incidence rates for GD CNVs are based on predominantly Caucasian study populations [3,4]. For the current study, we sought to use a comparison cohort of African Americans, similar to our GTP study cohort. A matrix of genome-wide identity by state (IBS) pairwise distances was used for visualization of population substructure for the GTP cohort, the 3 dbGaP cohorts (Prostate Cancer, HRS, and Breast Cancer), and 9 populations of Hapmap samples (Figure 4-1 A). As evident by the similar clustering of the GTP and three dbGaP samples, the study and comparison population assessed for CNVs are all of similar ancestry. Thus, geographical ancestry does not explain the difference in the prevalence of CNVs in our GTP cohort compared to the dbGaP comparison cohorts. The clusters of samples with GD CNVs show a similar pattern (CNV carriers in red and blue, Figure 4-1 B).

Next, we examined the differences in characteristics between individuals in our cohort who have a GD CNV (N=72) and study participants without these CNVs ($N \leq$ 4,028; some participants lack phenotype data) (Table 4-4). Only 16.9% of GD CNV carriers report having greater than a 12^{th} grade education or GED, compared to 38.2% of non-GD CNV carriers in the GTP ($p = 2.4 \times 10\text{-}4$). Conversely, a greater percent of GD CNV carriers attained less than a 12^{th} grade education compared to non-GD CNV carriers (31% compared to 21.8%, $p = 0.065$). Although this latter finding is not statistically

significant, it is consistent with the overall trend of lower educational attainment for GD CNV carriers. These results indirectly support our assumptions that some or all of these 72 CNV carriers may have somewhat impaired cognitive function, and replicate findings from a prior study in a different population [8]. We also found that GD CNV carriers are more likely to report being single or never married than non-GD CNV carriers (73.2% vs. 58.6%; p=0.013). The observed difference in relationship status may be attributed to functional impairments typically associated with GD CNVs, including intellectual disability and psychiatric disorders. This phenomenon has been previously reported for adults with the 22q11.2 deletion [28]. The number of females in the group with a pathogenic CNV (79.2%) is slightly higher in comparison to the other group (72.9%) but the difference is not significant. We attribute this difference to the fact that there were a greater percentage of study participants, most likely female, recruited from OB/GYN in the GD CNV group compared to the non-GD CNV group (55.4% vs. 50.7%, respectively.)

DISCUSSION

In this study, we found that the subjects recruited from an urban, impoverished, high-risk population are 2.2 times more likely to harbor a GD CNV than individuals from another population with similar ancestry who participated in other genetic studies. Many (38.2%) of the patients in the GTP cohort were recruited during a visit to their primary care doctor. This ascertainment strategy could introduce a source of bias into our study. Although the primary clinic typically sees patients for routine well visits or minor, acute illness, it is possible that recruitment from this location results in the ascertainment of a

non-representative sample of the population; one that is "sicker," more likely to be seeking medical care, and therefore more likely to be harboring a CNV. To assess that possibility, we compared rates of recruitment from the primary care clinic as well as patient status (patient vs. visitor) between groups with and without one of the 72 CNVs. We determined that there was no statistically significant difference between these groups relative to either primary care recruitment ($p=0.46$) or patient status ($p=0.45$) (Table 4-4), suggesting that being a CNV carrier is not associated with being a patient at the primary clinic.

We wanted to be certain that the enrichment of GD CNVs was not due to inclusion of related individuals and inheritance of certain CNVs. This is of particular concern for our cohort given that the majority resides in a localized, inner-city region of Atlanta, GA. Thus, to ensure that the CNV carriers described in the manuscript are not related, we have used genotypes across the genome to calculate the proportion of alleles shared identical by descent (IBD). The proportions generated between all pairs of the 72 CNV carriers range from (0.0 – 0.035), confirming that these individuals are indeed unrelated (Supplementary Table 4-7).

While we do not know the reason for the increase in clinically significant CNVs in our study population, we are confident that neither geographical ancestry nor relatedness contributes significantly to the burden of these rare and potentially pathogenic CNVs in the GTP cohort. One commonality among the study participants is that the majority live in poverty (57.8 % based on a monthly income of < $1,000), has a 12^{th} grade education or lower (62.2%), and is unemployed (68.4%). Each of these variables are factors that associate with socioeconomic status (SES) [31]. Thus, it is possible that

the health outcomes commonly associated with these CNVs, such as intellectual disabilities and neuropsychiatric disorders, indirectly increase the risk of living in a community of low socioeconomic status [32]. This downward-drift hypothesis is well supported by the scientific literature on psychiatric disorders, particularly severe disorders such as schizophrenia, which impair ability to perform basic skill of daily living [33–35]. Of further note, a recent population-based study that examined the association between genetic diversity in populations across the world and economic development found no statistically significant relationship. Thus it is unlikely that, alone, the shared genetic "make-up" of the GTP cohort, irrespective of the CNV burden, is responsible for their economic outcome [36].

Whether or not a relationship exists between CNV carrier status and SES, there is a well-documented correlation between carrier status and health outcomes among those of low SES. This phenomenon is known as the social causation hypothesis [33]. Research reveals that those with intellectual disability, an outcome common among the CNVs described here, and living in low SES are more likely to suffer poor health [37]. Additionally, the prevalence of schizophrenia (another outcome associated with the CNVs identified in our sample) is also higher among those from urban environments and of lower SES [38,39]. A recent report also shows a higher prevalence of mental disorders among low income children [40,41]. Put bluntly, those within our impoverished study population who harbor a genomic disorder CNV have an even greater likelihood of negative health outcomes compared to other members of this community with similar health disparities but less genetic risk.

Adults diagnosed with the genetic variants discussed in this manuscript may

benefit from therapeutic care. Additionally, carriers of the pathogenic CNVs have a 50% risk for passing the variants to their biological children, who could receive and benefit from early healthcare intervention. However, genetic tests that are standard-of-care and consistent with recommendations by the American College of Medical Genetics and Genomics (ACMG) may not be routinely deployed in this population [42–44]. Children with genetic disorders are eligible for state-run early intervention services (e.g., Babies Can't Wait) and additional services are available for special needs children in low-income households (Children's Medical Services). Programs also exist for transitioning special-needs individuals from pediatric to adult care (e.g., GA-PEACH-T). The individuals in this study, had they been diagnosed, could have accessed a range of services, some of which are specifically designed to provide healthcare access for those in low-resource settings. Similar services are available to children across the United States. Each state has a Title V program, funded by the Maternal and Child Health Bureau that provides services to children with special health care needs [45]. Thus, similar populations across the US who may also have an increased burden of genomic disorder CNVs, readily identifiable by access to genetic testing, are also eligible for services that can improve long term health.

Data from adults with fetal alcohol syndrome (FAS) show that early access to services can have a profound and lasting effect on outcome. In a recent study, individuals with FAS who were mildly affected as children and did not qualify for a diagnosis, scored worse as adults on tests of adaptive functioning and entry into adult roles than individuals with severe FAS [46]. Authors of these studies hypothesize that early diagnosis and access to services is protective against negative outcomes, paradoxically

leaving the less severely affected at higher risk for poor functioning later in life [47]. Similar studies of pathogenic CNVs do not yet exist, but it is likely that early access to services would improve long-term outcome for individuals with GDs. It is critical that in the vulnerable population represented by low SES communities, the standard-of-care and professional clinical recommendations are uniformly applied so that eligible individuals can receive services and have the best chance at improved outcomes.

In the current study, we do not have data about prior CNV diagnosis, nor do we ask about previous referrals for genetic testing. However, for a subset of participants (~20%) we collected a more extensive medical history, and none of these subjects self-reported a genetic diagnosis. We did not collect data allowing us to assess whether these individuals would have met the criteria for genetic testing referral. Since many of the genomic variants we identified have incomplete penetrance or variable expressivity, it is possible that some CNV carriers in our population may be asymptomatic or mildly symptomatic such that manifestation of the variant has gone undetected [48]. However, it is highly likely that these individuals may experience some negative health effect as a result of the CNV and are also unaware of their CNV status (supported by the non-increase in receipt of disability support among CNV carriers). Recently, Männik *et al* examined the prevalence of rare CNVs among 16,000 adults from Estonia, the United Kingdom, Italy and Minnesota (USA) [8]. This cohort had not been previously tested for genomic disorder CNVs. However, the study reports a statistically significant association between the presence of rare CNVs > 250 kb and level of education (less than a secondary education). Interestingly, our results suggest this key finding holds in the GTP study participants, as we too observed a measurable impact of rare CNVs on

educational attainment.

It has previously been documented that 22q11.2 deletion syndrome is under-diagnosed in African American (AA) populations [49]. These diagnostic challenges are likely due to a slightly different constellation of presenting clinical features compared to Caucasian populations [50] and diminished characteristic craniofacial dysmorphology of the syndrome in AA patients [49]. An absence of characteristic craniofacial dysmorphology for AA patients has also been noted for Prader-Willi syndrome [51], possibly leading to under diagnosis. The results of the current study suggest that this finding may extend to other CNV syndromes as well, and under-diagnosis of clinical genetic syndromes in AA patients may be more widespread than is currently appreciated. Prior studies of 22q11.2 deletion syndrome and Prader-Willi syndrome advanced the idea that for clinical purposes, the "clinicians threshold of suspicion…should be lower in AA individuals" [51] with some hallmark symptoms, even if facial features are not present. We believe this suggestion bears adopting for genetic testing of all syndromes in African American populations.

It is possible that CNV syndromes are less likely to be diagnosed among medically underserved populations living in low-resource settings in the United States. Recent advances in whole-genome analysis are leading to new understanding of the genetic risk factors associated with or causative of disease. Access to affordable technology has led to more in-depth genetic inquiry and a remarkable wealth of data. It is critical that these advances be uniformly applied, and that all populations receive access to genetic testing, particularly when early intervention can improve outcome and offer access to services. Genetic findings can be used to minimize existing health disparities

that are already high among minority populations of medically underserved inner-city individuals, but only if there is access to genetic testing for populations of low socioeconomic status.

The possible public misunderstanding or misuse of these findings is of some concern. Studies show that the public is not knowledgeable about genetics [52,53] and that there is a general public endorsement for genes as the causes of both health and social outcomes [54]**.** It is therefore important to consider the public's possible misunderstanding of data that could be misinterpreted to support the idea that genetic differences in a poorer population, especially when those differences are associated with developmental and intellectual deficits and mental illness, are the explanation for the poverty of the community examined. For that reason, it is important to emphasize that though these findings are statistically significant and clinically relevant, the CNVs are rare in an absolute sense even in the subject population, being identified in less than 2% of the subjects tested, and that the findings show no racial differences.

Table 4-1. Genomic disorder CNV regions assessed in the study and control cohorts

Chromosome Region	Coordinates (hg19) *	Phenotypes/Syndrome***	Reference(s)
1q21 (TAR)	chr1:145401254-145928123	Thrombocytopenia-absent radius	[233,245,278]
1q21.1	chr1:146577487-147394506	DD, ASD, congenital heart disease, SZ	[107,232,233,245,279]
2q11.2	chr2:96545351-98013866**	DD, dysmorphic features, ADHD	[246]
2q11.2-2q13	chr2:107100000-113065779**	DD, hypotonia, dysmorphic features	[246]
2q13	chr2:111442131-113065779**	Congenital heart defects, dysmorphic features, hypotonia, cognitive impairment, ASD	[246]
3q29	chr3:195756054-197344665	ASD, SZ, learning disabilities, heart defects, anxiety disorders and depression, GI disorders	[233,245,280–283]
5q35	chr5:175728978-177013961	Sotos syndrome	[232,233,245,284]
7q11.23	chr7:72744454-74142513	Williams-Beuren, SZ	[107,285,286]
8p23.1	chr8:8119295-11765719	ID, microcephaly, facial dysmorphisms, cardiac anomalies, behavioral problems	[233,245,287]
10q23	chr10:81641918-88828018	Infantile juvenile polyposis, ID, macrocephaly, gastrointestinal symptoms, early-onset colorectal cancer	[232,245,249,288]
15q11.2 (BP1-BP2)	chr15:22765628-23300287	SZ, other neuropsychiatric disorders	[107,232,247]
15q11.2-q13 (BP2-BP3)	chr15:23758390-28557186	Angelman/Prader-Willi syndromes, SZ	[107,232,233,245,289]
15q13.2-q13.3 (BP4-BP5)	chr15:31137104-32445408	DD, ID, epilepsy, autism, SZ	[107,232,233,245,290]

*Unless otherwise specified all coordinates were obtained from the ISCA database (dbsearch.clinicalgenome.org)

**Coordinates extrapolated from literature referenced

***Abbreviations: DD (developmental delay), ID (intellectual disability), SZ (schizophrenia)

Table 4-1. Genomic disorder CNV regions assessed in the study and control cohorts. *Cont.*

Chromosome Region	Coordinates (hg19) *	Phenotypes/Syndrome***	Reference(s)
16p13.11	chr16:15504454-16292268	DD, ID, epilepsy, ASD, congenital anomalies, SZ	[107,232,233,245,291]
16p12.1	chr16:21946524-22467284**	DD, ID, SZ	[107,232,245,292]
16p11.2 (distal)	chr16:28822499-29042499**	DD, ID, SZ	[107,232,245,293]
16p11.2	chr16:29649996-30199855	DD, ID, ASD, epilepsy, SZ	[107,232,233,245,294,295]
17p11.2	chr17:16757111-20219651	Smith-Magenis / Potocki-Lupski syndromes	[232,233,245,296,297]
17q11.2 (NF1)	chr17:29162822-30218667	Neurofibromatosis type I	[232,233,245,298]
17q12	chr17:34856056-36248918	Renal cysts and diabetes syndrome, DD, ID, epilepsy, SZ	[107,232,233,245,299]
17q21.31	chr17:43705165-44188442	DD, facial dysmophisms, epilepsy, heart defects	[232,233,245,300]
22q11.2 (proximal)	chr22:18661725-21561514	DiGeorge/Velo-cardio-facial syndrome, DD, ID, SZ	[107,232,233,245,301–304]
22q11.2 (distal)	chr22:22115848-23696229	DD, pre-term birth, growth restriction, cardiac defects, microcephaly, minor skeletal anomalies	[232,245,301,305]
22q13.3	chr22:51045517-51178945	Phelan-McDermid syndrome	[232,306]

*Unless otherwise specified all coordinates were obtained from the ISCA database (dbsearch.clinicalgenome.org)
**Coordinates extrapolated from literature referenced
***Abbreviations: DD (developmental delay), ID (intellectual disability), SZ (schizophrenia)

Table 4-2. Prevalence of assessed genomic disorder CNVs among unrelated African Americans in our study and comparison cohorts

Study cohort	Total genomic disorder CNVs	Genomic disorder CNV deletions	Genomic disorder CNV duplications
Grady Trauma Project (N=4100)	1.8% (N=72)	1.2% (N=49)	0.6% (N= 23)
Comparison cohorts			
Total (N=3,883)	0.8% (N=30)	0.6% (N=23)	0.2% (N=7)
A Multiethnic Genome-wide Scan of Prostate Cancer (phs000306.v4.p1) (N=2,031)	0.6% (N=13)	0.5% (N=11)	0.1% (N=2)
Health and Retirement Study (phs000428.v1.p1) (N=1,391)	0.9% (N=13)	0.6% (N=9)	0.3% (N=4)
Multiethnic Cohort (MEC) Breast Cancer Genetics (phs000517.v3.p1) (N=457)	0.9% (N=4)	0.7% (N=3)	0.2% (N=1)

Table 4-3. Odds ratios resulting from logistic regression analysis comparing CNV burden while controlling for ancestry

	GTP (N=4100)	Non-GTP (N=3883)	Odds Ratio	95% Conf. Interval	p-value
Total CNVs	72	30	2.22	1.45-3.49	3.2×10^{-4}
Deletions	49	23	1.99	1.21-3.37	8.0×10^{-3}
Duplications	23	7	2.94	1.31-7.46	0.01

Table 4-4. Comparison of demographics between GTP study participants with and without a deleterious CNV[a]

General Demographics (%)	GTP participants without a deleterious CNV (N = 4028)	GTP participants with a deleterious CNV (N = 72)
Female	72.9	79.2
Single or never married	58.6	73.2*
greater than a 12th grade education or GED	38.2	16.9**
Less than a 12th grade education	21.8	31.0
Unemployed	68.5	64.8
Monthly income below $1000/month	57.7	64.7
Monthly income below $500/month	31.3	36.8
Receiving disability support	18.8	23.9
Patients versus visitors	85.8	89.1
Patients visiting primary care doctor	38.3	33.8
Patients visiting OB/GYN	50.7	55.4

[a]Statistically significant by Pearson Chi-square between carriers and non-carriers of deleterious CNV: *p=0.013; **p =2.4×10^{-4}

Figure 4-1. Multidimensional scaling (MDS) of genome-wide IBS (identify by state) pairwise distances, identifying clusters of populations by ancestry. Population abbreviations in the legend: Prostate (dbGaP Prostate Cancer), HRS (dbGaP Health and Retirement Study), BreastCancer (dbGaP Breast Cancer Study), CEU (Utah residents with Northern and Western European ancestry), CHB (Han Chines in Beijing, China) CHD (Chines in Metropolitan Denver, Colorado), GIH (Gujarati Indians in Houston, Texas), JPT (Japanese in Tokyo, Japan), MEX (Mexican ancestry in Los Angeles, California), GTP (Grady Trauma Project), YRI (Yoruba in Ibadan, Nigeria), LWK (Luhya in Webuye, Kenya), ASW (African ancestry in Southwest, USA). **A.)** A comparison of ancestry between the GTP study cohort, dbGaP cohort and HapMap populations.

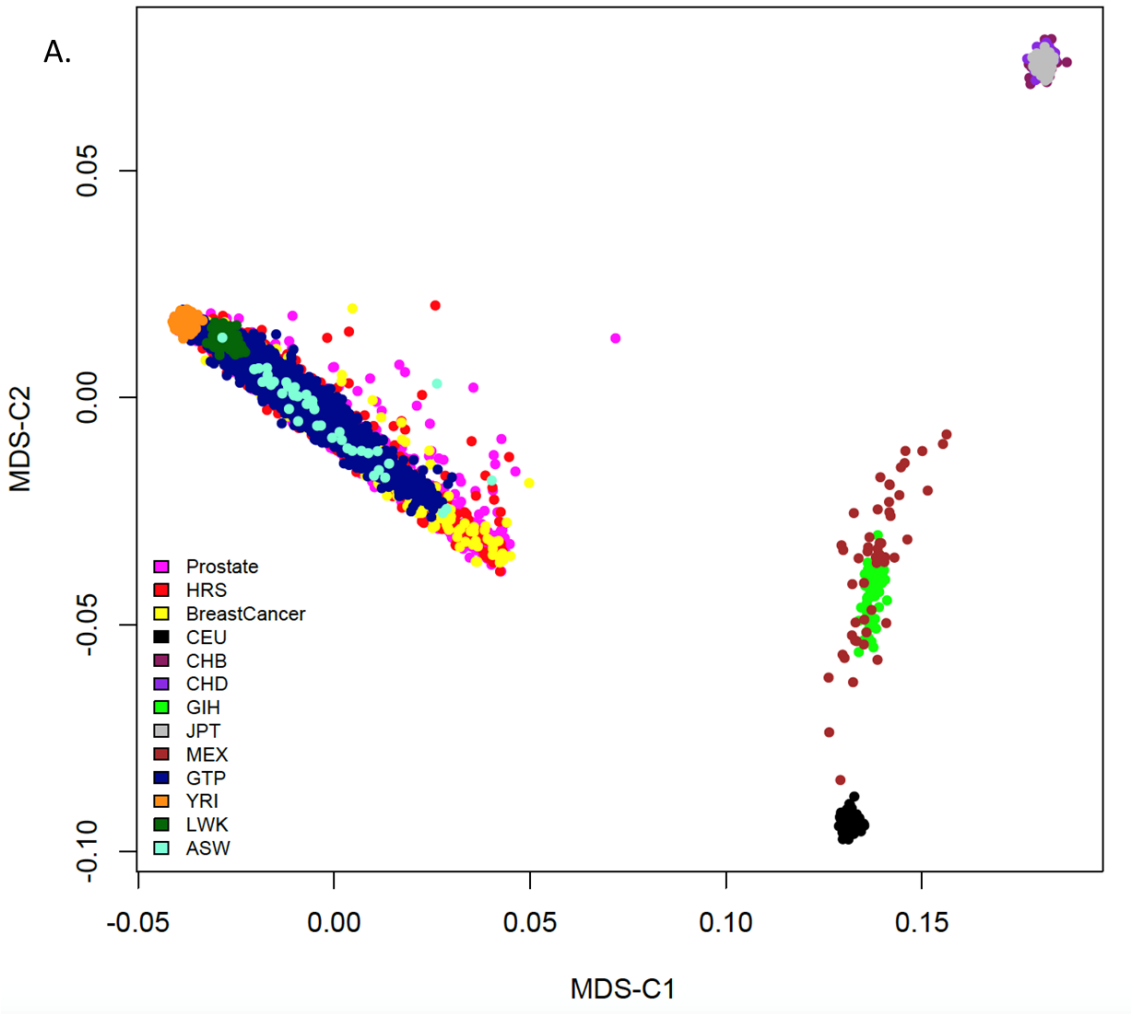
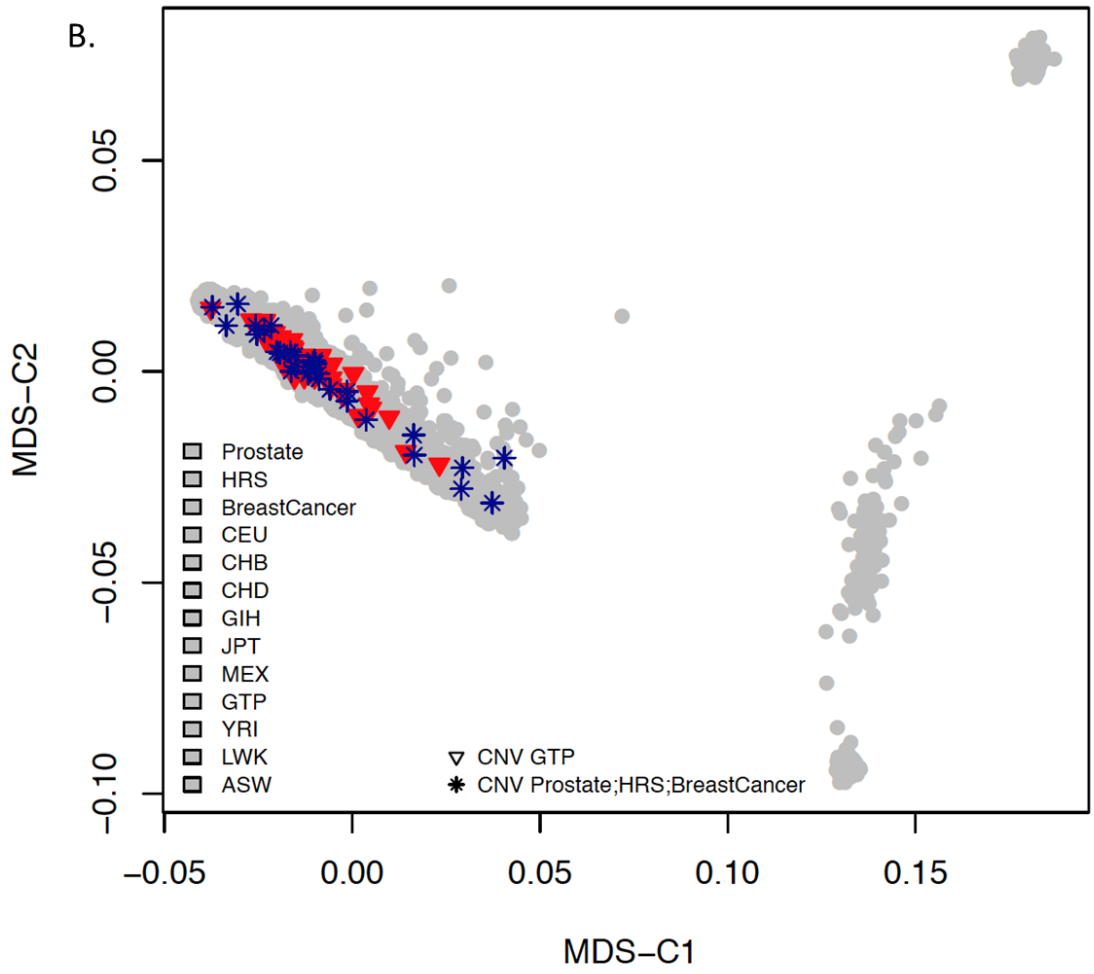


Figure 4-1. B) A comparison of ancestry between CNV carriers in the GTP study cohort (red triangles) and the dbGaP cohorts (red asterisks). All other populations are shown in gray.



Supplementary Table 4-1. Statistically significant CNVs identified in three separate studies

Cooper et al. (2011)	*Kaminsky et al. (2011)*	*Moreno-De-Luca et al. (2013)*
CNV-deletions (22)	**CNV-deletions (14)**	**CNV-deletions (19)**
1p36		
1q21.1	1q21.1	1q21.1
	1q21 (TAR)	1q21 (TAR)
2q37		
	3q29	3q29
4p16.3		
5q35	5q35	5q35
7q11.23	7q11.23	7q11.23
	8p23.1	8p23.1
9q34		
10q23		10q23
15q11.2 (*NIPA1*)		
15q13.3	15q13.2-q13.3 (BP4-BP5)	15q13.2-q13.3 (BP4-BP5)
15q11.2-q13 (BP2-BP3)	15q11.2-q13 (BP2-BP3)	15q11.2-q13 (BP2-BP3)
16p13.11	16p13.11	16p13.11
16p12.1		16p12.1
16p11.2 distal		16p11.2 distal
16p11.2	16p11.2	16p11.2
17p13.3		
17p11.2	17p11.2	17p11.2
		17q11.2 (NF1)
17q12	17q12	17q12
17q21.31	17q21.31	17q21.31
22q11.2	22q11.2	22q11.2
22q11.2 distal		22q11.2 distal
22q13.3		
CNV – duplications (11)	**CNV – duplications (7)**	**CNV – duplications (11)**
1p36		
1q21.1	1q21.1	1q21.1
2q13		
7q11.23	7q11.23	7q11.23
		8p23.1
15q11.2 - q13 (BP2-BP3)	15q11.2 - q13 (BP2-BP3)	15q11.2 - q13 (BP2-BP3)
		15q13.2-q13.3 (BP4-BP5)
15q13.3		
16p11.2 distal		16p11.2 distal
16p11.2	16p11.2	16p11.2
17p11.2	17p11.2	17p11.2
17q12	17q12	17q12
		22q11.2(distal)
22q11.2	22q11.2	22q11.2
cases/controls	**cases/controls**	**cases/controls**
(8,329/15,767)	(15,749/10,118)	(31,516/13,696)

Supplementary Table 4-2. Demographics of the Grady Trauma Project study participants (N=9,553)

Age (years)	
minimum	18
maximum	90
mean ± SD	40.0 ± 14.0
Gender (%)	
female	75.7
male	24.3
Self-reported ethnicity (%)	
African American	93
Hispanic	0.7
Asian	0.1
Caucasian	3.1
Mixed	1.8
Other	1.2
Highest grade completed (%)	
< 12th	21.7
12th or high school graduate	36
GED	5
some college or technical school	22.7
technical school graduate	4.5
college graduate	8.2
graduate school	1.8
Currently employed	
yes	32.9
no	67.1
House monthly outcome	
$ 0 - 249	21.3
$ 250 - 499	9.3
$ 500 - 999	26.5
$ 1000 - 1999	26.9
$ 2000 or more	16.1

Supplementary Table 4-2. Demographics of the Grady Trauma Project study participants (N=9,553). Cont.

Past drug or alcohol problem		
	yes	23.2
	no	76.8
Current drug or alcohol problem		
	yes	5.0
	no	95.0
Witnessed or experienced a traumatic event*		
	yes	90.7
	no	9.3
Experienced child abuse		
	yes	40.2
	no	59.8
Current PTSD		
	yes	23.4
	no	76.6
Lifetime PTSD		
	yes	48.9
	no	51.1
Current Major Depressive Disorder		
	yes	18.7
	no	81.3
Lifetime Major Depressive Disorder		
	yes	44.7
	no	55.3
Treated for Schizophrenia		
	yes	4.9
	no	95.1
Hopitalized for psychiatric disorder		
	yes	14.6
	no	85.4
Attempted suicide		
	yes	13.9
	no	86.1

*scored 1 or higher on Traumatic Events Inventory (TEI) excluding child abuse

Supplementary Table 4-3. Probe coverage by CNV region and microarray platform

CNV Region	Number Probes Human1M Duo	Number Probes HumanOmni 2.5
1q21.1 (TAR)	254	250
1q21.1	439	763
2q11.2	486	598
2q11.2-2q13	1906	3669
2q13	609	1015
3q29	722	1663
5q35	730	1058
7q11.23	635	795
8p23.1	1987	5073
10q23	2854	6048
15q11.2 (BP1-BP2)	162	312
15q11.2-q13 (BP2-BP3)	2328	4933
15q13.2-q13.3 (BP4-BP5)	486	1183
16p13.11	689	843
16p12.1	176	261
16p11.2	310	312
16p11.2 (distal)	162	155
17p11.2	1367	1950
17q11.2	454	735
17q12	681	1258
17q21.31	183	359
22q11.2 (proximal A-D)	1448	2219
22q11.2 (proximal A-B)	940	1506
22q11.2 (distal type I D-F)	906	1150
22q11.2 (distal type I D-E)	569	711
22q11.2 (distal type II E-F)	337	439
22q13.3	27	64

Supplementary Table 4-4. Odds ratios resulting from logistic regression analysis comparing CNV burden while controlling for ancestry using 10 principal components

	GTP (N=4100)	Non-GTP (N=3883)	Odds Ratio	95% Conf. Interval	p-value
Total CNVs	72	30	2.24	1.46-3.51	3.1×10^{-4}
Deletions	49	23	1.20	1.21-3.39	7.8×10^{-3}
Duplications	23	7	2.95	1.32-7.52	0.01

Supplementary Table 4-5. Comparison of demographics between GTP study participants in entire cohort and participants analyzed for CNVs

General Demographics (%)	All GTP participants (N = 9,553)	GTP participants analyzed for deleterious CNV (N = 4,100)
Female	75.7	73.0
Single or never married	58.4	58.8
greater than a 12th grade education or GED	62.7	62.2
Less than a 12th grade education	21.7	22.0
Unemployed	67.1	68.4
Monthly income below $1000/month	57.1	57.8
Monthly income below $500/month	30.6	31.4
Receiving disability support	19.8	18.9
Patients versus visitors	84.9	85.8
Patients visiting primary care doctor	38.8	38.2
Patients visiting OB/GYN	48.7	50.8

Supplementary Table 4-6. A list of the select CNVs assessed and the number identified among the four cohorts in the current study

Deletion Regions (26)	GTP AA (N=4100)	dbGAP Prostate Cancer AA (2031)	dbGAP Breast Cancer AA (461)	dbGAP HRS AA (1391)	dbGAP All AA (3883)
1q21 (TAR)	1	0	0	0	0
1q21.1	6	0	0	1	1
2q11.2	0	0	1	0	1
2q11.2-2q13	0	0	0	1	1
2q13	3	0	0	0	0
3q29	0	0	0	0	0
5q35	0	0	0	0	0
7q11.23	0	0	0	0	0
8p23.1	0	0	0	0	0
10q23	0	0	0	0	0
15q11.2 (BP1-BP2)	11	10	1	2	13
15q11.2-q13 (BP2-BP3)	0	0	0	0	0
15q13.2-q13.3 (BP4-BP5)	5	0	0	1	1
16p13.11	6	0	0	1	1
16p12.1	4	0	0	2	2
16p11.2 (distal)	0	0	0	0	0
16p11.2	7	1	1	0	2
17p11.2	0	0	0	0	0
17q11.2	0	0	0	0	0
17q12	3	0	0	0	0
17q21.31	0	0	0	0	0
22q11.2 (proximal A-B)	1	0	0	0	0
22q11.2 (proximal A-D)	1	0	0	0	0
22q11.2 (distal type I)	0	0	0	0	0
22q11.2 (distal type II; E-F)	1	0	0	1	1
22q13.3	0	0	0	0	0
Total deletions	**49**	**11**	**3**	**9**	**23**
% deletions	**1.2**	**0.5**	**0.7**	**0.6**	**0.6**

Supplementary Table 4-6. A list of the select CNVs assessed and the number identified among the four cohorts in the current study (Cont.)

Duplication Regions (17)	GTP AA (N=4100)	dbGAP Prostate Cancer AA (2031)	dbGAP Breast Cancer AA (461)	dbGAP HRS AA (1391)	dbGAP All AA (3883)
1q21.1	4	0	0	1	1
2q11.2	0	0	0	0	0
2q13	0	0	0	0	0
7q11.23	0	0	0	0	0
8p23.1	0	0	0	0	0
10q23	0	0	0	1	1
15q11.2-q13 (BP2-BP3)	1	0	0	0	0
15q13.2-q13.3 (BP4-BP5)	1	1	0	0	1
16p11.2 (distal)	2	0	0	0	0
16p11.2	6	0	1	0	1
17p11.2	0	1	0	0	1
17q12	2	0	0	0	0
22q11.2 (proximal A-B)	2	0	0	0	0
22q11.2 (proximal A-D)	3	0	0	1	1
22q11.2 (distal type I; D-E)	1	0	0	0	0
22q11.2 (distal type I; D-F)	1	0	0	0	0
22q11.2 (distal type III; F-H)	0	0	0	1	1
Total duplications	**23**	**2**	**1**	**4**	**7**
% duplications	**0.6**	**0.1**	**0.2**	**0.3**	**0.2**
Total CNVs	**72**	**13**	**4**	**13**	**30**

Supplementary Table 4-7. Complete list of all the pathogenic CNVs identified in the four study chohorts along with the start and end breakpoints predicted by PennCNV

CNV Region	CNV Type	Start coordinate *	End coordinate*	Approximate Size (kb)	Study ID	Study
1q21 (TAR)	Deletion	145394955	145762959	368	8636	Grady Trauma Project
1q21.1	Duplication	146035080	149205349	3170	7529	Grady Trauma Project
1q21.1	Duplication	146501348	147386452	885	10169	Grady Trauma Project
1q21.1	Duplication	146501348	147398560	897	7722	Grady Trauma Project
1q21.1	Duplication	146528706	147256257	728	5097	Grady Trauma Project
1q21.1	Duplication	146556821	147424344	868	5722643008_R03C01	Health and Retirement Study
1q21.1	Deletion	146501348	147531377	1030	10820	Grady Trauma Project
1q21.1	Deletion	146501348	147398560	897	12510	Grady Trauma Project
1q21.1	Deletion	146501348	147394004	893	12563	Grady Trauma Project
1q21.1	Deletion	146501348	147394004	893	12614	Grady Trauma Project
1q21.1	Deletion	146501348	147394004	893	6021	Grady Trauma Project
1q21.1	Deletion	146501348	147394004	893	8470	Grady Trauma Project
1q21.1	Deletion	146472911	147397041	924	5815065002_R03C01	Health and Retirement Study
2q11.2	Deletion	96517653	97871314	1354	EC014114	Breast Cancer
2q11-2q13	Deletion	107091864	113098440	6007	5617561229_R04C01	Health and Retirement Study
2q13	Deletion	111392259	112869041	1477	10440	Grady Trauma Project
2q13	Deletion	111392259	112869041	1477	12171	Grady Trauma Project
2q13	Deletion	111392259	113042446	1650	5202	Grady Trauma Project
10q23	Duplication	81577614	89112638	7535	5815225004_R02C01	Health and Retirement Study
15q11.2(BP1-BP2)	Deletion	22750305	23114440	364	5030	Grady Trauma Project
15q11.2(BP1-BP2)	Deletion	22750305	23109890	360	5121	Grady Trauma Project
15q11.2(BP1-BP2)	Deletion	22750305	23115431	365	7356	Grady Trauma Project
15q11.2(BP1-BP2)	Deletion	22750305	23096921	347	9516	Grady Trauma Project
15q11.2(BP1-BP2)	Deletion	22808811	23109890	301	8710	Grady Trauma Project
15q11.2(BP1-BP2)	Deletion	22832976	23086929	254	11387	Grady Trauma Project
15q11.2(BP1-BP2)	Deletion	22832976	23272733	440	12471	Grady Trauma Project
15q11.2(BP1-BP2)	Deletion	22832976	23086929	254	6097	Grady Trauma Project
15q11.2(BP1-BP2)	Deletion	22832976	23086929	254	8735	Grady Trauma Project
15q11.2(BP1-BP2)	Deletion	22835646	23086929	251	6820	Grady Trauma Project
15q11.2(BP1-BP2)	Deletion	22835646	23086929	251	6840	Grady Trauma Project
15q11.2(BP1-BP2)	Deletion	22586494	23227522	641	4778320159_R01C02	Prostate Cancer
15q11.2(BP1-BP2)	Deletion	22652214	23116232	464	4676956130_R01C02	Prostate Cancer

*hg19 coordinates

Supplementary Table 4-7. Complete list of all the pathogenic CNVs identified in the four study chohorts along with the start and end breakpoints predicted by PennCNV. *Cont*.

CNV Region	CNV Type	Start coordinate *	End coordinate*	Approximate Size (kb)	Study ID	Study
15q11.2(BP1-BP2)	Deletion	22652214	23164315	512	4676964057_R01C01	Prostate Cancer
15q11.2(BP1-BP2)	Deletion	22652214	23227522	575	4808568030_R01C02	Prostate Cancer
15q11.2(BP1-BP2)	Deletion	22750305	23227522	477	4676892101_R01C02	Prostate Cancer
15q11.2(BP1-BP2)	Deletion	22750305	23116232	366	4676956117_R01C01	Prostate Cancer
15q11.2(BP1-BP2)	Deletion	22750305	23288179	538	4716567005_R01C01	Prostate Cancer
15q11.2(BP1-BP2)	Deletion	22750305	23272733	522	4811441052_R01C02	Prostate Cancer
15q11.2(BP1-BP2)	Deletion	22736956	23116232	379	4808568106_R01C01	Prostate Cancer
15q11.2(BP1-BP2)	Deletion	22769771	23227522	458	4716567059_R01C01	Prostate Cancer
15q11.2(BP1-BP2)	Deletion	22750305	23288179	538	EC029183	Breast Cancer
15q11.2(BP1-BP2)	Deletion	22585470	23228712	643	5767817072_R04C01	Health and Retirement Study
15q11.2(BP1-BP2)	Deletion	22750305	23157625	407	5815170007_R04C01	Health and Retirement Study
15q11.2-q13 (BP2-B3)	Duplication	23683783	28746356	5063	6365	Grady Trauma Project
15q13.2-q13.3 (BP4-BP5)	Duplication	30941244	32850317	1909	11574	Grady Trauma Project
15q13.2-q13.3 (BP4-BP5)	Duplication	30913207	32515973	1603	4676948092_R01C02	Prostate Cancer
15q13.2-q13.3 (BP4-BP5)	Deletion	30941244	32515849	1575	10994	Grady Trauma Project
15q13.2-q13.3 (BP4-BP5)	Deletion	30941244	32515849	1575	11017	Grady Trauma Project
15q13.2-q13.3 (BP4-BP5)	Deletion	30941244	32444196	1503	11174	Grady Trauma Project
15q13.2-q13.3 (BP4-BP5)	Deletion	30941244	32444196	1503	12807	Grady Trauma Project
15q13.2-q13.3 (BP4-BP5)	Deletion	30941244	32850317	1909	6487	Grady Trauma Project
15q13.2-q13.3 (BP4-BP5)	Deletion	30943512	32515849	1572	5826338068_R01C01	Health and Retirement Study
16p13.11	Deletion	15487812	16291983	804	12130	Grady Trauma Project
16p13.11	Deletion	15487812	16303388	816	6005	Grady Trauma Project
16p13.11	Deletion	15487812	16291983	804	6949	Grady Trauma Project
16p13.11	Deletion	15493046	16291983	799	11983	Grady Trauma Project
16p13.11	Deletion	15493046	16291983	799	12240	Grady Trauma Project
16p13.11	Deletion	15487812	16291983	804	10357	Grady Trauma Project
16p13.11	Deletion	15126890	16305355	1178	5751882147_R04C01	Health and Retirement Study
16p12.1	Deletion	21949122	22421321	472	10516	Grady Trauma Project
16p12.1	Deletion	21949122	22421321	472	12004	Grady Trauma Project
16p12.1	Deletion	21949122	22421321	472	7753	Grady Trauma Project
16p12.1	Deletion	21949122	22421321	472	9016	Grady Trauma Project
16p12.1	Deletion	21839340	22440319	601	5885311043_R04C01	Health and Retirement Study
16p12.1	Deletion	21949122	22425409	476	5473229012_R03C01	Health and Retirement Study

*hg19 coordinates

Supplementary Table 4-7. Complete list of all the pathogenic CNVs identified in the four study chohorts along with the start and end breakpoints predicted by PennCNV. *Cont*.

CNV Region	CNV Type	Start coordinate *	End coordinate*	Approximate Size (kb)	Study ID	Study
16p11.2 distal	Duplication	28824400	29028905	205	10458	Grady Trauma Project
16p11.2 distal	Duplication	28488943	29182190	693	7084	Grady Trauma Project
16p11.2	Duplication	29652488	30192561	540	10009	Grady Trauma Project
16p11.2	Duplication	29652488	30187676	535	5175	Grady Trauma Project
16p11.2	Duplication	29652488	30192561	540	7643	Grady Trauma Project
16p11.2	Duplication	29677823	30187676	510	3100	Grady Trauma Project
16p11.2	Duplication	29768019	30147265	379	6581	Grady Trauma Project
16p11.2	Duplication	29825022	30187676	363	10939	Grady Trauma Project
16p11.2	Duplication	29752774	30198151	445	EC031135	Breast Cancer
16p11.2	Deletion	29644174	30203423	559	10880	Grady Trauma Project
16p11.2	Deletion	29644174	30187676	544	7726	Grady Trauma Project
16p11.2	Deletion	29647342	30198319	551	10631	Grady Trauma Project
16p11.2	Deletion	29647342	30192561	545	7010	Grady Trauma Project
16p11.2	Deletion	29652488	30192561	540	3092	Grady Trauma Project
16p11.2	Deletion	29652488	30198319	546	6231	Grady Trauma Project
16p11.2	Deletion	29652488	30198319	546	6968	Grady Trauma Project
16p11.2	Deletion	29595483	30199805	604	4808681049_R01C02	Prostate Cancer
16p11.2	Deletion	29752774	30198151	445	EC033513	Breast Cancer
17p11.2	Duplication	16422602	17701515	1279	4716559018_R01C01	Prostate Cancer
17q12	Duplication	34651852	36350137	1698	11473	Grady Trauma Project
17q12	Duplication	34814753	35749716	935	7074	Grady Trauma Project
17q12	Deletion	34812078	36279313	1467	9671	Grady Trauma Project
17q12	Deletion	34815551	36223325	1408	11931	Grady Trauma Project
17q12	Deletion	34815551	36274189	1459	6058	Grady Trauma Project
22q11.2 (proximal A-B)	Duplication	18878027	20295420	1417	12634	Grady Trauma Project
22q11.2 (proximal A-B)	Duplication	18650682	20416143	1765	11184	Grady Trauma Project
22q11.2 (proximal A-D)	Duplication	19019101	20378905	1360	10324	Grady Trauma Project
22q11.2 (proximal A-D)	Duplication	19024651	20307256	1283	8622	Grady Trauma Project
22q11.2 (proximal A-D)	Duplication	18954629	21465050	2510	5422	Grady Trauma Project
22q11.2 (proximal A-D)	Duplication	18874965	21464479	2590	5488560116_R04C01	Health and Retirement Study
22q11.2 (proximal A-B)	Deletion	19024651	20301438	1277	8110	Grady Trauma Project
22q11.2 (proximal A-D)	Deletion	18886915	20312668	1426	9729	Grady Trauma Project

*hg19 coordinates

Supplementary Table 4-7. Complete list of all the pathogenic CNVs identified in the four study chohorts along with the start and end breakpoints predicted by PennCNV. *Cont*.

CNV Region	CNV Type	Start coordinate *	End coordinate*	Approximate Size (kb)	Study ID	Study
22q11.2 (distal type I; D-F)	Duplication	21906437	23654993	1749	7068	Grady Trauma Project
22q11.2 (distal type I; D-E)	Duplication	21714103	22808604	1095	7116	Grady Trauma Project
22q11.2 (distal type II; E-F)	Deletion	22998337	23657766	659	6059804067_R03C01	Health and Retirement Study
22q11.2 (distal type II; E-F)	Deletion	22980351	23569063	589	12836	Grady Trauma Project
22q11.2 (distal type III; F-H)	Duplication	23664431	25069276	1405	5599235060_R03C01	Health and Retirement Study
22q11.2 (distal type III; F-H)	Duplication	23664431	24975488	1311	5624859219_R01C01	Health and Retirement Study

*hg19 coordinates

Supplementary Table 4-8. Pairwise comparison of relatedness using IBD. pi_hat values represent proportion of relatedness between samples with the same CNV.

CNV	Sample #1	Sample #2	pi_hat value
1q21.1 duplication	7529	10169	0.011
1q21.1 duplication	7529	7722	0.019
1q21.1 duplication	7529	5097	0.000
1q21.1 duplication	10169	7722	0.000
1q21.1 duplication	10169	5097	0.014
1q21.1 duplication	7722	5097	0.013
1q21.1 deletion	10820	12510	0.000
1q21.1 deletion	10820	12563	0.000
1q21.1 deletion	10820	12614	0.000
1q21.1 deletion	10820	6021	0.000
1q21.1 deletion	10820	8470	0.000
1q21.1 deletion	12510	12563	0.000
1q21.1 deletion	12510	12614	0.013
1q21.1 deletion	12510	6021	0.019
1q21.1 deletion	12510	8470	0.020
1q21.1 deletion	12563	12614	0.013
1q21.1 deletion	12563	6021	0.000
1q21.1 deletion	12563	8470	0.017
1q21.1 deletion	12614	6021	0.014
1q21.1 deletion	12614	8470	0.025
1q21.1 deletion	6021	8470	0.021
2q13 deletion	10440	12171	0.000
2q13 deletion	10440	5202	0.000
2q13 deletion	12171	5202	0.005

Supplementary Table 4-8. Pairwise comparison of relatedness using IBD. pi_hat values represent proportion of relatedness between samples with the same CNV. *Cont.*

CNV	Sample #1	Sample #2	pi_hat value
15q11.2(BP1-BP2) deletion	5030	5121	0.000
15q11.2(BP1-BP2) deletion	5030	7356	0.000
15q11.2(BP1-BP2) deletion	5030	9516	0.000
15q11.2(BP1-BP2) deletion	5030	8710	0.012
15q11.2(BP1-BP2) deletion	5030	11387	0.015
15q11.2(BP1-BP2) deletion	5030	12471	0.000
15q11.2(BP1-BP2) deletion	5030	6097	0.000
15q11.2(BP1-BP2) deletion	5030	8735	0.020
15q11.2(BP1-BP2) deletion	5030	6820	0.000
15q11.2(BP1-BP2) deletion	5030	6840	0.013
15q11.2(BP1-BP2) deletion	7356	9516	0.000
15q11.2(BP1-BP2) deletion	7356	8710	0.007
15q11.2(BP1-BP2) deletion	7356	11387	0.000
15q11.2(BP1-BP2) deletion	7356	12471	0.020
15q11.2(BP1-BP2) deletion	7356	6097	0.000
15q11.2(BP1-BP2) deletion	7356	8735	0.009
15q11.2(BP1-BP2) deletion	7356	6820	0.008
15q11.2(BP1-BP2) deletion	7356	6840	0.000
15q11.2(BP1-BP2) deletion	8710	11387	0.010
15q11.2(BP1-BP2) deletion	8710	12471	0.010
15q11.2(BP1-BP2) deletion	8710	6097	0.000
15q11.2(BP1-BP2) deletion	8710	8735	0.014
15q11.2(BP1-BP2) deletion	8710	6820	0.008
15q11.2(BP1-BP2) deletion	8710	6840	0.000
15q11.2(BP1-BP2) deletion	11387	12471	0.017
15q11.2(BP1-BP2) deletion	11387	6097	0.000
15q11.2(BP1-BP2) deletion	11387	8735	0.017
15q11.2(BP1-BP2) deletion	11387	6820	0.000
15q11.2(BP1-BP2) deletion	11387	6840	0.007
15q11.2(BP1-BP2) deletion	12471	6097	0.000
15q11.2(BP1-BP2) deletion	12471	8735	0.010
15q11.2(BP1-BP2) deletion	12471	6820	0.001
15q11.2(BP1-BP2) deletion	12471	6840	0.007
15q11.2(BP1-BP2) deletion	6097	8735	0.000
15q11.2(BP1-BP2) deletion	6097	6820	0.005
15q11.2(BP1-BP2) deletion	6097	6840	0.000

Supplementary Table 4-8. Pairwise comparison of relatedness using IBD. pi_hat values represent proportion of relatedness between samples with the same CNV. *Cont.*

CNV	Sample #1	Sample #2	pi_hat value
15q11.2(BP1-BP2) deletion	8735	6820	0.004
15q11.2(BP1-BP2) deletion	8735	6840	0.006
15q11.2(BP1-BP2) deletion	6820	6840	0.005
13.2-q13.3(BP4-BP5) deletion	10994	11017	0.026
13.2-q13.3(BP4-BP5) deletion	10994	11174	0.008
13.2-q13.3(BP4-BP5) deletion	10994	12807	0.026
13.2-q13.3(BP4-BP5) deletion	10994	6487	0.004
13.2-q13.3(BP4-BP5) deletion	11017	11174	0.012
13.2-q13.3(BP4-BP5) deletion	11017	12807	0.019
13.2-q13.3(BP4-BP5) deletion	11017	6487	0.012
13.2-q13.3(BP4-BP5) deletion	11174	12807	0.002
13.2-q13.3(BP4-BP5) deletion	11174	6487	0.000
13.2-q13.3(BP4-BP5) deletion	12807	6487	0.012
16p13.11 deletion	12130	12807	0.009
16p13.11 deletion	12130	6949	0.015
16p13.11 deletion	12130	11983	0.000
16p13.11 deletion	12130	12240	0.000
16p13.11 deletion	12130	10357	0.014
16p13.11 deletion	12807	6949	0.018
16p13.11 deletion	12807	11983	0.019
16p13.11 deletion	12807	12240	0.005
16p13.11 deletion	12807	10357	0.021
16p13.11 deletion	6949	11983	0.010
16p13.11 deletion	6949	12240	0.021
16p13.11 deletion	6949	10357	0.018
16p13.11 deletion	11983	12240	0.000
16p13.11 deletion	11983	10357	0.006
16p13.11 deletion	12240	10357	0.003

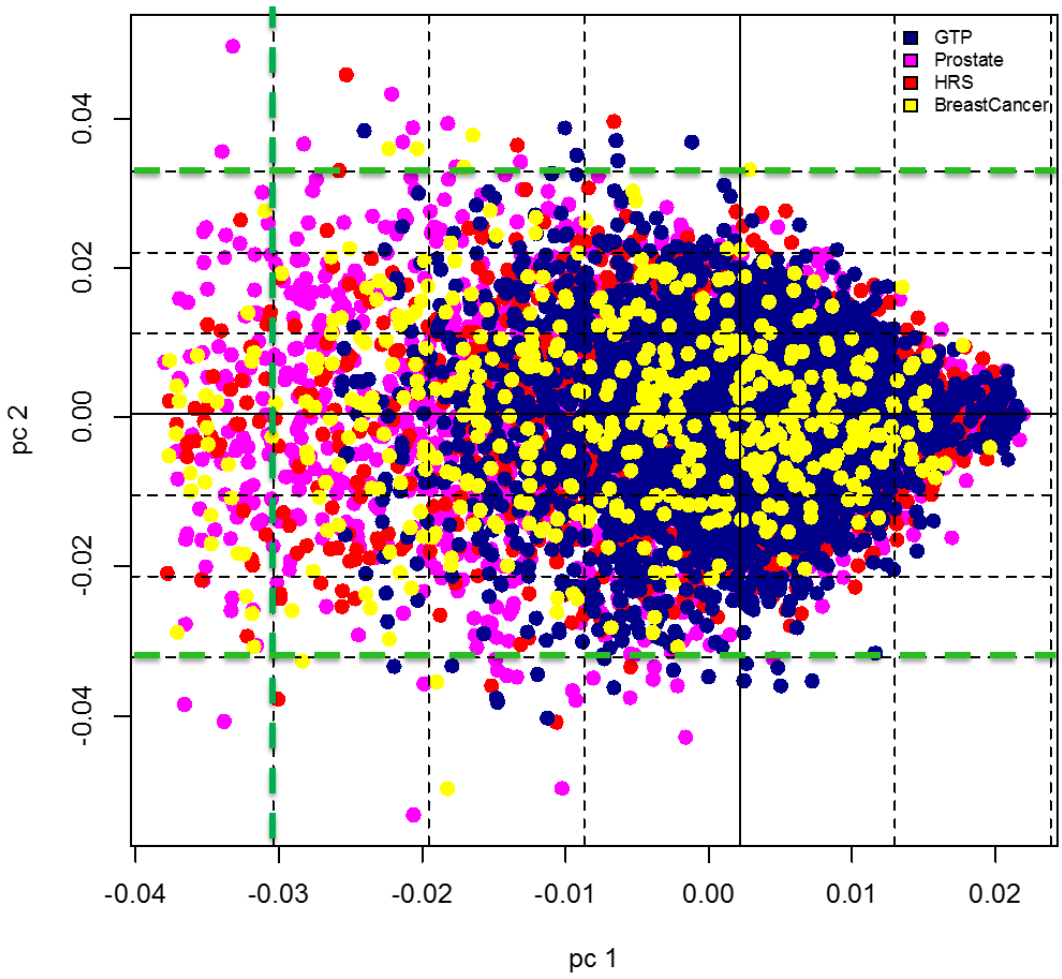
Supplementary Table 4-8. Pairwise comparison of relatedness using IBD. pi_hat values represent proportion of relatedness between samples with the same CNV. *Cont.*

CNV	Sample #1	Sample #2	pi_hat value
16p12.1 deletion	10516	12004	0.017
16p12.1 deletion	10516	7753	0.000
16p12.1 deletion	10516	9016	0.000
16p12.1 deletion	12004	7753	0.011
16p12.1 deletion	12004	9016	0.011
16p12.1 deletion	7753	9016	0.000
16p11.2 distal duplication	10458	7084	0.000
16p11.2 duplication	10009	5175	0.000
16p11.2 duplication	10009	7643	0.000
16p11.2 duplication	10009	3100	0.000
16p11.2 duplication	10009	6581	0.018
16p11.2 duplication	10009	10939	0.000
16p11.2 duplication	5175	7643	0.006
16p11.2 duplication	5175	3100	0.000
16p11.2 duplication	5175	6581	0.000
16p11.2 duplication	5175	10939	0.000
16p11.2 duplication	7643	3100	0.000
16p11.2 duplication	7643	6581	0.000
16p11.2 duplication	7643	10939	0.000
16p11.2 duplication	3100	6581	0.026
16p11.2 duplication	3100	10939	0.000
16p11.2 duplication	6581	10939	0.035
16p11.2 deletion	10880	7726	0.000
16p11.2 deletion	10880	10631	0.010
16p11.2 deletion	10880	7010	0.018
16p11.2 deletion	10880	3092	0.018
16p11.2 deletion	10880	6231	0.000
16p11.2 deletion	10880	6968	0.015
16p11.2 deletion	7726	10631	0.010
16p11.2 deletion	7726	7010	0.000
16p11.2 deletion	7726	3092	0.000
16p11.2 deletion	7726	6231	0.000
16p11.2 deletion	7726	6968	0.005

Supplementary Table 4-8. Pairwise comparison of relatedness using IBD. pi_hat values represent proportion of relatedness between samples with the same CNV. *Cont.*

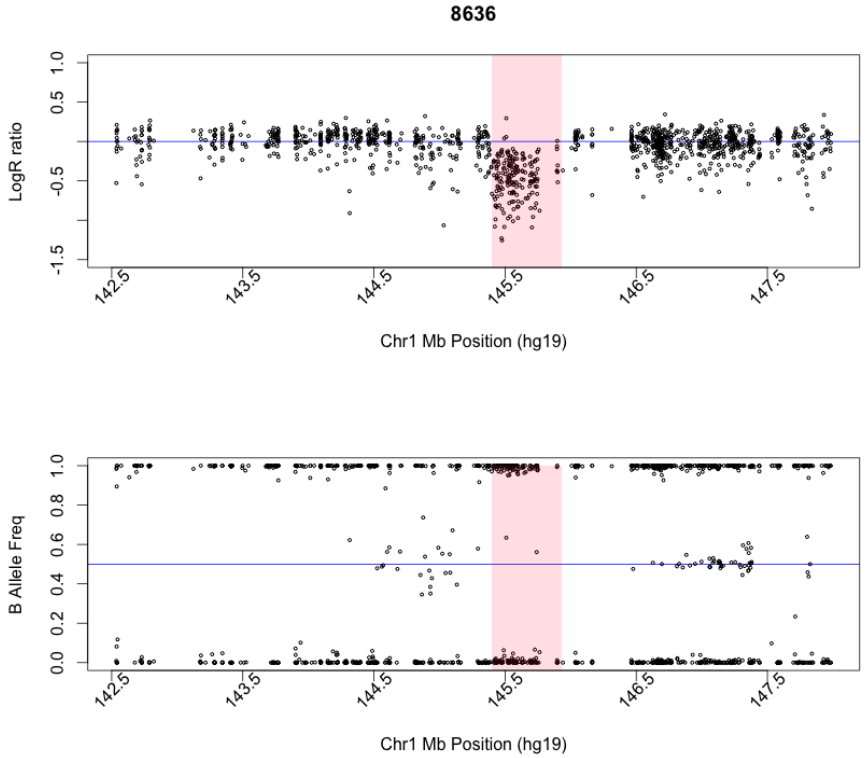
CNV	Sample #1	Sample #2	pi_hat value
16p11.2 deletion	10631	7010	0.009
16p11.2 deletion	10631	3092	0.011
16p11.2 deletion	10631	6231	0.005
16p11.2 deletion	10631	6968	0.021
16p11.2 deletion	7010	3092	0.012
16p11.2 deletion	7010	6231	0.000
16p11.2 deletion	7010	6968	0.016
16p11.2 deletion	3092	6231	0.000
16p11.2 deletion	3092	6968	0.000
16p11.2 deletion	6231	6968	0.000
17q12 duplication	11473	7074	0.014
17q12 deletion	9671	11931	0.000
17q12 deletion	9671	6058	0.023
17q12 deletion	11931	6058	0.000
22q11.2 proximal duplication	10324	8622	0.000
22q11.2 proximal duplication	10324	12634	0.013
22q11.2 proximal duplication	10324	11184	0.011
22q11.2 proximal duplication	10324	5422	0.003
22q11.2 proximal duplication	8622	12634	0.000
22q11.2 proximal duplication	8622	11184	0.000
22q11.2 proximal duplication	8622	5422	0.000
22q11.2 proximal duplication	12634	11184	0.008
22q11.2 proximal duplication	12634	5422	0.000
22q11.2 proximal duplication	11184	5422	0.006
22q11.2 proximal deletion	9729	8110	0.000
22q11.2 distal duplication	7068	7116	0.000

Supplementary Figure 4-1. A graph of the first and second principal components (pc1 and pc2) for all study participants in each of the cohorts: the Grady Trauma Project (**GTP**), a Multi-ethnic Genome-wide scan of Prostate Cancer (**Prostate)**, the Health and Retirement Study (**HRS**), and the Multiethnic Cohort Breast Cancer Genetics study (**BreastCancer**). Each dashed line represents 1 standard deviation from the median. All samples the fall outside of the 3 standard deviations (marked by the green lines) were removed as outliers.

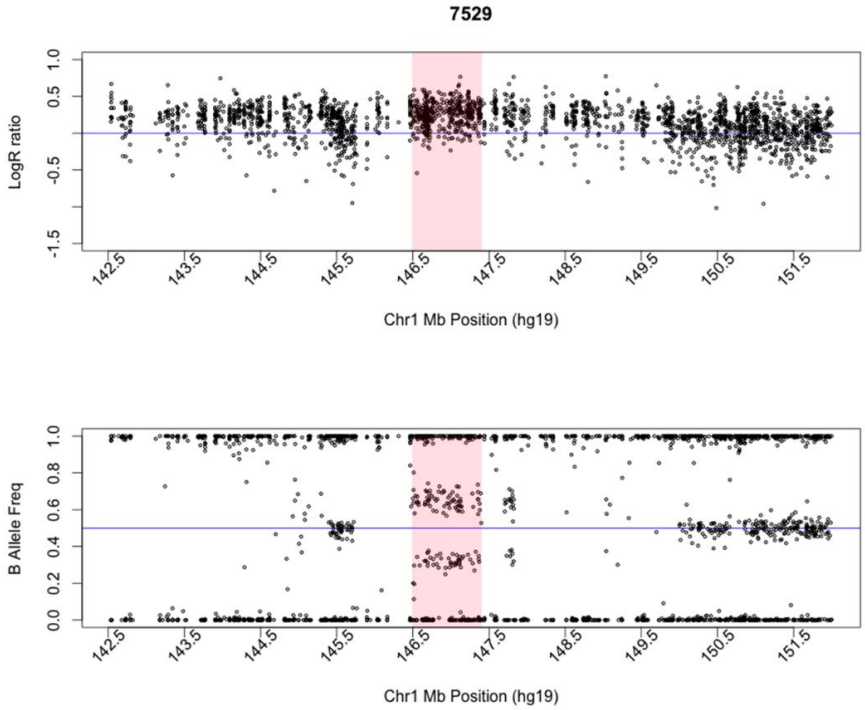


Supplementary Figure 4-2: Graphs of each GD CNV deletion or duplication by region

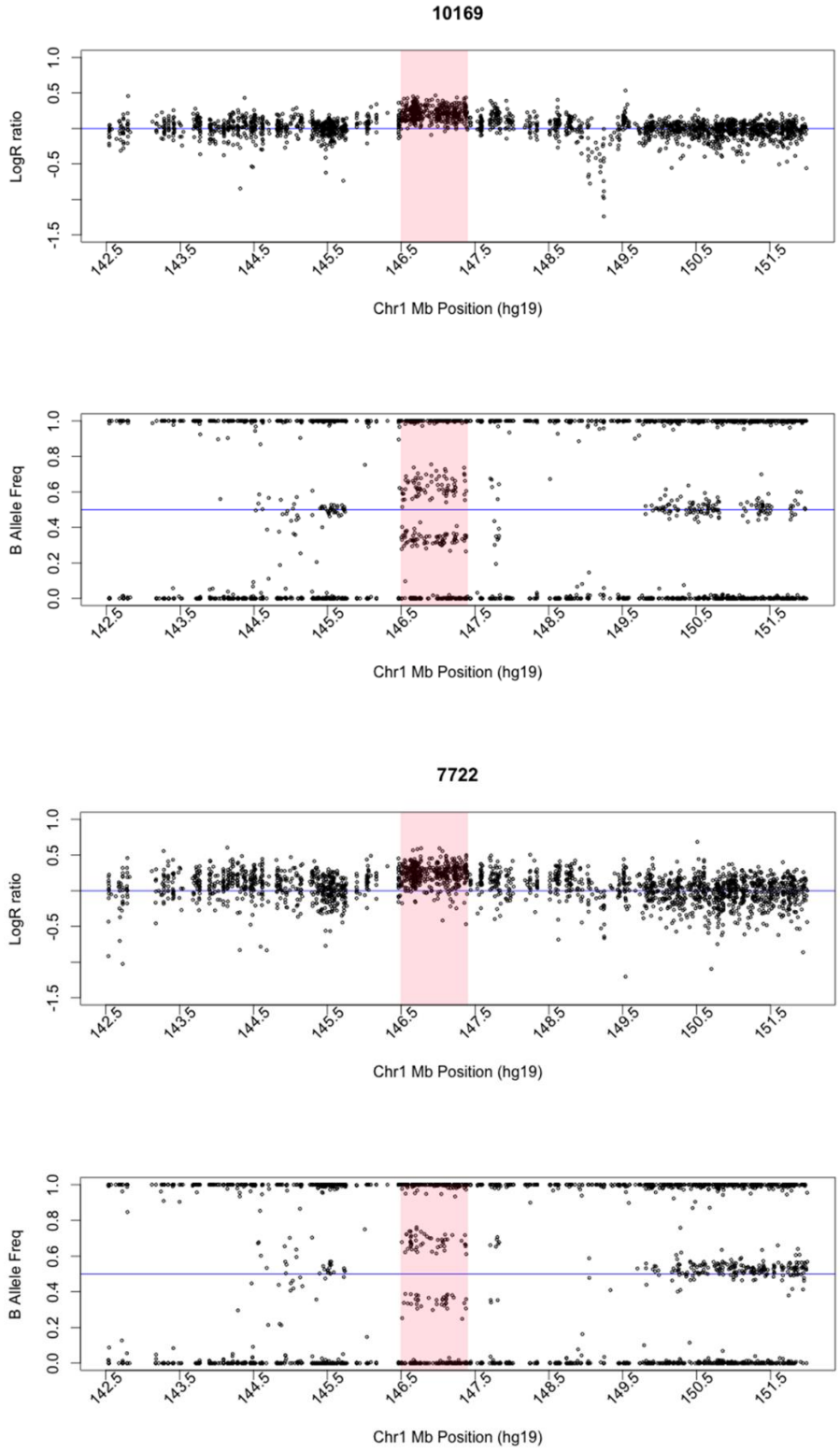
<u>1q21 (TAR) deletion</u>



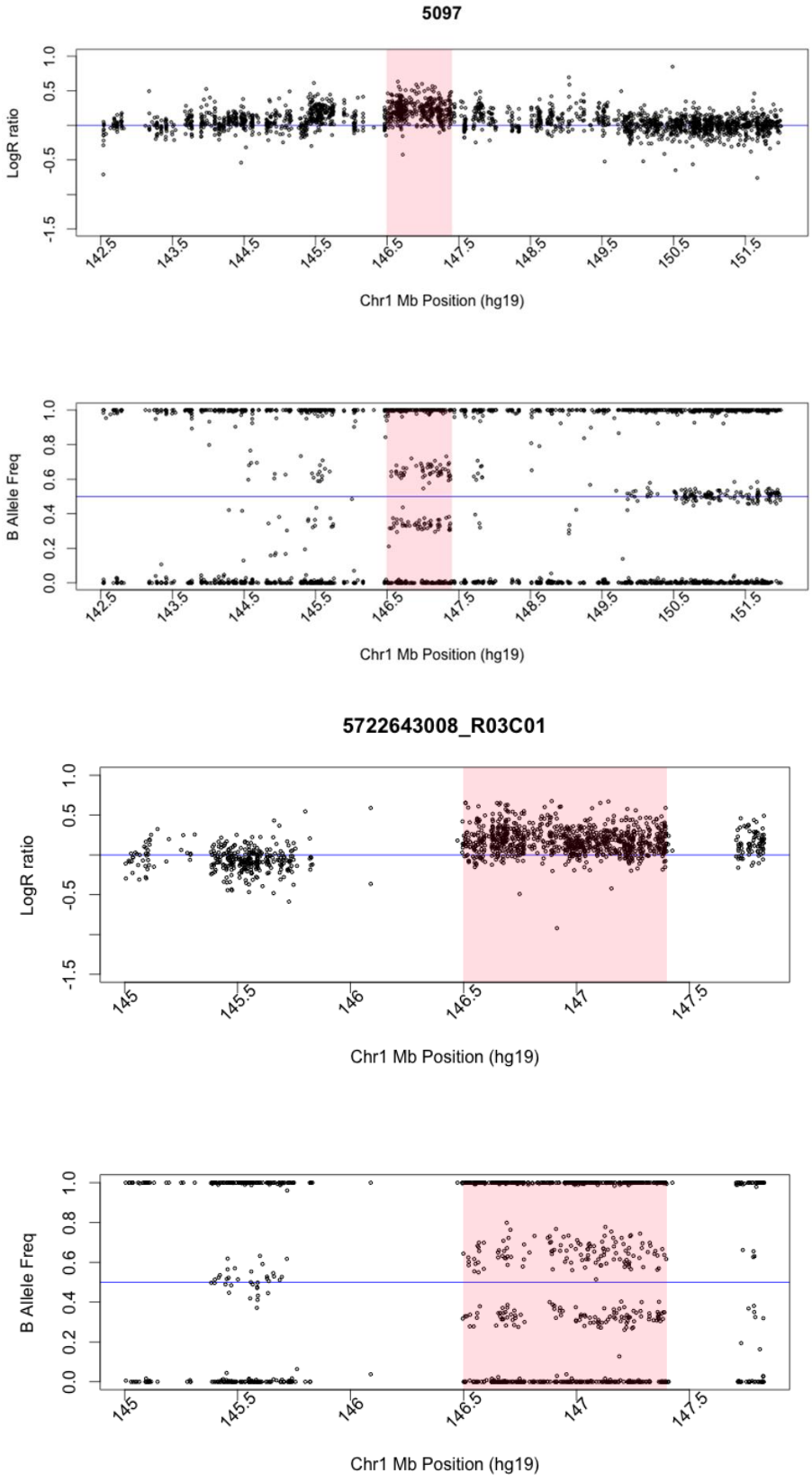
<u>1q21.1 duplications. *Cont.*</u>



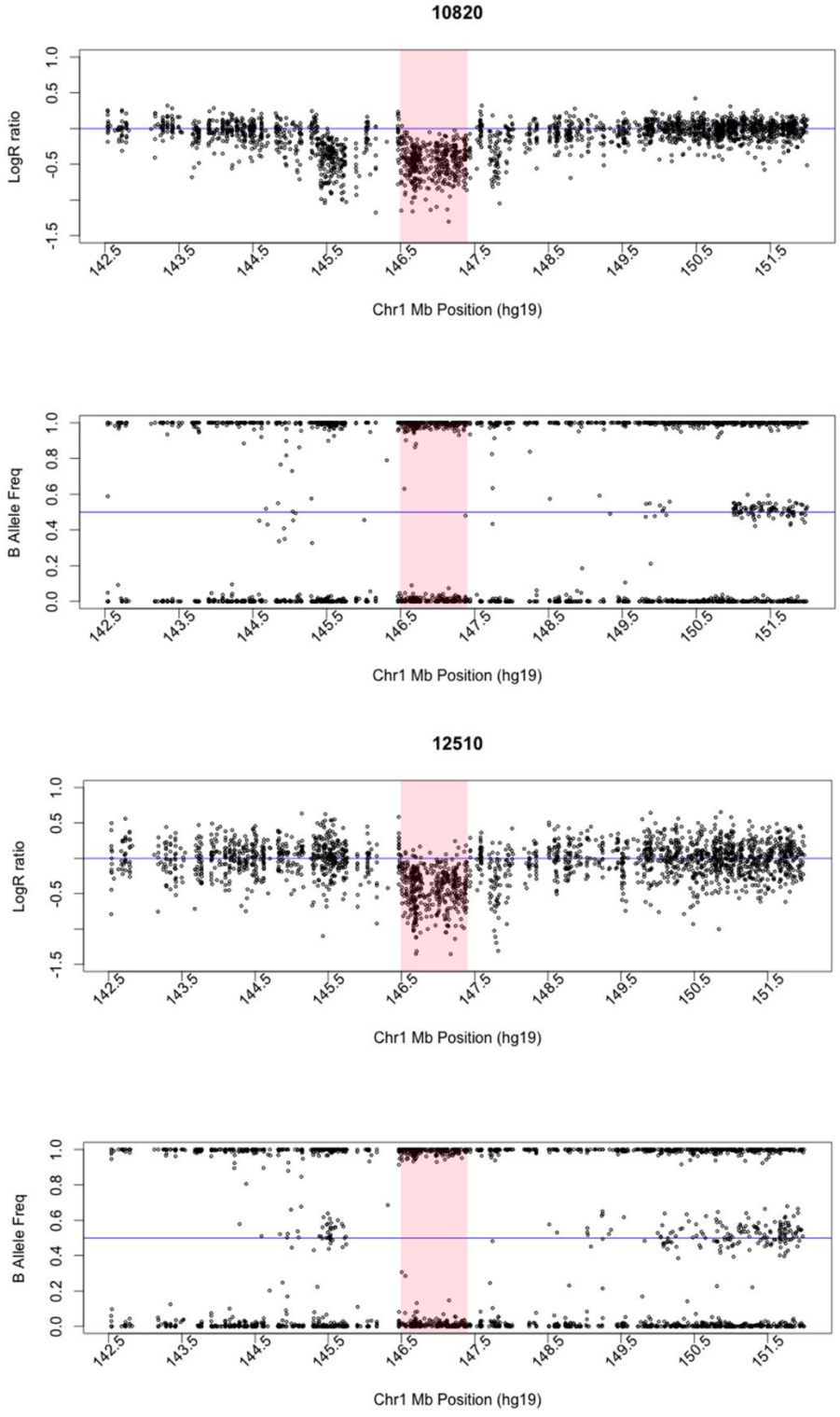
1q21.1 duplications. *Cont.*



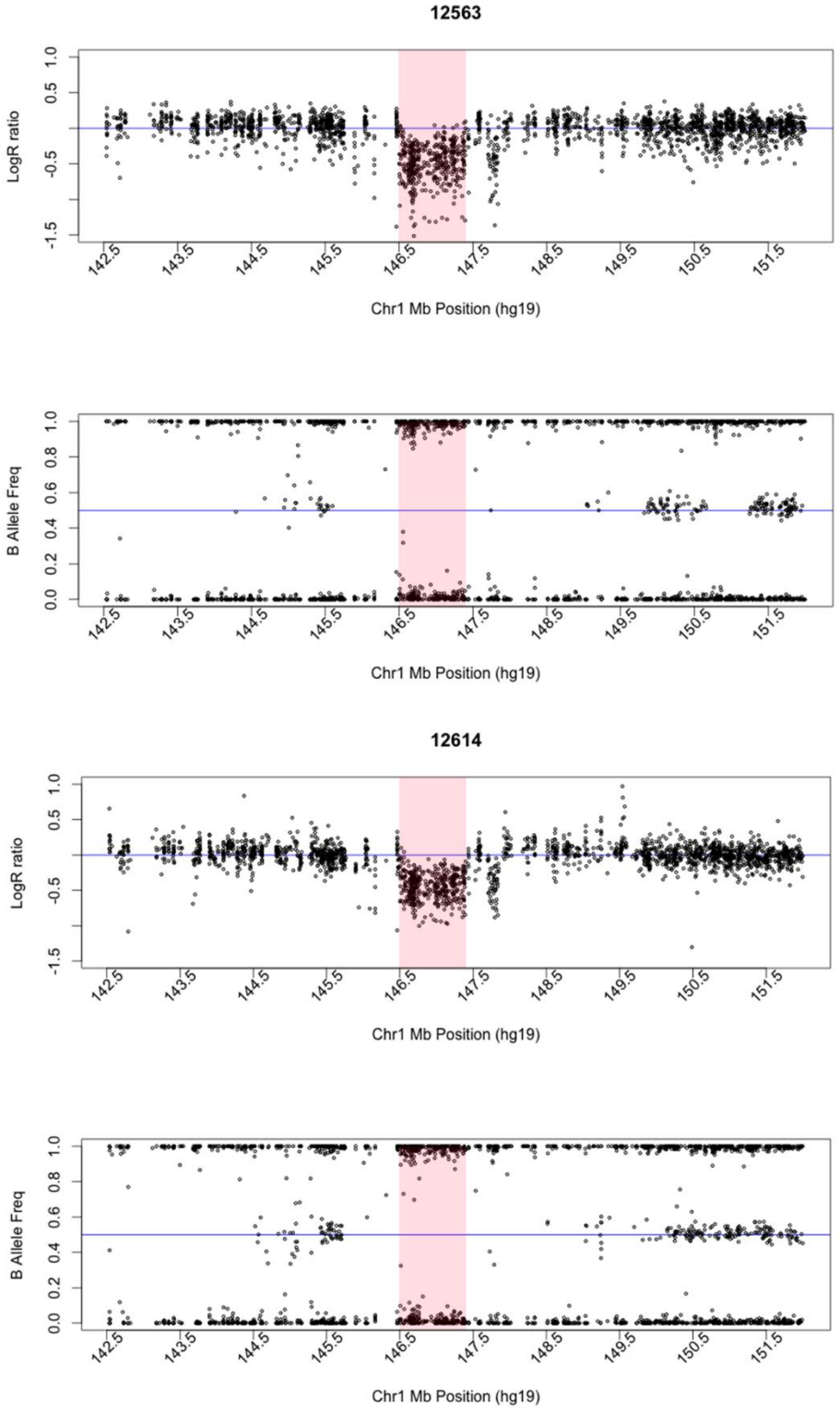
1q21.1 duplications. *Cont.*



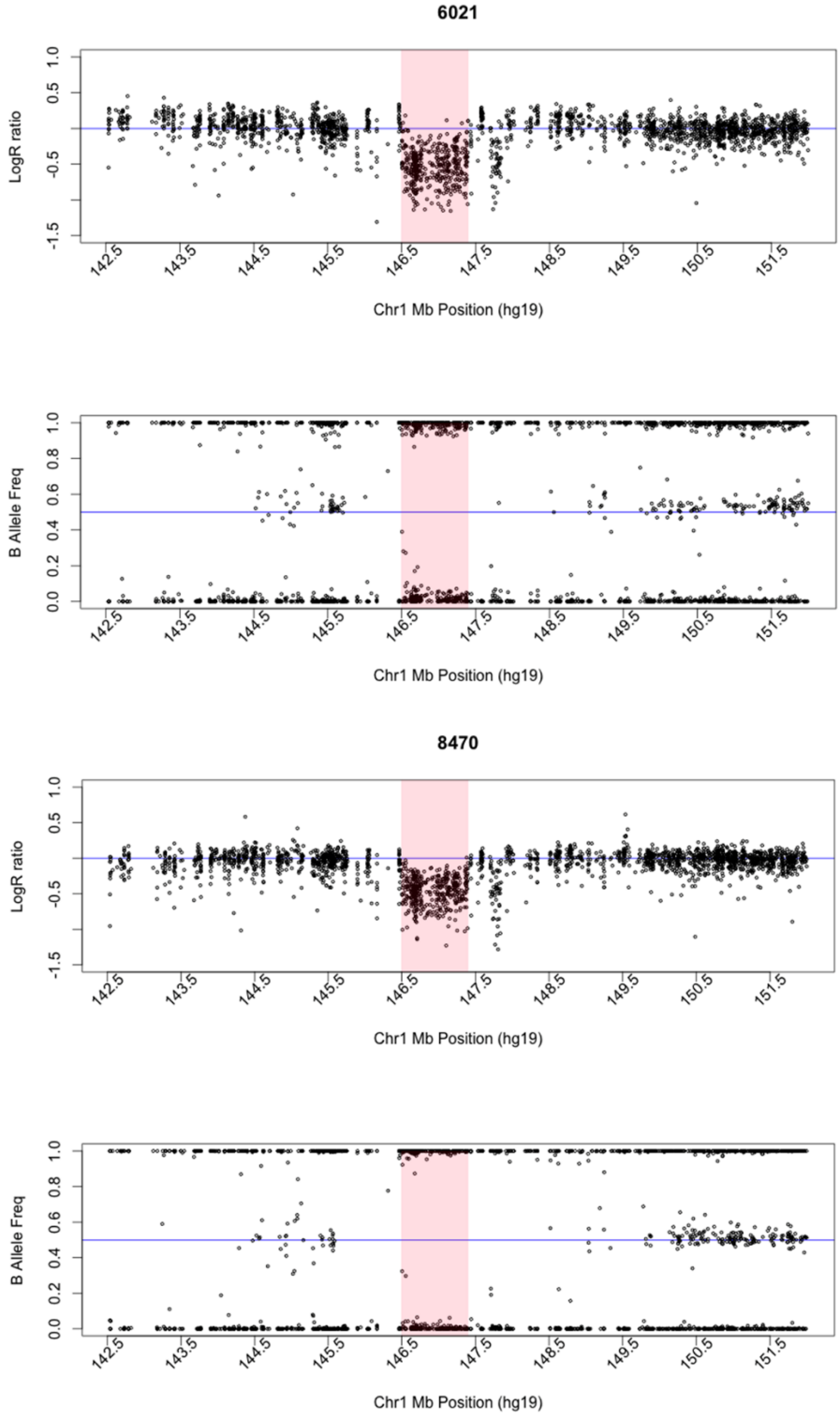
1q21.1 deletions



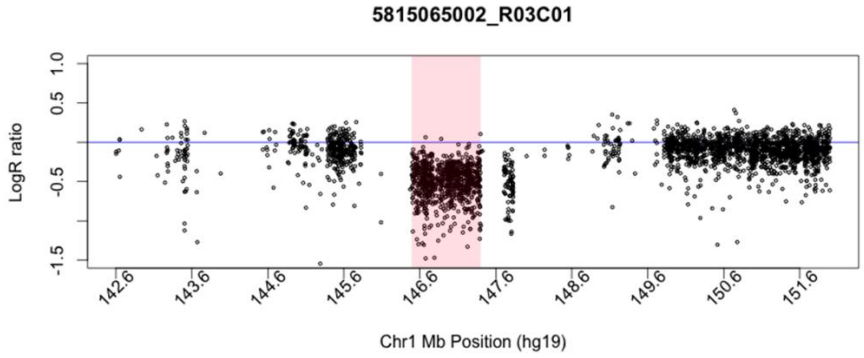
1q21.1 deletions Cont.

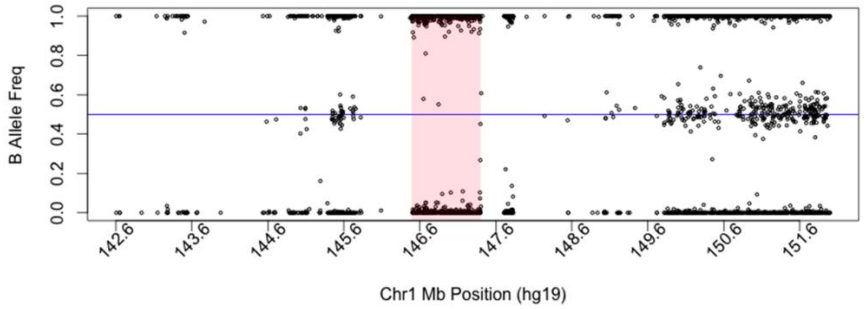


1q21.1 deletions Cont.

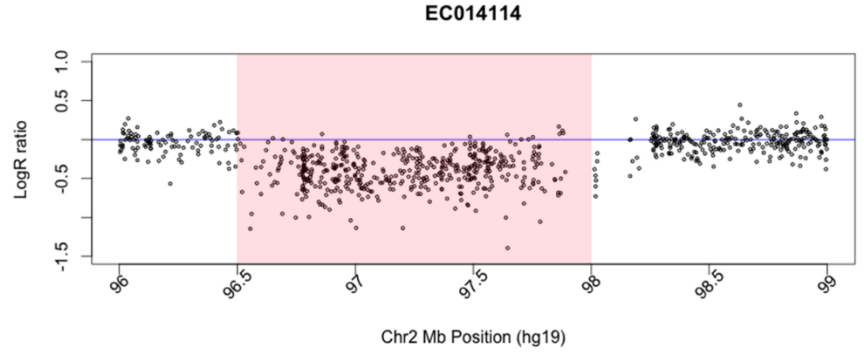


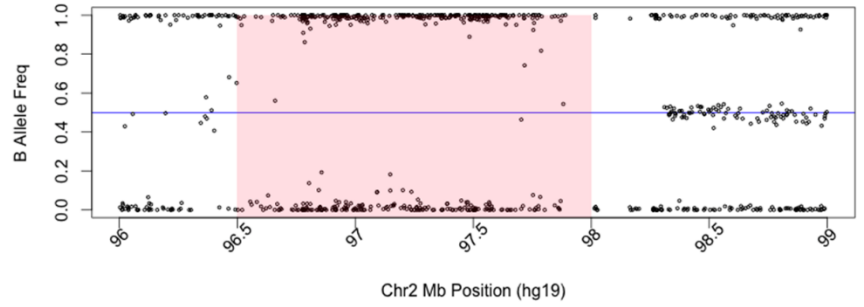
1q21.1 deletions Cont.



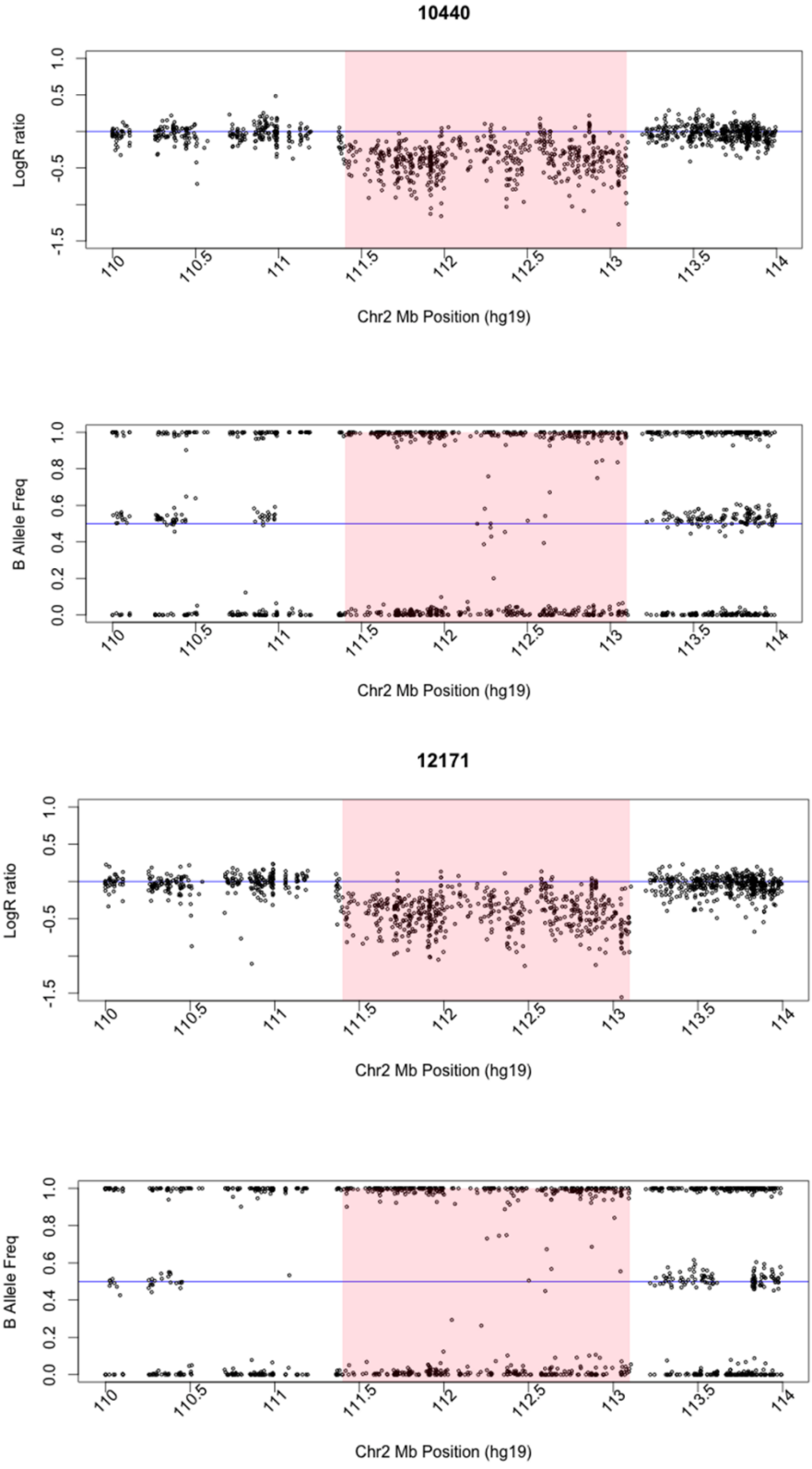


2q11.2 deletion

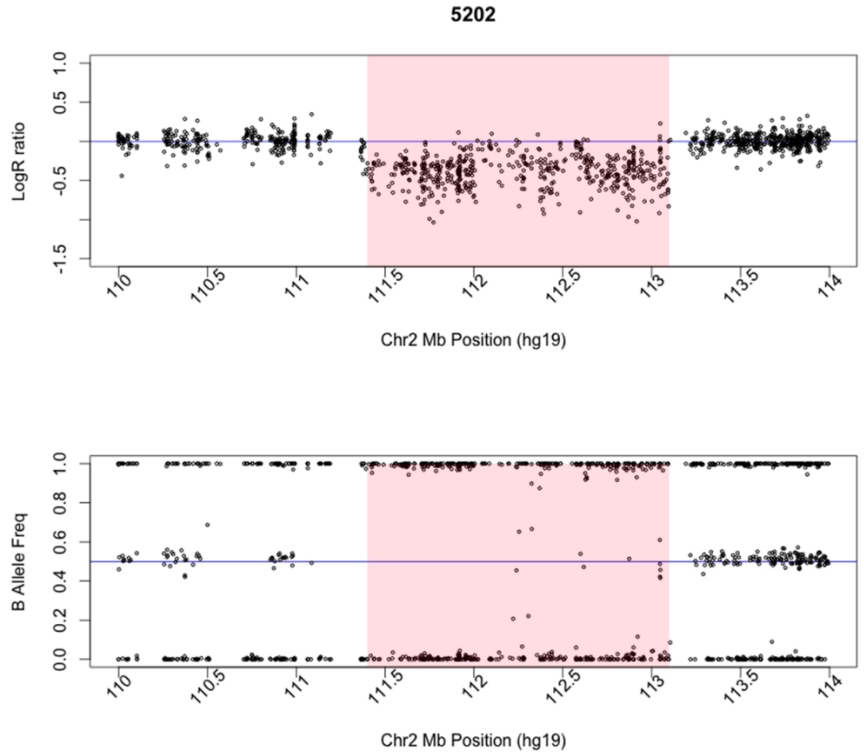




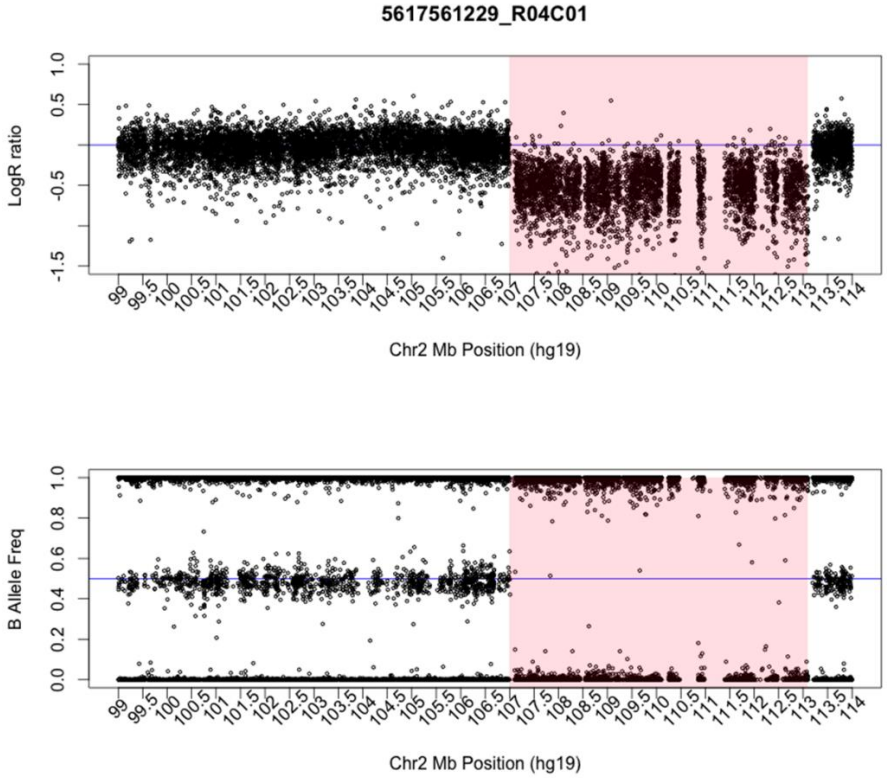
2q13 deletions



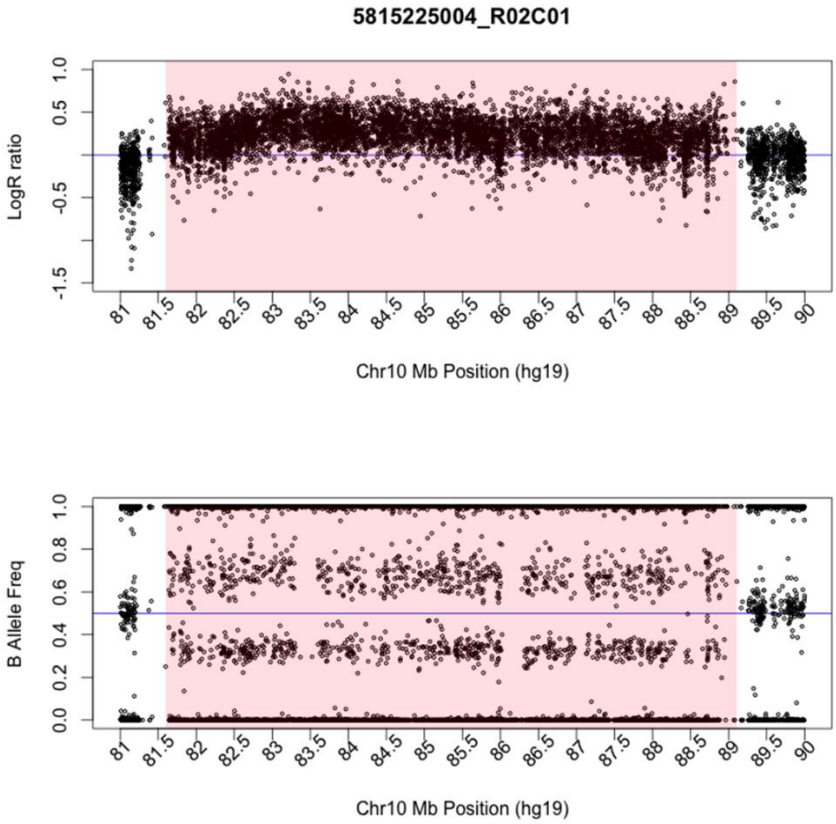
2q13 deletions. *Cont.*



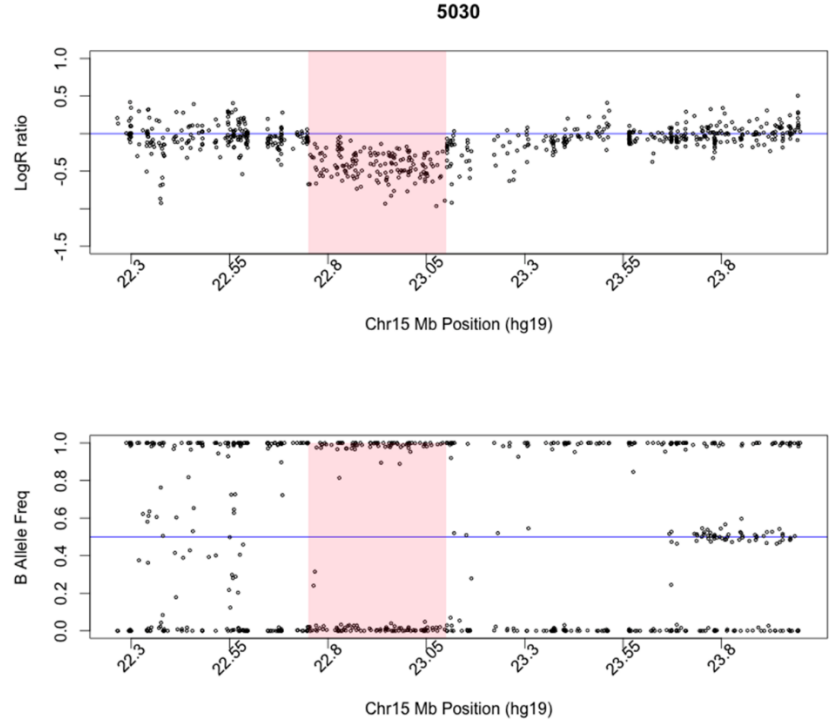
2q11-2q13 deletion



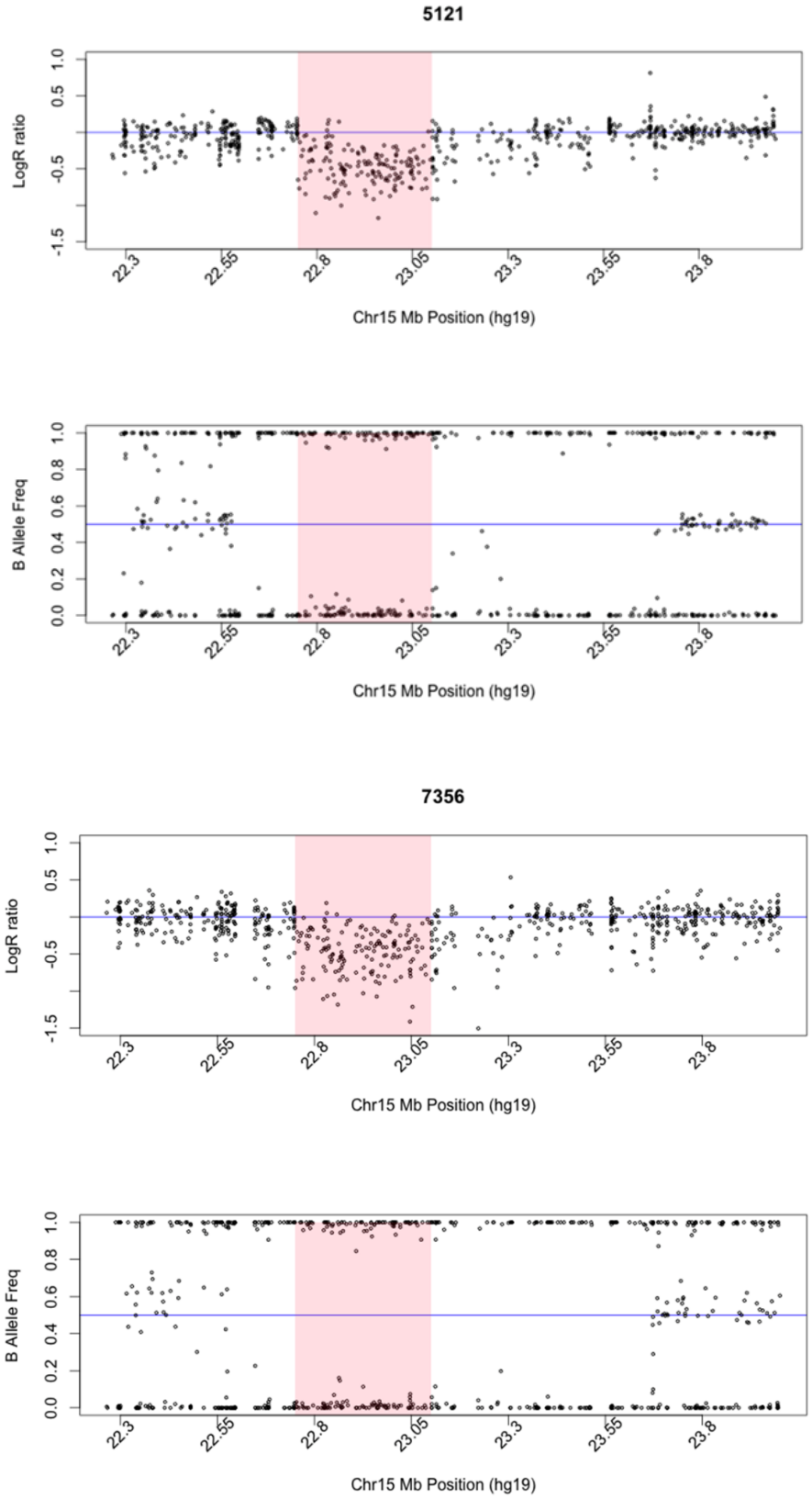
10q23 duplication



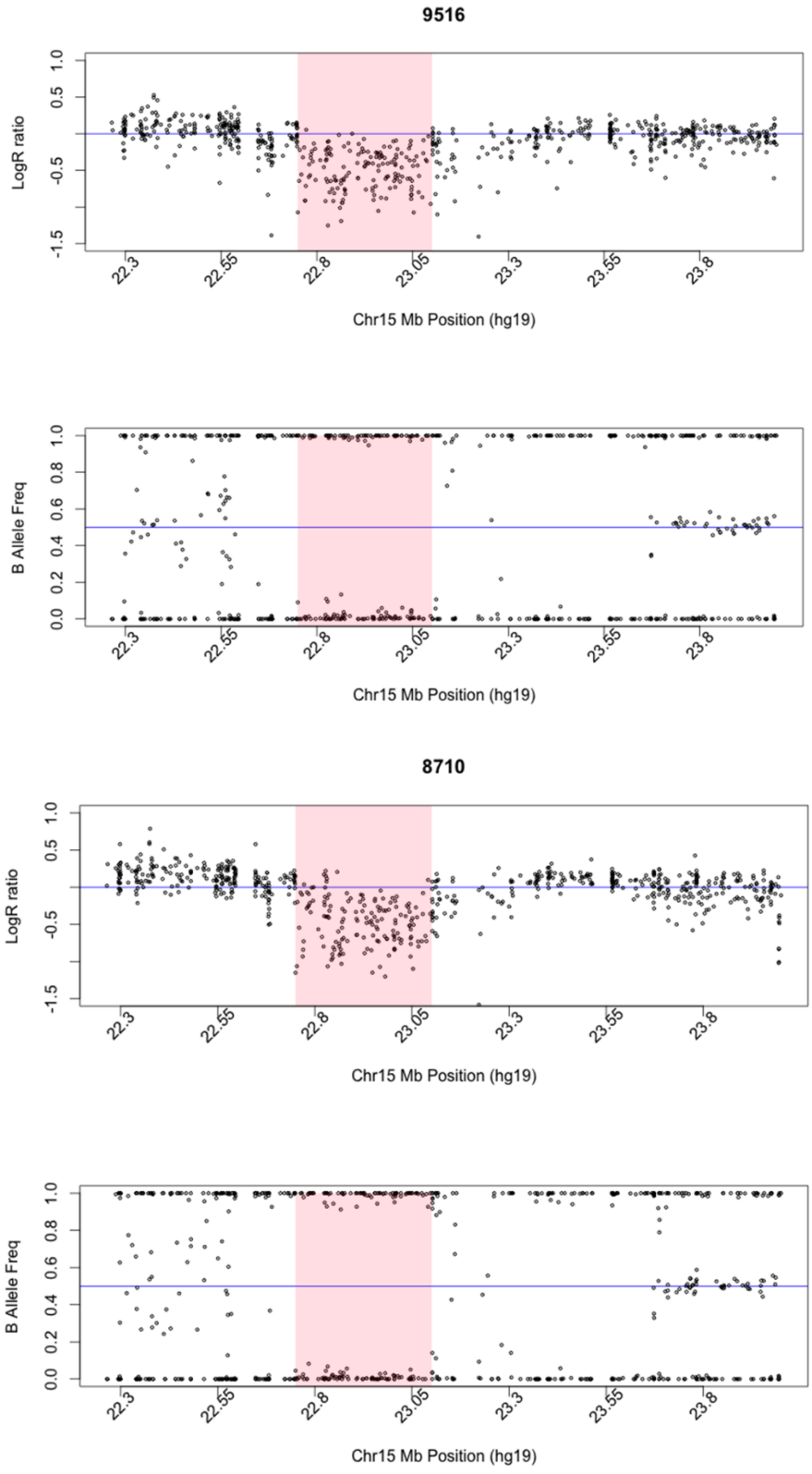
15q11.2 (BP1-BP2) deletions



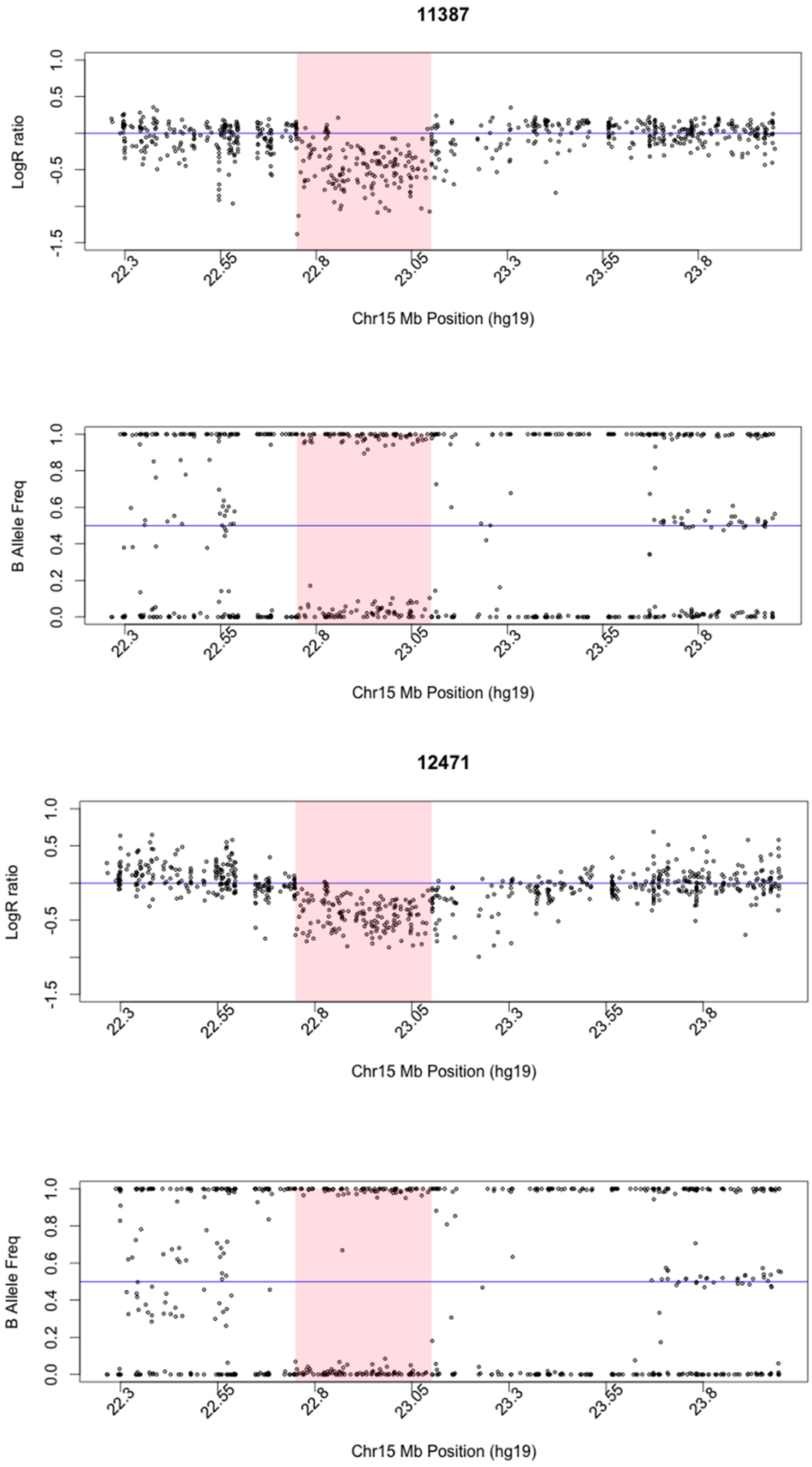
15q11.2 (BP1-BP2) deletions. *Cont.*



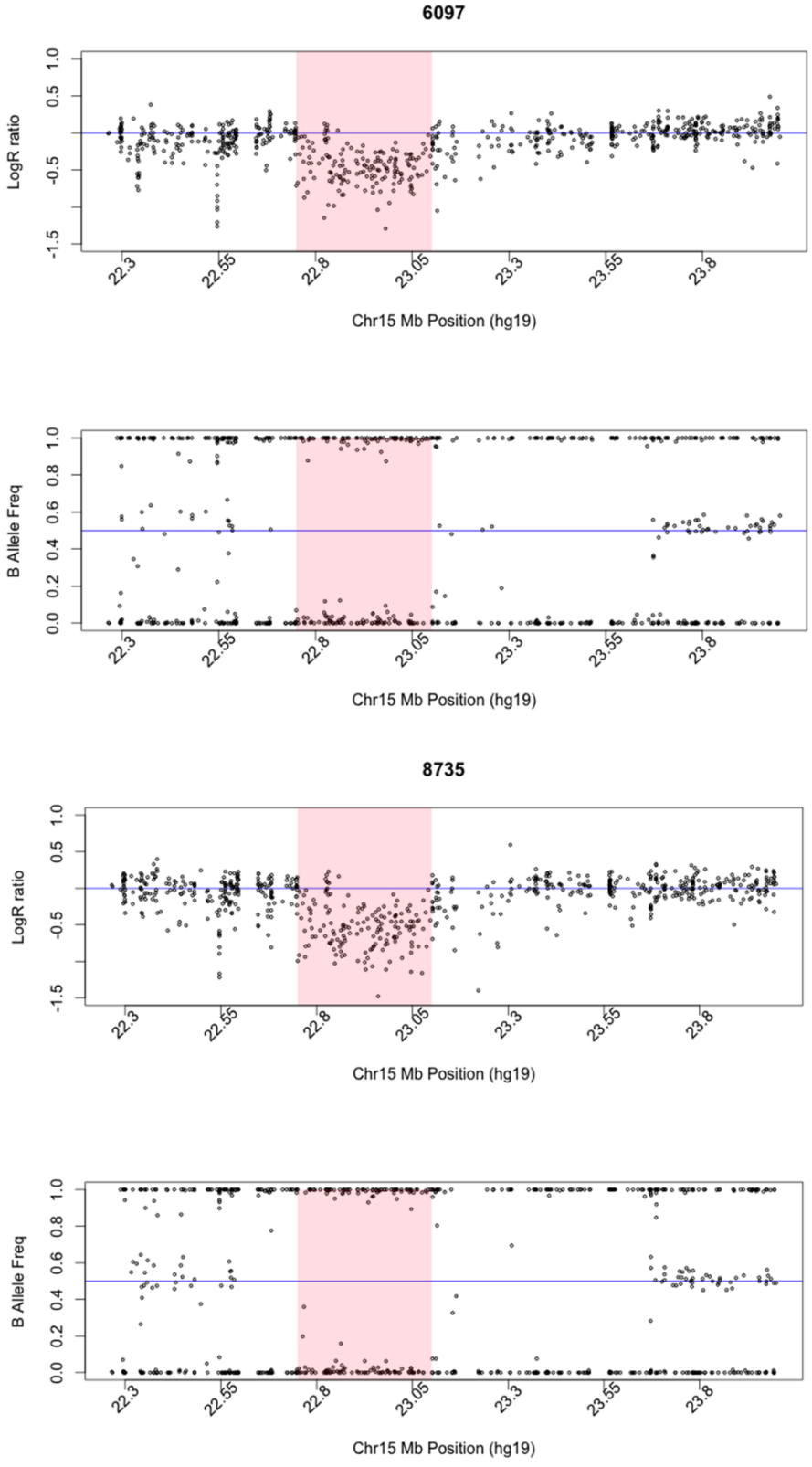
15q11.2 (BP1-BP2) deletions. *Cont.*



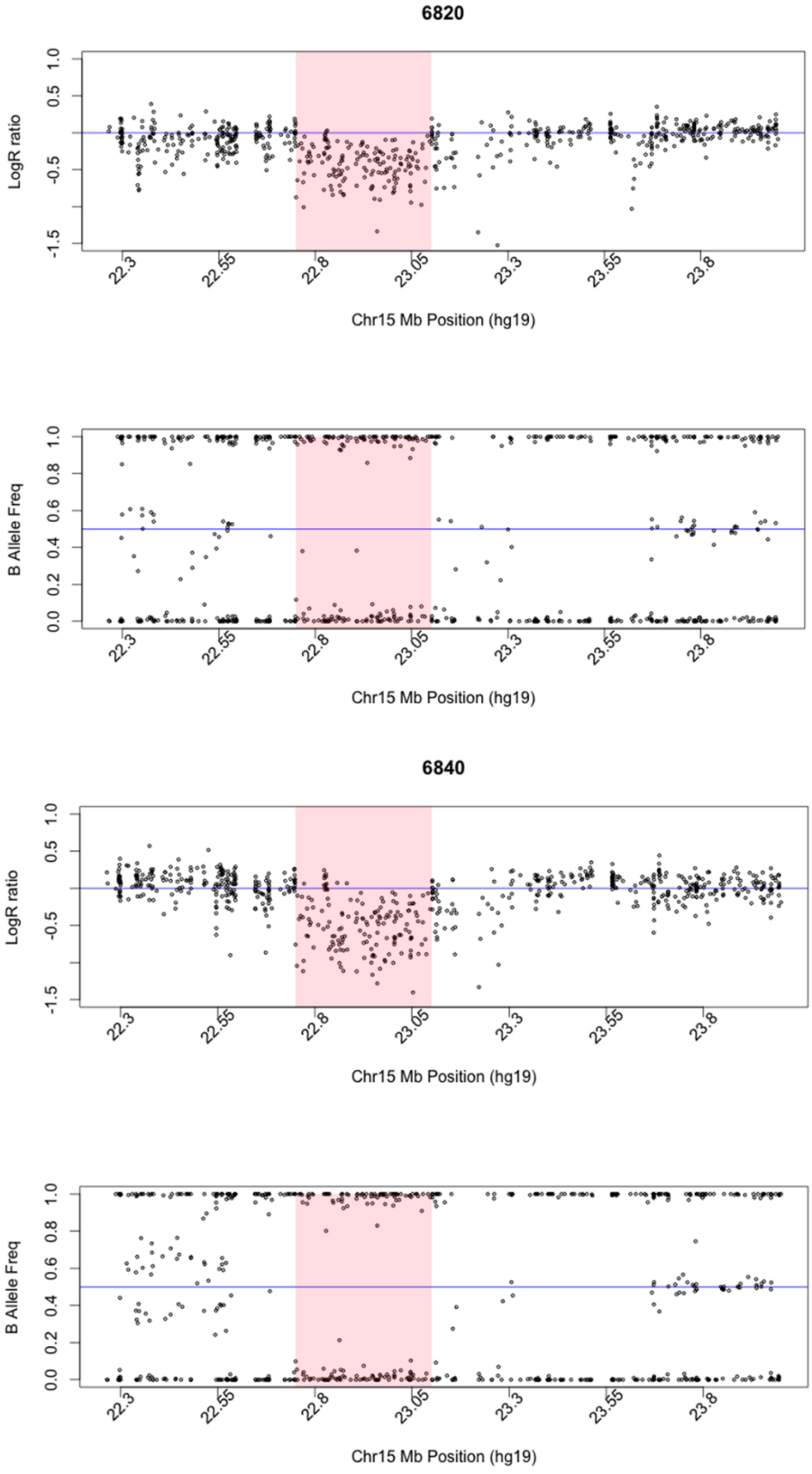
15q11.2 (BP1-BP2) deletions. *Cont.*



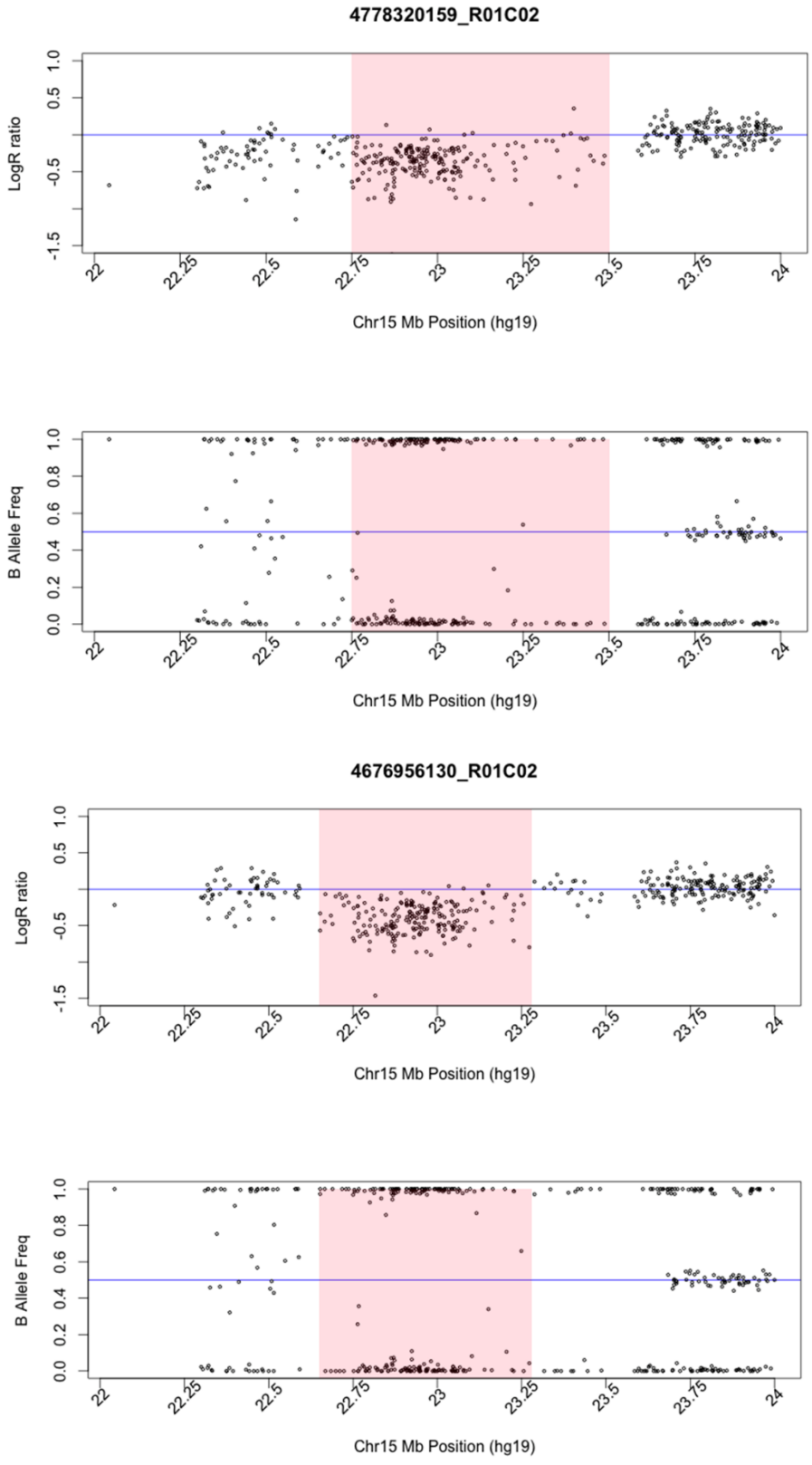
15q11.2 (BP1-BP2) deletions. *Cont.*



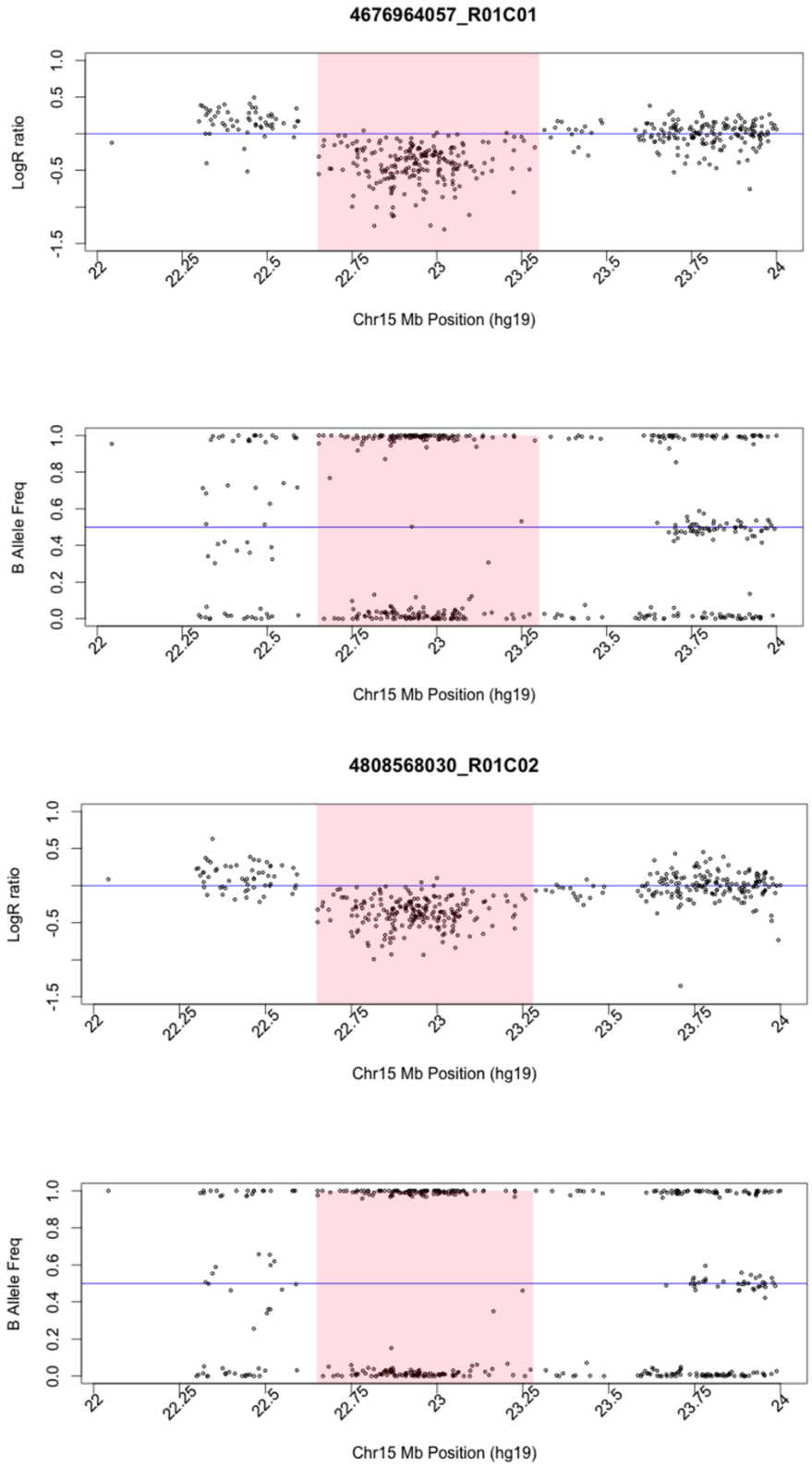
15q11.2 (BP1-BP2) deletions. *Cont.*



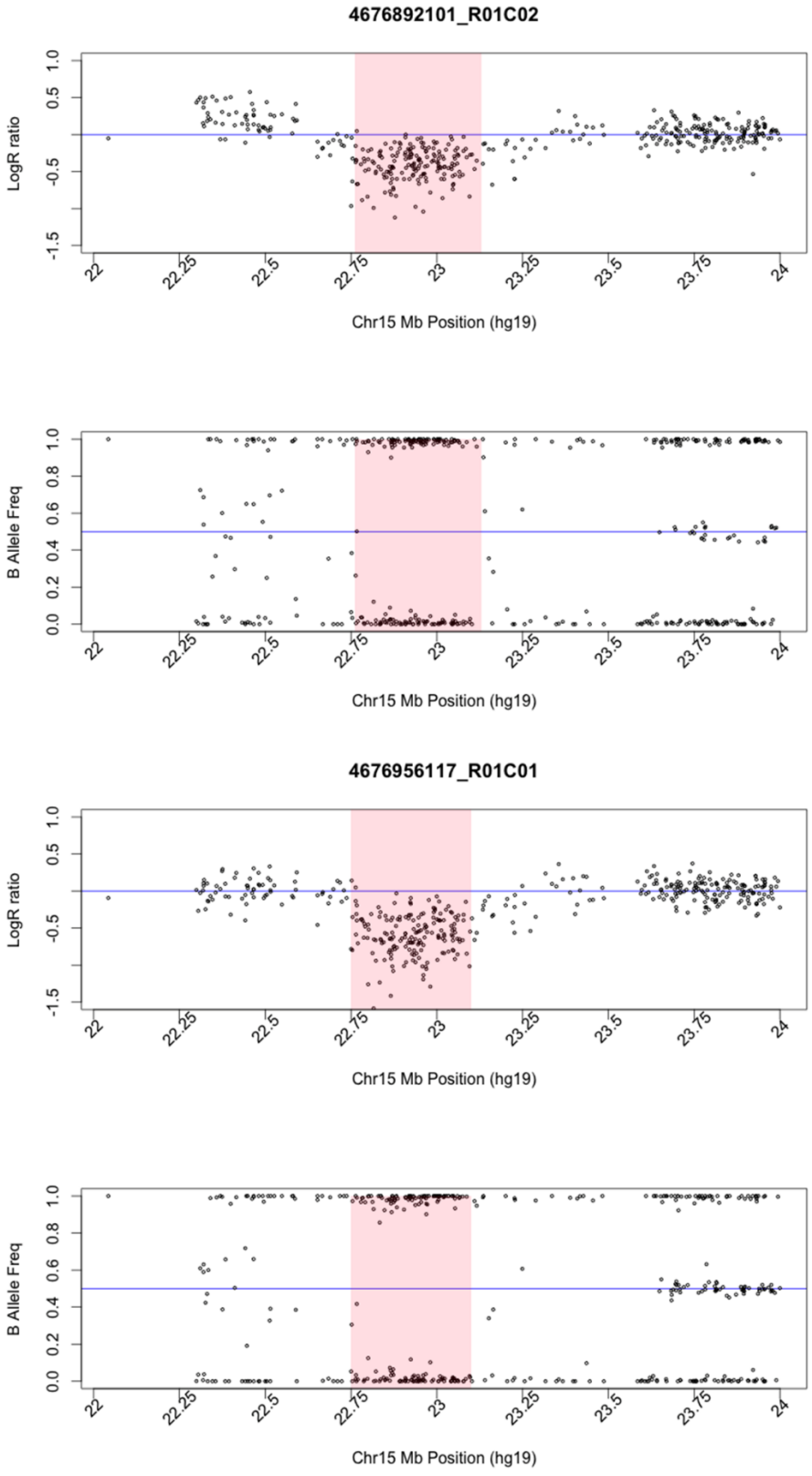
15q11.2 (BP1-BP2) deletions. *Cont.*



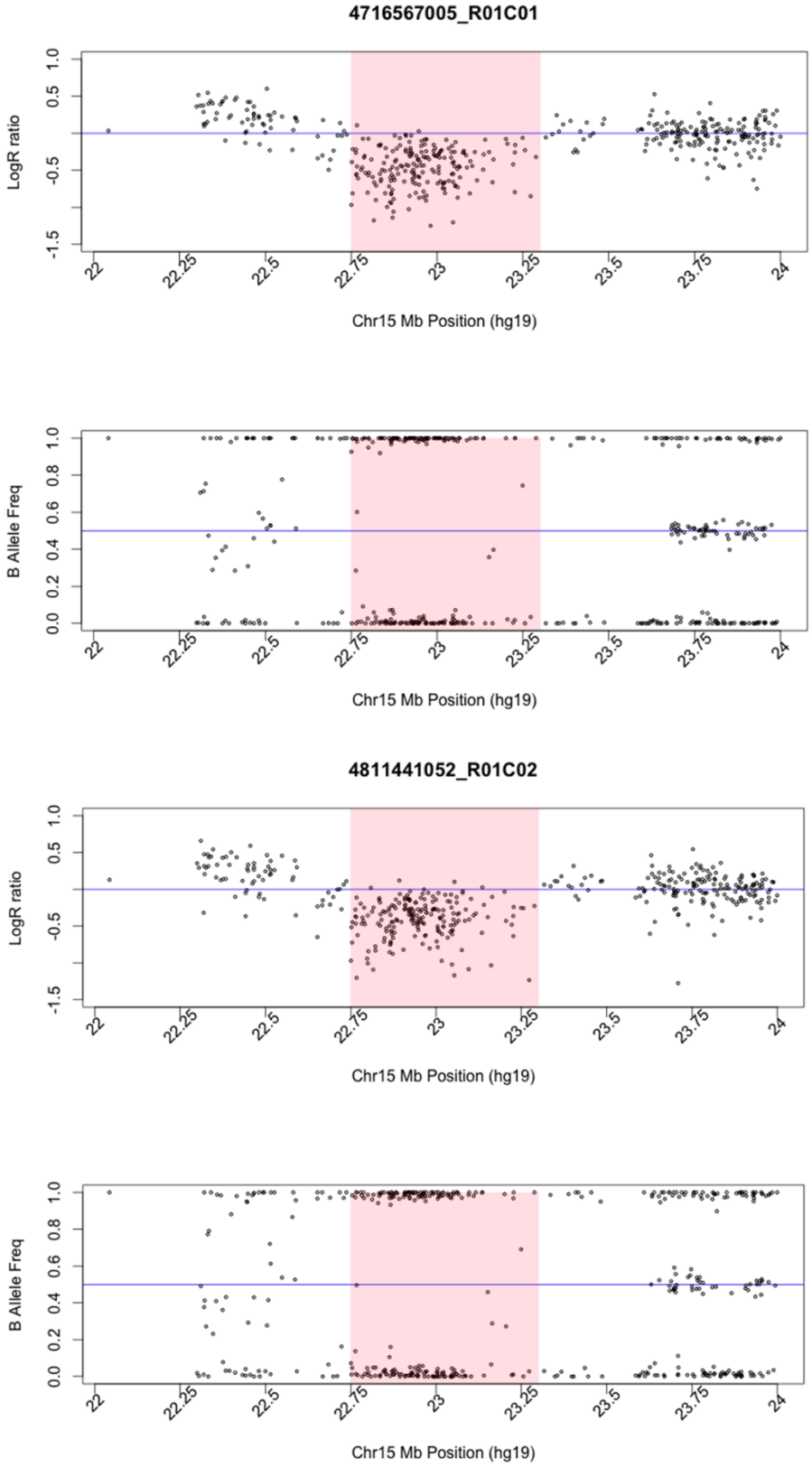
15q11.2 (BP1-BP2) deletions. *Cont.*



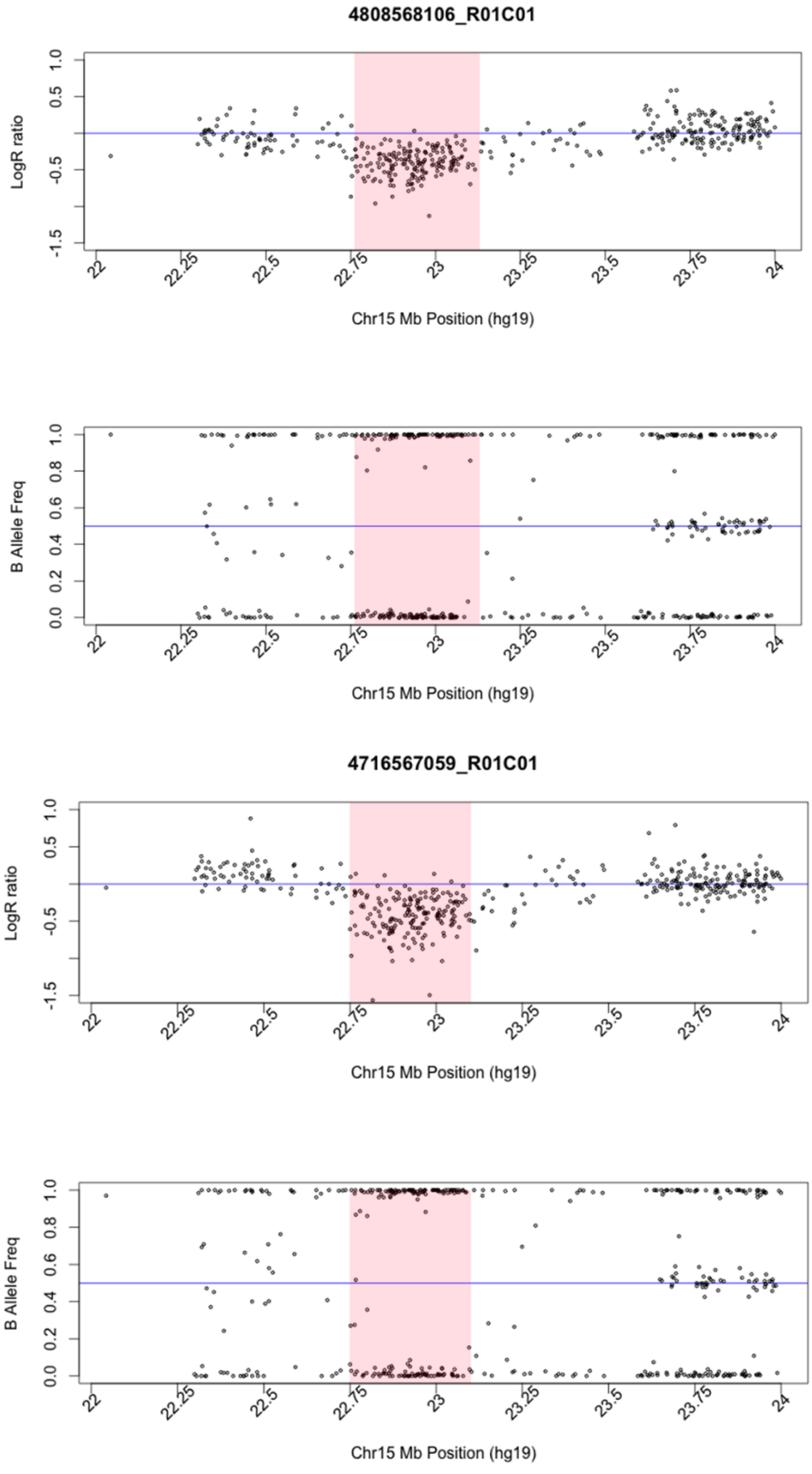
15q11.2 (BP1-BP2) deletions. *Cont.*



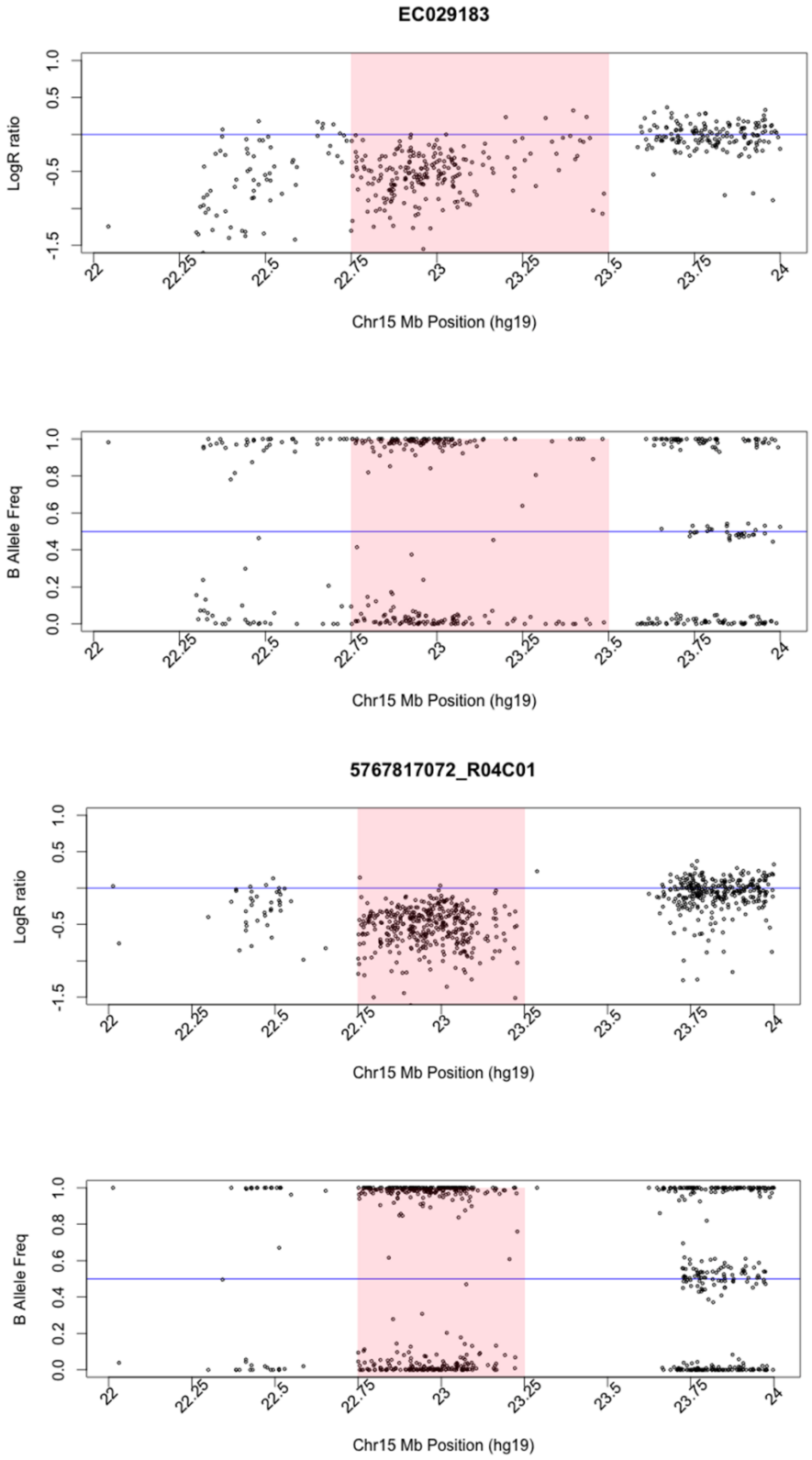
15q11.2 (BP1-BP2) deletions. *Cont.*



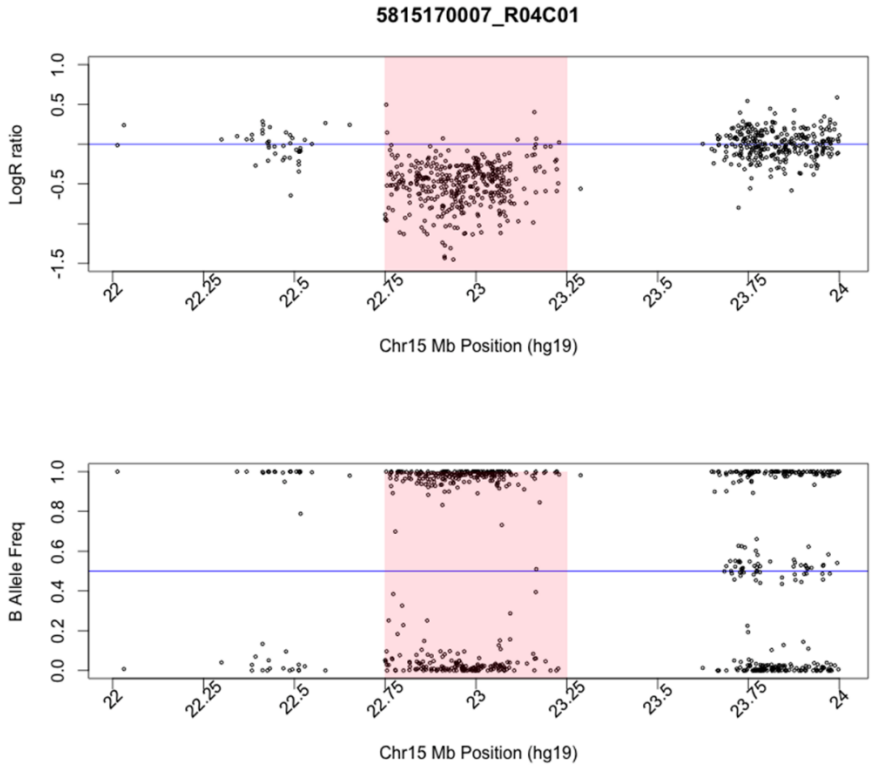
15q11.2 (BP1-BP2) deletions. *Cont.*



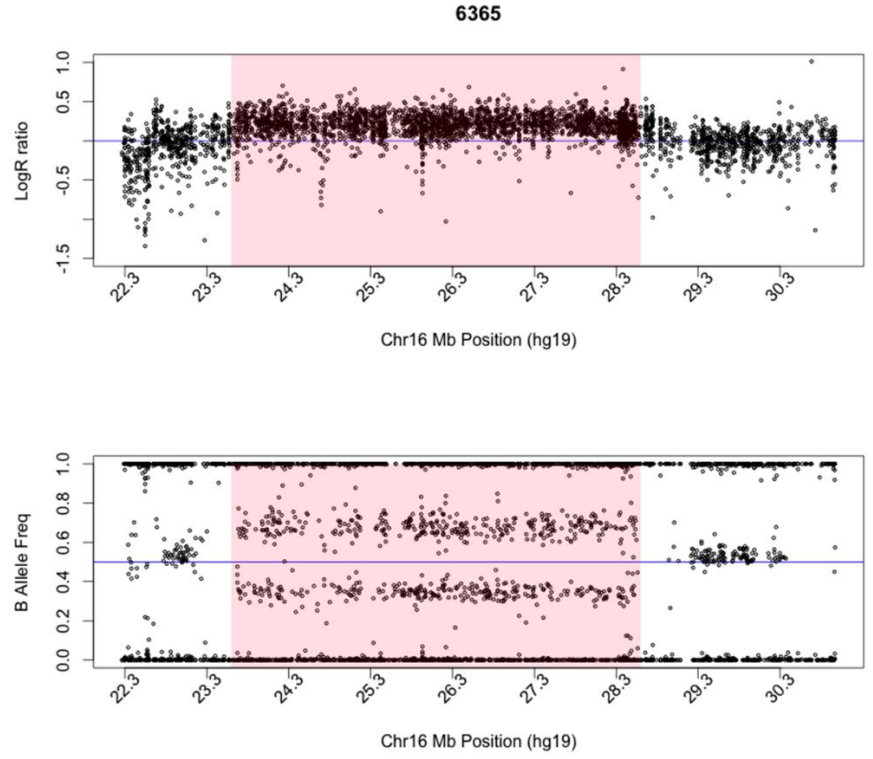
15q11.2 (BP1-BP2) deletions. *Cont.*



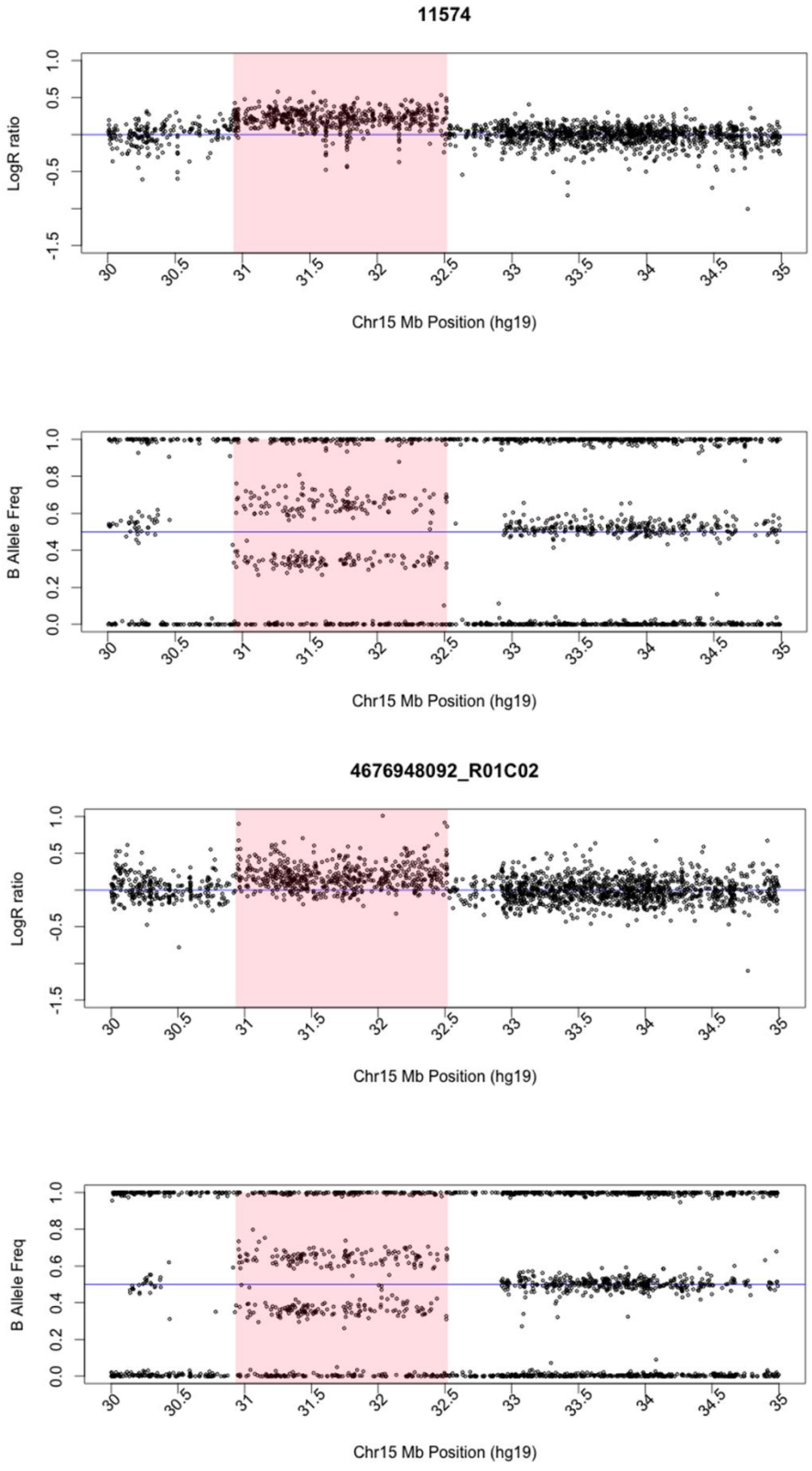
15q11.2 (BP1-BP2) deletions. *Cont.*



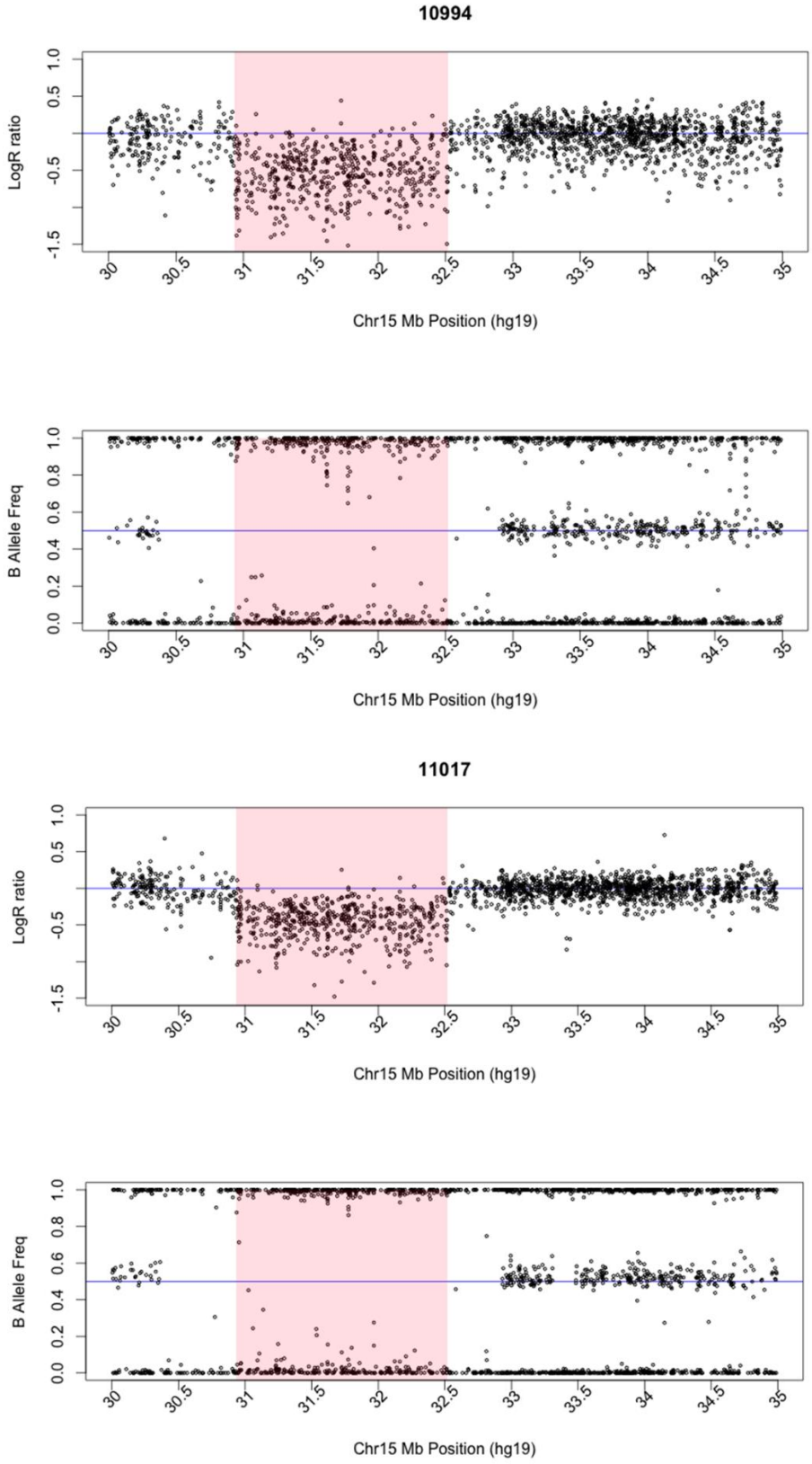
15q11.2-q13 (BP2-BP3) duplication



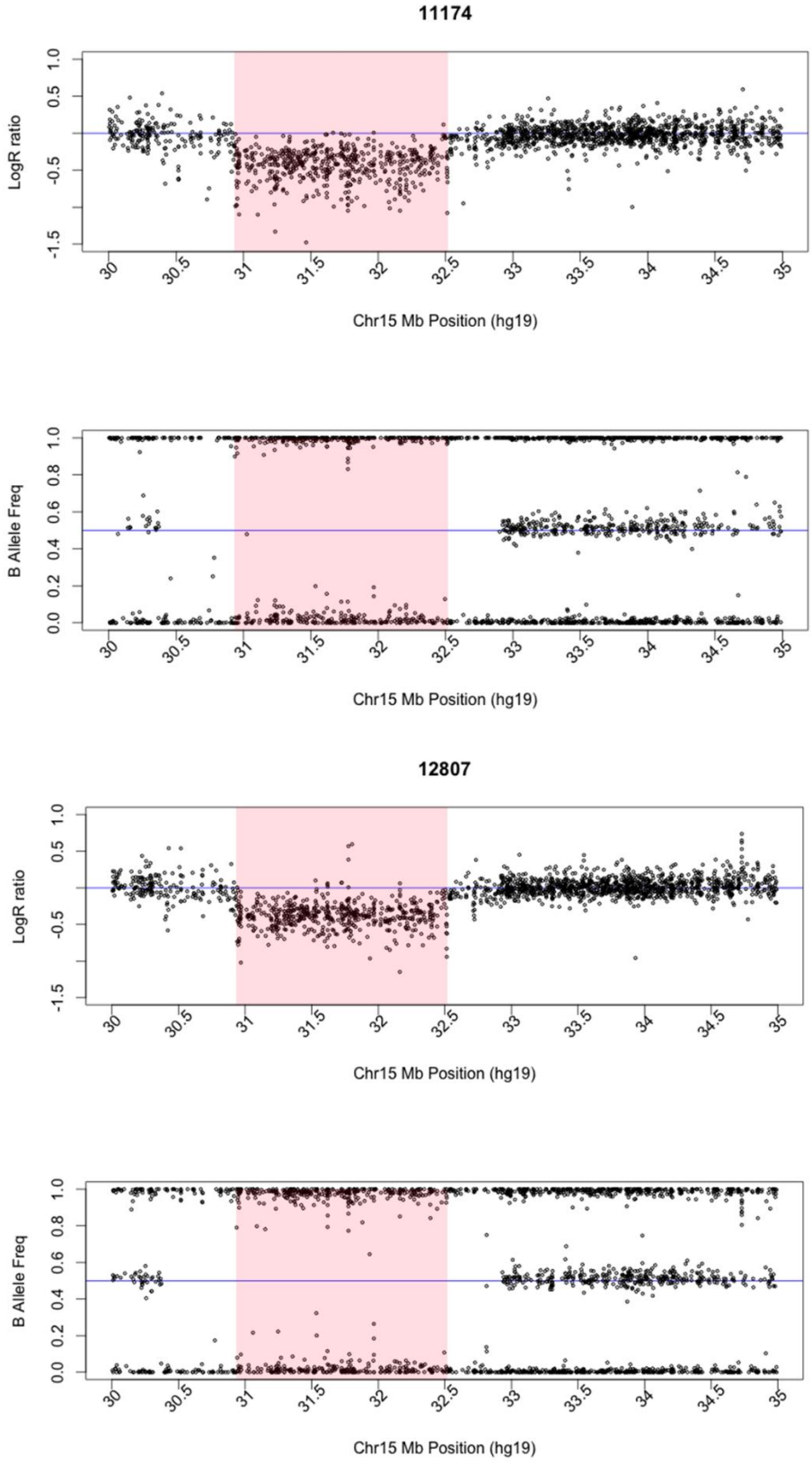
15q13.2-q13.3-q13 (BP4-BP5) duplications



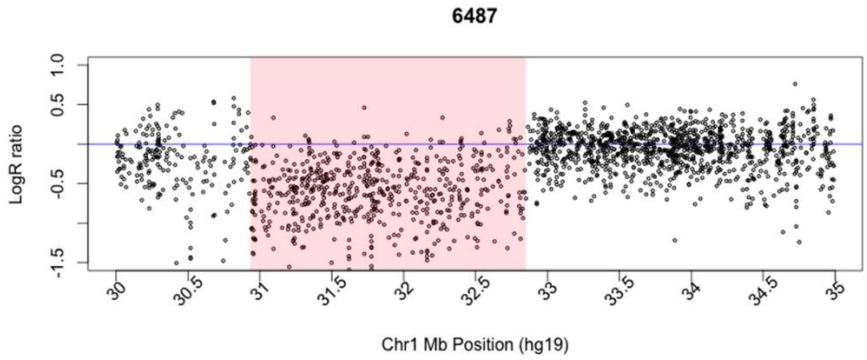
15q13.2-q13.3-q13 (BP4-BP5) deletions

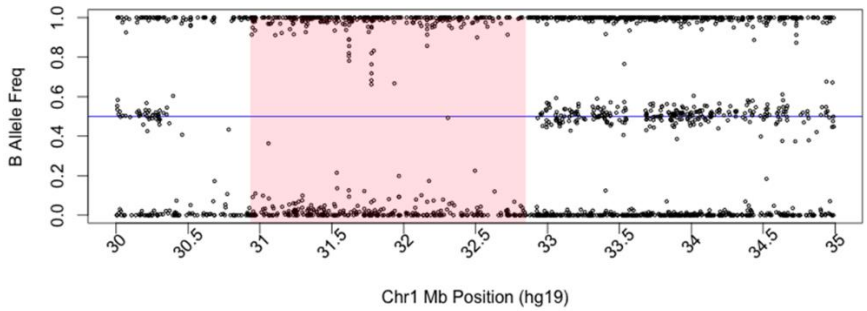


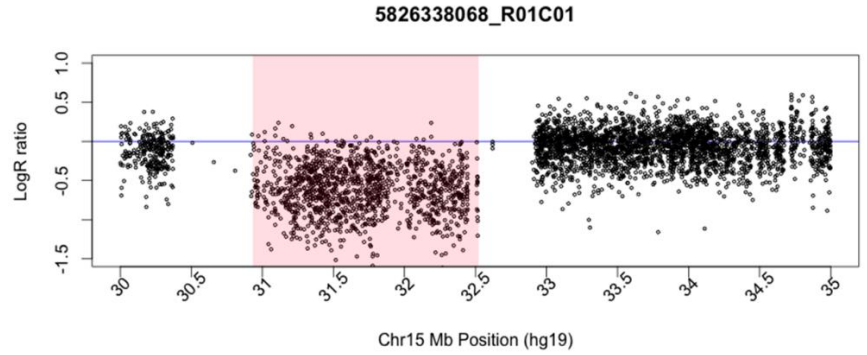
15q13.2-q13.3-q13 (BP4-BP5) deletions. *Cont.*

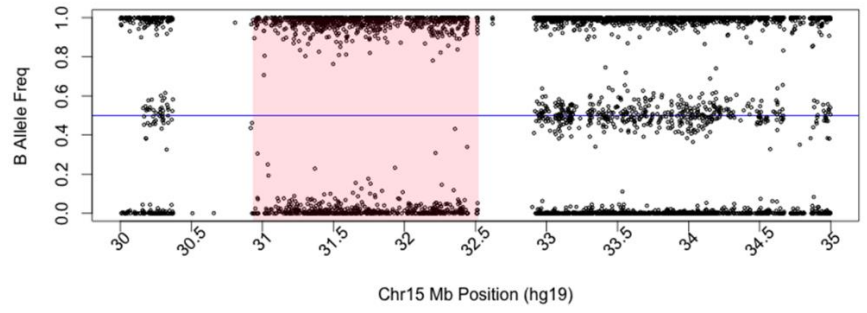


15q13.2-q13.3-q13 (BP4-BP5) deletions. *Cont.*

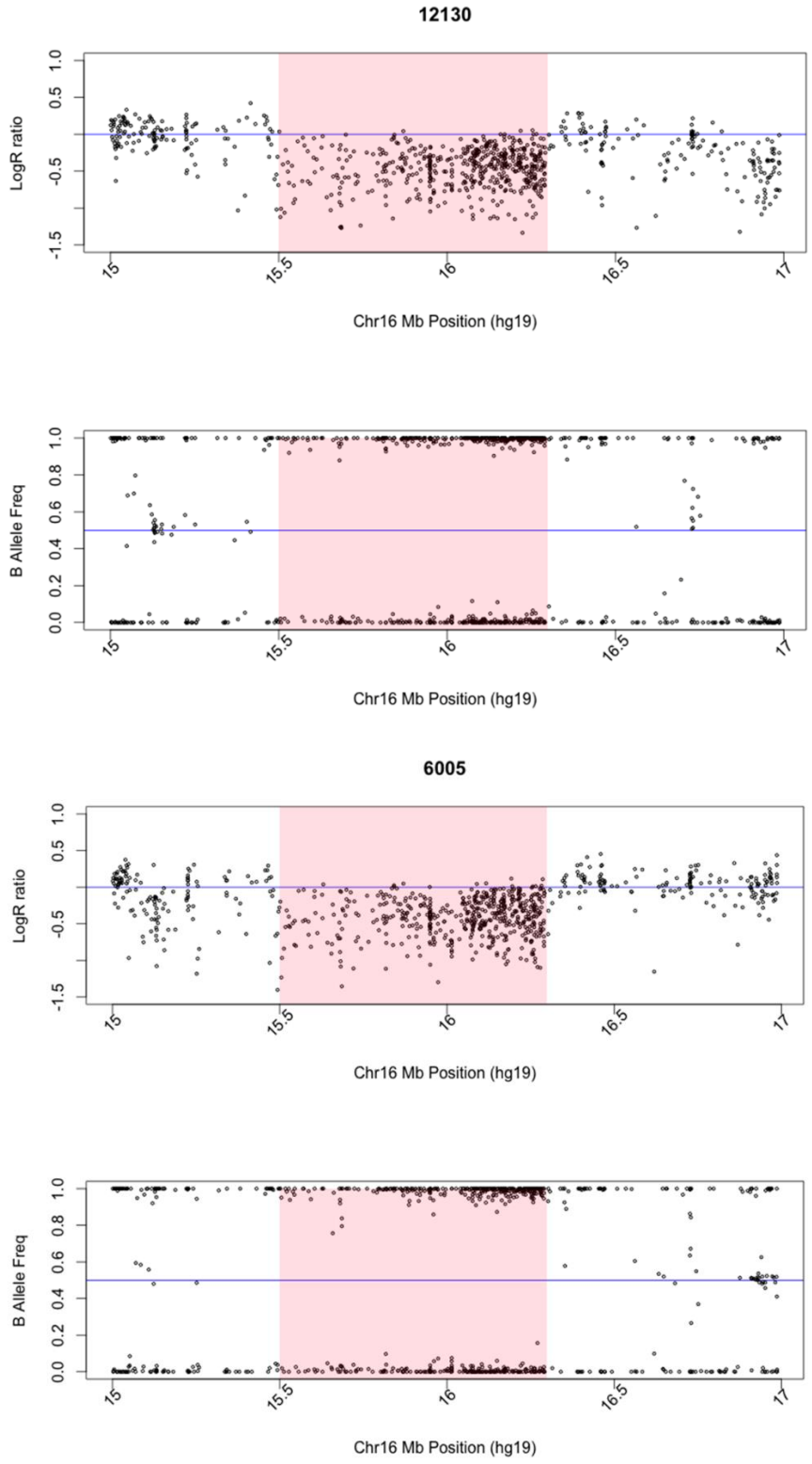




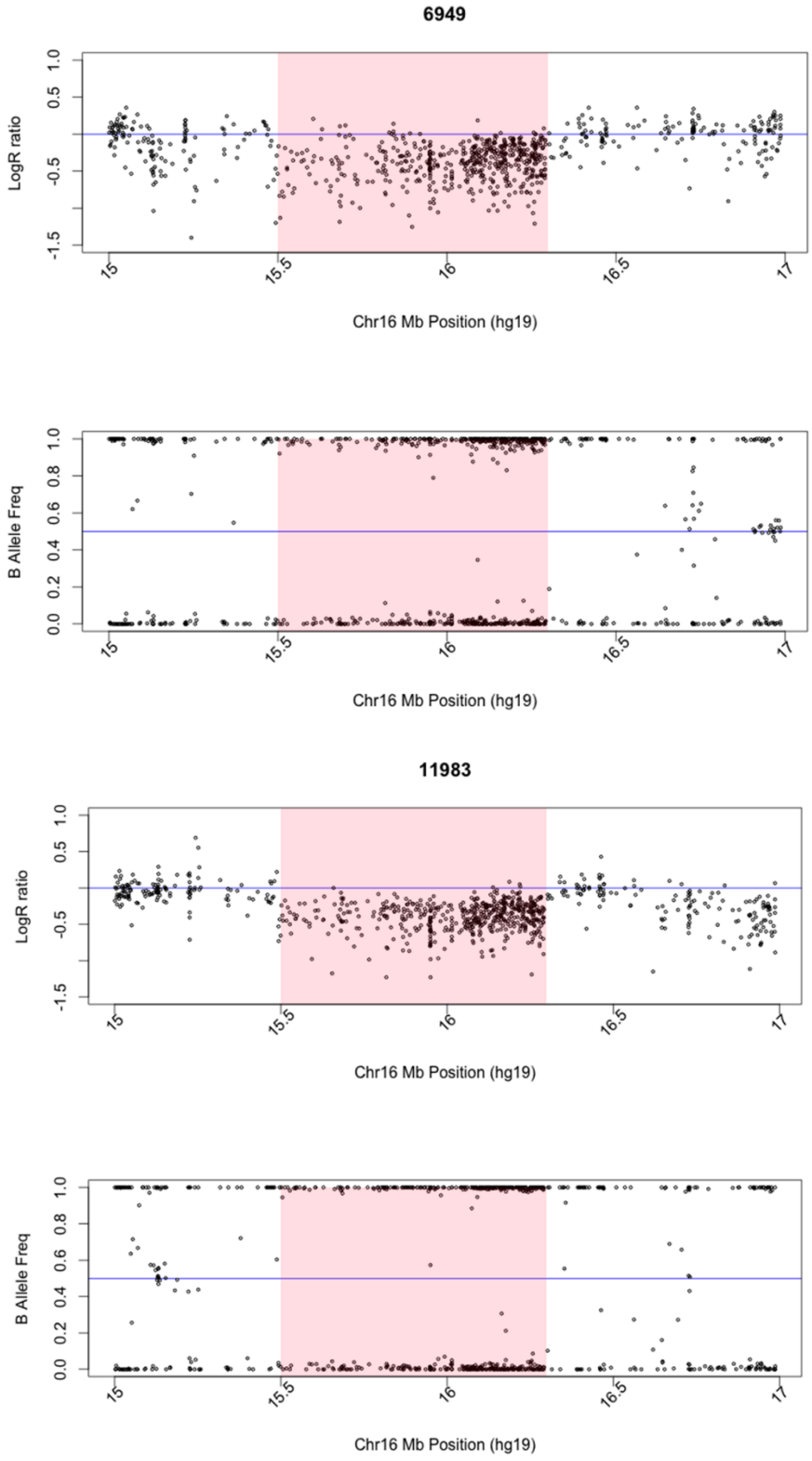




16p13.11 deletions

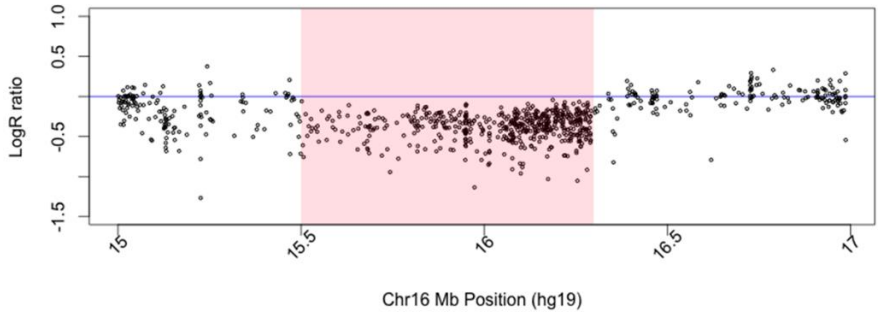


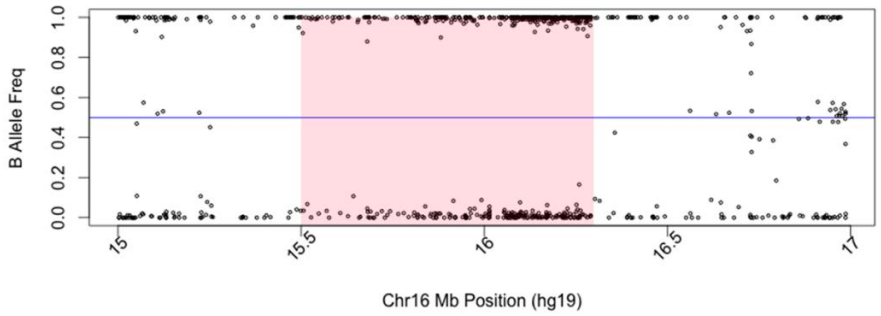
16p13.11 deletions. *Cont.*



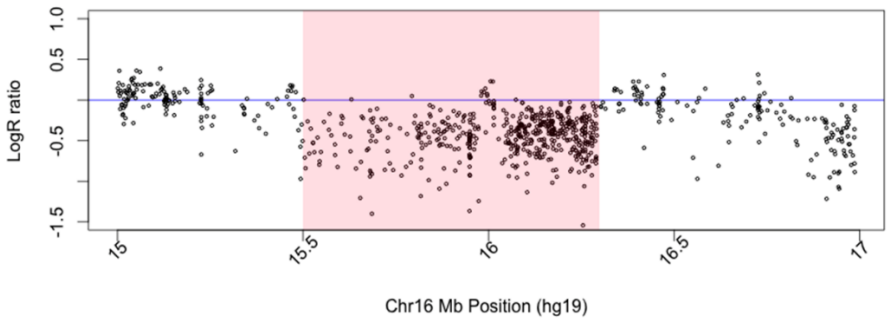
16p13.11 deletions. *Cont.*

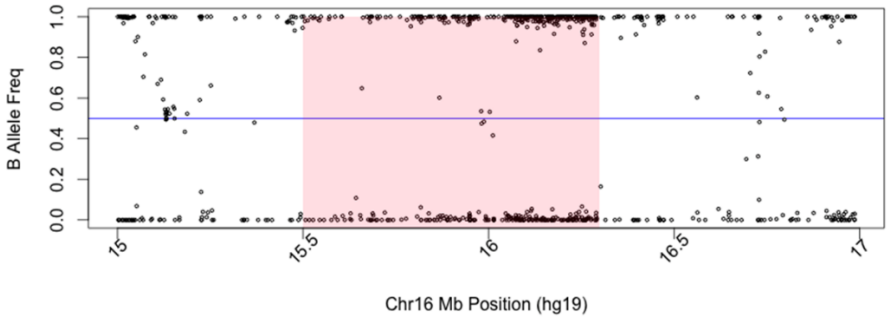
12240



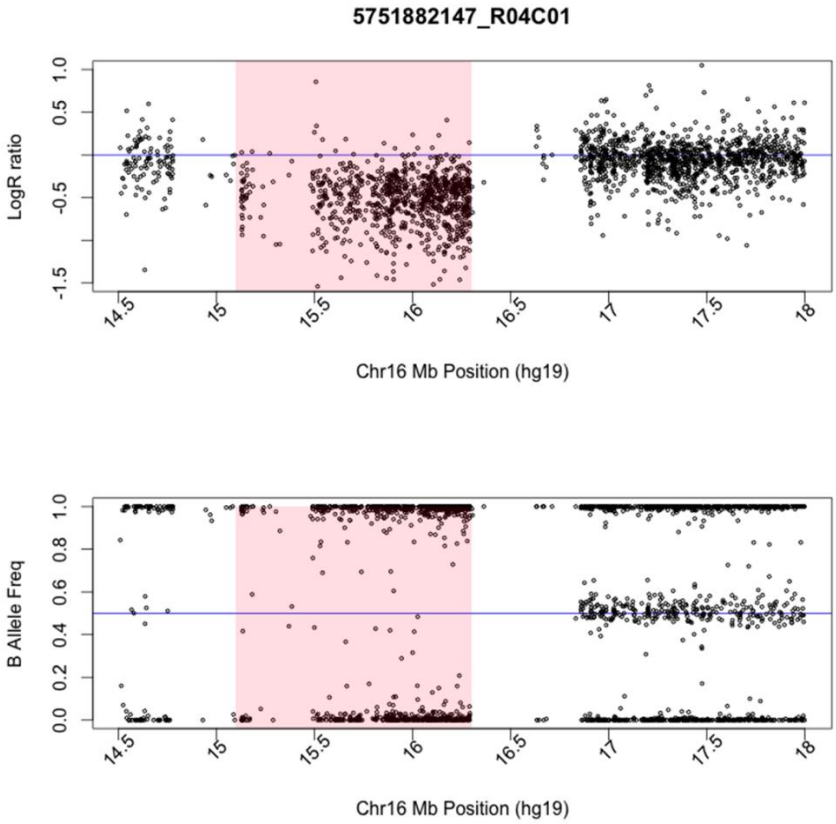


10357

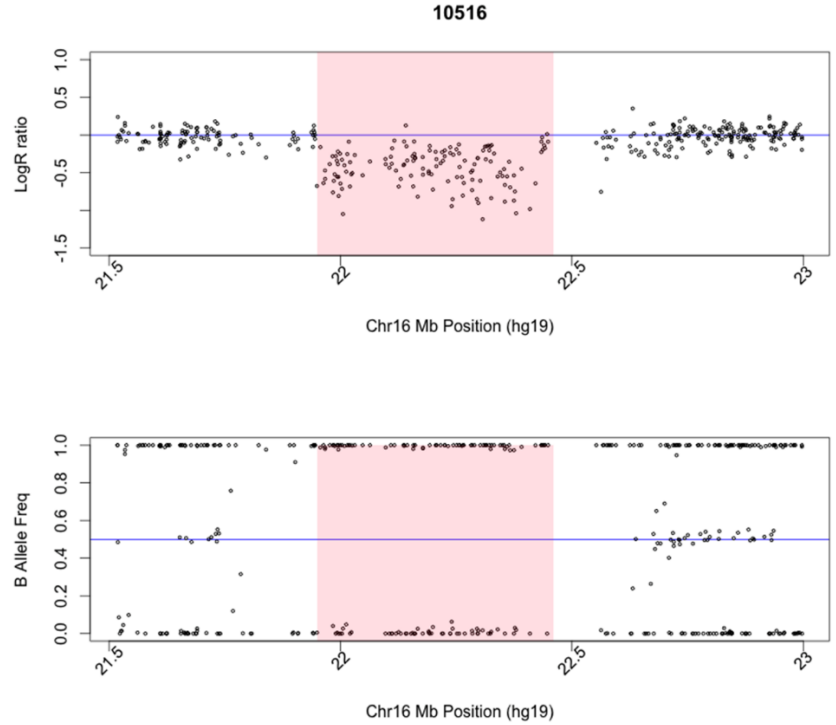




16p13.11 deletions. *Cont.*

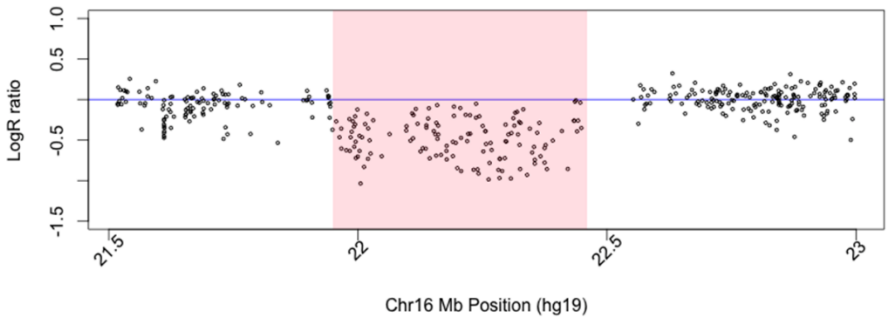


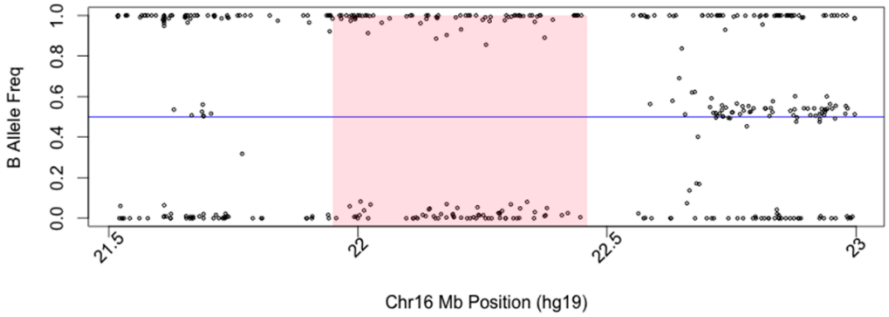
16p12.1 deletions



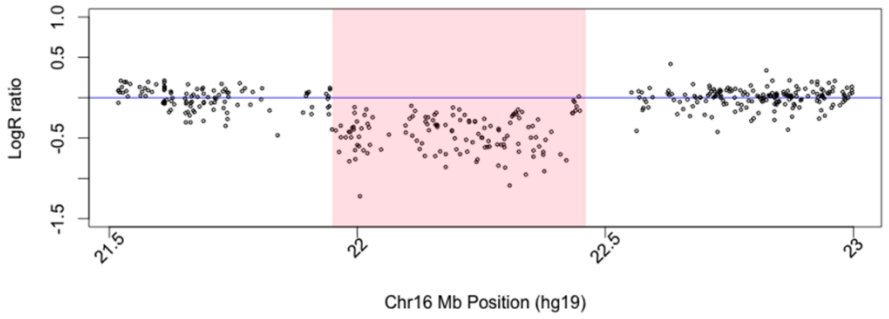
16p12.1 deletions. *Cont.*

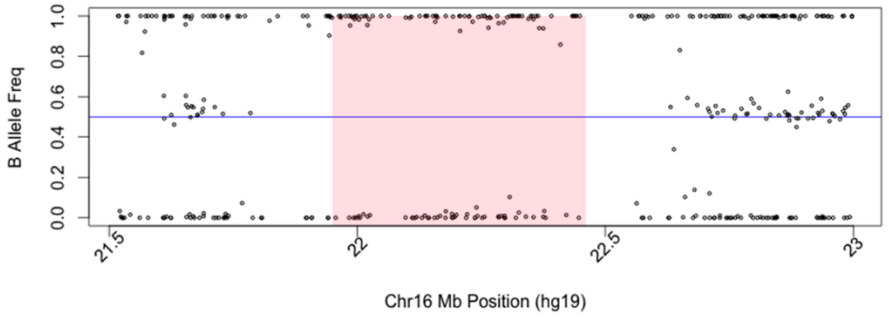
12004



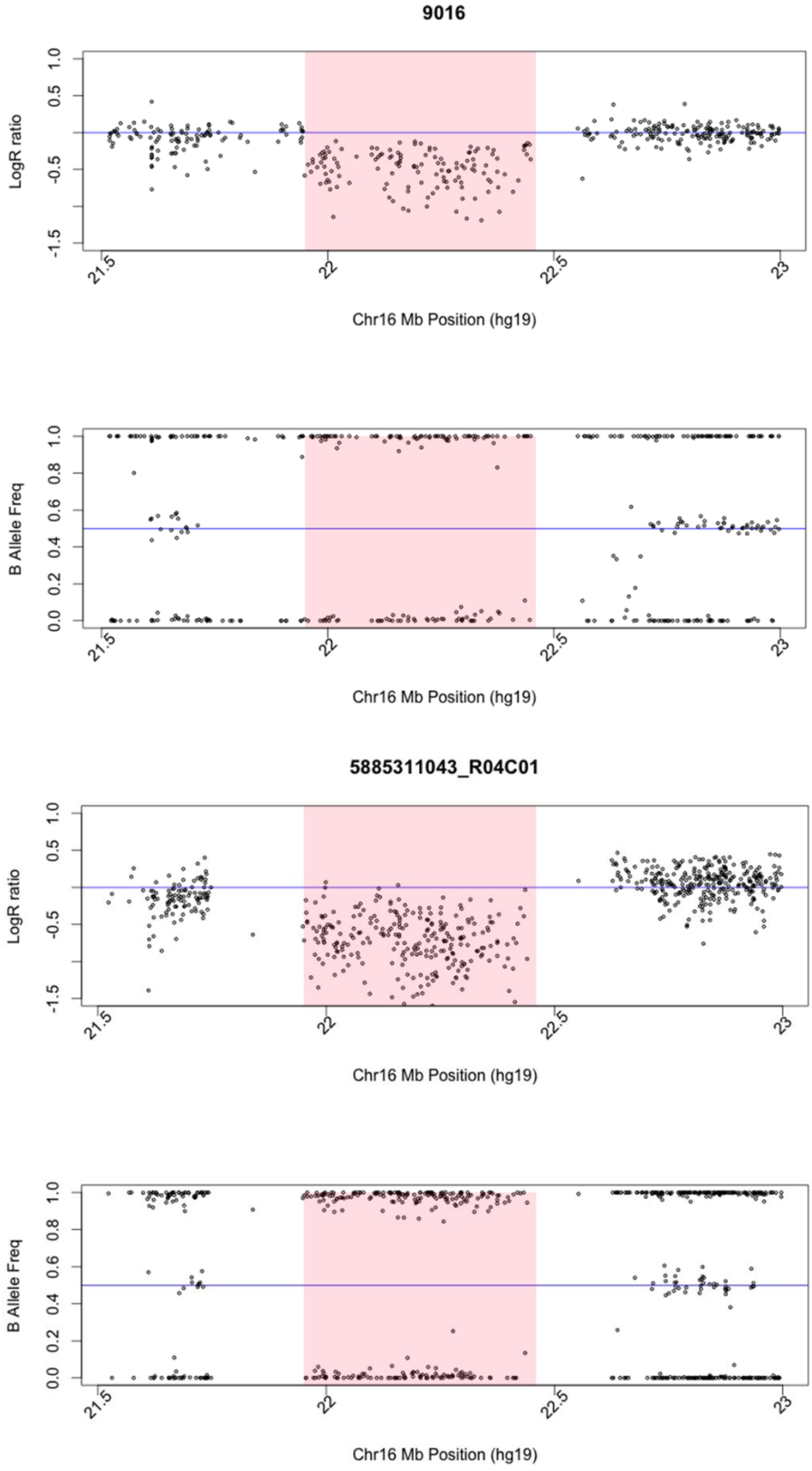


7753

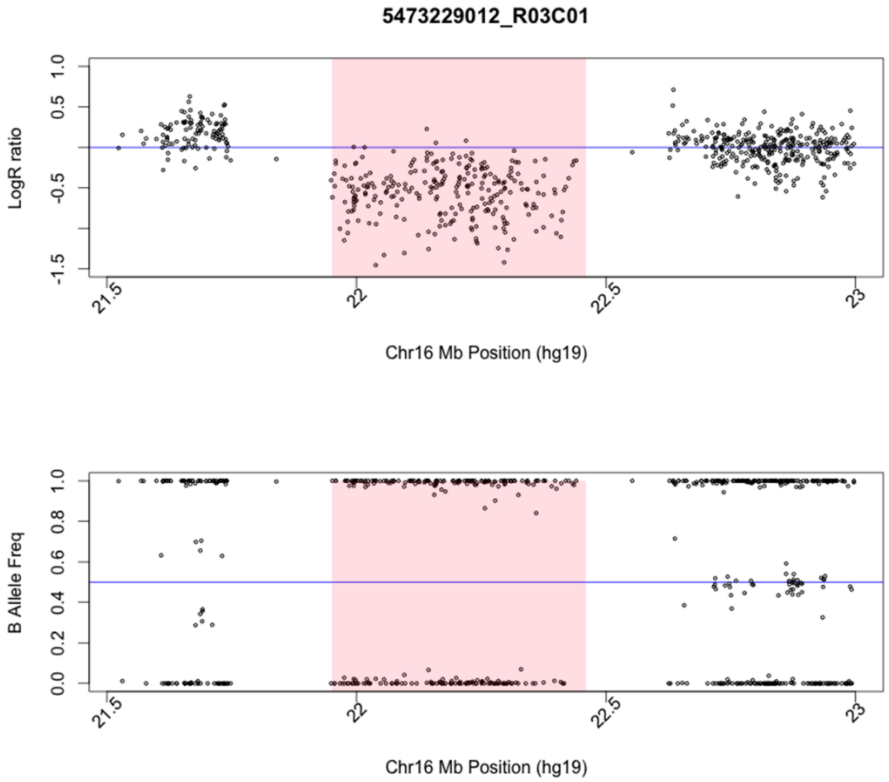




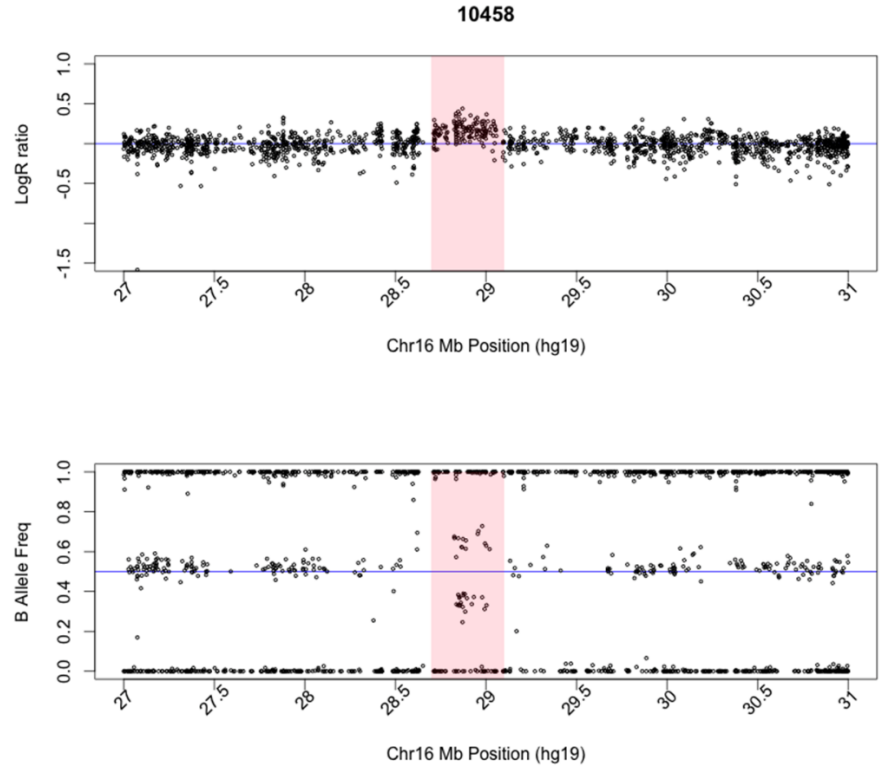
16p12.1 deletions. *Cont.*



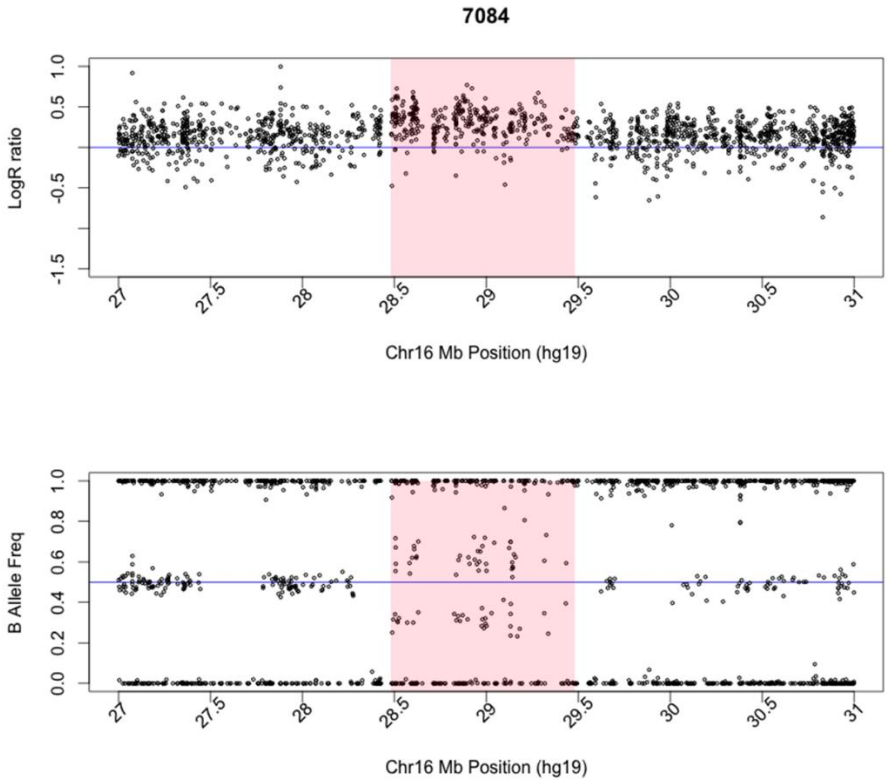
16p12.1 deletions. *Cont.*



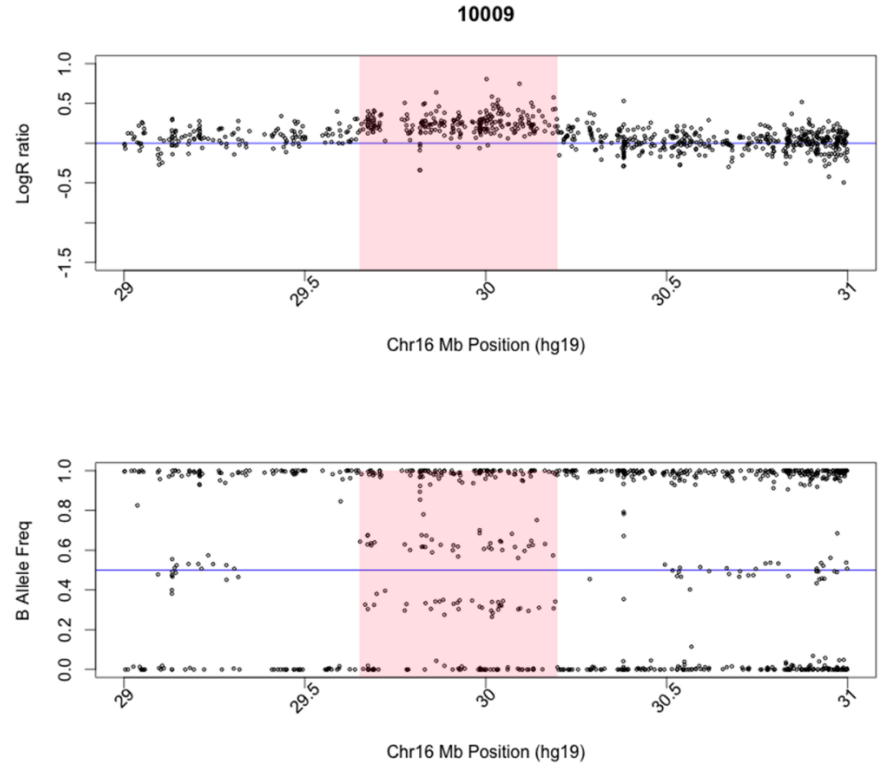
16p11.2 distal duplications



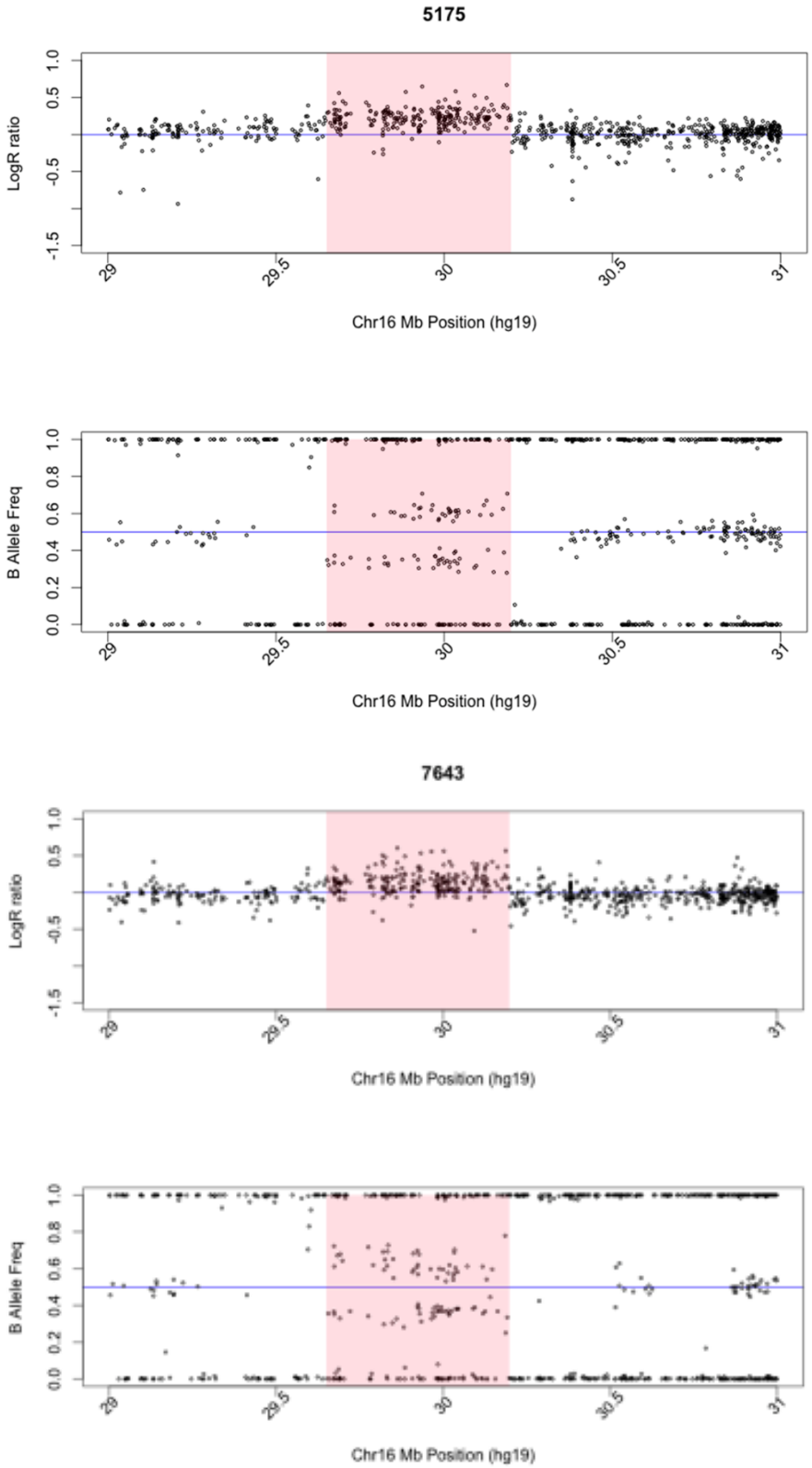
16p11.2 distal duplications. *Cont.*



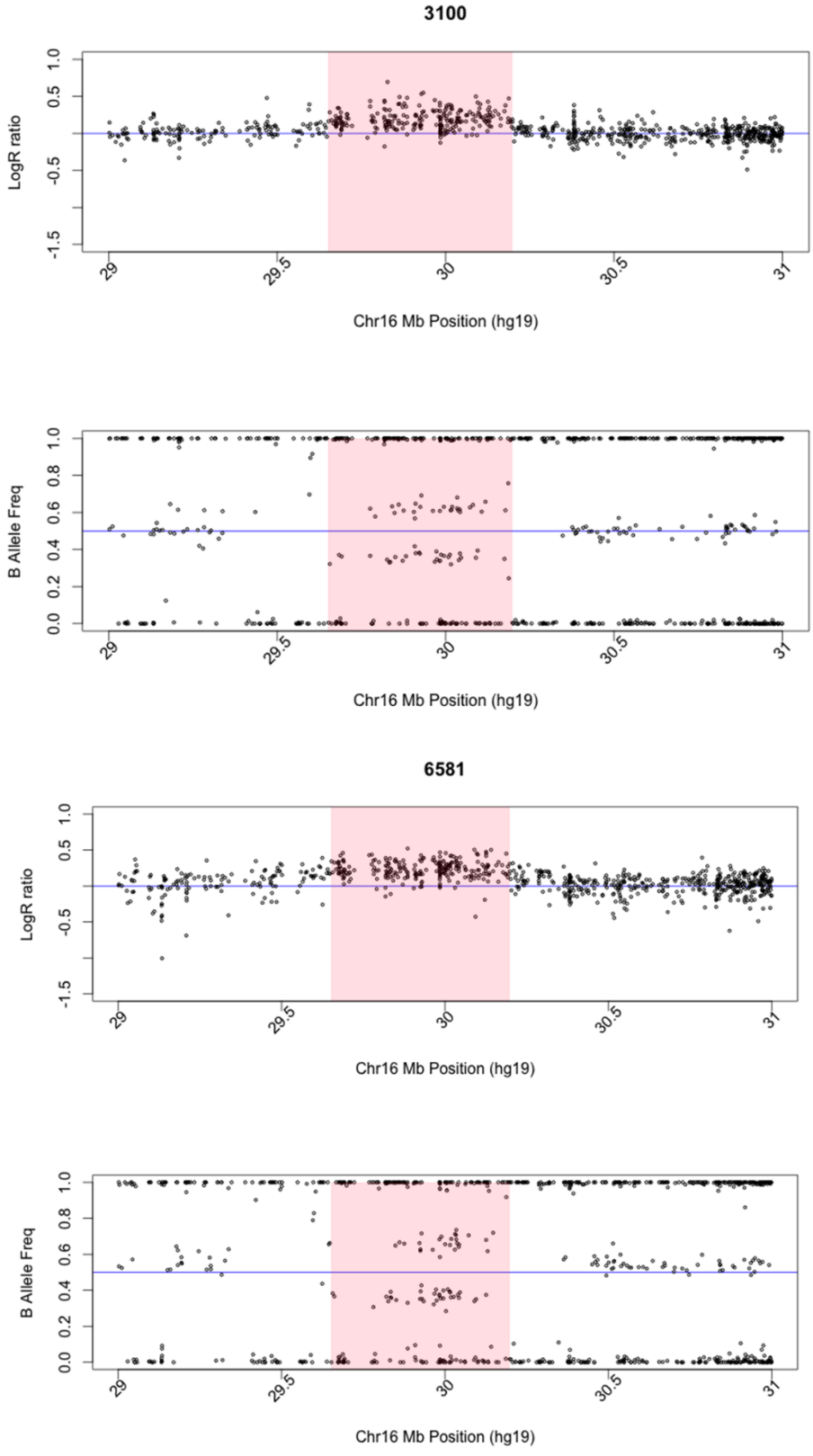
16p11.2 duplications



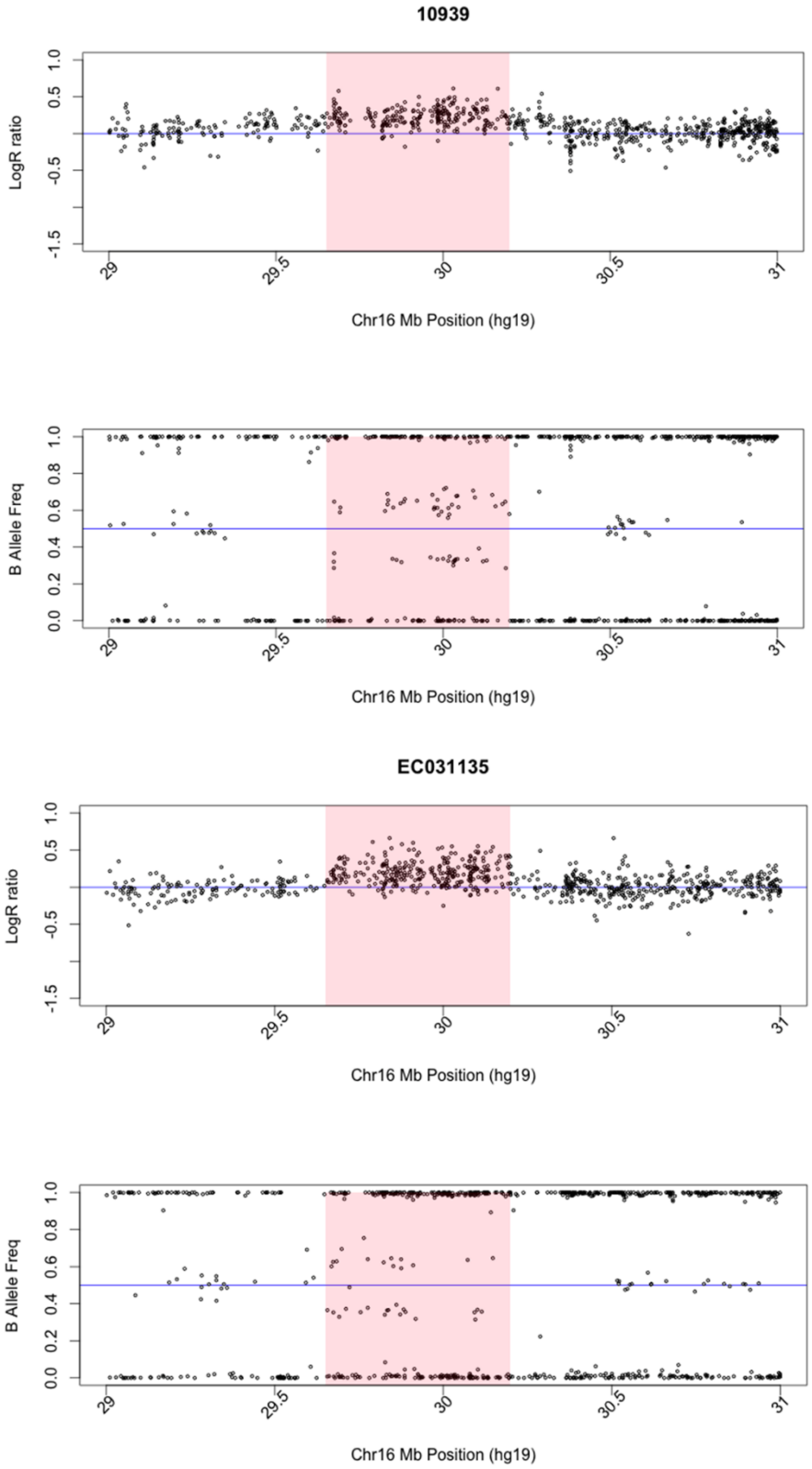
16p11.2 duplications. *Cont.*



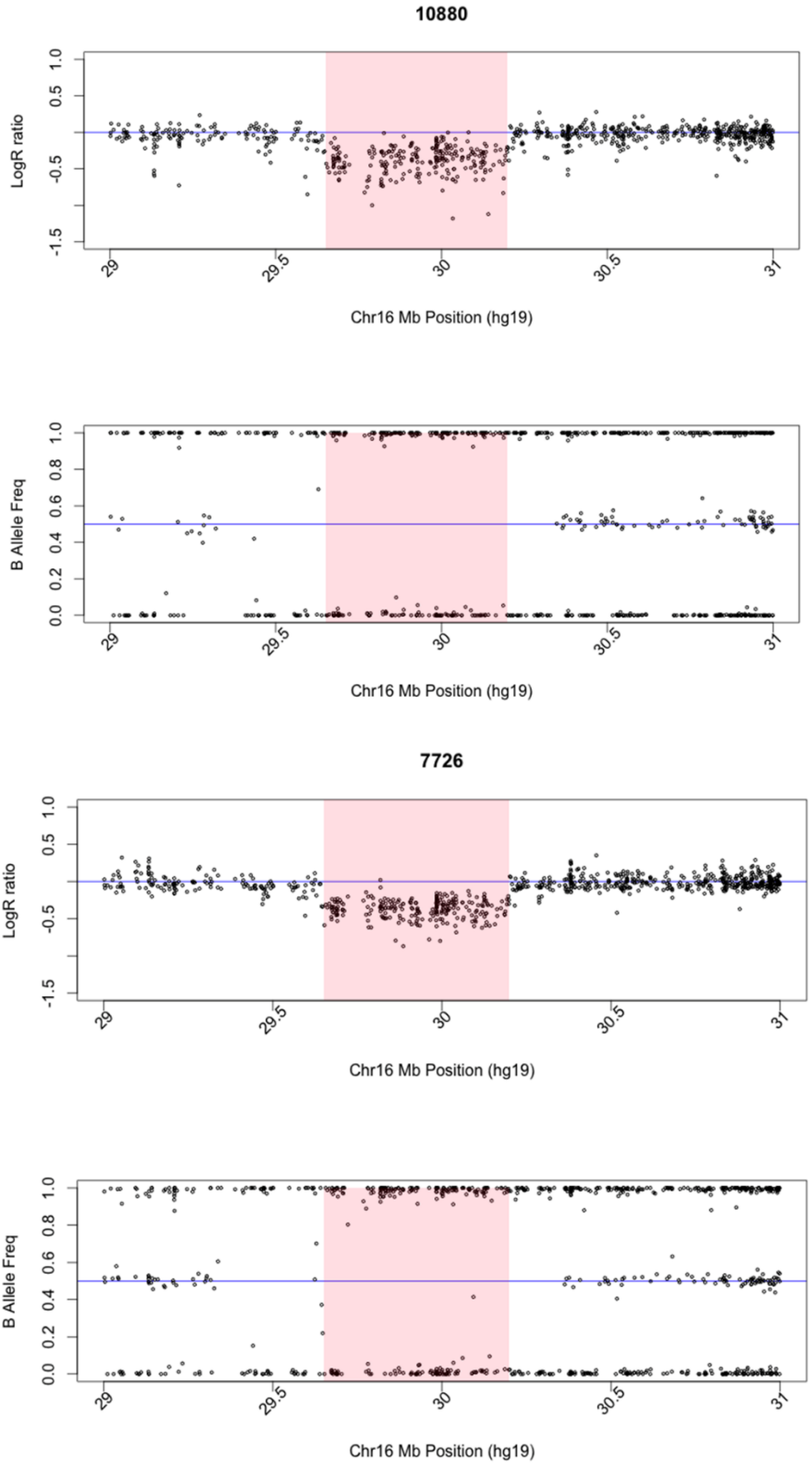
16p11.2 duplications. *Cont.*



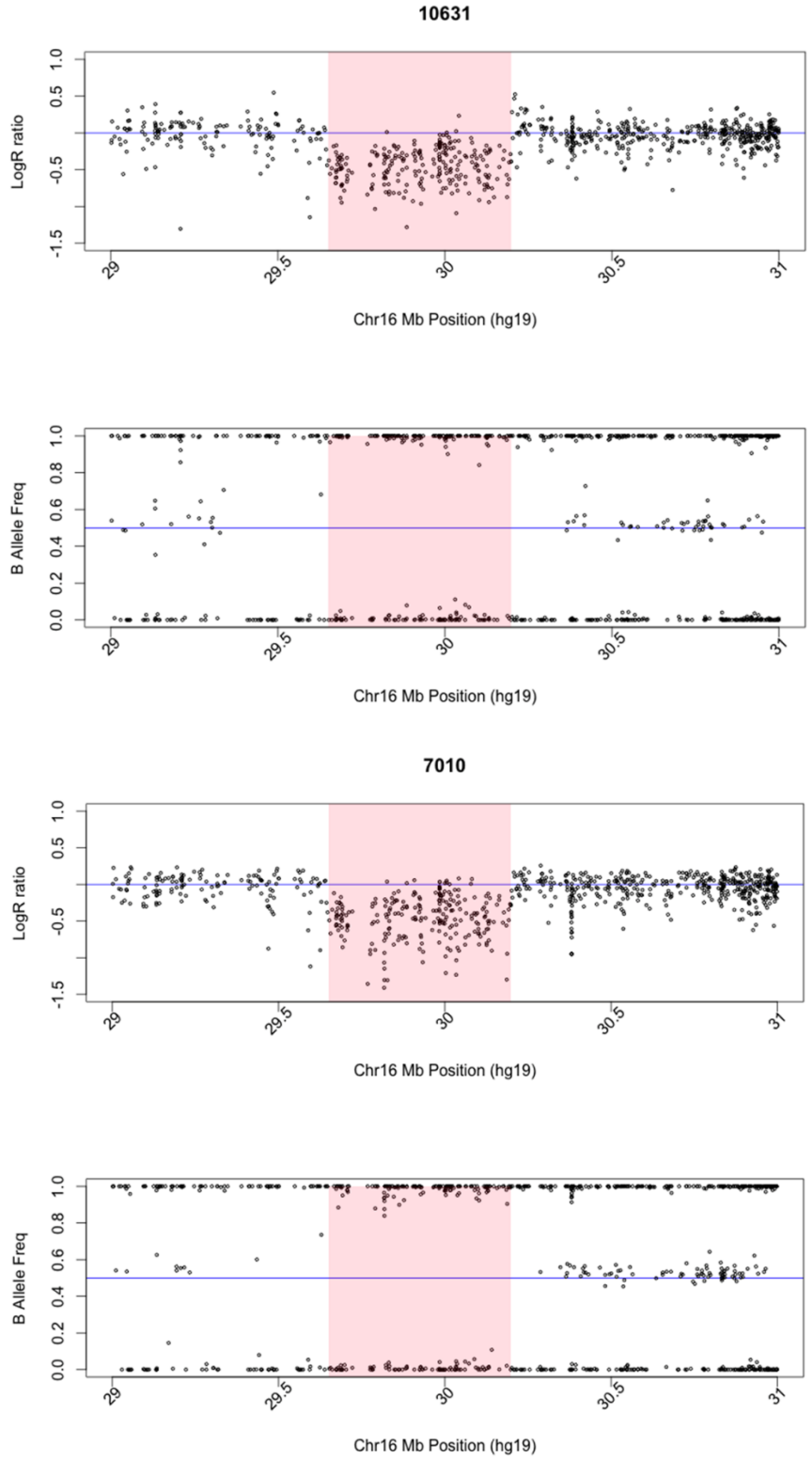
16p11.2 duplications. *Cont.*



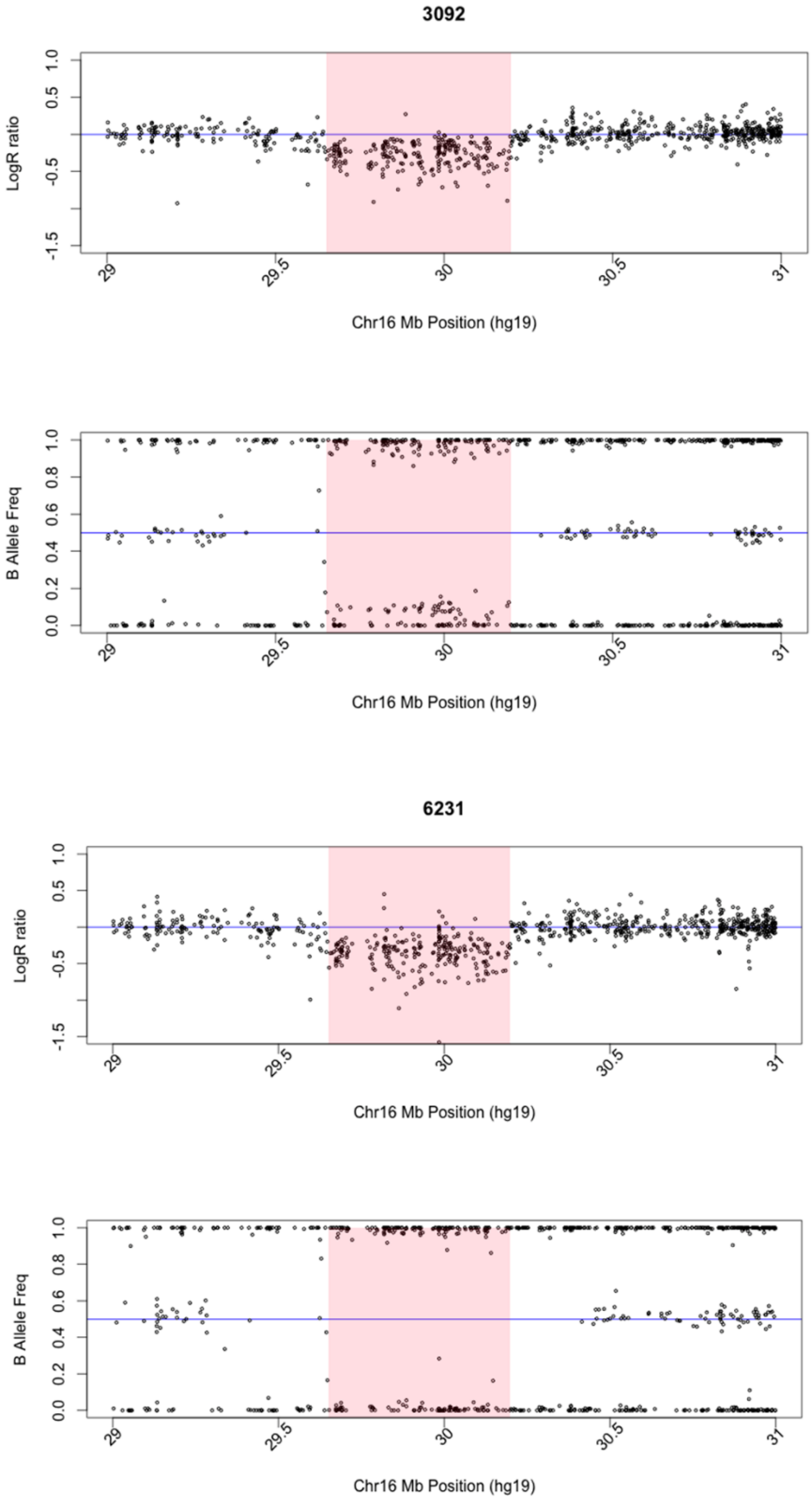
16p11.2 deletions



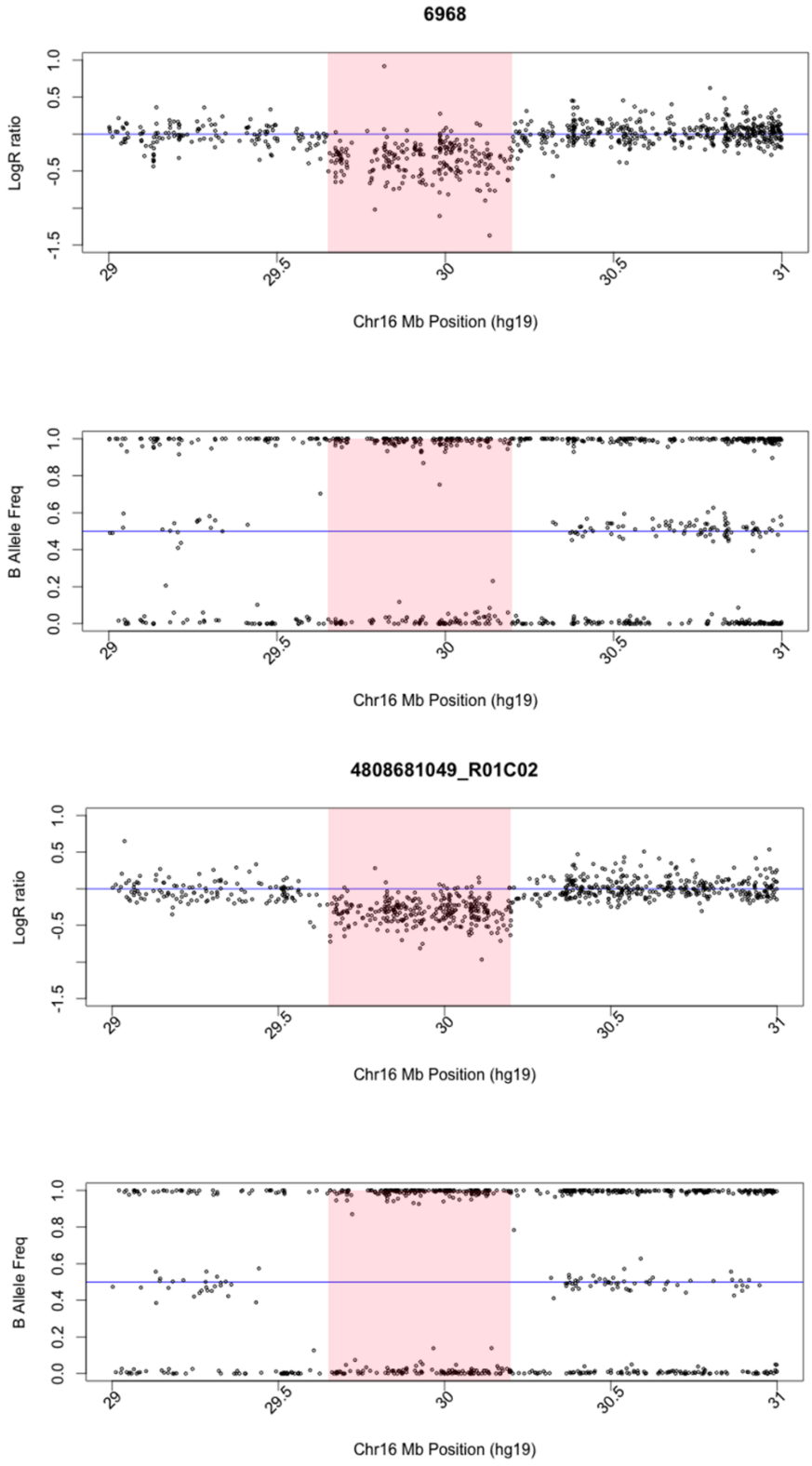
16p11.2 deletions. *Cont.*



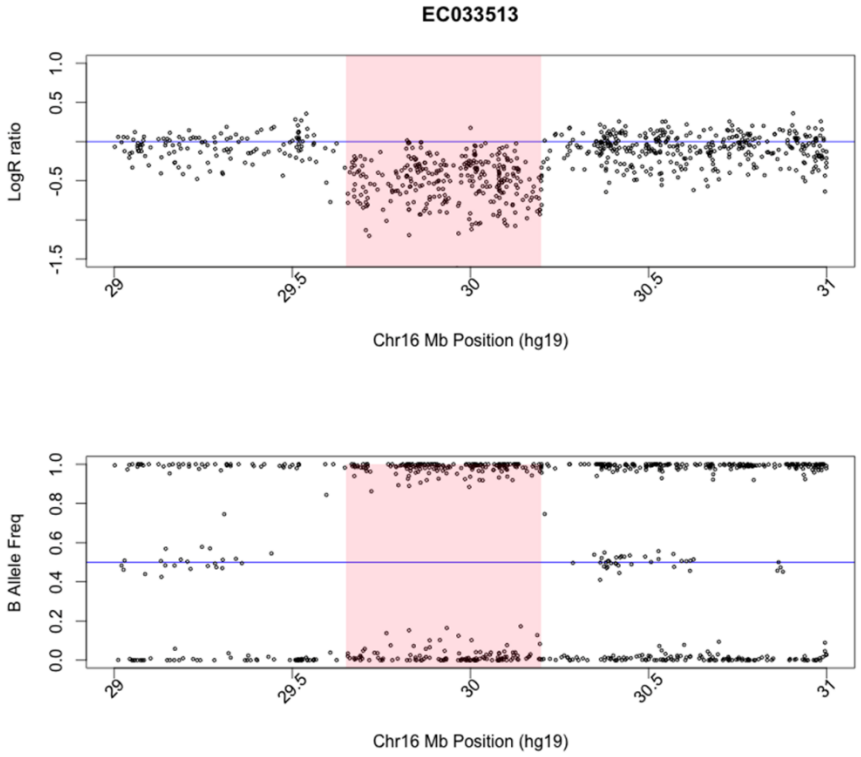
16p11.2 deletions. *Cont.*



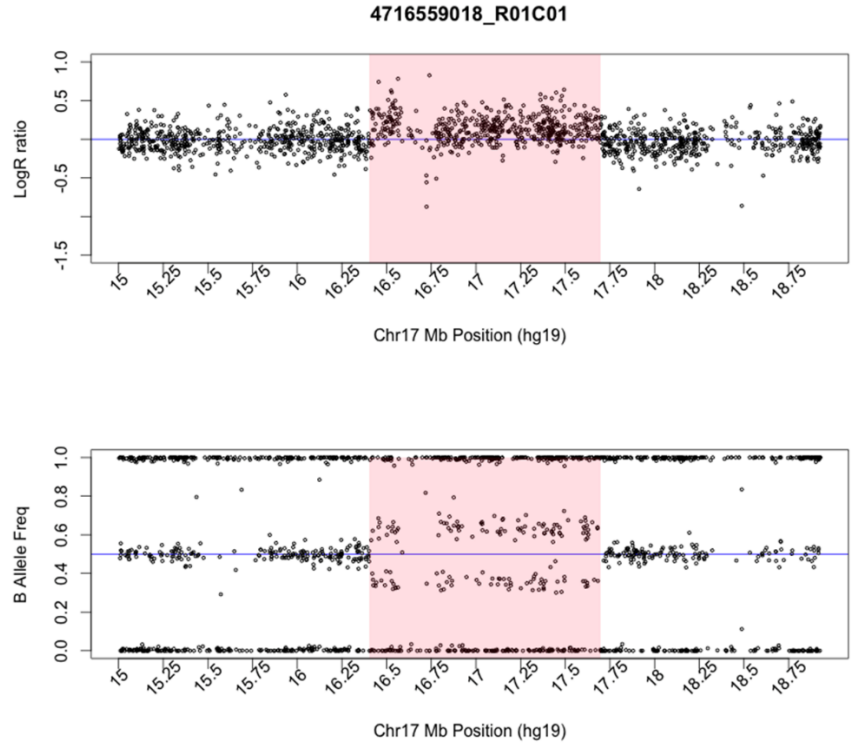
16p11.2 deletions. *Cont.*



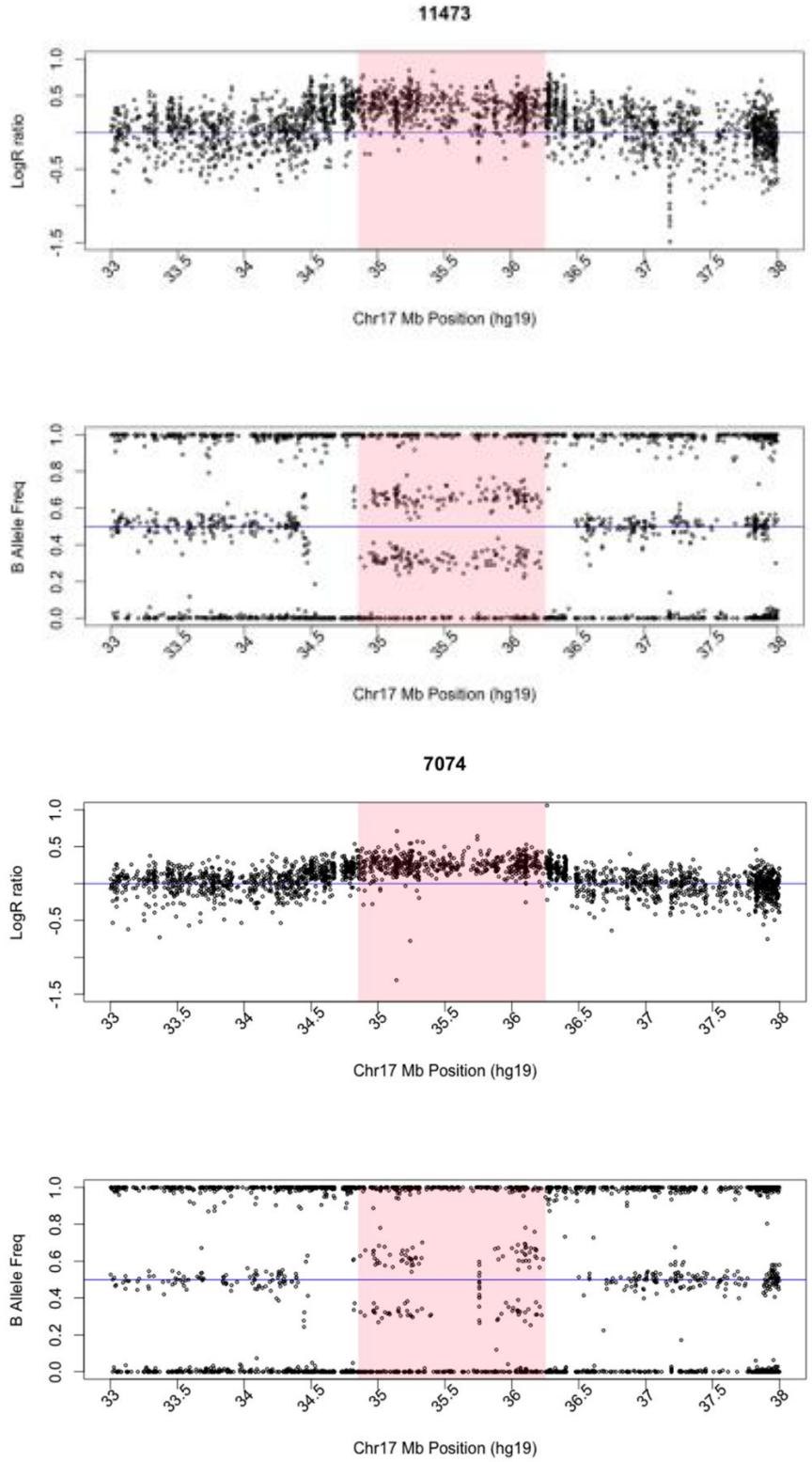
16p11.2 deletions. *Cont.*



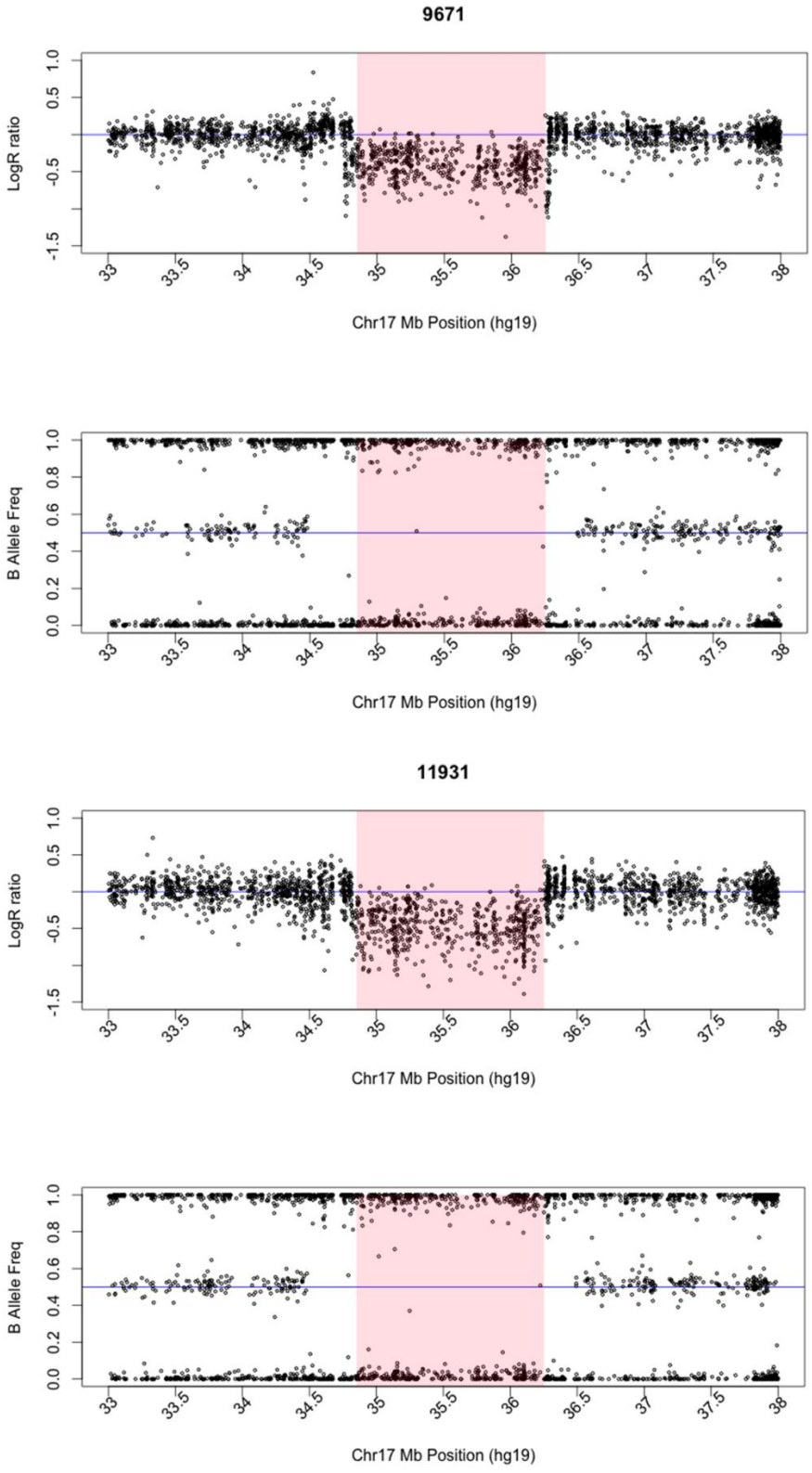
17p11.2 duplication



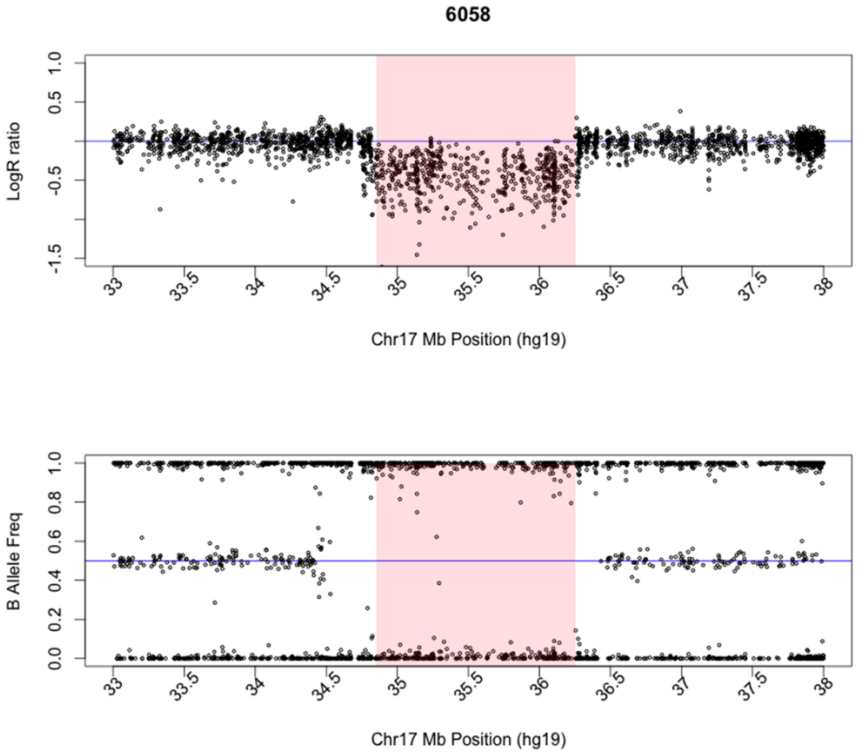
17q12 duplications



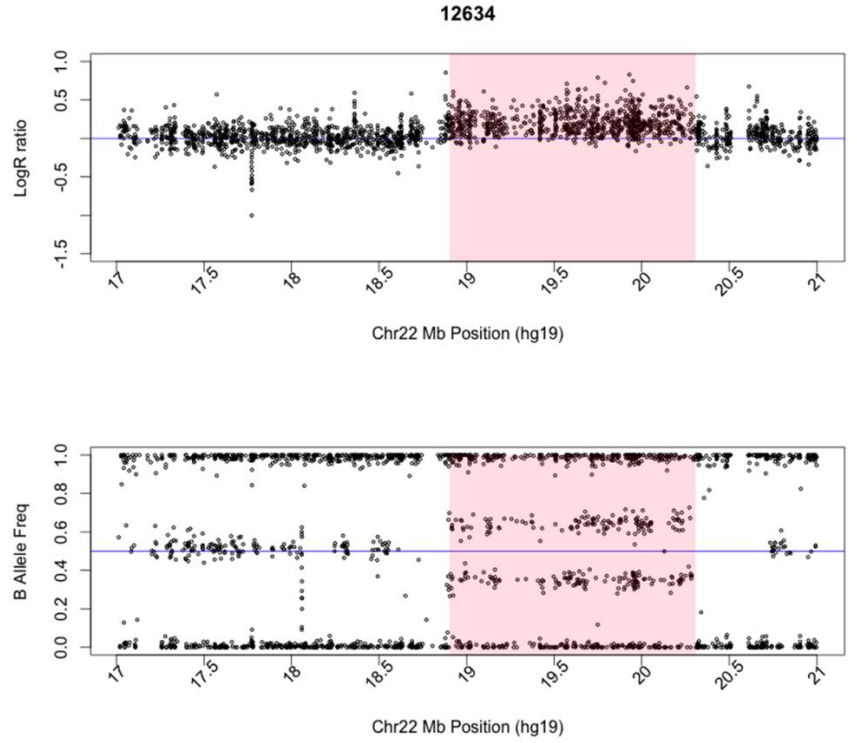
17q12 deletions



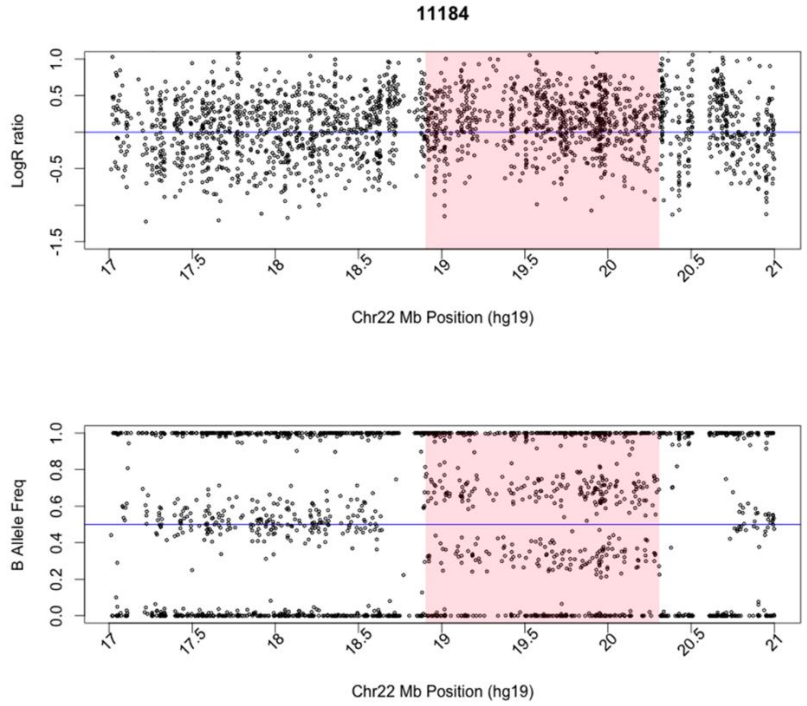
17q12 deletions. *Cont.*



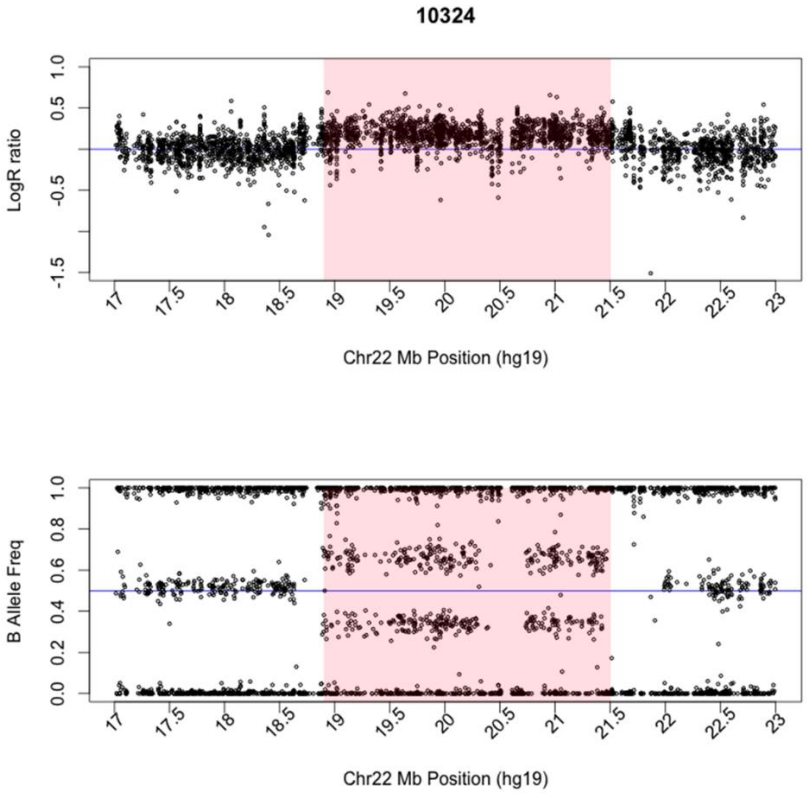
22q11.2 (proximal A-B) duplications



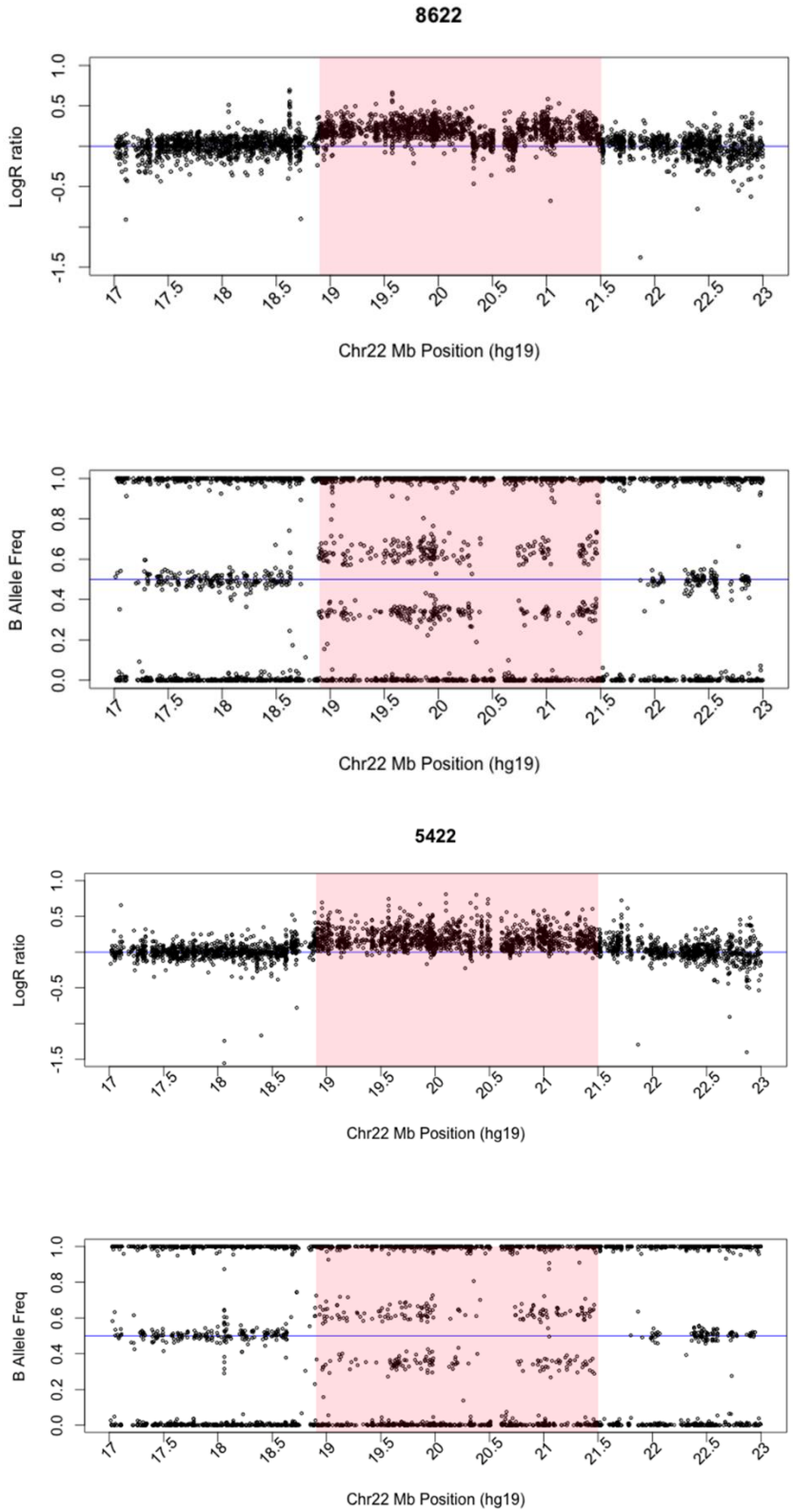
22q11.2 (proximal A-B) duplications. *Cont.*



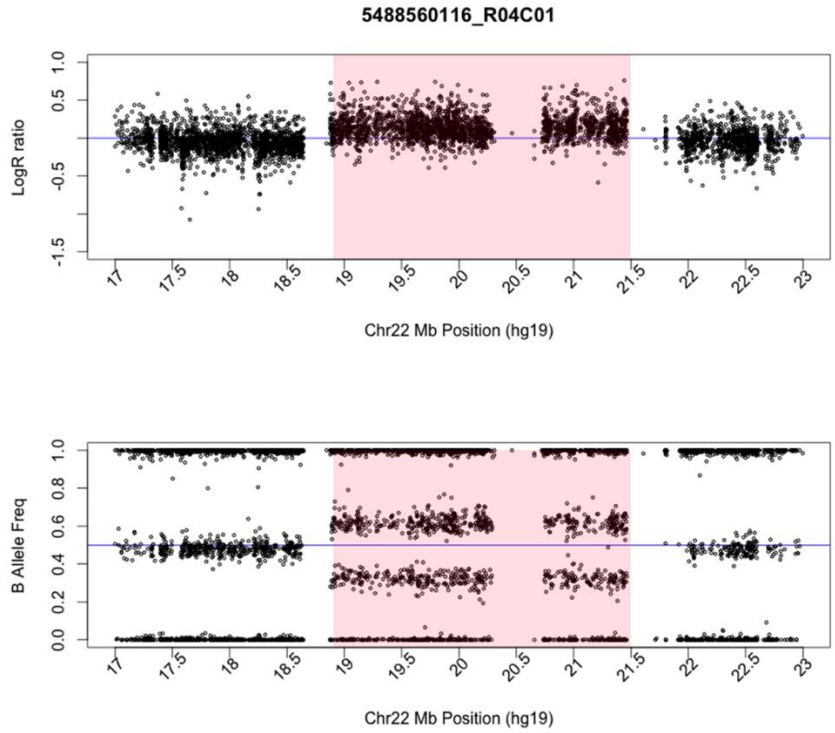
22q11.2 (proximal A-D) duplications



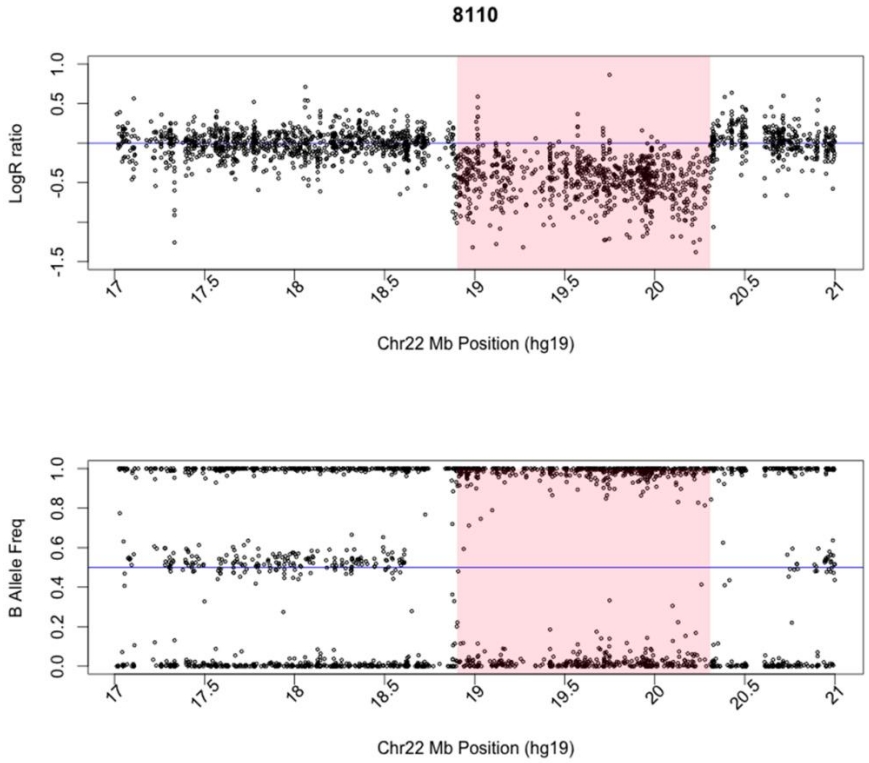
22q11.2 (proximal A-D) duplications. *Cont.*



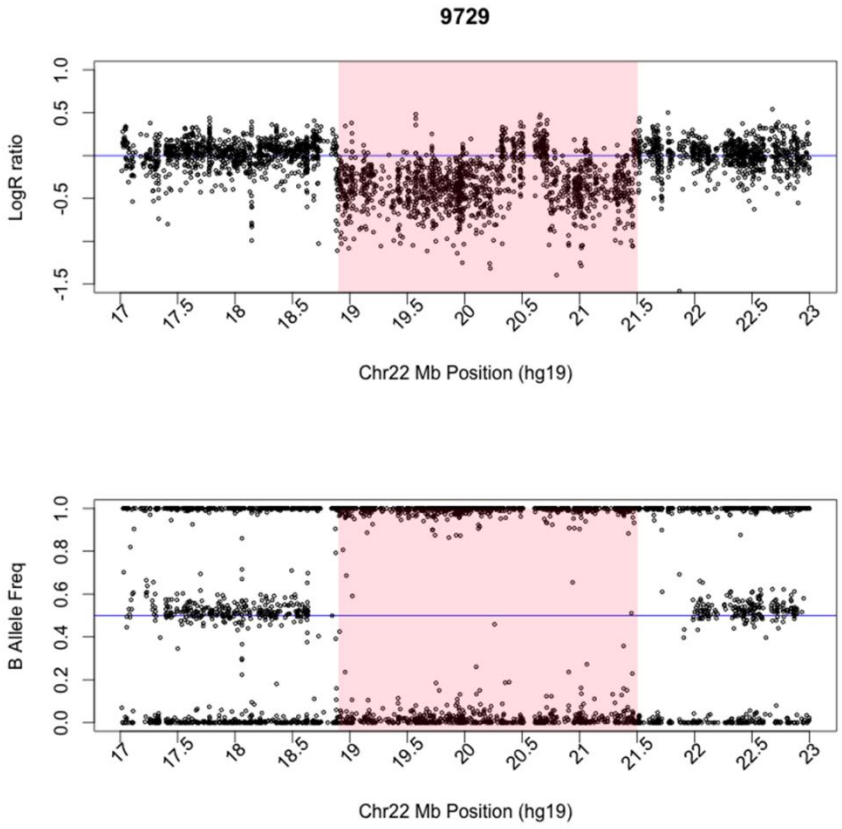
22q11.2 (proximal A-D) duplications. *Cont.*



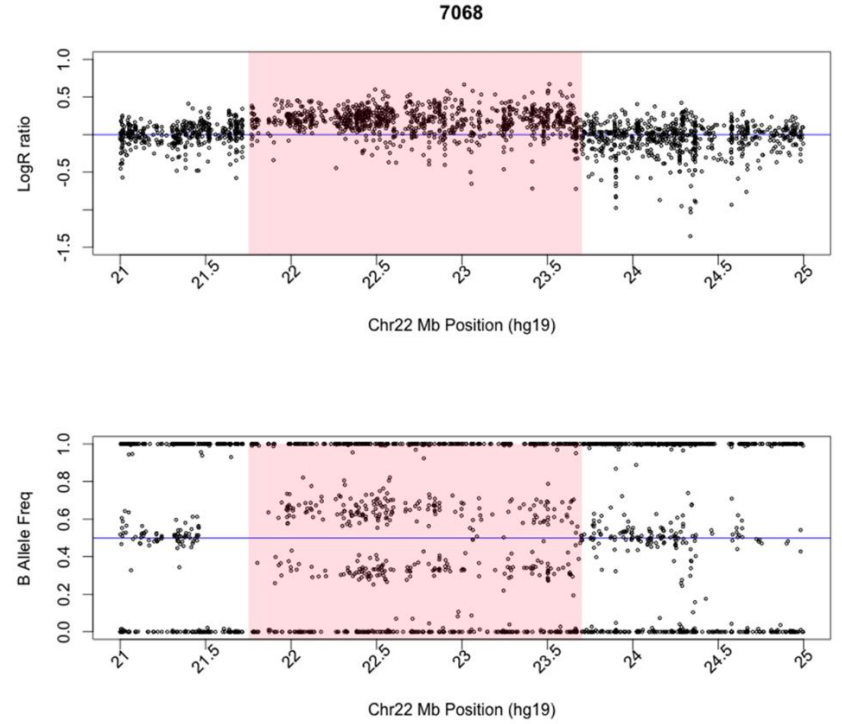
22q11.2 (proximal A-B) deletion



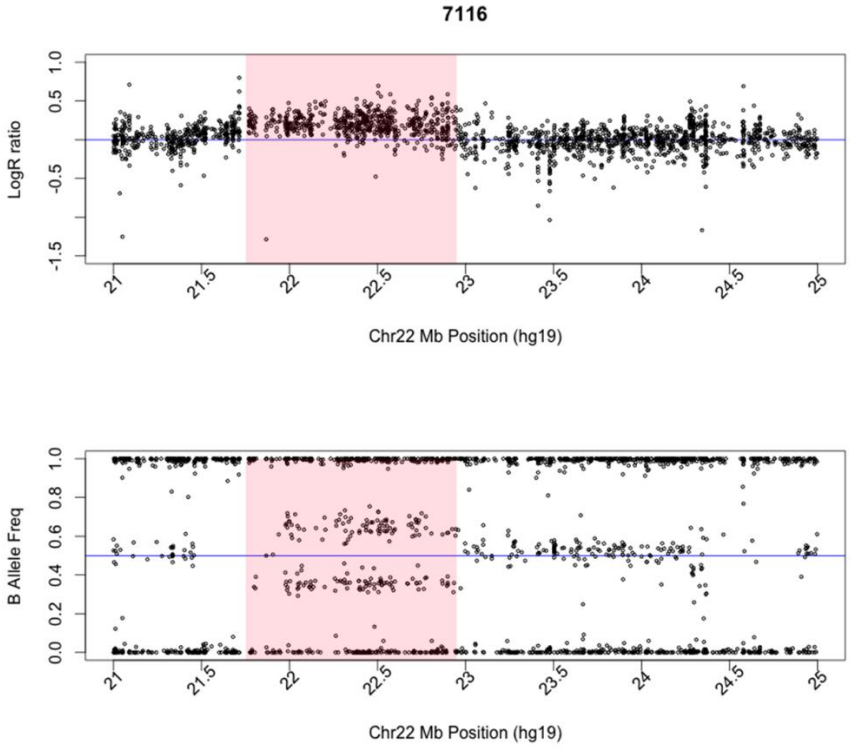
22q11.2 (proximal A-D) deletion



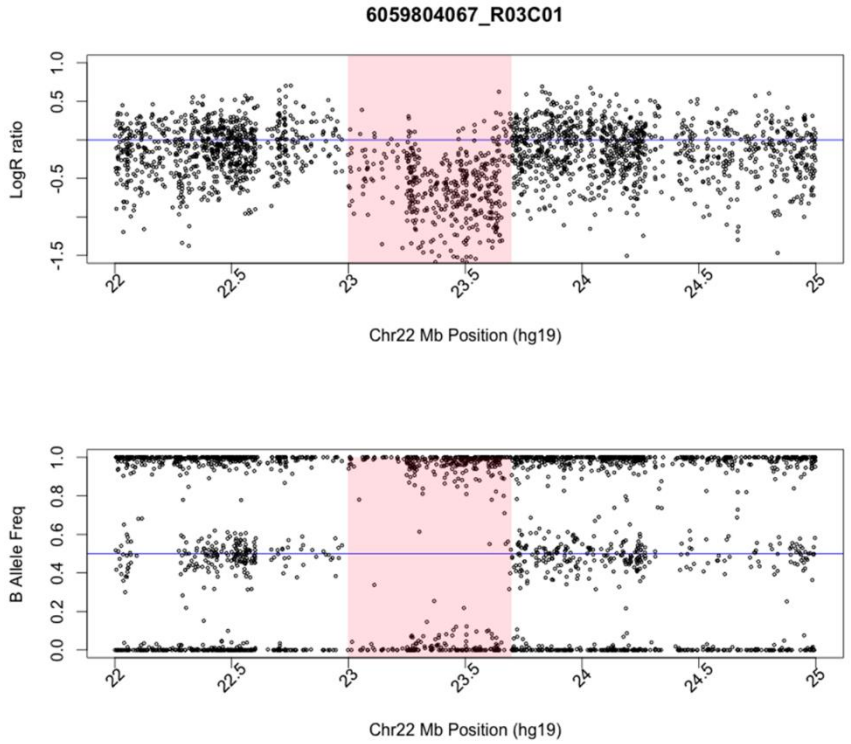
22q11.2 (distal type I D-F) duplication



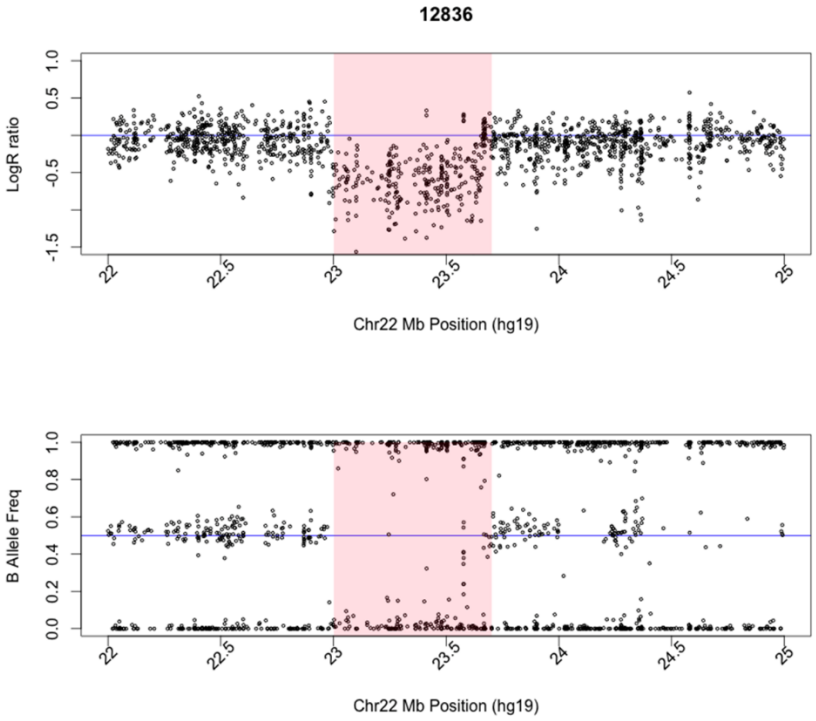
22q11.2 (distal type I D-E) duplication



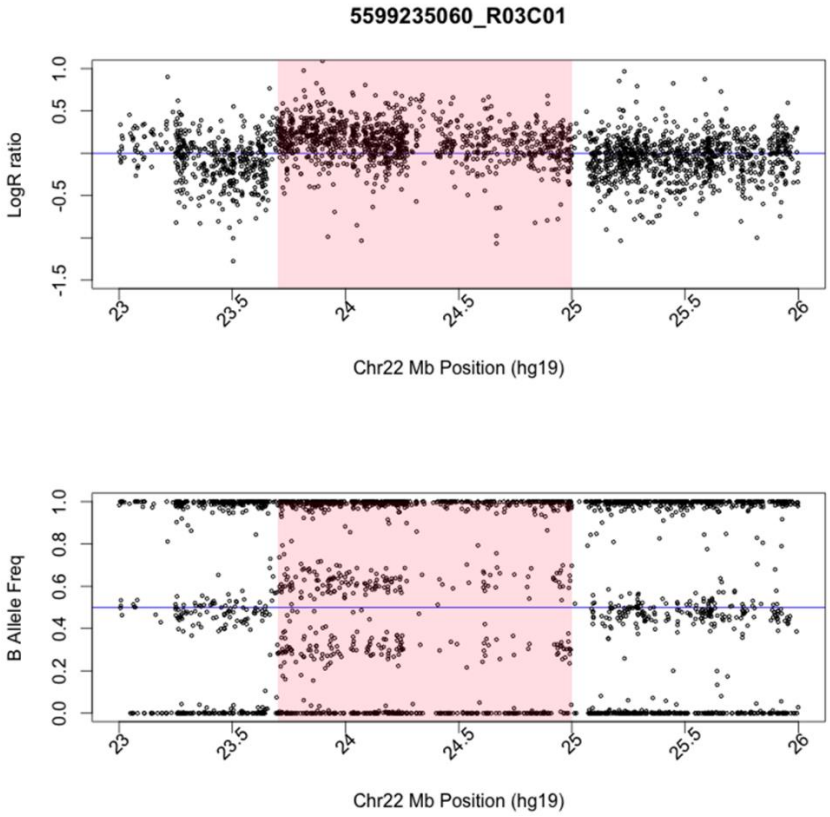
22q11.2 (distal type II E-F) deletions



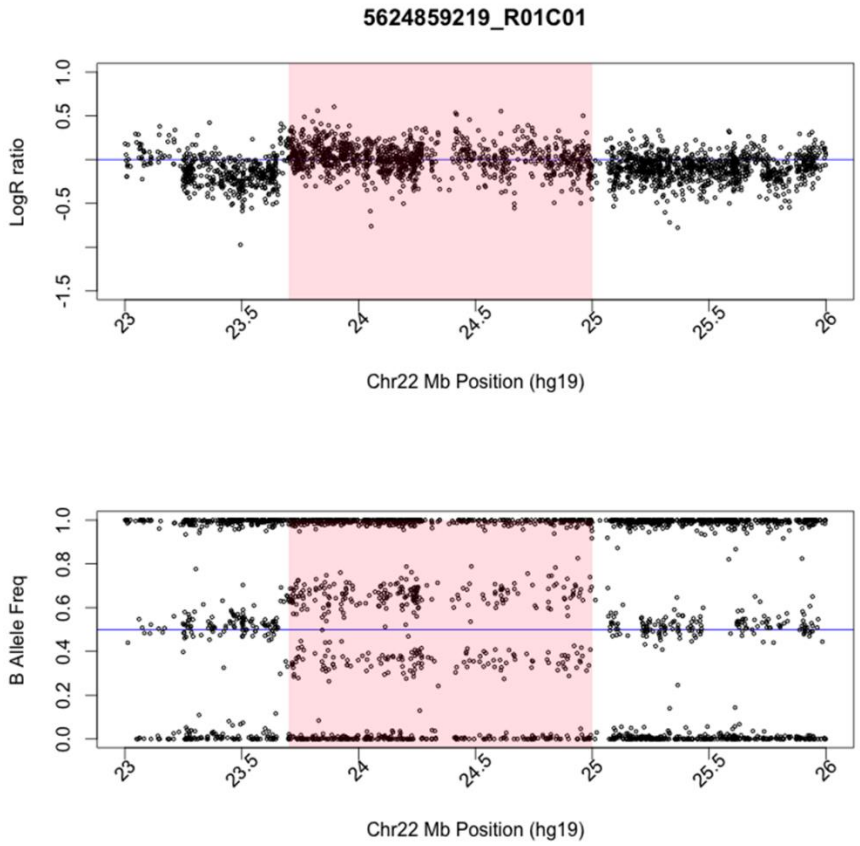
22q11.2 (distal type II E-F) deletions. *Cont.*



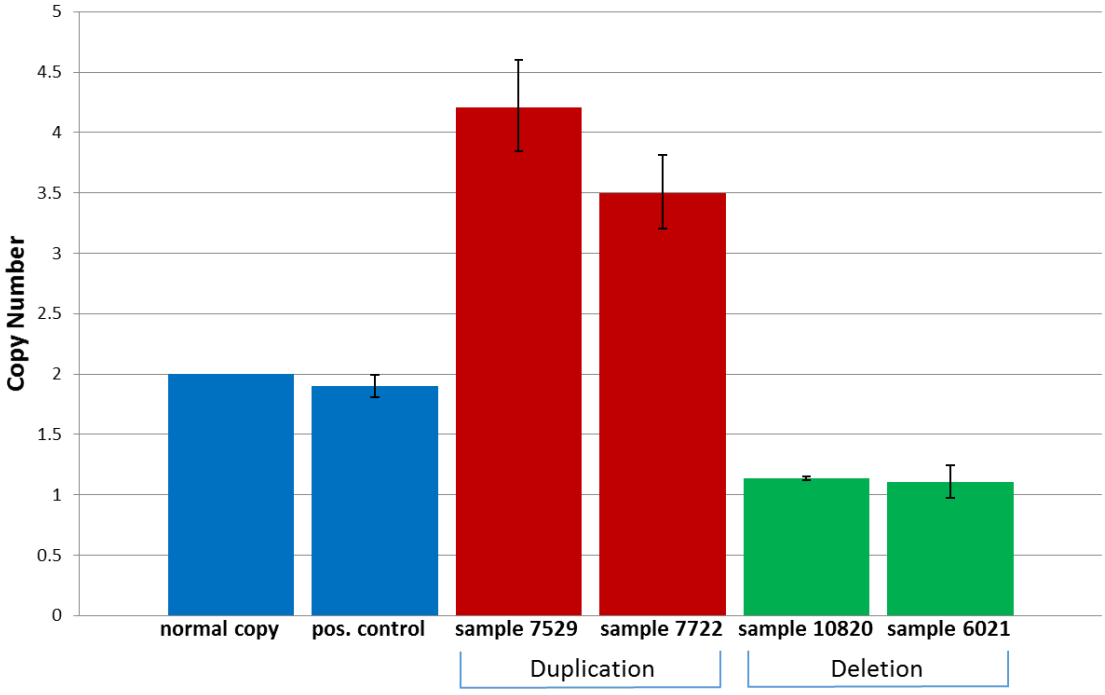
22q11.2 (distal type III F-H) duplications



22q11.2 (distal type III F-H) duplications. *Cont.*



Supplementary Figure 4-3. Validation of copy number at 1q21.1 by qPCR for two duplications and two deletions



References

1. Feuk L, Carson AR, Scherer SW: Structural variation in the human genome. Nat Rev Genet 2006 Feb;7:85–97.

2. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, et al.: Population analysis of large copy number variants and hotspots of human genetic disease. Am J Hum Genet 2009 Feb;84:148–161.

3. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al.: A copy number variation morbidity map of developmental delay. Nat Genet 2011 Sep;43:838–846.

4. Malhotra D, Sebat J: CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell 2012 Mar 16;148:1223–1241.

5. Gill M: Developmental psychopathology: The role of structural variation in the genome. Dev Psychopathol 2012 Nov;24:1319–1334.

6. Kotlar AV, Mercer KB, Zwick ME, Mulle JG: New discoveries in schizophrenia genetics reveal neurobiological pathways: A review of recent findings. Eur J Med Genet 2015 Dec;58:704–714.

7. Papolos DF, Faedda GL, Veit S, Goldberg R, Morrow B, Kucherlapati R, et al.: Bipolar spectrum disorders in patients diagnosed with velo-cardio-facial syndrome: does a hemizygous deletion of chromosome 22q11 result in bipolar affective disorder? Am J Psychiatry 1996 Dec;153:1541–1547.

8. Männik K, Mägi R, Macé A, Cole B, Guyatt AL, Shihab HA, et al.: Copy number variations and cognitive phenotypes in unselected populations. JAMA 2015 May 26;313:2044–2054.

9. Andero R, Brothers SP, Jovanovic T, Chen YT, Salah-Uddin H, Cameron M, et al.: Amygdala-dependent fear is regulated by Oprl1 in mice and humans with PTSD. Sci Transl Med 2013 Jun 5;5:188ra73.

10. Gillespie CF, Almli LM, Smith AK, Bradley B, Kerley K, Crain DF, et al.: Sex dependent influence of a functional polymorphism in steroid 5-α-reductase type 2 (SRD5A2) on post-traumatic stress symptoms. Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet 2013 Apr;162B:283–292.

11. Almli LM, Mercer KB, Kerley K, Feng H, Bradley B, Conneely KN, et al.: ADCYAP1R1 genotype associates with post-traumatic stress symptoms in highly traumatized African-American females. Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet 2013 Apr;162B:262–272.

12. Ressler KJ, Mercer KB, Bradley B, Jovanovic T, Mahan A, Kerley K, et al.: Post-traumatic stress disorder is associated with PACAP and the PAC1 receptor. Nature 2011 Feb 24;470:492–497.

13. Binder EB, Bradley RG, Liu W, Epstein MP, Deveau TC, Mercer KB, et al.: Association of FKBP5 polymorphisms and childhood abuse with risk of posttraumatic stress disorder symptoms in adults. JAMA J Am Med Assoc 2008 Mar 19;299:1291–1305.

14. Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, et al.: An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med Off J Am Coll Med Genet 2011 Sep;13:777–784.

15. Moreno-De-Luca D, Sanders SJ, Willsey AJ, Mulle JG, Lowe JK, Geschwind DH, et al.: Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. Mol Psychiatry 2013 Oct;18:1090–1095.

16. Riley KN, Catalano LM, Bernat JA, Adams SD, Martin DM, Lalani SR, et al.: Recurrent deletions and duplications of chromosome 2q11.2 and 2q13 are associated with variable outcomes. Am J Med Genet A 2015 Nov;167:2664–2673.

17. Torres F, Barbosa M, Maciel P: Recurrent copy number variations as risk factors for neurodevelopmental disorders: critical overview and analysis of clinical implications. J Med Genet 2015 Oct 26; DOI: 10.1136/jmedgenet-2015-103366

18. Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S, et al.: CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature 2014 Jan 16;505:361–366.

19. van Bon BWM, Balciuniene J, Fruhman G, Nagamani SCS, Broome DL, Cameron E, et al.: The phenotype of recurrent 10q22q23 deletions and duplications. Eur J Hum Genet EJHG 2011 Apr;19:400–408.

20. Gillespie CF, Bradley B, Mercer K, Smith AK, Conneely K, Gapen M, et al.: Trauma exposure and stress-related disorders in inner city primary care patients. Gen Hosp Psychiatry 2009 Dec;31:505–514.

21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al.: PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007 Sep;81:559–575.

22. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006 Aug;38:904–909.

23. Almli LM, Duncan R, Feng H, Ghosh D, Binder EB, Bradley B, et al.: Correcting Systematic Inflation in Genetic Association Tests That Consider Interaction Effects:

Application to a Genome-wide Association Study of Posttraumatic Stress Disorder. JAMA Psychiatry 2014 Oct 29; DOI: 10.1001/jamapsychiatry.2014.1339

24. Lin C-F, Naj AC, Wang L-S: Analyzing Copy Number Variation Using SNP Array Data: Protocols for Calling CNV and Association Tests [Internet]; in : Current Protocols in Human Genetics. John Wiley & Sons, Inc., 2001, [cited 2014 Feb 24].Available from: http://onlinelibrary.wiley.com/doi/10.1002/0471142905.hg0127s79/abstract

25. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al.: PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 2007 Nov 1;17:1665–1674.

26. Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K, et al.: High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome Res 2009 Sep;19:1682–1690.

27. Patterson N, Price AL, Reich D: Population structure and eigenanalysis. PLoS Genet 2006 Dec;2:e190.

28. Butcher NJ, Chow EWC, Costain G, Karas D, Ho A, Bassett AS: Functional outcomes of adults with 22q11.2 deletion syndrome. Genet Med Off J Am Coll Med Genet 2012 Oct;14:836–843.

29. Malhotra D, McCarthy S, Michaelson JJ, Vacic V, Burdick KE, Yoon S, et al.: High Frequencies of De Novo CNVs in Bipolar Disorder and Schizophrenia. Neuron 2011 Dec 22;72:951–963.

30. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al.: Strong association of de novo copy number mutations with autism. Science 2007 Apr 20;316:445–449.

31. Green LW: Manual for scoring socioeconomic status for research on health behavior. Public Health Rep 1970 Sep;85:815–827.

32. Emerson E: Poverty and people with intellectual disabilities. Ment Retard Dev Disabil Res Rev 2007;13:107–113.

33. Wadsworth ME, Achenbach TM: Explaining the link between low socioeconomic status and psychopathology: testing two mechanisms of the social causation hypothesis. J Consult Clin Psychol 2005 Dec;73:1146–1153.

34. Kendler KS, Eaves LJ: Models for the joint effect of genotype and environment on liability to psychiatric illness. Am J Psychiatry 1986 Mar;143:279–289.

35. Dohrenwend BP, Levav I, Shrout PE, Schwartz S, Naveh G, Link BG, et al.: Socioeconomic status and psychiatric disorders: the causation-selection issue. Science 1992 Feb 21;255:946–952.

36. Rosenberg NA, Kang JTL: Genetic Diversity and Societally Important Disparities. Genetics 2015 Sep 1;201:1–12.

37. Emerson E, Hatton C: Poverty, socio-economic position, social capital and the health of children and adolescents with intellectual disabilities in Britain: a replication. J Intellect Disabil Res JIDR 2007 Nov;51:866–874.

38. SARACENO B, LEVAV I, KOHN R: The public mental health significance of research on socio-economic factors in schizophrenia and major depression. World Psychiatry 2005 Oct;4:181–185.

39. Haddad L, Schäfer A, Streit F, Lederbogen F, Grimm O, Wüst S, et al.: Brain structure correlates of urban upbringing, an environmental risk factor for schizophrenia. Schizophr Bull 2015 Jan;41:115–122.

40. Committee to Evaluate the Supplemental Security Income Disability Program for Children with Mental Disorders, Board on the Health of Select Populations, Board on Children, Youth, and Families, Institute of Medicine, Division of Behavioral and Social Sciences and Education, The National Academies of Sciences, Engineering, and Medicine: Mental Disorders and Disabilities Among Low-Income Children [Internet]. Washington (DC), National Academies Press (US), 2015, [cited 2015 Dec 28].Available from: http://www.ncbi.nlm.nih.gov/books/NBK332882/

41. Lederbogen F, Haddad L, Meyer-Lindenberg A: Urban social stress--risk factor for mental disorders. The case of schizophrenia. Environ Pollut Barking Essex 1987 2013 Dec;183:2–6.

42. Manning M, Hudgins L, Professional Practice and Guidelines Committee: Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. Genet Med Off J Am Coll Med Genet 2010 Nov;12:742–745.

43. Shaffer LG, American College of Medical Genetics Professional Practice and Guidelines Committee: American College of Medical Genetics guideline on the cytogenetic evaluation of the individual with developmental delay or mental retardation. Genet Med Off J Am Coll Med Genet 2005 Dec;7:650–654.

44. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al.: Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet 2010 May 14;86:749–764.

45. http://www.amchp.org/Policy-Advocacy/MCHAdvocacy/Pages/StateProfiles.aspx

46. Lynch ME, Kable JA, Coles CD: Prenatal alcohol exposure, adaptive function, and entry into adult roles in a prospective study of young adults. Neurotoxicol Teratol 2015 Oct;51:52–60.

47. Streissguth AP, Barr HM, Kogan J, Bookstein FL: Understanding the occurrence of secondary disabilities in clients with fetal alcohol syndrome (FAS) and fetal alcohol effects (FAE) 1996;

48. Deak KL, Horn SR, Rehder CW: The evolving picture of microdeletion/microduplication syndromes in the age of microarray analysis: variable expressivity and genomic complexity. Clin Lab Med 2011 Dec;31:543–564, viii.

49. McDonald-McGinn DM, Minugh-Purvis N, Kirschner RE, Jawad A, Tonnesen MK, Catanzaro JR, et al.: The 22q11.2 deletion in African-American patients: an underdiagnosed population? Am J Med Genet A 2005 Apr 30;134:242–246.

50. Veerapandiyan A, Abdul-Rahman OA, Adam MP, Lyons MJ, Manning M, Coleman K, et al.: Chromosome 22q11.2 deletion syndrome in African-American patients: a diagnostic challenge. Am J Med Genet A 2011 Sep;155A:2186–2195.

51. Hudgins L, Geer JS, Cassidy SB: Phenotypic differerencss in African Americans with Prader-Willi syndrome. Genet Med Off J Am Coll Med Genet 1998 Dec;1:49–51.

52. Bates BR, Lynch JA, Bevan JL, Condit CM: Warranted concerns, warranted outlooks: a focus group study of public understandings of genetic research. Soc Sci Med 1982 2005 Jan;60:331–344.

53. Lanie AD, Jayaratne TE, Sheldon JP, Kardia SLR, Anderson ES, Feldbaum M, et al.: Exploring the public understanding of basic genetic concepts. J Genet Couns 2004 Aug;13:305–320.

54. Shostak S, Freese J, Link BG, Phelan JC: The Politics of the Gene: Social Status and Beliefs about Genetics for Individual Outcomes. Soc Psychol Q 2009 Mar;72:77–93.

Chapter 5: Summary of Findings and Future Directions

Summary

The key to preventing disease relies greatly on our knowledge of cause but without elucidating mechanism our understanding of disease remains incomplete. Unfortunately, for most disease there isn't a single cause or a single mechanism that can be attributed to its development. This is particularly the case for complex diseases, which are often the result of multiple genetic and environmental factors, which each contribute to some fraction of disease risk. Fortunately, as scientists, we don't allow this complexity to thwart our efforts to better understand why and how disease develops. Every discovery of risk lends a piece to the puzzle and is a significant contributor to the picture as a whole. Until a viable method of prevention or cure becomes available, research must carry on to identify genetic risk variants and, most importantly, increase efforts to characterize the mechanistic outcomes of the genetic variation that associates with disease.

In this dissertation I have provided evidence that the rs2267735 risk allele disrupts estradiol-mediated activation of *ADCYAP1R1* gene expression, which leads to reduced levels of *ADCYAP1R1* and increased symptoms of posttraumatic stress disorder (PTSD). The results of this work build upon the original finding by Ressler et al. [1] of an association between PTSD and the *ADCYAP1R1* SNP (rs2267735), providing key insight into the molecular mechanisms responsible for this association. As hypothesized, the *ADCYAP1R1* SNP is located within a functional estrogen response element (ERE) that binds with estradiol activated estrogen receptor alpha (ERα) to influence expression of the PAC1 receptor gene, *ADCYAP1R1*. These findings have led to the development of a

model in which we predict that the SNP risk allele interferes with normal, stress-induced activation of *ADCYAP1R1* [2,3] and results in a dysregulated stress response that leads to PTSD. This work is the first to identify an inverse relationship between expression of the PAC1 receptor gene and PTSD symptoms. However, this finding is consistent with research involving PAC1 (-/-) knockout mice that have revealed PTSD like phenotypes, such as fear memory dysfunction, in the null animals [4,5].

An investigation of *ADCYAP1R1* expression in postmortem brain tissue of PTSD patients will be necessary to confirm the relationship between transcript levels and disease outcome, particularly in the context of genotype at rs2267735. One caveat that must be considered in order to obtain useful experimental data is that *ADCYAP1R1* is not highly expressed in all regions of the brain (unpublished work). Thus, sampling of whole brain for the purpose of measuring *ADCYAP1R1* gene expression would necessitate the careful extraction of specific brain regions such as the BNST where we have observed differential expression in our experiments with mice. Further research is vital to elucidate the role estradiol plays in PTSD susceptibility and explore its potential as a target for pharmacological intervention. To begin this line of inquiry, a cohort of female emergency room (ER) patients who had just experienced a severe trauma could be recruited and followed prospectively. Blood drawn in the ER immediately after the traumatic event could be used to measure levels of estrogen. PTSD symptoms could then be assessed at various time points to determine if estrogen levels at time of trauma predict the development and or persistence of PTSD.

In addition to identify mechanisms it is also important to continue exploring other PTSD associated genetic variation, particularly given the low effect sizes of currently

identified SNPs that don't fully explain the heritability estimates of PTSD. Copy number variants (CNVs), which are deletions or duplications of the genome, have only recently been investigated for a role in PTSD risk. Given their size and potential impact on gene function, we speculated that PTSD associated CNVs might confer an increased odds of disease risk several fold greater than that for SNPs; a phenomenon that has been well documented for genetic risk of schizophrenia [6]. Leveraging currently available SNP microarray data from genome wide association studies (GWAS), we identified CNVs among study participants of two PTSD cohorts: the Grady Trauma Project (N=2,396) and the Army STARRS Project (N=7,190). Deviations from the expected allele frequencies and signal intensities at SNP loci were used by CNV calling algorithms to detect CNVs across the genome. After taking measures to remove poor quality samples and reduce false positive CNVs calls, we ran statistical tests to determine if CNVs associated with risk for PTSD. Burden analyses were run to detect an increased prevalence of varying sized deletions and duplications in PTSD cases versus controls. We also performed association tests to identify specific CNVs that increase genetic susceptibility to PTSD.

For African American only GTP study participants, there was no statistically significant burden in the number of deletions or duplications, of any size (>50kb) among individuals with PTSD. We did, however, identify a statistically significant association between PTSD diagnosis and a 750kb duplication, which contains two genes within the putative breakpoints of the CNV gain. Of the two genes only one was expressed at detectable levels in whole brain mRNA. This gene, *CNTN5*, has been implicated in neuropsychiatric and brain-related disorders including schizophrenia, autism, ADHD, anorexia nervosa, and Alzheimer's [7–12]. Additionally, the *CNTN5* protein, NB-2

localizes to the amygdala in humans, a region of the brain critical for fear learning [13], and is involved in synaptic plasticity, which may be important in the regulation of fear memory consolidation [14]. Disruption of processes involved in fear response can lead to PTSD [15]. Thus, aberrant expression of *CNTN5*, which may result from the duplication, could feasibly have an impact of the development of PTSD.

In the STARRS cohort, we performed the same analyses but limited our investigation to a study cohort of self-identified, non-Hispanic white study participants. Our burden analysis of CNVs in this study population did reveal a statistically significant increase in prevalence of deletions greater than 50kb among PTSD cases compared to unaffected controls. When analyzed for larger deletions (>500kb) the burden remained statistically significant. A more careful assessment of the genomic regions uncovered by deletions of these various sizes should be performed and compared between cases and controls to identify genes or gene networks that may be differentially influenced by CNV burden. In this study cohort, we also identified a CNV that associated with PTSD. This CNV is a deletion, which is variably sized among the PTSD cases. For the regions deleted among carriers of this CNV, there is 20kb of shared overlap that falls within the gene, *IMMP2L*. *IMMP2L* has been implicated in neurodevelopment phenotypes [16], ADHD [17], and Tourette syndrome [18]. It is also in a region of the genome that has been identified as an autism spectrum locus [19]. Similar to *CNTN5*, for which there are several published manuscripts linking the gene to brain-related disorders, *IMMP2L* is also involved in biologically relevant mechanisms that support a hypothesis for a role of *IMMP2L* in risk for PTSD.

Other analyses need to be performed in order to confirm the preliminary results of the CNV by PTSD analyses that have been summarized. As discussed in the introduction of this dissertation, several single nucleotide polymorphisms (SNPs) have been identified, which confer some level of risk to the development of PTSD among individuals exposed to trauma. For the majority of these SNPs however, the association with disease outcome is largely influenced by environment or may have only been identified among specific cohorts. Gene by environmental interaction and population stratification within sample datasets has largely contributed to the irreproducibility of associations in PTSD genetic research. This issue applies equally to PTSD association analyses with CNVs. The analyses we performed to identify CNVs that increase risk for PTSD were done in PLINK which did not allow us to account for environmental effects on association. CNV analysis will need to be performed in R (statistical software) in order to include covariates in future analyses [20,21]. For the GTP data we used principal component analysis to identify a cohort of ancestrally similar participants and only these individuals were included in our analyses. We did not control for population stratification in our analyses using the STARRS data but, rather, chose samples based only on self-identified race. These analyses should be redone using principal components analysis to account for ancestry.

With the help and direction of research scientists and bioinformaticists that have been involved in the successful investigation of CNVs that associate with schizophrenia, a workgroup has been established, under the leadership of Dr. Caroline Nievergelt, to analyze CNV data across multiple PTSD cohorts. As we have observed for GWAS, increased sample numbers are key to the success of revealing statistically significant

disease associations. These future analyses will hopefully result in the replication of some of the associations or CNV burden presented here. Also, with larger numbers and more careful analyses, there is an increased likelihood of identifying additional CNVs that confer risk for PTSD, revealing more secrets to the mechanisms of PTSD development.

During the investigation of CNVs in the genomes of African American GTP study participants for the purpose of CNV by PTSD association analysis, we noticed a significant number of large, rare genomic disorder CNVs. To quantitate the prevalence of these events among our study population we applied the same methodology used to make the GTP CNV calls to data obtained from dbGAP. We specifically chose dbGAP datasets with a large number of self-identified African American participants and raw intensity data derived from Illumina whole-genome SNP microarrays data with equal or greater probe coverage to that of the Illumina array used for GTP CNV calling. After quality control measures were employed to remove samples with poor quality CNV calls, related individuals, and individuals with dissimilar ancestry, we were able to compare the number of genomic disorder CNVs between our GTP cohort (N=4100) and a comparison cohort of similar size and ancestry comparison cohort (N=3883). We assessed the frequency of CNVs for 24 chromosome regions, which harbor a duplication and/or deletion associated with outcomes such as developmental delay, intellectual disability, congenital anomalies, epilepsy, autism and schizophrenia. For these particular CNVs, we identified 72 in our GTP cohort (1.8%) and only 30 in the comparison cohort (0.8%). Using logistic regression and controlling for race in our analysis, we tested this finding for statistical significance. Our results indicated that the odds of having a genomic disorder CNV are 2.2 times greater among GTP study participants than non-GTP study

participants from the comparison cohort, a relationship that is statistically significant and not driven by relatedness or racial differences between the groups. Because the GTP study participants were recruited from an impoverished population, we hypothesize that the increase in prevalence of pathogenic CNVs is due to a phenomena known as downward drift, where individuals with intellectual disabilities or neuropsychiatric disorders are at a higher risk for living in a community of low socioeconomic status (SES). Regardless of the reason for the burden of these CNVs in the medically underserved GTP population, it is likely that standard of care genetic testing is not readily available to this community, adding to the health disparities they already face. We propose that special attention be given to communities of similar, low SES where burden of CNVs may be higher and genetic testing may provide significant benefit.

To better elucidate the reason for the increased prevalence of CNVs in the GTP cohort, it would be important to determine is other low SES populations have similar burden. This could be readily tested if a GWAS dataset with income, education and employment demographics was available for general research use through dbGAP. To make a conclusion regarding association between SES and the prevalence of genomic disorder CNVs, at least two or more datasets would need to be analyzed. Provided such datasets exist, genotypes could be used to infer ancestry and relatedness, and probe intensity values would be used along with PennCNV to make CNV calls (as described elsewhere).

To identify if there is a true discrepancy in referral for genetic testing within low SES and medically underserved communities, I recommend developing a survey instrument. This survey could be administered at both Grady Memorial Hospital and

Emory University Hospital(s) to obtain assessments from health care settings that serve individuals with a broad range of household incomes, educational obtainment and employment status. As part of this instrument, questions should assess the following: 1) knowledge of genetic testing, 2) whether or not they or someone they know has ever been referred for genetic testing, 3) interest in receiving genetic testing for themselves or their children- in that event that is was recommended, and 4) if no interest in getting genetic testing, what are some of the reasons.

Conclusion

In conclusion, the work presented in this dissertation has contributed to our knowledge of genetic risk for PTSD. I have contributed data that clarifies the molecular mechanism by which a PTSD associated SNP increases PTSD susceptibility. I have also provided preliminary data from which further investigation into CNV by PTSD associations can be launched. Lastly, incidental genetic findings resulting from my research into the genetics of PTSD have uncovered a potential relationship between socioeconomic status and burden of pathogenic CNVs. Hopefully, this finding will be researched further and greater attention will be given to individuals within similar, medically underserved populations to improve access to genetic testing.

References

1. Ressler KJ, Mercer KB, Bradley B, Jovanovic T, Mahan A, Kerley K, et al.: Post-traumatic stress disorder is associated with PACAP and the PAC1 receptor. Nature 2011 Feb 24;470:492–497.

2. Hammack SE, Cheung J, Rhodes KM, Schutz KC, Falls WA, Braas KM, et al.: Chronic stress increases pituitary adenylate cyclase-activating peptide (PACAP) and brain-derived neurotrophic factor (BDNF) mRNA expression in the bed nucleus of the stria terminalis (BNST): roles for PACAP in anxiety-like behavior. Psychoneuroendocrinology 2009 Jul;34:833–843.

3. Roman CW, Lezak KR, Hartsock MJ, Falls WA, Braas KM, Howard AB, et al.: PAC1 receptor antagonism in the bed nucleus of the stria terminalis (BNST) attenuates the endocrine and behavioral consequences of chronic stress. Psychoneuroendocrinology 2014 Sep;47:151–165.

4. Otto C, Kovalchuk Y, Wolfer DP, Gass P, Martin M, Zuschratter W, et al.: Impairment of mossy fiber long-term potentiation and associative learning in pituitary adenylate cyclase activating polypeptide type I receptor-deficient mice. J Neurosci Off J Soc Neurosci 2001 Aug 1;21:5520–5527.

5. Takuma K, Maeda Y, Ago Y, Ishihama T, Takemoto K, Nakagawa A, et al.: An enriched environment ameliorates memory impairments in PACAP-deficient mice. Behav Brain Res 2014 Oct 1;272:269–278.

6. Kotlar AV, Mercer KB, Zwick ME, Mulle JG: New discoveries in schizophrenia genetics reveal neurobiological pathways: A review of recent findings. Eur J Med Genet 2015 Dec;58:704–714.

7. Biffi A, Anderson CD, Desikan RS, et al: GEnetic variation and neuroimaging measures in alzheimer disease. Arch Neurol 2010 Jun 1;67:677–685.

8. Daalen E van, Kemner C, Verbeek NE, Zwaag B van der, Dijkhuizen T, Rump P, et al.: Social responsiveness scale-aided analysis of the clinical impact of copy number variations in autism. neurogenetics 2011 Nov 1;12:315–323.

9. Lionel AC, Crosbie J, Barbosa N, Goodale T, Thiruvahindrapuram B, Rickaby J, et al.: Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. Sci Transl Med 2011 Aug 10;3:95ra75.

10. Maziade M, Raymond V, Cliche D, Fournier JP, Caron C, Garneau Y, et al.: Linkage results on 11Q21-22 in Eastern Quebec pedigrees densely affected by schizophrenia. Am J Med Genet 1995 Dec 18;60:522–528.

11. Nakabayashi K, Komaki G, Tajima A, Ando T, Ishikawa M, Nomoto J, et al.: Identification of novel candidate loci for anorexia nervosa at 1q41 and 11q22 in Japanese by a genome-wide association analysis with microsatellite markers. J Hum Genet 2009 Sep;54:531–537.

12. St Clair D, Blackwood D, Muir W, Carothers A, Walker M, Spowart G, et al.: Association within a family of a balanced autosomal translocation with major mental illness. Lancet 1990 Jul 7;336:13–16.

13. Kamei Y, Takeda Y, Teramoto K, Tsutsumi O, Taketani Y, Watanabe K: Human NB-2 of the contactin subgroup molecules: chromosomal localization of the gene (CNTN5) and distinct expression pattern from other subgroup members. Genomics 2000 Oct 1;69:113–119.

14. Shimoda Y, Koseki F, Itoh M, Toyoshima M, Watanabe K: A cis-complex of NB-2/contactin-5 with amyloid precursor-like protein 1 is localized on the presynaptic membrane. Neurosci Lett 2012 Feb 29;510:148–153.

15. Parsons RG, Ressler KJ: Implications of memory modulation for post-traumatic stress and fear disorders. Nat Neurosci 2013 Feb;16:146–153.

16. Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH, et al.: Relative burden of large CNVs on a range of neurodevelopmental phenotypes. PLoS Genet 2011 Nov;7:e1002334.

17. Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT, et al.: Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. Mol Psychiatry 2010 Jun;15:637–646.

18. Petek E, Windpassinger C, Vincent JB, Cheung J, Boright AP, Scherer SW, et al.: Disruption of a novel gene (IMMP2L) by a breakpoint in 7q31 associated with Tourette syndrome. Am J Hum Genet 2001 Apr;68:848–858.

19. Maestrini E, Pagnamenta AT, Lamb JA, Bacchelli E, Sykes NH, Sousa I, et al.: High-density SNP association study and copy number variation analysis of the AUTS1 and AUTS5 loci implicate the IMMP2L-DOCK4 gene region in autism susceptibility. Mol Psychiatry 2010 Sep;15:954–968.

20. Subirana I, Diaz-Uriarte R, Lucas G, Gonzalez JR: CNVassoc: Association analysis of CNV data using R. BMC Med Genomics 2011 May 24;4:47.

21. Lin C-F, Naj AC, Wang L-S: Analyzing Copy Number Variation Using SNP Array Data: Protocols for Calling CNV and Association Tests [Internet]; in : Current Protocols in Human Genetics. John Wiley & Sons, Inc., 2001, [cited 2014 Feb 24].Available from: http://onlinelibrary.wiley.com/doi/10.1002/0471142905.hg0127s79/abstract