ACOSTA, GLEN HOWEL G., M.S. Single-Nucleotide Variations and Linkage Disequilibrium Patterns in Three Candidate Genes for Attention Deficit Hyperactivity Disorder (AD/HD). (2009)
Directed by Dr. David L. Remington. 81pp.

Allelic variations in the genes involving the dopaminergic system, particularly the dopamine transporter (*DAT1/SLC6A3*), dopamine receptor 4 (*DRD4*) and Catecol-O-Methyltransferase (*COMT*) genes have been associated with Attention Deficit Hyperactivity Disorder (AD/HD). However, the results of these studies have been variable and inconclusive in part due to the inconsistencies of experimental and statistical methodologies, phenotypic heterogeneity, low penetrance of the genes implicated, and population stratification. Genetic association studies based on linkage disequilibrium (LD) offer a promising approach to the study of common complex diseases. This study characterized LD patterns in three human populations (CEU, YRI, CHB+JPT) in the three genes mentioned above and identified factors affecting the inconsistencies of genetic association studies in AD/HD. We used the HapMap database and the Haploview program to evaluate linkage disequilibrium patterns of SNPs in these genes. The regression results suggest that there is a trend toward poorer capturing of rare SNPs, which would mean lesser detection of AD/HD associations if rare SNPs were causative on these genes. However, sparse sampling of tag SNPs in these genes (*COMT* and *DAT1*/*SLC6A3*) does not capture the other SNPs well and that a denser tag SNP set is needed to further test our results. The significant reductions of the $r^2$ in the other two populations relative to the CEU supports the contention that associations on these genes

varies between populations and suggests that prior assessment of tag SNPs using the HapMap is an essential step in the design of genetic association studies.

# SINGLE-NUCLEOTIDE VARIATIONS AND LINKAGE DISEQUILIBRIUM PATTERNS IN THREE CANDIDATE GENES FOR ATTENTION DEFICIT HYPERACTIVITY DISORDER (AD/HD)

by

Glen Howel G. Acosta

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2009

Approved by

Dr. David L. Remington

_____
Committee Chair

APPROVAL PAGE

This thesis has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____
Dr. David L. Remington

Committee Members _____
Dr. Dennis R. LaJeunesse

_____
Dr. Malcolm D. Schug

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

# CHAPTER I

## INTRODUCTION

Attention deficit hyperactivity disorder (AD/HD; American Psychiatric Association, 1994) is characterized by chronic, pervasive, and developmentally inappropriate patterns of inattention, impulsivity, and/or hyperactivity (Anastopoulos, 1999; Anastopoulus and Shelton, 2001). It is a neurobehavioral disorder with a childhood onset and may persist in adulthood. The prevalence of AD/HD may vary considerably depending on the diagnostic methods used and it is estimated to occur in 5-10% school age children worldwide (Shastry, 2004). It is also been found to be more frequent in boys than girls with a 3-4:1 ratio (Ogdie et al., 2003). The *Diagnostic and Statistical Manual of Mental Disorders IV* (DSM-IV) of the American Psychiatric Association (APA) is the currently used diagnostic scheme for AD/HD (Heiser et al., 2004). It distinguishes among diagnostic subtypes characterized by *predominantly inattentive* (avoids tasks, forgetful, fails to finish tasks and usually loses things), *predominantly hyperactive/impulsive* (fidgety, talks a lot, blurts out answers to question and often interrupts), and the *combined type*. Each subtype has nine observable behaviors and at least six of these have to be present in school or play activities for at least six months for diagnosis (Millichap, 1999). These observable behaviors interfere with the individual's peer and

family relationships as well as his/her performance in school and/or work. Although the etiology of AD/HD is still unknown, current studies support the concept that AD/HD is a neurobehavioral disorder with multiple causes (Cooper and Bilton, 1999; Faraone and Biederman, 1998). It has been strongly supported that AD/HD is highly heritable (Cooper and Bilton, 1999; Lopez, 1965; Millichap, 1999; Faraone et al., 2005). Genes involved in the neurotransmission such as neurotransmitters, its receptors and transporters are possible targets to further study the genetics AD/HD. (Biederman and Spencer, 1999). Although candidate gene studies and genomic screen analyses provide reputable evidence on the genetic influence of AD/HD (Acosta et al., 2003; Shastry, 2004; DiMaio et al., 2003; Fisher et al., 2002; Ogdie et al., 2004;) their results are variable and there is limited progress in identifying the specific genetic factors that may contribute in the presentation of the disorder.

The candidate gene studies suggest that multiple genes such as, *DAT1/SLC6A3, DRD4,* and *COMT* genes may be associated in AD/HD. However, the results of these studies searching for specific variants on these genes are contradictory (Acosta et al., 2003; DiMaio et al., 2003; Faraone et al., 2001) The inconsistencies on these studies may be due to small contribution of a single polymorphism in the inheritance of AD/HD and genetic association studies may be at risk with spurious association (Pritchard et al., 2000). The inability to replicate studies because of poor methodologies and issues in detecting association precludes a definitive conclusion on the genetics of AD/HD. Besides an accurate phenotype characterization, it is critical to have a well-developed

methodology (such as genetic marker selection, sampling and genotyping methods) in order to find true associations to better understand the genetics of complex disorders.

Association studies are one of the study designs that can define the genetic correlates of common complex disorders such as AD/HD. Although it is a straightforward method, it has been unsuccessful in identifying replicable associations for human complex diseases (Sullivan, 2007; Hirschhorn et al., 2002). The International HapMap project is aimed to catalog millions of SNPs in the human genome and currently have genotype data in four human populations. It is essential for any candidate gene association studies to evaluate SNPs of interest before genotyping. Based on the idea that specific polymorphisms in the DNA sequence may reflect relationships between genotype and phenotype, these DNA sequence variations (such as SNPs) can be employed to infer the genetic basis of complex traits and/or predict the evolution of these variations (Chu et al., 2009); and that these variations are considered common (Frazer et al., 2009; Doris, 2002; Chen et al., 2006; Collins et al., 1999; Risch and Merikangas, 1996; Lander, 1996). In the presence of linkage disequilibrium (LD), polymorphisms (such as SNPs or haplotypes) that are in physical proximity to a polymorphism that causes the phenotype can show a difference in frequency between the affected and unaffected individuals (Barnes, 2006). However, LD is generally affected by evolutionary factors such as population admixture and the age of the allele (Goldstein and Weale, 2001; Barnes, 2006). Previous genetic association studies on the *SLC6A3/DAT1* and *COMT* genes have used 4-5 SNP haplotype to detect association for AD/HD

phenotypes using large sample populations (Genro et al., 2008; Halleland et al., 2008). These studies have shown that modest SNPs could be used to detect associations with a thorough consideration of the SNP selection and population samples.

***Specific Aims and Assumptions***

We used SNPs and their linkage disequilibrium (LD) patterns to evaluate single nucleotide genetic variation among these three candidate genes in three human populations. The main purpose of this project is to evaluate the properties and predictive power of modest set of tag SNPs that are on the three candidate genes implicated in AD/HD, namely *DAT-1/SLC6A3*, *DRD-4*, and *COMT* genes on three populations genotyped on the HapMap database. (CEU – Utah residents with Northern and Western European ancestry from the CEPH (Centre d'Etude du Polymorphisme Humain) collection; YRI – Yoruba in Ibadan, Nigeria; and CHB+JPT – Han Chinese in Beijing, China + Japanese in Tokyo, Japan). This project had three specific aims:

1. Evaluate the practical usefulness of the HapMap database by obtaining genotyping data on the candidate genes and selecting up to three tag SNPs according to their minor allele frequency (MAF) using the CEU population.

2. Use a low-throughput genotyping method (PCR-RFLP) to evaluate the selected SNPs on each gene using a panel of DNA samples in order to confirm the alleles called on the HapMap, follow their Mendelian inheritance pattern

on each gene. Evaluate the limitations of this PCR-RFLP and compare it with a high-throughput method of genotyping (SNuPe).

3. Test whether sparse sampling of SNPs contributes to inconsistent detection of potentially causative genetic variants among human populations. Specifically, we hypothesized:

    a. The use of three or fewer SNPs per gene will not have consistently high LD ($r^2$) with other SNPs detected. This was tested by evaluating the highest $r^2$ among the set assayed tag SNPs with each of the other HapMap SNPs in each gene.

    b. The detection of associations will be best for causative SNPs that are common, per the common-disease common-variant (CDCV) hypothesis. This was tested by doing a regression of SNP-tag set $r^2$ values on MAF for each gene, with the expectation that the regression coefficient will be significantly positive.

    c. The tag SNPs selected in one population are less useful in other populations. This was tested by comparing $r^2$ of each identified SNP with the tag SNP set in the CEU with that in the other two populations. Significant reduction in the other two populations relative to the CEU would support our hypothesis.

If some SNPs show poor association with the assayed SNPs, if associations vary by population, or if potentially-causative rare alleles have poor associations with more common assayed SNP alleles, these factors could partially explain the inconsistent results of AD/HD association studies in the current literature. This project will provide information to the broader genetics community on the advantages and disadvantages of SNP genotyping methods and tools for evaluating genetic markers that may help better design association studies. These steps will further the study of genetic complex traits such as AD/HD and will advance our understanding of their pathogenesis.

# CHAPTER II

# REVIEW OF THE LITERATURE

## *AD/HD Phenotype and Comorbid Disorders*

Comorbidity refers to diseases or disorders that occur together, implying overlapping symptom patterns or, etiological commonalities, with implications for treatment, or increasing risk of negative outcomes (Lilienfeld et al, 1994). The presence of AD/HD in children increases their chances of having externalizing disorders such as oppositional defiant disorder (ODD) and conduct disorder (CD). They are also at risk for internalizing disorder such as depression and anxiety, and learning disabilities (Biederman et al, 1991). Figure 1 shows the prevalence rate of these behavioral problems (Biederman, 2005).

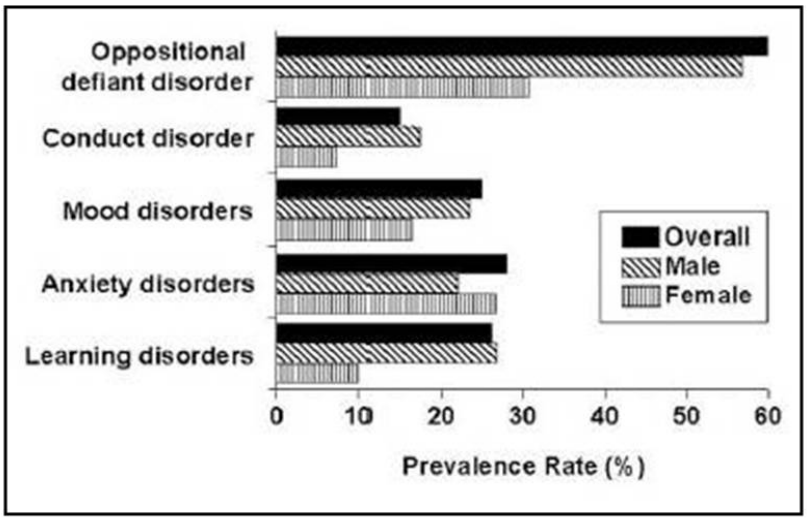


**Figure 1. AD/HD and Comorbid Disorders**. Approximate prevalence of comorbid conditions in children with AD/HD (from Biederman, 2005).

Children with AD/HD can display variability in terms of the presence or absence of comorbid conditions. AD/HD has been shown to be associated with an antisocial behavior trajectory that starts with oppositional defiant disorder (ODD) and is followed by conduct disorder (CD) (American Psychiatric Association, 1994). Jensen et al reviewed studies of AD/HD and its comorbid conditions. They found that the available data suggest that comorbidity between AD/HD and CD/ODD is higher (42%-93%) than between AD/HD and internalizing disorder (13%-50.8%) (1997). A higher incidence of psychiatric disorder such as, ODD, CD, antisocial personality and criminal behavior, has been observed among parents and other biological relatives of children with ODD or CD. Also, a review of twin studies found that there is higher concordance of CD/ODD among identical twins than fraternal twins (Altepeter and Korger, 1999). These studies suggest a genetic component that may be involved in the development of ODD/CD.

Oppositional Defiant Disorder (ODD) and Conduct disorder (CD) are often discussed together as antisocial behavior. Although ODD and CD share common features of inattentive, impulsive, and overactive behaviors, these two are distinct from each other. The primary features of ODD include patterns of negativistic, defiant, noncompliant and uncooperative behaviors. In contrast, CD involves patterns of behavior in which the basic rights of others are violated. According to the American Psychiatric Association (APA, 1994), CD is distinctive due to the recurrent violation of the rights of others and or the societal norms and rules (Altepeter and Korger, 1999).

Oppositional defiant disorder (ODD) is diagnosed in 5-25% of school age children (Altepeter and Korger, 1999). The DSM-IV of the APA (1994) categorizes ODD as a pattern of oppositional, negative, and hostile behavior that is observed at least six months or longer. These behaviors can affect the individual's social, occupational, and academic functioning. In addition, four or more behaviors such as losing one's temper, arguing with adults, deliberately annoying others, blaming others for one's misdeeds and being angry and resentful, are present during this period of time (Altepeter and Korger , 1999).

Angold et al reviewed literature from community studies that used the DSM-III, DSM-III-R or DSM IV as a diagnostic guide and have reported rates of comorbidity between AD/HD and CD/ODD. They have calculated the odds ratio (OR) and the confidence interval (CI). Furthermore, they assess the variability between study effects by performing Chi-square tests for each OR. They found that AD/HD-CD/ODD comparison to be homogeneous. The result of their meta-analysis suggests that the presence of AD/HD in children increases the odds of having ODD/CD by 7.7 – 14.8 fold (Mean 10.7) with a 95% confidence interval (1999). A two-year longitudinal study also supports the hypothesis that the hyperactive-impulsive behaviors seen in AD/HD influence the development of ODD (Burns and Walsh, 2002). Although current literature suggests biological causes of ODD, they have not clarified whether genetic factors predispose some children with AD/HD to develop ODD. The

phenotypic similarities of these comorbid disorders, particularly ODD with AD/HD, may suggest that the same genes may also be implicated.

### *AD/HD and the Brain*

Cognitive research studies have observed the presence of a dysfunctional response system in the frontal lobe of the brain (where the neuropsychological mechanism is located), which causes the impulsive behavior seen in AD/HD (Cooper and Bilton, 1999). The researchers characterize the affected individuals to be experiencing considerably greater problems in inhibiting or delaying response than most people. In addition, the neuroimaging research studies have found that the impulsive behavior seems to line with the damages in the pre-frontal cortex of the brain (Cooper and Bilton, 1999). Magnetic Resonance Imaging (MRI), Electrocephalographic examination (EEG) and Positron Emission Tomography (PET) are the procedures used to view the anatomy and activity of the brain.

Furthermore, there have been eleven structural imaging studies of children, adolescents, and adults with AD/HD. They have used computerized tomography (CT) and MRI. Ten of these studies have found evidence of structural abnormalities in the brain of AD/HD probands. Four of these studies specifically found abnormalities in the frontal cortex (Faraone and Biederman, 1998). Comparatively, PET scans, which assess regional cerebral blood flow (rCBF) or glucose metabolism, show evidence for brain dysfunction among these AD/HD patients. Both structural and functional

neuroimaging studies suggest that the fronto-subcortical system in the brain is implicated in pathophysiology of attention deficit hyperactivity disorder (Frarone and Biederman, 1998). Neuroimaging studies are ideal for testing hypotheses about the dysfunctional locus in the brain that might be implicated in AD/HD. These studies directly assess the brain structure and function. However, they are expensive and are only applied to small samples. This will affect the statistical power of these studies and may not give a definitive conclusion on the pathology of AD/HD (Faraone and Biederman, 1998). Nonetheless, these studies provide increasing support for the concept of AD/HD as a brain based disorder that has a biological basis.

***Heritability of AD/HD***

Many research studies support the presence of a genetic component of AD/HD (Cooper and Bilton, 1999; Lopez, 1965; Millichap, 1999, Faraone et al., 2005). These studies include family studies which have shown that the disorder is more common in biological relatives of children with AD/HD than it is in biological relatives of children without AD/HD (Cooper and Bilton, 1999). Furthermore, twin studies provide direct evidence on the heritability of AD/HD. Lopez (1965) studied hyperactivity in twins. Among the ten pairs of twins in which (at least one exhibits the hyperactive subtype), all four pairs of MZ (monozygotic), all boys, were concordant (i.e. both affected). Only one of the six DZ (dizygotic) showed concordance (Millichap, 1999). Faraone et al (2005) have compiled heritability estimates from 20 twin studies from the United

States, Australia, Scandinavia, and the European Union. The heritability estimate is found to be 76%, and suggests that AD/HD is among the most heritable of psychiatric disorders. Figure 2 is a summary of these studies. These studies provide strong evidence for a genetic basis contributing to the behavioral symptoms of AD/HD.

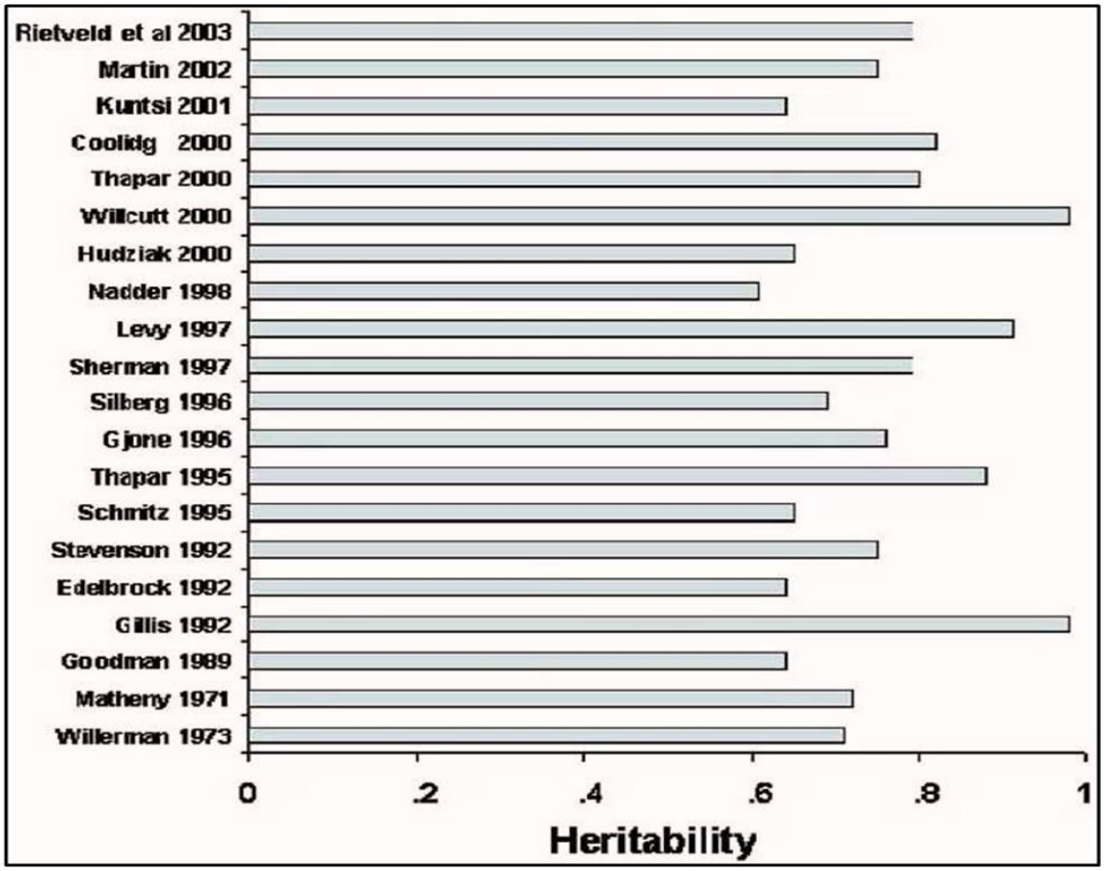


**Figure 2. Heritability of AD/HD.** Estimated heritability of AD/HD compiled from 20 twin studies (from Farone et al., 2005).

### *Genome-wide Scans and AD/HD*

A complementary strategy for pinpointing genetic risk factors involved in AD/HD susceptibility is to perform a systematic genome wide scan of affected subjects. A genome-wide scan by Fisher et al (2002) found that 17p11 (maximum LOD Score, MLS=2.98) region could harbor risk genes for AD/HD. The chromosomal regions 5p12, 10q26, 12q23, and 16p13 are also of interest. In addition, a fine mapping study by Smalley et al (2002), using same sample as Fisher et al, suggests the 16p13 (MLS=4.2) region may contain possible genes contributing to AD/HD symptoms. Furthermore, Ogdie et al extended the sample size used by Fisher et al and performed a fine mapping study. They found that 6q12 (MLS=3.30), 17p11 (MLS 3.63) and 5p13 (MLS=2.25) are likely to harbor susceptibility genes for AD/HD. Interestingly, the regions 16p13, 17p11 and 5p13 are also regions of interest in autism (Ogdie eat al, 2004; Smalley et al., 2002). These studies serve as foundation for subsequent investigations using association methods to detect risk genes of moderate effect size.

### *Candidate Genes and AD/HD*

Attention deficit and hyperactivity disorder, like most psychiatric disorders, is considered to be genetically complex. Its phenotype ranges from mildly to severely affected. The disorder exhibits familial clustering, but its transmission does not clearly show a classical Mendelian segregation (Acosta et al., 2003). Most molecular genetic

studies of AD/HD center on the candidate gene approach. Through observations based on using psychostimulant drugs, animal models and by theoretical considerations, the genes in the dopaminergic, serotinergic, and adrenergic systems are possible targets (Shastry, 2004). There have been a number of cumulative studies that have found evidence supporting the presence of major genes that may be implicated to the susceptibility of AD/HD. Researchers have used different genetic approaches such as family-based and case-control studies to test for association and/or linkage to these candidate genes. However, these studies show variable results. Some studies found linkage while others did not (Acosta et.al. 2003). Therefore, it is important to do further testing of association on susceptibility genes and dissect each gene for its small genetic effect and possibly its interactions with other candidate genes. There are eight genes implicated in the dopaminergic system, which includes its receptors, transporters and the metabolizing enzymes.

### *The DAT1/SLCA3 Gene*

Gene knock-out mice experiments show that a mouse that lacks the dopamine transporter gene (*DAT1* or *solute carrier family 6, member 3* (*SLC6A3*)) has increased locomotor activity (Giros et al, 1996). The disruption of *DAT1* gene inhibits the proper delivery of dopamine. This makes the knock-out mice display behavior such as hyperactivity like that observed in AD/HD. Methylphenidate, the major pharmacological treatment of AD/HD, and other psychostimulant drugs, target the

dopamine transporter and increase synaptic levels of dopamine. These drugs help control the symptoms seen in AD/HD (DiMaio et al, 2003).

*DAT1/SLC6A* has been localized to chromosome 5p15.3. DiMaio et al (2003) compiled studies of association between the 480-bp VNTR (variable number tandem repeat) or 10R allele of the dopamine transporter gene and AD/HD (Table 1). Most of these studies have used case-control and family based association designs. They used

**Table 1. Association Studies between AD/HD and the *DAT1/SLC6A3* 480-bp VNTR Allele (from DiMaio et al., 2003).**

| Study | Location | No of probands | Diagnostic system | Test of association | Linkage | Statistic | p-value |
|---|---|---|---|---|---|---|---|
| Barr et al[27] | Canada | 102 | DSM-IV | TDT | - | $\chi^2 = 2.6$ | 0.06 |
| Roman et al[28] | Brazil | 81 | DSM-IV | HHRR | - | $\chi^2 = 0.02$ | 0.88 |
| Curran et al[29] | Turkey | 111 | DSM-IV | TDT | - | $\chi^2 = 0.93$ | 0.34 |
| Curran et al[29] | United Kingdom | 66 | DSM-IV | TDT | + | $\chi^2 = 8.97$ | 0.001 |
| Holmes et al[30] | United Kingdom | 137 | ICD-10, DSMIV and DSM-III-R | TDT | - | OR = 0.89 | 0.59 |
| Palmer et al[31] | United States | 209 | DSM-IV and DSM-III-R | TDT | - | OR = 0.88 | 0.4 |
| Daly et al[32] | Ireland | 118 | DSM-IV | HHRR | + | RR = 1.2 | 0.006 |
| Waldman et al[33] | United States | 122 | DSM-IV | TDT | +* | OR =1.63 | 0.06 |
| Cook et al[34] | United States | 49 | DSM-III-R | HHRR | + | OR = 3.17 | 0.01 |

Note: DSM = Diagnostic and Statistical Manual of Mental Disorders; ICD-10 = International Statistical Classification of Diseases and Related Health Problems TDT = Transmission Disequilibrium Test; HHRR = Haplotype-Based Relative Risk; OR = Odds Ratio; *For combined type only

statistical analysis, such as HHRR test (Haplotype based Haplotype Relative Risk and TDT (Transmission Disequilibrium Test). Two out of the six studies using the TDT statistic have identified linkage and two out of the three studies using the HHRR test also found linkage. The causes of the discrepancies among these studies are unknown, but linkage in the TDT studies may be difficult to detect if sample sizes are insufficient in each group. Consequently, Ouelle-Morin et al did not find association in the VNTR

region of this gene with AD/HD related phenotypes in a Canadian population-based sample of same-age twins but found positive association for the rs27072 polymorphism (2007). This suggests that other polymorphisms in this gene may exist as possible markers for AD/HD and different polymorphisms may be associated in different population groups. However, due to the alteration of dopamine transporter regulation by the psychostimulant drugs and its contribution to the AD/HD symptom relief, *DAT1/SLC6A3* is still a good candidate gene for AD/HD.

### *The DRD4 Gene*

The *dopamine receptor 4 gene* (*DRD4*) is in chromosome 11p15.5. This gene is one of the five that code for receptor proteins for the dopamine neurotransmitter. Its diversity is primarily due to the length and single-nucleotide polymorphism (SNP) variation in a 48-bp VNTR (variable number of tandem repeat) in exon 3 of the gene. This region encodes the third intracellular loop of this receptor (Chang et al., 1996). There are about 11 *DRD4* VNTR repeat alleles found worldwide in the human population. The 4-, 7-, and 2- repeat alleles are the most prevalent (Ding et al., 2002; Wang et al., 2004). The 7-repeat allele has been of interest because of its high frequency in the Americas (where AD/HD is more prevalent), but its frequency is low in Asian populations where AD/HD has low prevalence (Grady et al., 2003; Leung et al., 1996).

*DRD4* is also a good candidate gene because of the pharmacological treatments (such as amphetamine, methylphenidate, and pemoline) for AD/HD that target primarily the dopamine system (Swanson et al, 1998). It is observed that DRD4 mRNA is localized in the frontal and prefrontal cortical regions of the brain, suggested regions where attention is regulated. Also, this gene is associated with the novelty seeking personality that may be related to the expression of behaviors seen in AD/HD (Swanson et al, 1998). *DRD4* gene-knockout mice show reduce novelty seeking behavior and increased locomotor sensitivity to ethanol, cocaine, and methylamphetamine (Bobb et al, 2004).

A study using a family-based approach provides further evidence that the *DRD4* gene is associated with a refined phenotype of AD/HD (Swanson et al., 1998). Faraone et al. (2001) used meta-analysis of both case-control (8 studies) and family-based (14 studies) association studies between AD/HD and the 7R allele of the 48bp VNTR polymorphism. Five of the eight case-control studies found positive association with a combined estimate of OR (odds of ratio) 1.9. In the family based studies, nine studies showed positive linkage with a combined estimate OR of 1.4. Their meta-analysis study highly support the association of 7R allele of the *DRD4* gene with AD/HD. In addition, DiMaio et al. (2003) reviewed association studies between AD/HD and *DRD4* 7-repeat allele (Table 2). These studies either used case-control association or family-based association but reported variable results. Furthermore, a study found that volumetric abnormalities in the dorsolateral prefrontal cortex and

cerebellum may represent an intermediate neuroanatomical phenotype between *DRD4* risk alleles and the clinical expression of adult AD/HD (Monuteaux et al, 2008). This makes the *DRD4* gene a good candidate for AD/HD, and perhaps other polymorphisms are also implicated in the presentation of the disorder.

### *The COMT Gene*

The *Catechol-O-Methyltransferase* (*COMT*) gene is an attractive candidate for AD/HD because the enzyme encoded by this gene is involved in the degradation of dopamine, norepinephrine, and epinephrine (Faraone et al., 2005). It is not as well studied as the *DAT1* and *DRD4* genes. The gene that encodes for the COMT enzyme has been localized to the chromosomal region 22q11.1-q11.2. Some studies have observed the involvement of *COMT* in several AD/HD-related behaviors, substance abuse cases and novelty-seeking personality (Bobb et al., 2004). DiMaio et al. also compiled association studies between AD/HD and *COMT* polymorphisms (Table 3), but only one of these studies shows evidence of linkage (2003). Due to the inconsistent results, further molecular studies in defining markers (such as SNPs) on this gene are needed in order to test for association with AD/HD.

In addition, the *COMT* gene contains functional polymorphisms such as the Val158Met, rs4680 which can affect the activity of the enzyme to process dopamine in the prefrontal cortex (Lachman et al., 1996; Chen et al., 2004). A study on an ethnically homogenous

**Table 2. Association Studies between AD/HD and the *DRD4* 7-Repeat Allele from DiMaio et al., 2003).**

| Study | Location | No of probands | Diagnostic system | Test of association | Linkage | Statistic | p-value* |
|---|---|---|---|---|---|---|---|
| ***Case-Control Association Studies*** | | | | | | | |
| Mill et al[58] | United Kingdom | 132 | DSM-IV | $\chi^2$ | + | OR = 6.2 | 0.01 |
| Holmes et al[30] | United Kingdom | 129 | ICD-10, DSM-IV and DSM-III-R | $\chi^2$ | + | OR = 1.9 | 0.001 |
| Muglia et al[59] | Canada | 66 | DSM-IV | $\chi^2$ | + | OR = 2.5 | 0.01 |
| Comings et al[60] | United States | 52 | DSM-III-R and DSM-IV | $\chi^2$ | + | $\chi^2$ = 6.64 | 0.01 |
| Rowe et al[61] | United States | 70 | DSM-IV | $\chi^2$ | + | $\chi^2$ = 5.9 | < 0.05 |
| Swanson et al[62] | United States | 39 | DSM-IV and ICD-10 | $\chi^2$ | + | $\chi^2$ = 4.65 | < 0.035 |
| Castellanos et al[63] | United States | 41 | DSM-III-R | $\chi^2$ | - | $\chi^2$ = 0.06 | 0.81 |
| La Hoste et al[64] | United States | 39 | DSM-IV | $\chi^2$ | + | OR = 3.0 | 95% CI = 1.3-7.1 |
| ***Family-Based Association Studies*** | | | | | | | |
| Roman et al[28] | Brazil | 81 | DSM-IV | HHRR | - | $\chi^2$ = 0.37 | 0.54 |
| Payton et al[52] | United Kingdom | 103 | ICD-10, DSM-IV and DSM-III-R | TDT | - | N/A | 0.75 |
| Mill et al[58] | United Kingdom | 85 | DSM-IV | TDT, HHRR | - | N/A | |
| McCracket et al[65] | United States | 371 | DSM-IV | TDT | + | $\chi^2$ = 5.4 | 0.02 |
| Barr et al[66] | Canada | 82 | DSM-IV | TDT | + | $\chi^2$ = 15.68 | <0.016 |
| Holmes et al[30] | United Kingdom | 110 | ICD-10, DSM-IV and DSM-III-R | TDT | - | OR = 0.95 | 95% CI = 0.6-1.5 |
| Kotler et al[67] | Israel | 49 | DSM-IV | HHRR | - | LR = 7.94 | 0.16 |
| Muglia et al[59] | Canada | 66 | DSM-IV | TDT | + | z = 1.41 | 0.07 |
| Hawi et al[68] | Ireland | 78 | DSM-IV | HHRR | - | $\chi^2$ = 0.00 | 0.95 |
| Tahir et al[69] | Turkey | 104 | DSM-IV | TDT | + | $\chi^2$ = 2.79 | 0.05 |
| Eisenberg et al[70] | Israel | 49 | DSM-IV | HHRR | - | $\chi^2$ = 0.14 | 0.71 |
| Farone et al[71] | United States | 54 | DSM-IV | TDT | + | $\chi^2$ = 7.4 | 0.007 |
| Rowe et al[61] | United States | 70 | DSM-IV | TDT | - | $\chi^2$ = 0.03 | N/A |
| Smalley et al[72] | United States | 129 | DSM-III-R and DSM-IV | TDT | + | $\chi^2$ = 4.85 | 0.03 |
| Swanson et al[62] | United States | 52 | DSM-IV | HHRR | + | $\chi^2$ = 4.65 | < 0.035 |

Note: CI = Confidence interval; LR = Likelihood Ration; *Unless otherwise indicated. Other abbreviations see footnotes on Table 1.

and Caucasian origin sample population found that the Val allele on this SNP is more frequent in AD/HD group than in the control population (Kereszturi et al., 2008). In addition, the Val allele (homozygous) was found to be associated with AD/HD (predominantly inattentive) subjects with comorbid ODD in a Chinese population sample group. They also looked into variants on the *MAOA* gene but found no positive association, suggesting that there is no epistatic effect of *MAOA* to the *COMT* gene (Qian et al, 2009).

**Table 3. Association Studies between AD/HD and *COMT* Val158Met Polymorphism (from DiMaio et al., 2003)**

| Study | Location | No of probands | Diagnostic system | Test of association | Linkage | Statistic | p-value* |
|---|---|---|---|---|---|---|---|
| Payton et al[52] | United Kingdom | 98 | ICD-10, DSM-IV and DSM-III-R | TDT | - | -- | -- |
| Manor et al[94] | Israel | 70 | DSM-IV | HHRR | - | LR = 1.74 | 0.19 |
| Tahir et al[95] | Turkey | 72 | DSM-IV | TDT | - | $\chi^2 = 0.93$ | NS |
| | | | | HHRR | | $\chi^2 = 2.2$ | NS |
| Hawi et al[96] | Ireland | 94 | DSM-IV | HHRR | - | $\chi^2 = 0.18$ | 0.67 |
| Eisenberg et al[97] | Israel | 48 | DSM-IV | HHRR | + | $\chi^2 = 4.72$ | 0.03 |
| Barr et al[98] | Canada | 77 | DSM-IV | TDT | - | $\chi^2 = 1.25$ | 0.26 |

Note: NS = Not Significant, for other abbreviations, see footnotes of Tables 1 and 2.

### *Genetic Association Studies, SNPs as Genetic Markers, Common-Disease Common-Variant (CDCV) Hypothesis and Linkage Disequilibrium*

Genetic association studies are one of the study designs that can define the genetic correlates of complex disorders such as AD/HD. Although it is a straightforward method, it has been unsuccessful in identifying replicable associations for human complex diseases (Sullivan, 2007; Hirshhorn et al., 2002). Association studies test for a particular genetic marker that could potentially be implicated with the disease of interest

across the population rather than within families, which allow these studies better power to detect the effects of common variants compared to linkage studies (Hirschhorn, 2005; Risch, 1990; Risch and Merikangas, 1996). Efficient recognition of genetic variations among individuals in different populations is an essential step for a successful association study to further understand the genetics of AD/HD and may help resolve the varying results of these studies and/or in any other association studies (DiMaio et al, 2003; Cardon and Bell, 2001). The completion of the Human Genome Project has led to the development of numerous highly polymorphic markers (microsatellites and SNPs). Genetic markers are defined as any polymorphic Mendelian character that can be used to follow a chromosomal segment through a pedigree. A good genetic marker has to be polymorphic or have a high heterozygosity index, meaning that a randomly selected individual in a given population has a considerable chance of being heterozygous for that polymorphism of interest (Strachan and Read, 2004).

SNPs are the preferred markers for association studies because they are estimated to occur at a frequency of 1in every 1,000 bases on average in the human genome (Syvanen, 2001). Their abundance helps define linkage disequilibrium islands and can be scored through various high-throughput methods (Sobrino et al, 2004). Figure 3 shows steps of SNP genotyping. They are less mutable than microsatellite markers and common SNPs could be found in different populations, but their allele frequencies may vary (Romualdi et al, 2002). With the assumption that the susceptibility to common diseases is mainly determined by common ancestral DNA

variants, a more stable marker like SNPs should be more valuable for identifying ancestral haplotypes or sets of SNP alleles occurring together on a chromosome (Strachan and Read, 2004). This has led to the common-disease common-variant (CDCV) hypothesis which argues that the genetic variation underlying susceptibility to common heritable disease/trait existed within the founding population of contemporary human population and that these genetic variants have escaped selective pressure and have small to modest effect on the disease/trait (Doris, 2002; Chen et al., 2007; Collins et al., 1999; Risch and Merikangas, 1996; Lander, 1996). Furthermore, four arguments support the CDCV hypothesis (Hirshhorn, 2005; Lohmueller et al., 2003; Reich and Lander, 2001):

1. Common diseases/traits are not as evolutionarily disadvantageous as single-gene diseases/trait. Monogenic traits usually cause early death and/or limit reproductive capability.
2. Variants for monogenic traits are highly penetrant whereas multiple variants are often implicated in common complex traits. This decreases the selective pressure of the genetic variants on common traits.
3. Monogenic traits are rare and polygenic traits are more common. This leads to the hypothesis that the causal genetic variants for common complex traits should have a high frequency in the population due to demographic history of the human population.

4. Empirical experiments supports that common genetic variants contribute to the risk of common/complex traits.

These assumptions may also explain the varying results found in the association studies found on AD/HD (DiMaio et al., 2003; Acosta et al., 2003; Hirschhorn et al., 2002).

Genetic association studies are based on linkage disequilibrium (LD) between marker alleles and causative alleles for a trait or disease. LD is defined as the association of alleles on a chromosome. It is a measurement of the difference between the combined observed allele frequencies of a two-locus allelic combination and the product of their individual allele frequencies (expected frequencies). A low LD means that alleles at the two loci tend to be randomly associated. Human variation studies have shown that nearby SNPs typically show high LD levels and these form highly variable segments (haplotypes) in the genome (Gabriel et al, 2002). The linkage disequilibrium pattern of SNPs enables the capture of common genetic variants by genotyping subsets of SNPs ('tag' SNPs, Haplotype-tagging) across the genes of interests or chromosomes implicated in a disease or disorder (Haiman and Stram, 2008). The knowledge of LD and common haploytpe patterns in disease association can improve the cost effectiveness of these studies because it guides in the selection of informative `tag` SNPs (Bakker et al, 2005). The International HapMap Consortium uses $D'$ and $r^2$ as the main statistical measures for LD (Barnes, 2006). The calculation of LD is derived from consideration of two loci A (*A*, *a* alleles) and B (*B*, *b* alleles) and four possible haplotypes with their allele

frequencies as p. $D' = (p_{AB} - p_A p_B)/D_{max}$; $D_{max}$ is the maximum value of $|p_{AB}\text{-}p_A p_B|$ possible allele frequencies; $D = p_{AB}p_{ab}\text{-}p_{aB}$. If $D'$=1 then the two SNPs have not been separated by recombination during the history of the sample. The $D'$ measurement is useful within studies but not between different populations (Barnes, 2006; Pritchard and Przeworski, 2001; Mueller, 2004). $r^2$ or $\Delta^2 = (p_{AB}\text{-}p_A p_B)^2/(p_A p_a p_B p_b)$ (Starchan and Read, 2004). $r^2$ is the preferred choice measure of LD when the focus is on the predictability of one polymorphism given the other and therefore it is often used in power studies for genetic association study designs (Chen et al., 2006). The inverse value of $r^2$, $(1/ r^2)$ provides a practical estimate of magnitude by which the sample size must be increased in a study design to detect association between the genetic marker and the disease/trait of interest (Pritchard and Przeworski, 2001). The International HapMap consortium has provided a public database of common variation in the human genome, particularly SNPs. They have genotyped 269 DNA samples from four different populations (CEU, YRI, and CHB+JPT) and characterized LD patterns to provide recombination hotspots in the human genome (The International HapMap Consortium, 2005). This has become an essential resource that can guide study designs (SNP selection and defining LD patterns and haplotypes) for future genetic association studies.

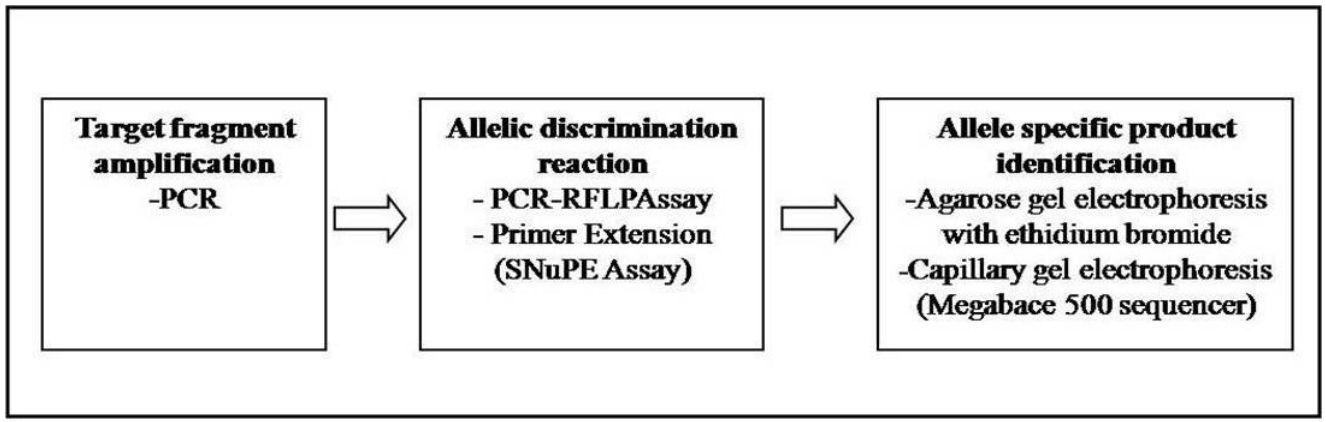


**Figure 3. SNP Genotyping Methods.** The first step of any genotyping method is PCR, amplification of the region containing the SNP. Further assays such as SNuPe and PCR-RFLP are some of the SNP genotyping methods to determine the SNP of interest; Megabace Sequencer and ethidium bromide agarose gels respectively, are assays used for detection. This figure is modified from Chen and Sullivan (2003).

# CHAPTER III

# MATERIALS AND METHODS

## *SNP Genotyping: SNuPe*

A panel of DNA samples was used to optimize this genotyping method. The selection of SNPs for *DAT1/SLC6A3, DRD4* and *COMT* genes was gathered from database searches through Ensemble, dbSNP, HGbase, and TSC (Nelson, personal communication, data not shown). These database searches allowed us to identify SNPs in the candidate genes and derive the primary sequence of the SNP region for each gene. For this method, primers were designed to amplify the region containing the SNP of interest using Primer 3 software (http://frodo.wi.mit.edu/). The PCR products were approximately 300 base pairs (bp). After PCR amplification, the SNP in that region was assayed using a single base extension (SNuPe) reaction. SNuPe involved binding an oligonucleotide primer immediately 5' of the SNP and then extending it by a single base using differently fluorescently-labeled dideoxy terminators (ddNTPs) for each base; this labeled the candidate SNP gene site with flurochromes**.** The SNPs were then scored by capillary gel electrophoresis on a MegaBace 500 sequencer. The MegaBace SNuPe genotyping kits and protocols by Amersham Biosciences were followed accordingly (Appendices A-C).

### *DNA Samples for PCR-RFLP*

Five family trio DNA samples were ordered from National Institute of General Medical Science (NIGMS) DNA Bank. These were used to assess the Mendelian inheritance for each SNP and give information about the polymorphism value for these SNPs. Table 4 shows sample information. The DNA samples were quantified using the Nanodrop Spectrophometer ND1000.

**Table 4. DNA Information**. Samples were from the National Institute of General Medicine Science (NIGMS) DNA Bank.

| No. | Catalog # | Pedigree | Description | ng/ul | 260/280 |
|---|---|---|---|---|---|
| 1 | NA12335 | 1330 | father | 326.84 | 1.88 |
| 2 | NA12336 | 1330 | mother | 319.18 | 1.89 |
| 3 | NA12339 | 1330 | son | 331.6 | 1.87 |
| 4 | NA10846 | 1334 | father | 285.5 | 1.85 |
| 5 | NA10847 | 1334 | mother | 364.97 | 1.86 |
| 6 | NA12143 | 1334 | son | 287.89 | 1.86 |
| 7 | NA07029 | 1340 | father | 280.26 | 1.88 |
| 8 | NA07019 | 1340 | mother | 267.75 | 1.86 |
| 9 | NA07027 | 1340 | son | 259.33 | 1.85 |
| 10 | NA07048 | 1341 | father | 318.97 | 1.88 |
| 11 | NA06991 | 1341 | mother | 338.53 | 1.89 |
| 12 | NA07010 | 1341 | daughter | 283.3 | 1.86 |
| 13 | NA07349 | 1345 | father | 286.11 | 1.88 |
| 14 | NA07348 | 1345 | mother | 220.96 | 1.84 |
| 15 | NA07356 | 1345 | son | 263.27 | 1.84 |

From the National Institute of General Medical Science (NIGMS) DNA Bank

***SNP Selection***

The SNPs (tag SNPs) selected did not have a particular functional polymorphism that caused a change in the amino acid sequence. SNPs were first screened according to their minor allele frequencies (MAF, at least 0.2 and greater if possible), and unique polymorphism patterns on the CEU (Centre d'Etude du Polymorphisme Humain, individuals from Utah, USA with European descent) population. We utilized the genotype data available on this population in the HapMap database. SNPs were then further screened according to the availability of restriction enzymes. Three SNPs were selected for *COMT* (rs737866, rs5993882, and rs4633) and *DAT1/SLC6A3* (rs10052016, rs1042098, and rs463379) genes. However, there were only two SNPs genotyped on the *DRD4* gene on the HapMap database (accessed August 2008) and there were no restriction enzymes available for these SNPs. The two SNPs assayed in the *DRD4* gene were selected based on the heterozygosity score on the NCBI website and/or their location in the gene (Table 5). A study using a common 5-SNP (rs2550948, rs11564750, rs261759, rs2652511, rs2975223) haplotype located in the 5' region of *SLC6A3/DAT1* detected association with AD/HD (243 families with 186 parent proband trios and 57 parent-child duos, Brazil population) (Genro et al., 2008). In addition, using a 4-SNP haplotype (rs6269, rs4633, rs4818, rs5680) on the *COMT* gene showed a trend for association with their hyperactivity/impulsivity sample group (435 probands and 245 controls from all parts of Norway) (Halleland et al., 2008). Our SNP

selection method (using 2-3 SNPs as tags) was thus not as robust as those used in these recent studies.

***SNP-Genotyping: PCR-RFLP***

The Primer 3 program was used to design primers to amplify the region containing the SNPs of interest. The FASTA sequences (from NCBI) were used as the input for the program. The same sequences were used on the NEBcutter website to select for restriction enzymes and made sure that the restriction enzymes cut specifically at the SNPs and not at other sites within the PCR product. Table 5 summarizes the tag SNPs on each gene, restriction enzymes, primer sets and their annealing temperatures.

We used 25 nanograms (ng) of DNA per sample to perform PCR. A 10x PCR buffer (Promega), left and right primers, $MgCl_2$, dNTPs and deionized water were added accordingly with a twenty five microliter (25µl) total volume. A thermal cycling protocol was used adjusting the denaturation, primer annealing and DNA synthesis temperatures (see Appendix D). PCR products were confirmed by running gel electrophoresis using ethidium bromide treated agarose gels (80-90 volts for 1-2 hrs). After the presence of the product was confirmed, 10 µl of the PCR product was used for PCR-RFLP. The incubation temperature of PCR-RFLP was based on the information that came with the restriction enzyme. The PCR-RFLP products were electrophoresed on

ethidium bromide treated agarose gel with the same voltage. SNP genotypes were called according to their banding patterns (Table 5).

***'Tag' SNPs Analysis***

The Haploview software was used to assess the linkage disequilibrium scores for the SNPs selected on each gene (http://www.broad.mit.edu/mpg/haploview/). We used the assayed SNPs (PCR-RFLP) as our 'tag' SNPs and compared the LD values (expressed as the squared correlation of allelic occurrences, $r^2$) of this set of SNPs with all other identified SNPs in the three populations (CEU, YRI and CHB+JPT). LD values were evaluated using the Haploview tagger program and setting the $r^2$ threshold to 0.01 in order to capture all the SNPs in each gene. The Haploview (Tagger) assigned the SNPs to any of the tag SNPs where they had the highest $r^2$ value. A Wilcoxon Signed Ranks Test was used to compare the distribution of LD values ($r^2$) in the CEU population from which the tag SNPs were initially chosen to those in the other populations. A regression analysis was used to test the relationship of the SNPs' $r^2$ values with their MAF values within each population and plot their distribution (SPSS program).

**Table 5. Tag SNPs Information on each gene.** This table includes primer sets used for amplification, their annealing temperatures and the SNPs restriction enzyme information.

| Gene | Chromosome Location | NCBI Ref SNP ID | Allele | Oligo Primer | Annealing Temp (°C) | PCR product (bp) | Restriction Enzyme | Enzyme Incuabtion Temp (°C) | Digestion product |
|---|---|---|---|---|---|---|---|---|---|
| DAT1/SLC6A3 solute carrier family 6 (neurotransmitter transporter, dopamine),member 3 | 5p15.3 | rs1042098 | C/T | LEFT PRIMER AGACGAAGACCCCAGGAAGT | 60.11 | | | | TT = 167bp |
| | | | | RIGHT PRIMER GCACGGAAAGGTGTAAACAG | 58.28 | 167 | AciI | 37 | CT = 167bp, 142bp and 25bp |
| | | | | | | | | | CC = 142bp and 25bp |
| | | rs10052016 | A/G | LEFT PRIMER cagttgagatttcagtacccctt | 59.93 | | | | AA = 150bp |
| | | | | RIGHT PRIMER atgtatctgctgtctttggttg | 58.33 | 150 | NlaIII | 37 | AG = 150bp, 100bp and 50bp |
| | | | | | | | | | GG = 100bp and 50bp |
| | | rs463379 | C/G | LEFT PRIMER CGTCATTTTGGAGAGGAGGA | 60.19 | | | | CC = 213bp |
| | | | | RIGHT PRIMER CAAAGGCACCTACTGGCTGT | 60.31 | 213 | AciI | 37 | CG = 213bp, 171bp and 42bp |
| | | | | | | | | | GG = 171 bp and 42 bp |
| DRD4 dopamine receptor D4 | 11p15.5 | rs936463 | C/T | LEFT PRIMER GGTGGGTTCCTGTCCTGAG | 60.51 | | | | TT = 208bp |
| | | | | RIGHT PRIMER GTAGCCCAGCCAGGTGAC | 59.21 | 208 | | | CT = 208bp, 127bp and 81bp |
| | | | | | | | | | CC = 127bp and 81bp |
| | | | | LEFT PRIMER GTGGTGGTCGGTGGGTTC | 62.32 | | MboII | 37 | TT = 175bp |
| | | | | RIGHT PRIMER GGAGCAGGCAGGACACAG | 60.58 | 175 | | | CT = 175 bp, 136bp and 39bp |
| | | | | | | | | | CC = 136bp and 39bp |
| | | rs936461 | A/G | LEFT PRIMER GTCACCCCTCTTTGGTGAAG | 59.55 | | | | AA = 190bp |
| | | | | RIGHT PRIMER AAGACCGTGAGCTAGGTAGGC | 59.92 | 190 | BfuAI | 50 | AG = 190bp, 128bp and 62bp |
| | | | | | | | | | GG = 128bp and 62bp |
| COMT catechol-O-methyltransferase | 22q11.21-q11.23; 22q11.21 | rs737866 | A/G | LEFT PRIMER catggagccagcttttctc | 59.96 | | | | AA = 245bp |
| | | | | RIGHT PRIMER GGACCACGTGGGAATGTTAG | 60.23 | 245 | BpmI | 37 | AG = 245bp,170bp and 75bp |
| | | | | | | | | | GG = 170bp and 75bp |
| | | rs5993882 | G/T | LEFT PRIMER cagaccagaatactttagccagt | 60.08 | | | | TT = 239bp |
| | | | | RIGHT PRIMER TTGTGTTTTGTTAATTGCGTTTG | 59.96 | 239 | TseI | 65 | TG = 239bp, 159bp and 80bp |
| | | | | | | | | | GG = 159bp and 80bp |
| | | rs4633 | C/T | LEFT PRIMER ACAACCTGCTCATGGGTGAC | 61 | | | | TT = 222bp |
| | | | | RIGHT PRIMER TCCTGTAAGGGCTTTGATGC | 60.21 | 222 | PmlI (C) | 37 | CT = 22bp, 175bp and 47bp |
| | | | | | | | | | CC = 175bp and 47bp |

# CHAPTER IV

# RESULTS

## *SNP Genotyping: SNuPe*

With the DNA samples provided, we ran PCR using the *DRD4* gene primer sets with 25 nanograms (ng) of DNA. We used two different concentrations of $MgCl_2$ for comparison and adjusted the annealing temperatures accordingly. A 50-Enhanced PCR cycling temperature program (courtesy of Li-cor Biotechnology) was used to amplify the regions containing the SNPs of interest. PCR products were electrophrosed on ethidium bromide treated agarose gels. Figure 4 shows that using a lesser concentration of $MgCl_2$ (2.5mM) was better for these primer sets. We then used the same protocols for the primer sets on *DAT-1/SLC6A3* and *COMT* genes (data not shown). These results showed that there is an ~80% success rate of amplifying PCR products (regions of interest) on each gene using this PCR protocol, and the 50 Enhanced thermal cycling program (See Appendix for protocols). Robust amplification of these regions is essential for the downstream steps involved in SNP genotyping assays.

After amplifying these regions in each gene, the SNuPe genotyping kits and protocols (provided by Amersham Biosciences) were followed accordingly. The products were run on a Megabace 500 Sequencer which showed

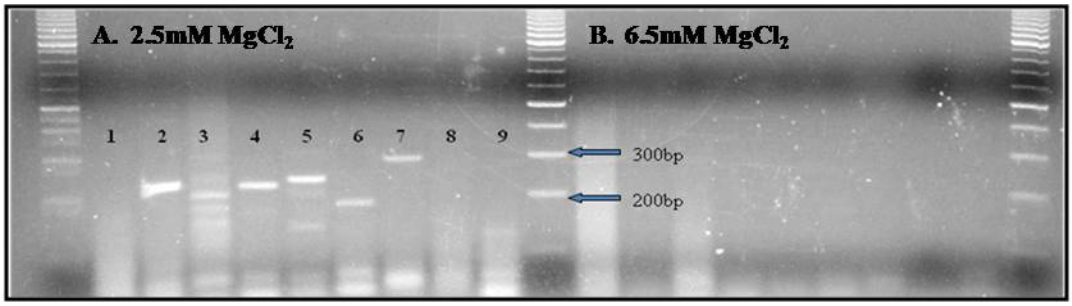


**Figure 4. PCR products using one set of *DRD4* gene primers with varying $MgCl_2$ (mM, milliMolar) concentrations.** This shows that 2.5mM of $MgCl_2$ works better with this primer set (A). Products were between 200bp (base pairs) and 300bp. The numbers indicate individual DNA samples.

electropherogram peaks of the SNPs and provided quality scores indicating the confidence of the SNP calls. A quality score of 5 to 10 was suggested to be reliable by Amersham Biosciences. During the final spin (clean-up step, sephadex plate) of SNuPe products, some of the samples were not recovered, which may have resulted in our poor quality scores. An alcohol clean up method was used, but the SNuPe results were worse than with the sephadex clean up method. After several trials of making sephadex plates and using water to optimize this clean up step, I was able to recover 90% of my samples. This is yet to be tested on real samples. Furthermore, in order to confirm the SNPs that are read by the MegaBACE sequencer, I sequenced the regions of interest using a Licor DNA Analyzer protocols (provided by Li-Cor Biotechnology). The sequencing results showed the same allele reading on the MegaBACE. Table 6 shows our sequencing results compared to the SNuPe results (see Appendices A-C for protocol information).

**Table 6. SNuPe and sequencing genotyping assay results from one *DRD4* SNP, two *DAT1/SLC6A3* SNPs and three *COMT* SNPs.** The number besides the SNuPe genotypes indicate the quality score of the SNP calls. A value from 5-10 is considered confident as suggested by Amersham Biosciences. N/D indicates no data.

| DNA# | DRD4-3740649 (C/T) | | DAT-1460700 (A/G) | | DAT-1461753 (C/T) | |
|---|---|---|---|---|---|---|
| | *SNuPe* | *Seq* | *SNuPe* | *Seq* | *SNuPe* | *Seq* |
| 10 | CC - 10 | CC | AG - 0.1 | AA | CT - 6.1 | CT |
| 11 | CC - 2.7 | CC | AA - 0.1 | AA | CC - 4.1 | CC |
| 12 | CC - 0.3 | CC | AA - 0.1 | AA | CC - 3.4 | CC |
| 13 | CC - 0.1 | CC | AA - 0.1 | AA | CC - 4.6 | CC |
| 14 | N/D | CC | AG - 0.1 | AA | CT - 5.9 | CT |
| 15 | N/D | N/D | AA - 0.1 | N/D | CC - 0.5 | N/D |
| 9 | N/D | CC | N/D | AA | N/D | CC |
| **DNA#** | **COMT-740603 (A/G)** | | **COMT-165656 (C/G)** | | **COMT-769224 (A/G)** | |
| | *SNuPe* | *Seq* | *SNuPe* | *Seq* | *SNuPe* | *Seq* |
| 6 | GG - 0.7 | N/D | CG - 0.2 | N/D | AG - 0.6 | N/D |
| 9 | AA - 4.7 | N/D | CG - 0.4 | CG | AG - 0.3 | GG |
| 10 | AG - 0.6 | GG | CG - 0.3 | CG | AG - 0.7 | GG |
| 11 | AA - 1.3 | AA | CG - 0.3 | CG | AG - 0.3 | N/D |
| 12 | AA - 0.3 | AA | N/D | CG | AG - 0.3 | N/D |
| 13 | AA - 2.7 | AA | CG - 0.2 | CG | AG - 0.5 | GG |
| 14 | AG - 0.4 | AG | N/D | CG | AG - 0.5 | AG |
| 15 | N/D | N/D | N/D | N/D | N/D | N/D |

ND = No Data

### *SNP Genotyping: PCR-RFLP*

Due to the unreliability of the SNuPe assay and budget issues, we adopted to the PCR-RFLP approach to assay SNPs. The candidate SNPs selected for the SNuPe assay did not have restriction enzymes, therefore, we chose SNPs accordingly. This was the limiting factor for this method (See methods for selecting tag SNPs). Some results of the PCR-RFLP gels were unclear due to incomplete digestion of the PCR products by the restriction enzymes and clarity of the bands on the agarose gels. However comparing the allele calls from the HapMap database and the inference of Mendelian allelic

inheritance of each family trio helped in the identification of the right genotype for each individual. Consequently, the primer sets for one SNP (rs9364643) on the *DRD4* gene did not amplify the specific product. Multiple bands were produced and therefore we did not proceed with PCR-RFLP. Perhaps another genotyping method may be needed to assay this SNP. Figures 5, 6 and 7 show PCR-RLFP results for *COMT* gene. The PCR-RFLP results for the *DAT1/SLC6A3* and *DRD4* genes were not included. Tables 7 (COMT) and 8 (SLC6A3/DAT1) show the SNP genotype calls on our PCR-RFLP results and HapMap genotype calls for selected individuals.

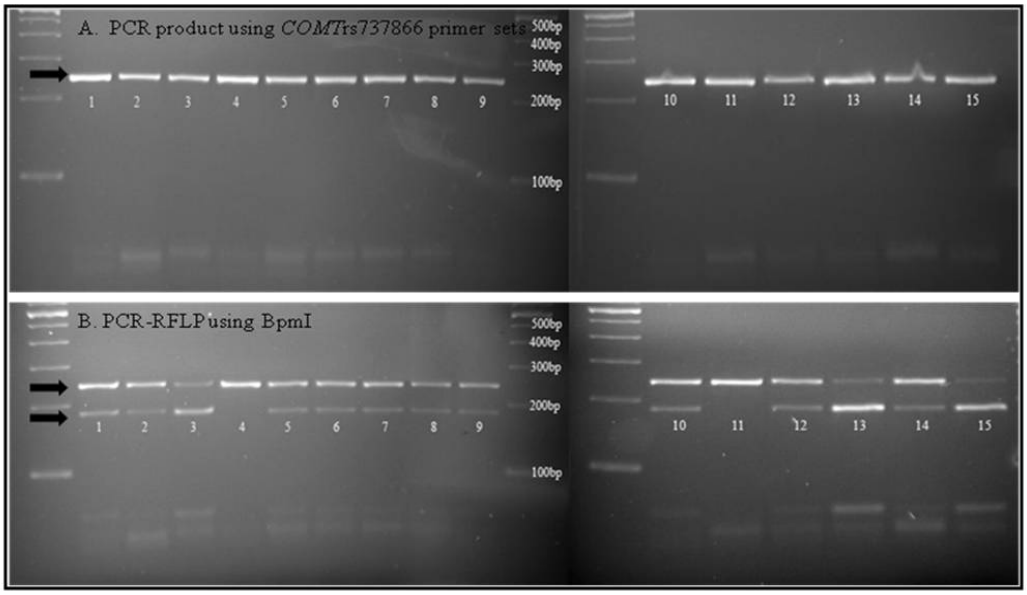


**Figure 5. PCR-RFLP genotyping results for *COMT* rs737866 SNP.** Figure A shows the PCR product and figure B shows the digested product using the BpmI restriction enzyme, arrows indicate the bands recognized.

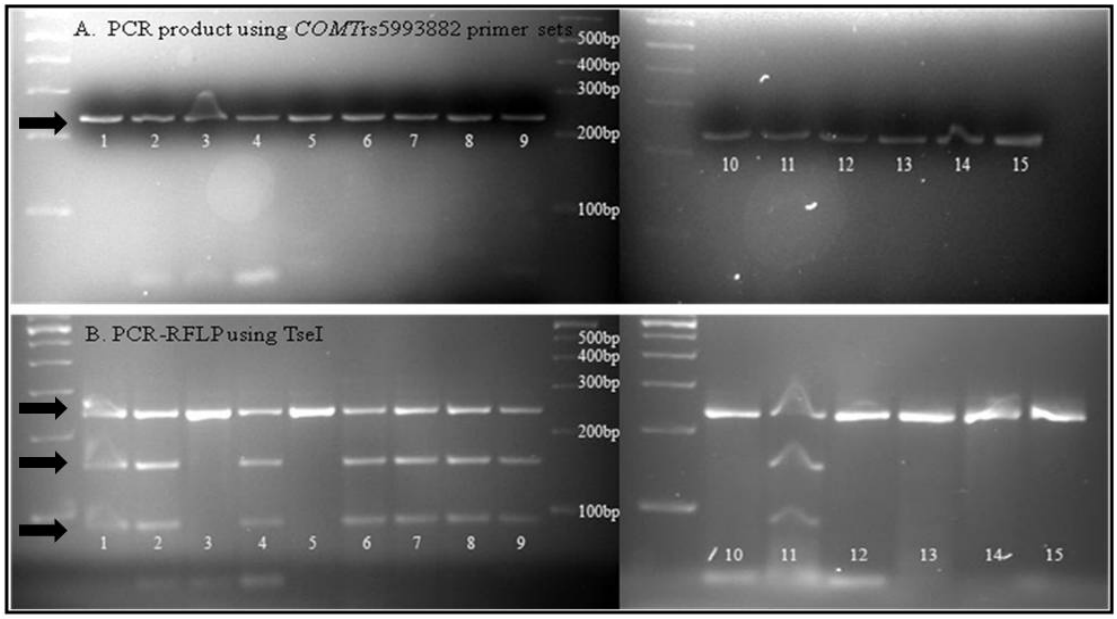


**Figure 6. PCR-RFLP genotyping results for *COMT* rs5993882 SNP.** Figure A shows the PCR product and figure B shows the digested product using the TseI restriction enzyme, arrows indicate the bands recognized.

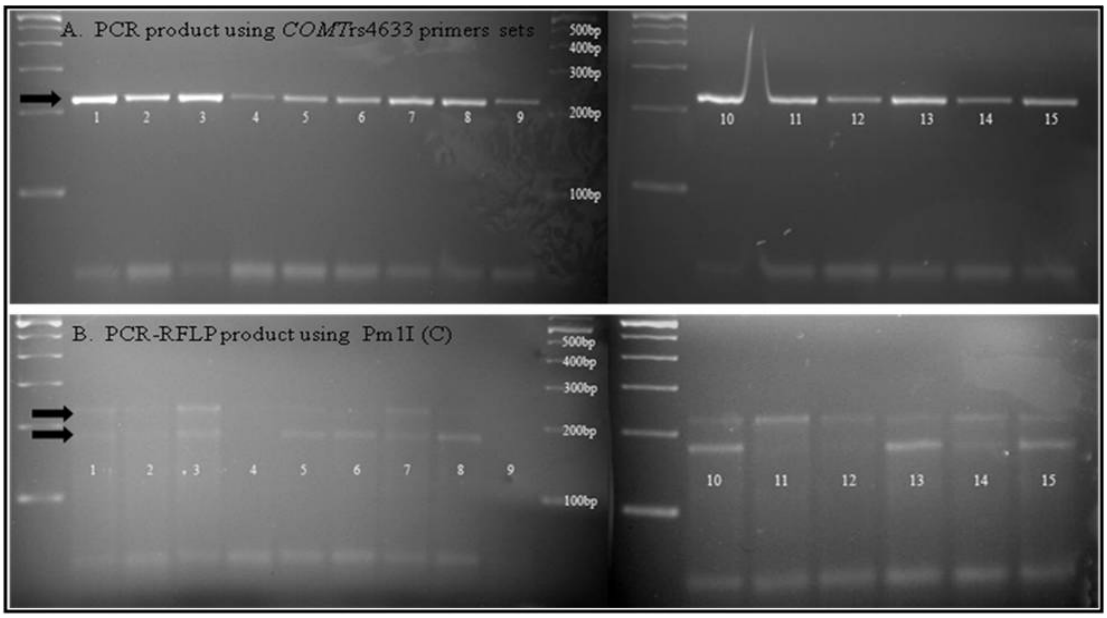


**Figure 7. PCR-RFLP genotyping results for *COMT* rs4633 SNP.** Figure A shows the PCR product and figure B shows the digested product using the Pm1I restriction enzyme, arrows indicate the bands recognized.

**Table 7. Genotype summary of the three SNPs (tag SNPs) on the *COMT* gene.** This compares our genotype calls from our PCR-RFLP assays and the genotype on the HapMap Database for our samples. N/D; No Data

| ***COMT* SNP ID:** | | **rs737866** | | **rs5993882** | | **rs4633** | |
|---|---|---|---|---|---|---|---|
| Individual # | Group | HapMAP genotype | PCR-RFLP genotype | HapMAP genotype | PCR-RFLP genotype | HapMAP genotype | PCR-RFLP genotype |
| 1 | father | N/D | AG | N/D | GT | N/D | CT |
| 2 | mother | N/D | AG | N/D | GT | N/D | CT |
| 3 | son | N/D | AG | N/D | TT | N/D | CT |
| 4 | father | AA | AA | GT | GT | CT | CT |
| 5 | mother | AG | AG | GT | TT | CC | CT |
| 6 | son | N/D | AG | N/D | GT | N/D | CC |
| 7 | father | AG | AG | GT | GT | CT | CT |
| 8 | mother | AG | AG | GT | GT | CC | CC |
| 9 | son | N/D | AG | N/D | GT | N/D | CT |
| 10 | father | AG | AG | TT | TT | CC | CC |
| 11 | mother | AA | AA | GT | GT | TT | TT |
| 12 | daughter | N/D | AG | N/D | TT | N/D | CT |
| 13 | father | N/D | AG | N/D | TT | N/D | CC |
| 14 | mother | AG | AG | GT | TT | CT | CT |
| 15 | son | N/D | AG | N/D | TT | N/D | CC |

ND = No Data

**Table 8. Genotype summary of the three SNPs (tag SNPs) on the *DAT1/SLC6A3* gene.** This compares our genotype calls from our PCR-RFLP assay and the genotype on the HapMap Database for our samples. N/D; No Data.

| ***DAT1/SLC6A3* SNP ID:** | | **rs1042098** | | **rs10052016** | | **rs463379** | |
|---|---|---|---|---|---|---|---|
| Individual # | Group | HapMAP genotype | PCR-RFLP genotype | HapMAP genotype | PCR-RFLP genotype | HapMAP genotype | PCR-RFLP genotype |
| 1 | father | N/D | TT | N/D | AA | N/D | CC |
| 2 | mother | N/D | TT | N/D | AG | N/D | CC |
| 3 | son | N/D | TT | N/D | AG | N/D | CC |
| 4 | father | TT | TT | AA | AA | CG | CG |
| 5 | mother | CC | CT | AG | AG | CC | CC |
| 6 | son | N/D | CT | N/D | AG | N/D | CG |
| 7 | father | TT | TT | AA | AA | CG | CG |
| 8 | mother | CT | CT | AA | AA | CG | CG |
| 9 | son | N/D | CT | N/D | AA | N/D | CC |
| 10 | father | TT | TT | AA | AA | CG | CG |
| 11 | mother | CT | CT | GG | GG | CC | CC |
| 12 | daughter | N/D | TT | N/D | AG | N/D | CC |
| 13 | father | N/D | TT | N/D | AA | N/D | CG |
| 14 | mother | TT | TT | AA | AA | CG | CG |
| 15 | son | N/D | TT | N/D | AA | N/D | CC |

ND = No Data

### *Allele Frequency and Linkage Disequilibrium*

We evaluated the performance of HapMap tag SNPs from the CEU population by comparing allele frequencies and LD patterns to the YRI and CHB+JPT populations. These comparisons helped us assess our hypothesis that sparse sampling of SNPs contributes to inconsistent detection of potentially causative genetic variants among human populations. In general, our results show differences on the SNPs minor allele frequency distribution in each gene on the different populations; and some SNPs (captured) have strong $r^2$ with the tag SNPs however, most of them have low $r^2$ with the tag SNPs. These observations may have limited our power to detect trait association with potential causative variants. Figure 8 shows the minor allele frequency distribution of the tag SNPs and captured SNPs on the *COMT* gene amongst the three populations. Figure 9 shows the MAF distribution on the *DAT1/SLC6A3* using 3 tag SNPs and on two populations only since there was no data on the HapMap database for one of the tag SNPs (rs463379) on the CHB+JPT populations. A two tag SNPs analysis was then executed using 2 tag SNPs (rs1042098, rs10052016) to evaluate the MAF patterns of the three populations (Figure 10). Although, allele frequencies of SNPs (markers) in each population represent only sample estimates of some underlying population parameters (Mueller, 2004), their values have an effect on the calculation of LD. Figures 11, 12 and 13 show the LD ($r^2$) distributions of the captured SNPs on these two genes in each population.

***MAF and LD ($r^2$) relationship***

The relationship between MAF and $r^2$ of captured SNPs with the tag SNPs for each population on each gene was estimated using a regression analysis. Overall, our analysis showed a positive relationship between $r^2$ and MAF (i.e. a tag SNP captures a SNP that is close or equal to its MAF, the closer the tag SNPs (MAF) the higher their LD on a particular SNP). This is evident with the results from the *COMT* gene and *DAT1/SLC6A3* (3 tag SNPs) analysis (Figures 14 and 15). However, this relationship may have been confounded by the tendency of some captured SNPs to have high $r^2$ (close to 1) with the tag SNPs of similar MAF on the *DAT1/SLC6A*3 gene. The data for *DAT1/SLC6A3* gene using 2 tag SNPs may suggest that using more tag SNPs are more efficient to better capture LD with the remaining SNPs (Figure 16). Table 9 *COMT* gene, Table 10 *SLC6A3/DAT1* gene (3 tag SNPs) and Table 11 *SLC6A3/DAT1* gene (2 tag SNPs) show MAF and $r^2$ values for all the SNPs in each gene in each population. Our ANOVA tests (one-tailed) showed that in the *COMT* gene there is a marginally significant regression effect of $r^2$ on MAF for the CEU population ($p$=0.024), but not for the YRI population ($p$=0.08) and for CHB+JPT ($p$=0.173). Even though these tag SNPs captured the same number of SNPs in each population, the inconsistent trends are probably due in part to the different MAFs of the tag SNPs amongst the three populations. In summary, Figure 14 shows the relationship plot between MAF and $r^2$ for each population on the *COMT* gene and Table 9 shows the $r^2$ and MAF values. The MAF and $r^2$ patterns of all the populations on this gene seemed dispersed but the linear regressions showed a positive slope. For the *DAT1/SLC6A3* gene, using 3 tag SNPs

showed significant effect of MAF on $r^2$ for each of the populations (CEU, $p$=0.003; YRI, $p$=0.0005). The regression patterns showed a positive slope, but clusters of captured SNPs that have an absolute LD ($r^2$ = 1) are apparent in both populations (Figure 15, Table 10). However, there was no significant effect of $r^2$ on MAF using the 2 tag SNPs (that can be scored in all populations) analysis on this gene (CEU, $p$=0.390; YRI, $p$=0.241; CHB_JPT, $p$=0.002). The CEU and YRI regressions showed positive slope, but no significant effect of $r^2$ on MAF may be due to the clusters of SNPs with high $r^2$. The significant regression results on the CHB+JPT population (using two tag SNPs) do not support the hypothesized relationship because of the scattered distribution of the plots and the slope is negative (Figure 16, Table 11).

***Linkage Disequilibrium ($r^2$) Patterns***

We measured the LD structure for the captured SNPs using the tag SNPs for each gene. We used the highest $r^2$ value between each SNP and the tag SNPs as our measure for LD and compared them amongst populations. We used the Wilcoxon Sign Test (one-tailed test) because we were only interested in the hypothesis that the $r^2$ values are lower on the in the YRI and CHB+JPT populations when each is compared to the CEU population. Our tests showed a significant difference in $r^2$ in the *COMT* gene between CEU and YRI ($p$=0.0265); and a suggestive difference between CEU and CHB+JPT populations ($p$=0.07). Figure 11 shows the $r^2$ distribution for each population on the *COMT* gene. On the *DAT1/SLC6A3* gene using three tag SNPs, there was significant

difference between CEU vs. YRI, $p$=0.022. Figure 12 shows the $r^2$ distribution for each population. In addition, using 2 tag SNPs in this gene also showed the same trend, CEU vs. YRI ( $p$=0.018); and for CEU vs. CHB_JPT, ($p$=0.040). The key differences between the data sets for this gene (2 or 3 tag SNPs) were not merely the number of tag SNPs but also the number of SNPs shared among CEU and YRI populations only versus those shared among all three populations. Figure 13 shows the $r^2$ distribution for each population.

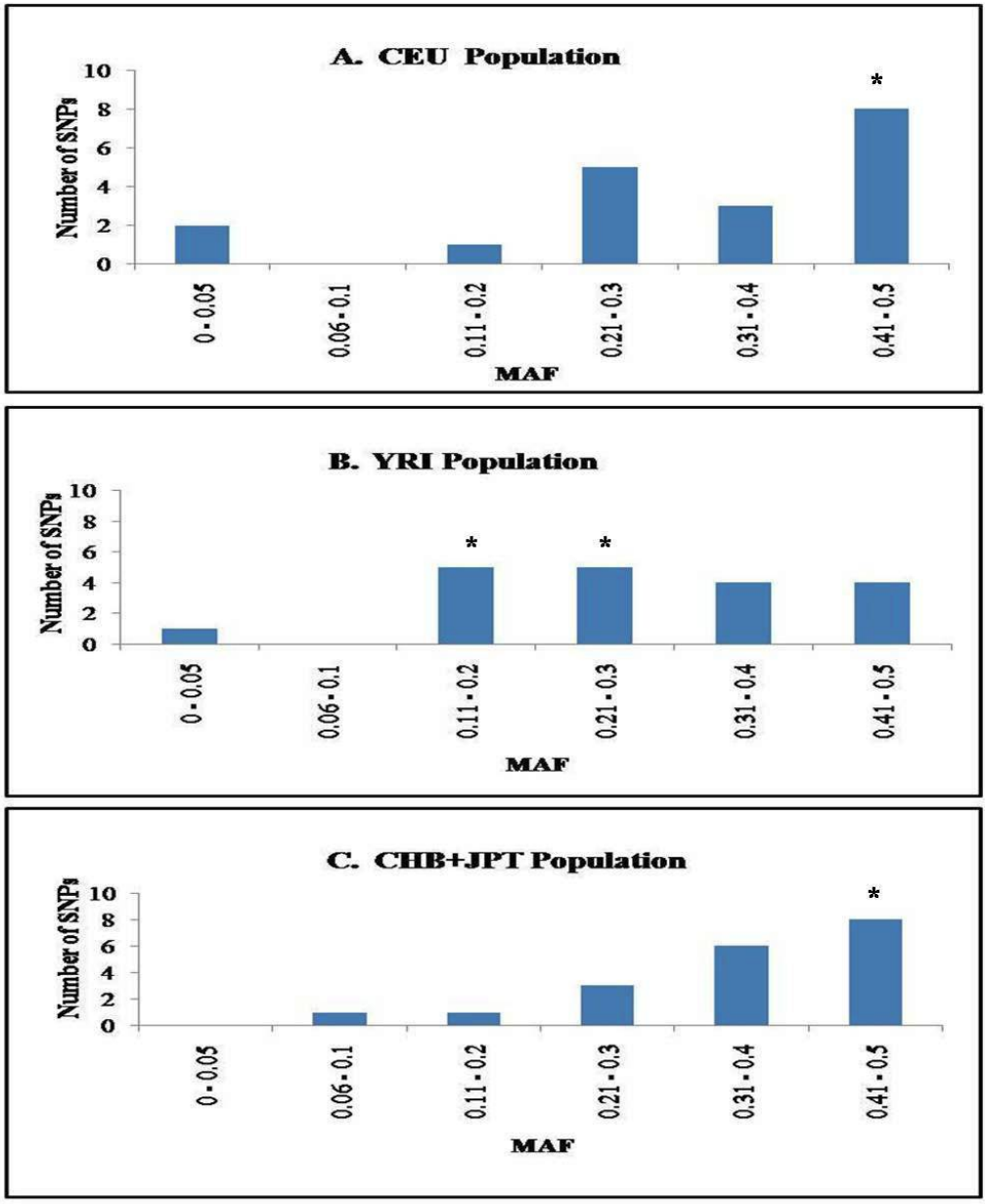


**Figure 8. Minor Allele Frequency (MAF) distributions of the SNPs on the three populations on the *COMT* gene.** MAF values are of the tag SNPs (rs737866, rs5993882, and rs4633) and the captured SNPs on this gene. The asterisks (*) indicate the most SNP that share the same MAF category on each population.

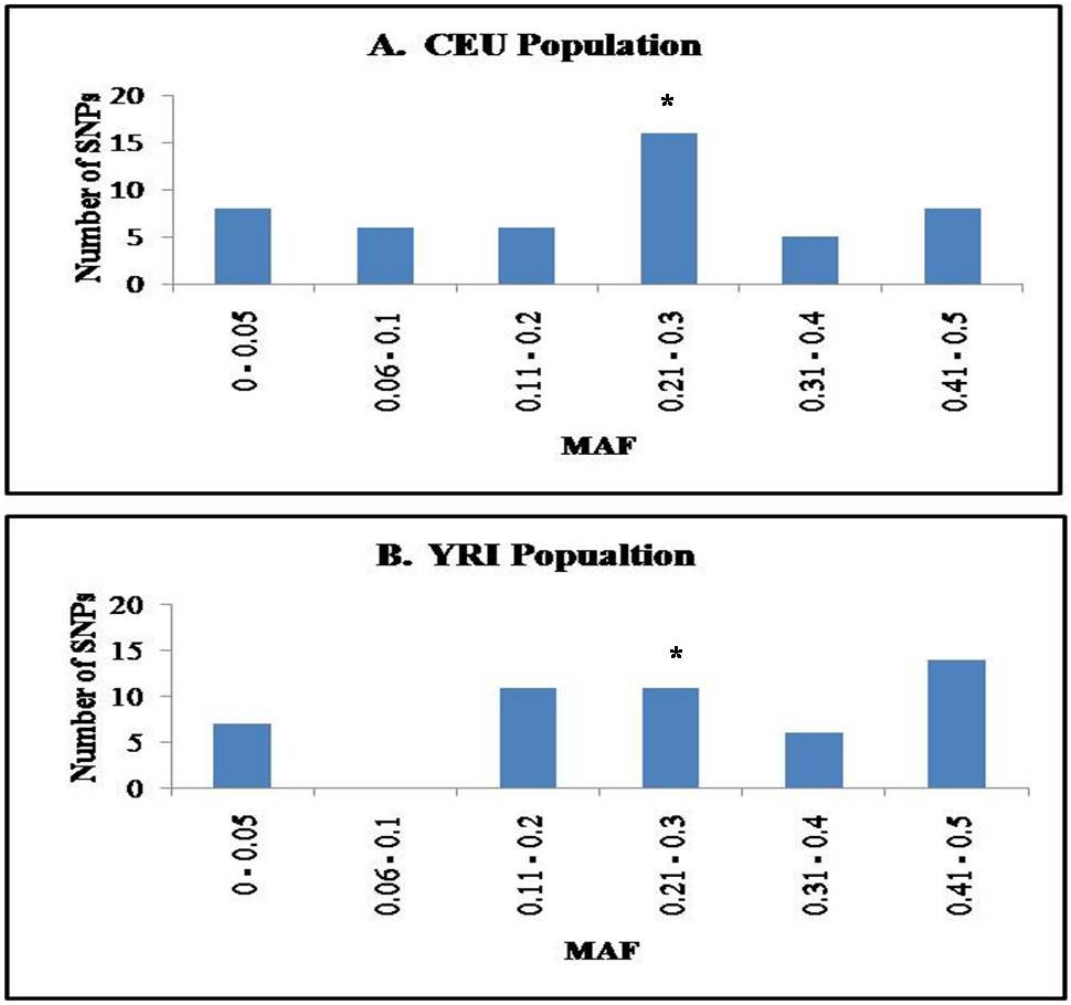


**Figure 9. Minor Allele Frequency (MAF) distributions of CEU and YRI populations on the *DAT1/SLC6A3* gene.** MAF values are of the 3 tag SNPs (rs1042098, rs10052016, and rs463379) and the captured SNPs on this gene. The asterisks (*) indicate the most SNP that share the same MAF category on each population.

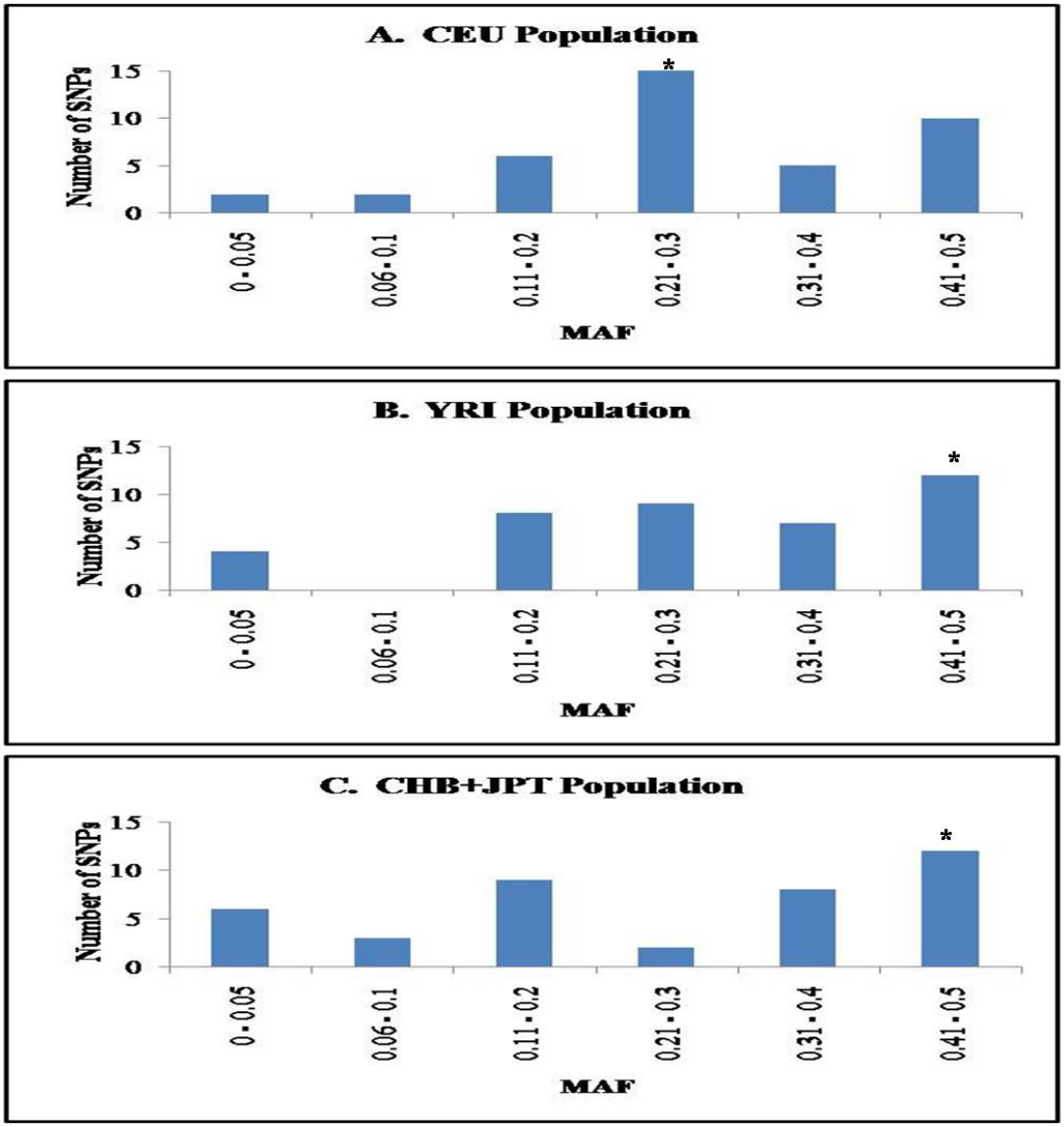


**Figure 10. Minor Allele Frequency (MAF) distributions of the three populations on the *DAT1/SCL6A3* gene.** MAF values are of the two tag SNPs (rs1042098 and rs1002016) and the captured SNPs on this gene. The asterisks (*) indicate the most SNP that share the same MAF category on each population.

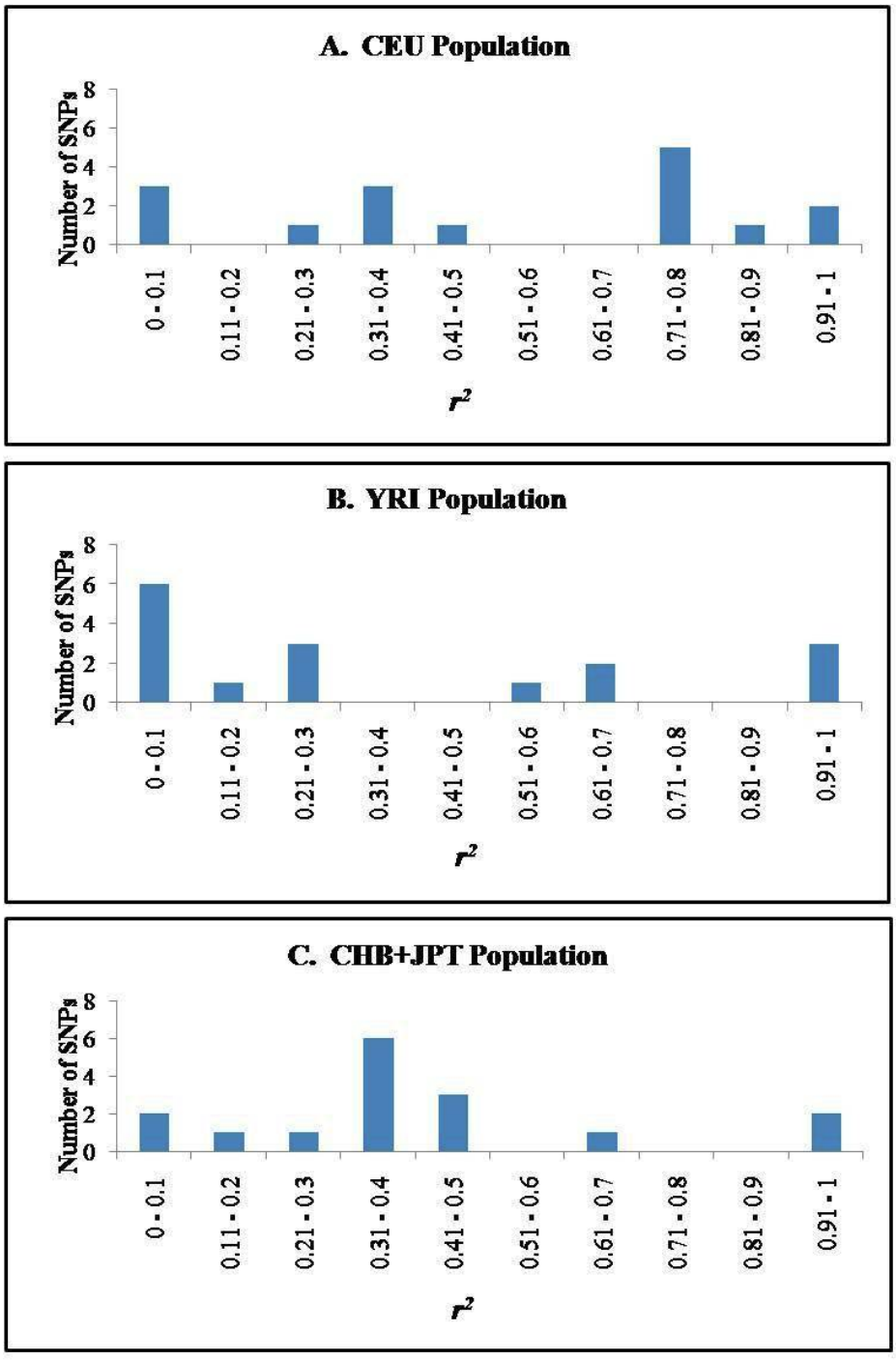


**Figure 11. Linkage disequilibrium ($r^2$) distributions of the SNPs captured by the best of the three tag SNPs on the *COMT* gene on each population.**

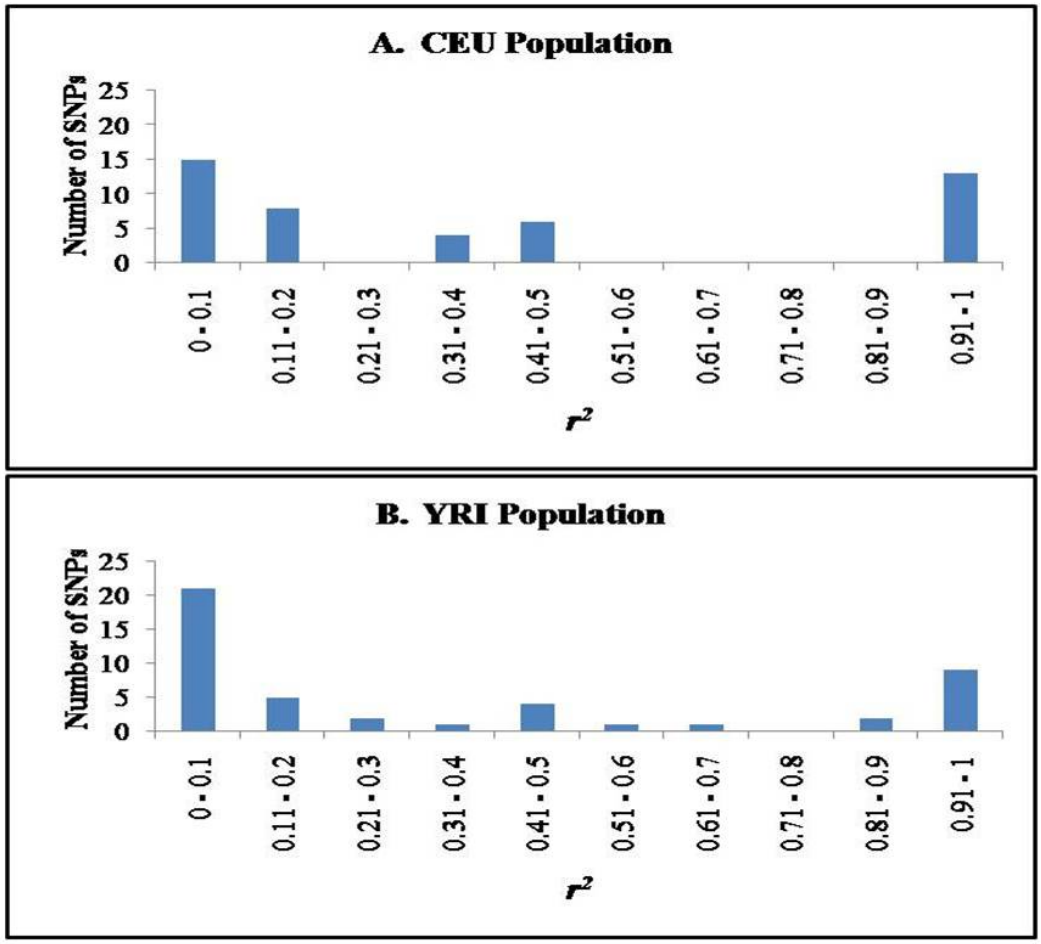


**Figure 12. Linkage disequilibrium ($r^2$) distributions of the SNPs captured by the best of the three tag SNPs on the *DAT1/SLC6A3* on the CEU and YRI populations.** SNP rs463379 is not currently genotyped in the CHB+JPT population , therefore we could not evaluate this group using the three tag SNPs.

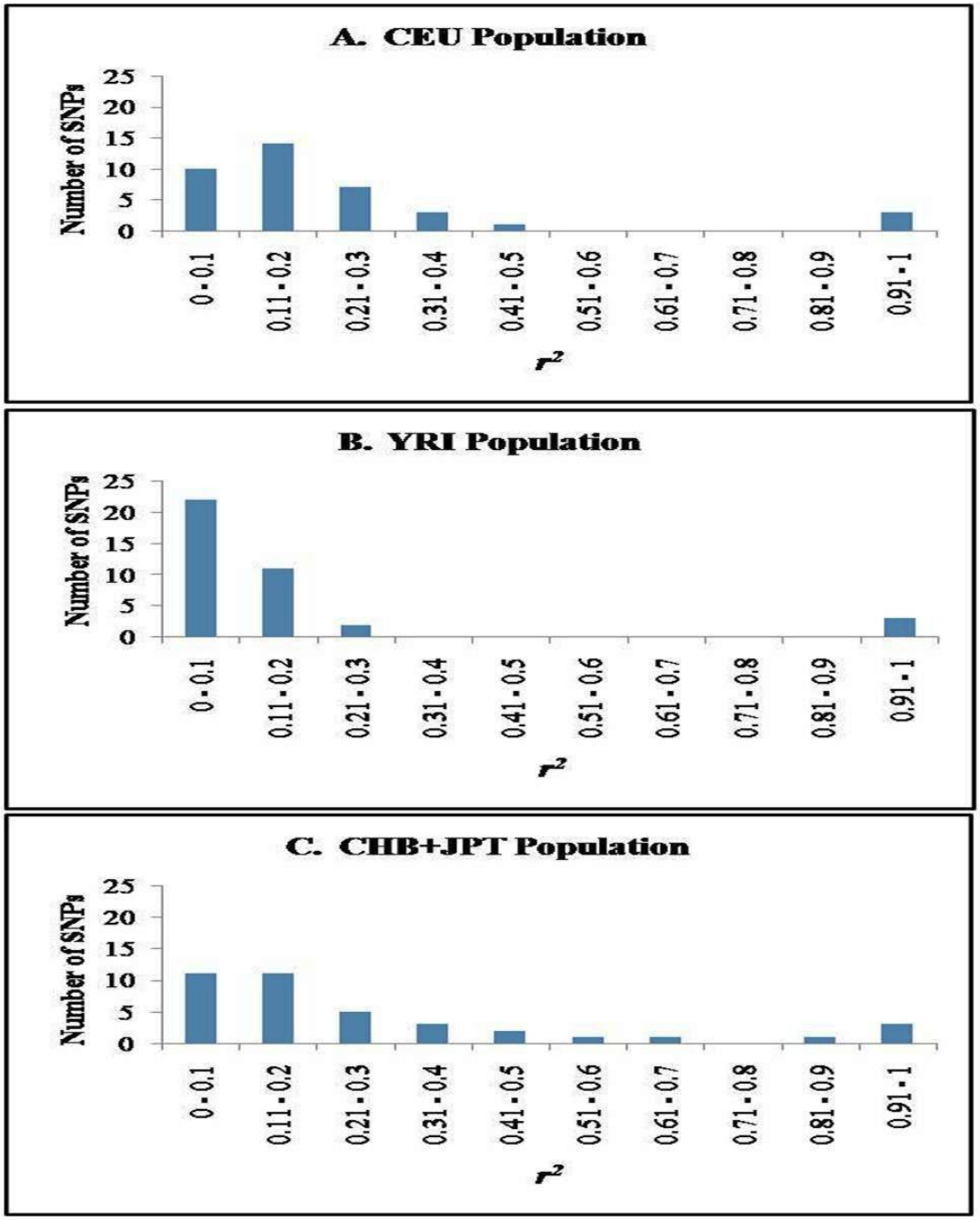


**Figure 13. Linkage disequilibrium value ($r^2$) distributions of the SNPs captured by the best of the two tag SNPs (rs1042098 and rs10052016) on the *DAT1/SLC6A3* gene on each population.**

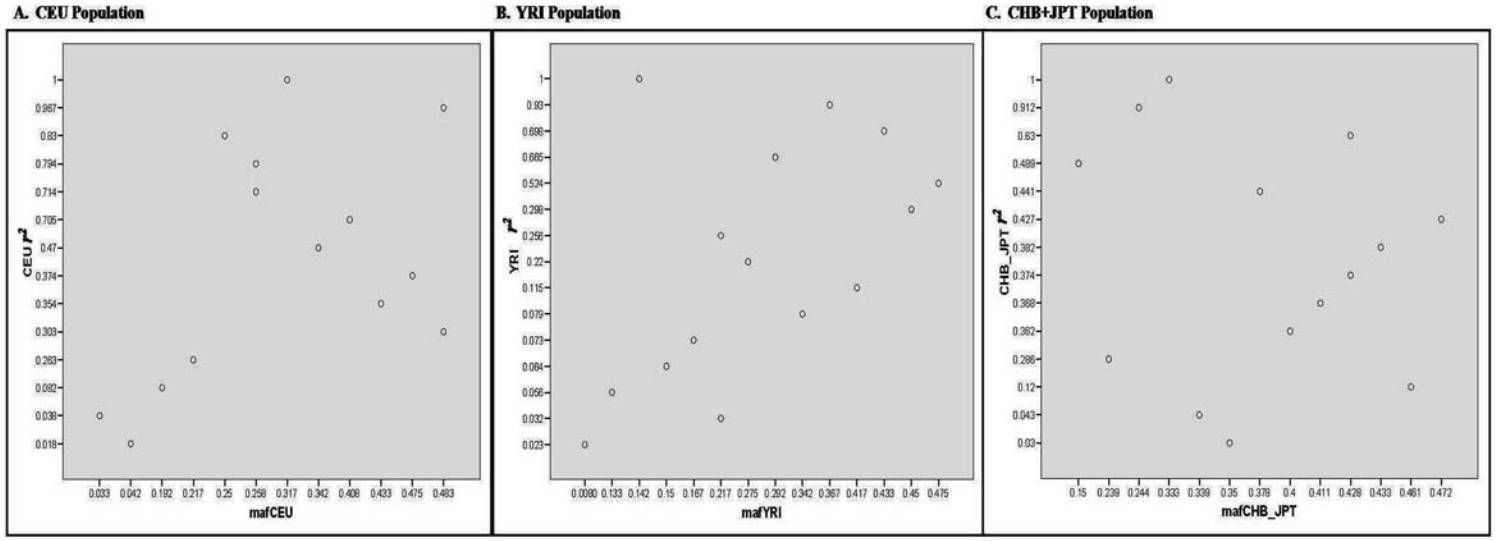


**Figure 14. $r^2$ and MAF regression plots of the three populations on the *COMT* gene.** The tag SNPs were excluded from the plots. The ANOVA (one-tail) test showed significant effect of MAF on $r^2$ in the CEU population (* $p$=0.024); but not in YRI ($p$=0.08) and the CHB+JPT ($p$=0.17) populations.

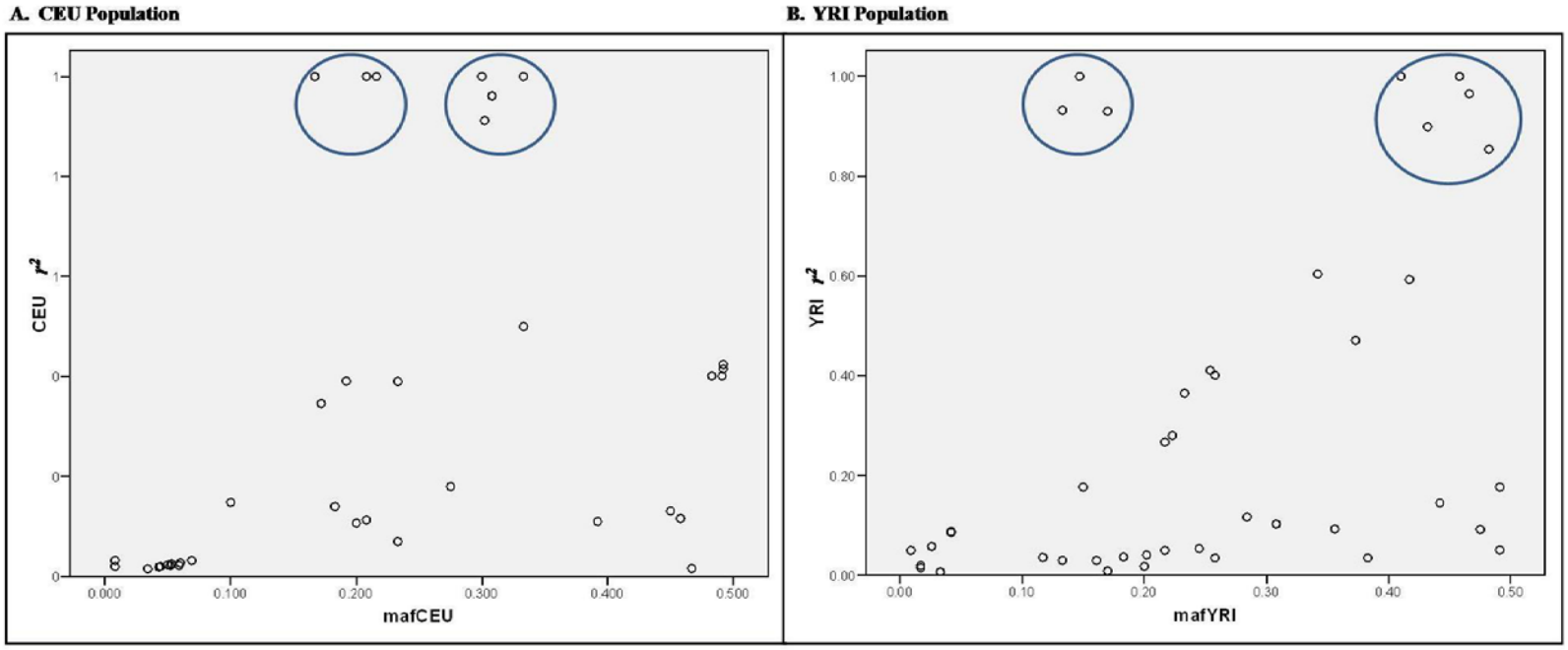


**Figure 15. $r^2$ and MAF regression plots of the two populations on the *DAT1/SLC6A3* gene using the three tag SNPs.** The tag SNPs were excluded from the plots. The ANOVA (one-tail) test showed significant effect of MAF on $r^2$ in these two populations (CEU, $p$=0.003; YRI, $p$=0.0005). The clustering of the captured SNPs due to its strong LD to one of the tag SNPs (circled) may be having a large effect on the regression.

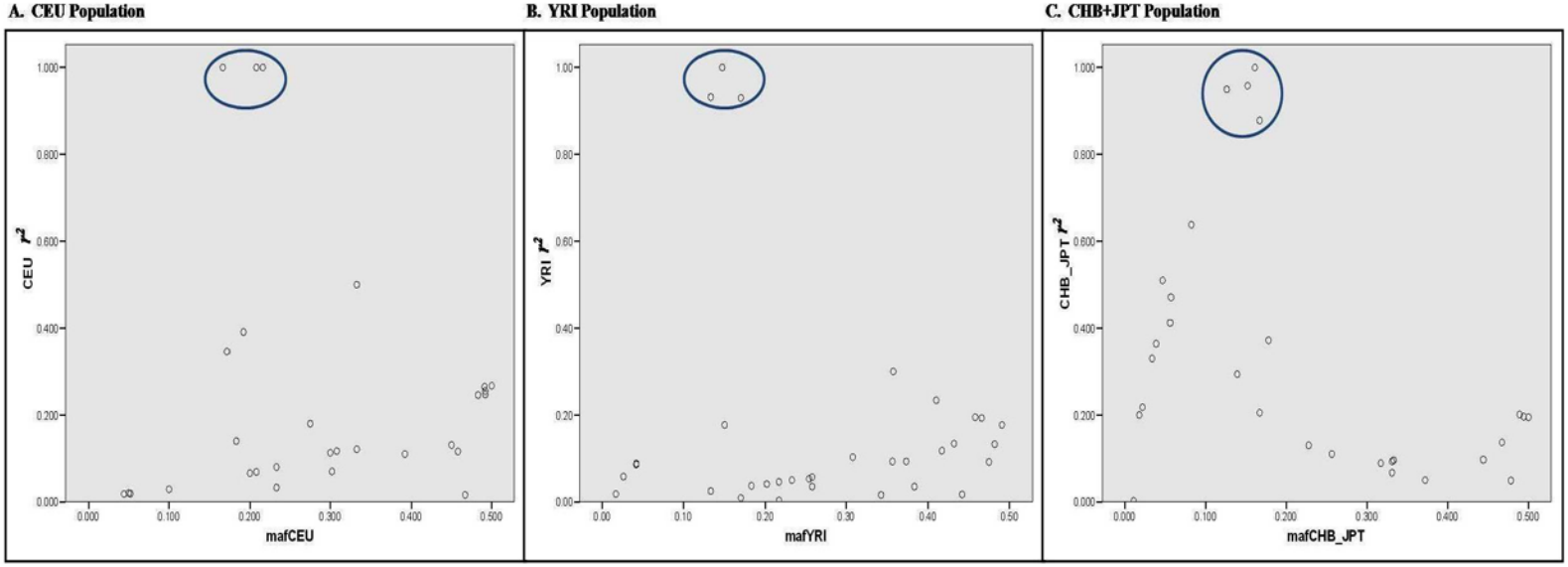


**Figure 16. $r^2$ and MAF regression plots of the three populations on the *DAT1/SLC6A3* gene using only two tag SNPs (rs1042098 and rs10052016).** The tag SNPs were excluded from the plots. The ANOVA (one-tail) test showed significant effect of MAF on $r^2$ in the CHB+JPT population only ($p$=0.002). However, the scattered distribution of the plots made the slope of the linear regression negative, making the significance irrelevant. The plots are SNPs with strong LD with the tag SNPs and may be affecting the regression analysis.

**Table 9. Tag SNPs (highlighted) and their captured SNPs on the *COMT* gene for each populations.** It includes the LD ($r^2$) and MAF values for each SNPs; SNPs are ordered according to their $r^2$ (highest to lowest).

**A. CEU Population**

| *SNP ID* | $r^2$ | *MAF* | *SNP ID* | $r^2$ | *MAF* | *SNP ID* | $r^2$ | *MAF* |
|---|---|---|---|---|---|---|---|---|
| rs737866 | 1 | 0.317 | rs5993882 | 1 | 0.217 | rs4633 | 1 | 0.475 |
| rs737865 | 1 | 0.317 | rs933271 | 0.83 | 0.25 | rs4680 | 0.967 | 0.483 |
| rs5993883 | 0.374 | 0.475 | rs174675 | 0.794 | 0.258 | rs2239393 | 0.705 | 0.408 |
| rs1544325 | 0.354 | 0.433 | rs174674 | 0.714 | 0.258 | rs4646312 | 0.705 | 0.408 |
| rs740603 | 0.303 | 0.483 | | | | rs4818 | 0.705 | 0.408 |
| rs174696 | 0.082 | 0.192 | | | | rs165774 | 0.47 | 0.342 |
| | | | | | | rs4646316 | 0.263 | 0.217 |
| | | | | | | rs174697 | 0.038 | 0.033 |
| | | | | | | rs174699 | 0.018 | 0.042 |

**B. YRI Population**

| *SNP ID* | $r^2$ | *MAF* | *SNP ID* | $r^2$ | *MAF* | *SNP ID* | $r^2$ | *MAF* |
|---|---|---|---|---|---|---|---|---|
| rs737866 | 1 | 0.142 | rs5993882 | 1 | 0.367 | rs4633 | 1 | 0.267 |
| rs737865 | 1 | 0.142 | rs174674 | 0.93 | 0.367 | rs4680 | 0.665 | 0.292 |
| | | | rs174675 | 0.93 | 0.367 | rs2239393 | 0.298 | 0.45 |
| | | | rs933271 | 0.698 | 0.433 | rs165774 | 0.256 | 0.217 |
| | | | rs5993883 | 0.524 | 0.475 | rs174696 | 0.079 | 0.342 |
| | | | rs1544325 | 0.22 | 0.275 | rs4818 | 0.073 | 0.167 |
| | | | rs740603 | 0.115 | 0.417 | rs4646316 | 0.064 | 0.15 |
| | | | | | | rs4646312 | 0.056 | 0.133 |
| | | | | | | rs174697 | 0.032 | 0.217 |
| | | | | | | rs174699 | 0.023 | 0.008 |

**C. CHB+JPT Population**

| *SNP ID* | $r^2$ | *MAF* | *SNP ID* | $r^2$ | *MAF* | *SNP ID* | $r^2$ | *MAF* |
|---|---|---|---|---|---|---|---|---|
| rs737866 | 1 | 0.333 | rs5993882 | 1 | 0.1 | rs4633 | 1 | 0.228 |
| rs737865 | 1 | 0.333 | | | | rs4680 | 0.912 | 0.244 |
| rs5993883 | 0.63 | 0.428 | | | | rs165774 | 0.489 | 0.15 |
| rs4646316 | 0.441 | 0.378 | | | | rs1544325 | 0.286 | 0.239 |
| rs740603 | 0.427 | 0.472 | | | | | | |
| rs933271 | 0.382 | 0.433 | | | | | | |
| rs174674 | 0.374 | 0.428 | | | | | | |
| rs174675 | 0.374 | 0.428 | | | | | | |
| rs2239393 | 0.368 | 0.411 | | | | | | |
| rs4646312 | 0.368 | 0.411 | | | | | | |
| rs4818 | 0.362 | 0.4 | | | | | | |
| rs174696 | 0.12 | 0.461 | | | | | | |
| rs174697 | 0.043 | 0.339 | | | | | | |
| rs174699 | 0.03 | 0.35 | | | | | | |

**Table 10. Tag SNPs (highlighted) and their captured SNPs on the *DAT1/SLC6A3* gene for two populations.** It includes the LD ($r^2$) and MAF values for each SNPs; SNPs are ordered according to their $r^2$ (highest to lowest).

**A. CEU Population**

| *SNP ID* | $r^2$ | *MAF* |
|---|---|---|
| rs1042098 | 1 | 0.267 |
| rs11133767 | 0.5 | 0.333 |
| rs8179029 | 0.391 | 0.192 |
| rs3776511 | 0.346 | 0.172 |
| rs3776512 | 0.346 | 0.172 |
| rs6347 | 0.18 | 0.275 |
| rs40184 | 0.131 | 0.45 |
| rs11564767 | 0.022 | 0.059 |
| rs2937639 | 0.016 | 0.467 |

| *SNP ID* | $r^2$ | *MAF* |
|---|---|---|
| rs10052016 | 1 | 0.208 |
| rs10040882 | 1 | 0.216 |
| rs10053602 | 1 | 0.208 |
| rs2042449 | 1 | 0.167 |
| rs27072 | 0.14 | 0.183 |
| rs27048 | 0.116 | 0.458 |
| rs2617605 | 0.11 | 0.392 |
| rs2963253 | 0.032 | 0.008 |
| rs410481 | 0.032 | 0.008 |

| *SNP ID* | $r^2$ | *MAF* |
|---|---|---|
| rs463379 | 1 | 0.3 |
| rs393795 | 1 | 0.3 |
| rs409588 | 1 | 0.3 |
| rs456082 | 1 | 0.3 |
| rs458860 | 1 | 0.3 |
| rs460000 | 1 | 0.3 |
| rs460700 | 1 | 0.3 |
| rs250682 | 1 | 0.3 |
| rs464528 | 1 | 0.333 |
| rs456774 | 0.961 | 0.308 |
| rs464061 | 0.912 | 0.302 |
| rs250686 | 0.424 | 0.492 |
| rs250681 | 0.415 | 0.492 |
| rs464049 | 0.415 | 0.492 |
| rs37020 | 0.401 | 0.491 |
| rs420422 | 0.401 | 0.483 |
| rs37022 | 0.39 | 0.233 |
| rs6350 | 0.148 | 0.1 |
| rs403636 | 0.113 | 0.208 |
| rs40358 | 0.107 | 0.2 |
| rs466630 | 0.07 | 0.233 |
| rs11564771 | 0.032 | 0.069 |
| rs11564772 | 0.027 | 0.06 |
| rs11564764 | 0.025 | 0.053 |
| rs6869645 | 0.023 | 0.05 |
| rs6876225 | 0.022 | 0.052 |
| rs10072058 | 0.02 | 0.008 |
| rs11564773 | 0.02 | 0.044 |
| rs6348 | 0.02 | 0.008 |
| rs11133768 | 0.019 | 0.043 |
| rs11564766 | 0.015 | 0.034 |

**B. YRI Population**

| *SNP ID* | $r^2$ | *MAF* |
|---|---|---|
| rs1042098 | 1 | 0.345 |
| rs11133767 | 0.177 | 0.491 |
| rs40184 | 0.092 | 0.475 |
| rs3776511 | 0.088 | 0.042 |
| rs8179029 | 0.086 | 0.042 |
| rs3776512 | 0.058 | 0.026 |
| rs11564767 | 0.051 | 0.491 |
| rs27048 | 0.035 | 0.258 |
| rs6347 | 0.035 | 0.383 |
| rs11564773 | 0.009 | 0.17 |

| *SNP ID* | $r^2$ | *MAF* |
|---|---|---|
| rs10052016 | 1 | 0.142 |
| rs10053602 | 1 | 0.147 |
| rs2042449 | 0.932 | 0.133 |
| rs10040882 | 0.93 | 0.17 |
| rs2617605 | 0.177 | 0.15 |
| rs2937639 | 0.103 | 0.308 |
| rs420422 | 0.093 | 0.356 |
| rs11133768 | 0.054 | 0.245 |
| rs410481 | 0.05 | 0.009 |
| rs6876225 | 0.041 | 0.202 |
| rs6869645 | 0.037 | 0.183 |
| rs11564772 | 0.03 | 0.161 |

| *SNP ID* | $r^2$ | *MAF* |
|---|---|---|
| rs463379 | 1 | 0.466 |
| rs458860 | 1 | 0.458 |
| rs460000 | 1 | 0.458 |
| rs456082 | 1 | 0.458 |
| rs464528 | 1 | 0.41 |
| rs409588 | 0.965 | 0.466 |
| rs456774 | 0.965 | 0.466 |
| rs250682 | 0.899 | 0.432 |
| rs464061 | 0.854 | 0.482 |
| rs250686 | 0.604 | 0.342 |
| rs393795 | 0.593 | 0.417 |
| rs460700 | 0.471 | 0.373 |
| rs37020 | 0.411 | 0.254 |
| rs250681 | 0.401 | 0.258 |
| rs464049 | 0.401 | 0.258 |
| rs40358 | 0.365 | 0.233 |
| rs6348 | 0.28 | 0.223 |
| rs403636 | 0.267 | 0.217 |
| rs37022 | 0.145 | 0.442 |
| rs10072058 | 0.117 | 0.284 |
| rs466630 | 0.05 | 0.217 |
| rs2963253 | 0.036 | 0.117 |
| rs27072 | 0.03 | 0.133 |
| rs6350 | 0.02 | 0.017 |
| rs11564771 | 0.018 | 0.2 |
| rs11564766 | 0.015 | 0.017 |
| rs11564764 | 0.007 | 0.033 |

**Table 11. Tag SNPs (highlighted) and their captured SNPs on the *DAT1/SLC6A3* gene for three populations.** It includes the LD ($r^2$) and MAF values for each SNPs; SNPs are ordered according to their $r^2$ (highest to lowest). This is only using two tag SNPs on this gene (rs1042098 and rs10052016).

**A. CEU Population**

| SNP ID | $r^2$ | MAF |
|---|---|---|
| rs1042098 | 1 | 0.267 |
| rs11133767 | 0.5 | 0.333 |
| rs8179029 | 0.391 | 0.192 |
| rs3776511 | 0.346 | 0.172 |
| rs3776512 | 0.346 | 0.172 |
| rs6347 | 0.18 | 0.275 |
| rs40184 | 0.131 | 0.45 |
| rs466630 | 0.033 | 0.233 |
| rs6869645 | 0.02 | 0.05 |
| rs6876225 | 0.019 | 0.052 |
| rs11564773 | 0.018 | 0.044 |
| rs2937639 | 0.016 | 0.467 |

| SNP ID | $r^2$ | MAF |
|---|---|---|
| rs10052016 | 1 | 0.208 |
| rs2042449 | 1 | 0.167 |
| rs10040882 | 1 | 0.216 |
| rs10053602 | 1 | 0.208 |
| rs2975292 | 0.267 | 0.5 |
| rs37020 | 0.265 | 0.491 |
| rs464049 | 0.255 | 0.492 |
| rs250681 | 0.255 | 0.492 |
| rs458334 | 0.255 | 0.492 |
| rs250686 | 0.247 | 0.492 |
| rs420422 | 0.246 | 0.483 |
| rs27072 | 0.14 | 0.183 |
| rs464528 | 0.121 | 0.333 |
| rs456774 | 0.117 | 0.308 |
| rs27048 | 0.116 | 0.458 |
| rs250682 | 0.113 | 0.3 |
| rs393795 | 0.113 | 0.3 |
| rs460700 | 0.113 | 0.3 |
| rs456082 | 0.113 | 0.3 |
| rs409588 | 0.113 | 0.3 |
| rs458860 | 0.113 | 0.3 |
| rs460000 | 0.113 | 0.3 |
| rs2617605 | 0.11 | 0.392 |
| rs37022 | 0.08 | 0.233 |
| rs464061 | 0.07 | 0.302 |
| rs403636 | 0.069 | 0.208 |
| rs40358 | 0.066 | 0.2 |
| rs6350 | 0.029 | 0.1 |

**B. YRI Population**

| SNP ID | $r^2$ | MAF |
|---|---|---|
| rs1042098 | 1 | 0.345 |
| rs11133767 | 0.177 | 0.491 |
| rs40184 | 0.092 | 0.475 |
| rs3776511 | 0.088 | 0.042 |
| rs8179029 | 0.086 | 0.042 |
| rs3776512 | 0.058 | 0.026 |
| rs6347 | 0.035 | 0.383 |
| rs27048 | 0.035 | 0.258 |
| rs37022 | 0.017 | 0.442 |
| rs11564773 | 0.009 | 0.17 |

| SNP ID | $r^2$ | MAF |
|---|---|---|
| rs10052016 | 1 | 0.142 |
| rs10053602 | 1 | 0.147 |
| rs2042449 | 0.932 | 0.133 |
| rs10040882 | 0.93 | 0.17 |
| rs2975292 | 0.3 | 0.357 |
| rs464528 | 0.234 | 0.41 |
| rs456082 | 0.195 | 0.458 |
| rs458860 | 0.195 | 0.458 |
| rs460000 | 0.195 | 0.458 |
| rs409588 | 0.193 | 0.466 |
| rs456774 | 0.193 | 0.466 |
| rs2617605 | 0.177 | 0.15 |
| rs250682 | 0.134 | 0.432 |
| rs464061 | 0.133 | 0.482 |
| rs393795 | 0.118 | 0.417 |
| rs2937639 | 0.103 | 0.308 |
| rs420422 | 0.093 | 0.356 |
| rs460700 | 0.093 | 0.373 |
| rs250681 | 0.057 | 0.258 |
| rs458334 | 0.057 | 0.258 |
| rs464049 | 0.057 | 0.258 |
| rs37020 | 0.053 | 0.254 |
| rs40358 | 0.05 | 0.233 |
| rs403636 | 0.046 | 0.217 |
| rs6876225 | 0.041 | 0.202 |
| rs6869645 | 0.037 | 0.183 |
| rs27072 | 0.025 | 0.133 |
| rs6350 | 0.018 | 0.017 |
| rs250686 | 0.016 | 0.342 |
| rs466630 | 0.003 | 0.217 |

**C. CHB+JPT Population**

| SNP ID | $r^2$ | MAF |
|---|---|---|
| rs1042098 | 1 | 0.101 |
| rs6347 | 0.638 | 0.082 |
| rs3776512 | 0.51 | 0.047 |
| rs3776511 | 0.471 | 0.057 |
| rs8179029 | 0.412 | 0.056 |
| rs6869645 | 0.364 | 0.039 |
| rs11564767 | 0.33 | 0.034 |
| rs6876225 | 0.218 | 0.022 |
| rs11564773 | 0.2 | 0.018 |

| SNP ID | $r^2$ | MAF |
|---|---|---|
| rs10052016 | 1 | 0.161 |
| rs10053602 | 1 | 0.161 |
| rs2975292 | 0.958 | 0.152 |
| rs2042449 | 0.95 | 0.126 |
| rs10040882 | 0.878 | 0.167 |
| rs2937639 | 0.372 | 0.178 |
| rs2617605 | 0.294 | 0.139 |
| rs27048 | 0.205 | 0.167 |
| rs456774 | 0.201 | 0.489 |
| rs464528 | 0.201 | 0.489 |
| rs250682 | 0.196 | 0.494 |
| rs393795 | 0.196 | 0.494 |
| rs409588 | 0.196 | 0.494 |
| rs456082 | 0.196 | 0.494 |
| rs458860 | 0.196 | 0.494 |
| rs460700 | 0.196 | 0.494 |
| rs464061 | 0.195 | 0.5 |
| rs460000 | 0.137 | 0.467 |
| rs40184 | 0.13 | 0.228 |
| rs27072 | 0.11 | 0.256 |
| rs37022 | 0.097 | 0.444 |
| rs250681 | 0.096 | 0.333 |
| rs250686 | 0.096 | 0.333 |
| rs458334 | 0.096 | 0.333 |
| rs464049 | 0.096 | 0.333 |
| rs37020 | 0.093 | 0.331 |
| rs403636 | 0.089 | 0.317 |
| rs420422 | 0.067 | 0.331 |
| rs40358 | 0.05 | 0.372 |
| rs466630 | 0.049 | 0.478 |
| rs6350 | 0.002 | 0.011 |

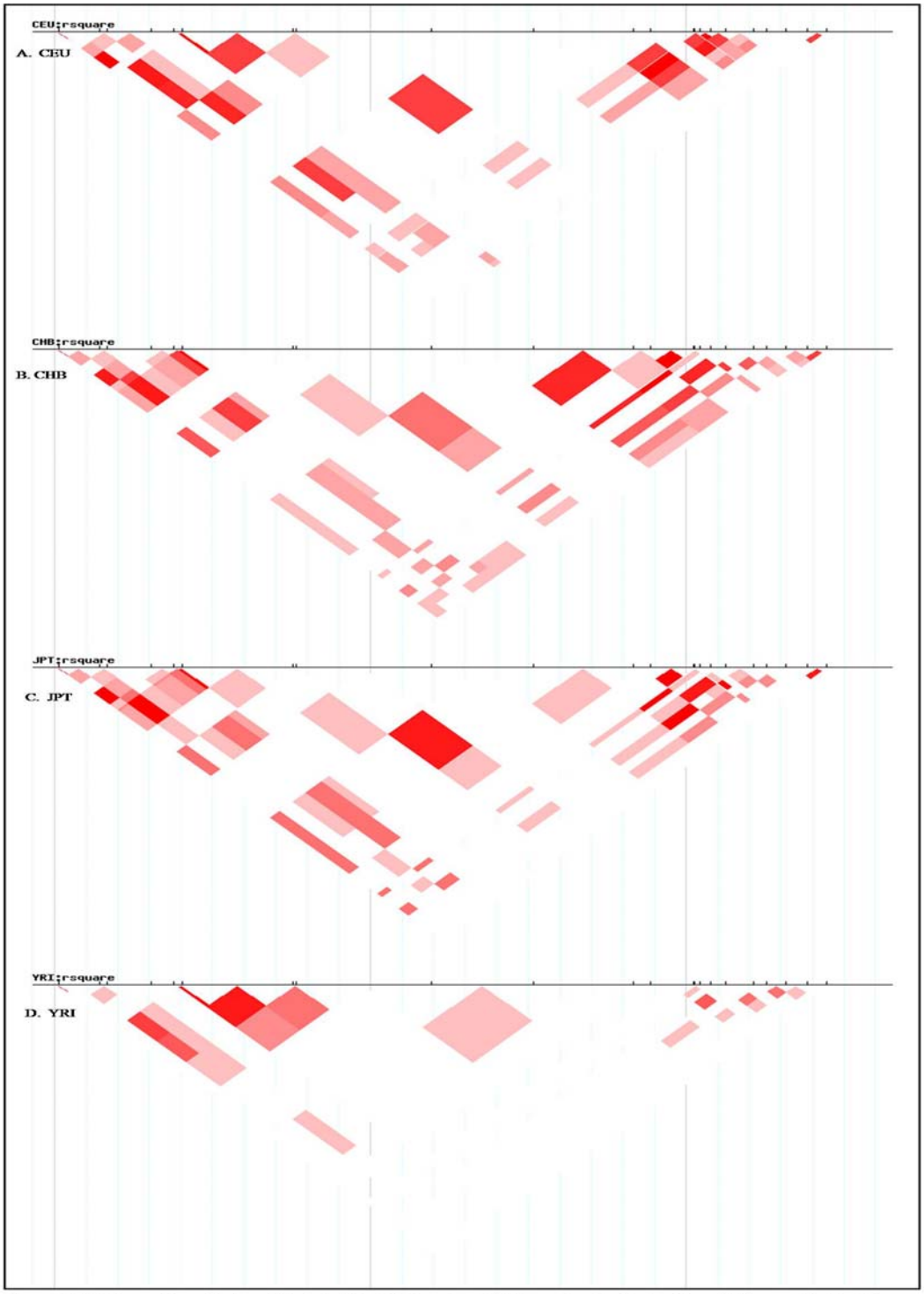


**Figure 17. Linkage disequilibrium ($r^2$) plots of the SNPs on the *COMT* gene for the different populations on the HapMap.** This is using the Genome Browser Phase 1 and 2 full data sets on the database. Due to the similarity of the CHB and JPT populations, their data were combined. The red regions represents strong LD, $r^2$=1 and intermediate shades of red represents intermediate-strength LD.

# CHAPTER V

# DISCUSSION

## *Utilizing the HapMap Database*

We evaluated the practical usefulness of the HapMap database by obtaining genotyping data on the three candidate genes (*COMT*, *DAT1/SLC6A3 and DRD4*) using the CEU population as reference. Due to costs and limitations of our genotyping method, we were unable to select 3 SNPs per LD block as envisioned for the use of the HapMap data. We selected tag SNPs (2-3 per gene) that have fairly good minor allele frequencies (0.2 -0.47, Tables 5, 9 and 10) and have polymorphism patterns distinct from each other. The database is user-friendly and provides detailed information on MAF and LD patterns of SNPs that facilitates in the design for genetic association studies in identifying common genetic variants that may contribute to complex common disorders. Currently, the International HapMap consortium provides a catalog of common genetic variations (SNPs) in four different populations (CEU, YRI, CHB+JPT) (The International HapMap Consortium, 2005; accessed October, 2008). The analysis of common SNPs is currently the approach to study the genetic bases of complex diseases or traits. This approach may allow us to identify risk factors and find genetic markers (on specific gene products) that affect the biological presentation of the diseases/traits of interest (Gonzalez-Niera, et al., 2006). However, the transferability of common SNPs (tag SNPs) amongst the human

population is still unclear. The database is continually being updated and still is incomplete as we have encountered with the *DRD4* gene in which there were only two genotyped SNPs on this population and were not informative (accessed October 2008). The database once complete or updated will provide preliminary data of the LD patterns of SNPs that may help an investigator tailor markers for its particular sample population for an effective association study design.

### *SNP Genotyping Assays*

One of the essential components of any genetic association studies is generating genotypes for the genetic markers, particularly SNPs, to test for association with the common trait of interest. We evaluated two SNP genotyping methodologies namely, PCR-RLFP and SNuPe. PCR is the first step for any genotyping methods (Chen and Sullivan, 2003). The PCR assay can be affected by several factors such as primer designs, thermal cycling (Tm of the primers) and reagents on the PCR. We have used the Primer 3 program (available online) to design primers for the two SNP genotyping methods (primer designs for the SNuPe by Sara Nelson) and selected primer pairs that are of similar melting temperatures. A touch-down PCR (50 Enhanced, see Appendices A-C) thermal cycling program was used to amplify the products for the SNuPe method and then visualize with ethidium bromide treated agarose gels. Our results showed that $MgCl_2$ is an important factor in PCR assays. It is suggested that lower concentration of $Mg^{2+}$ were desirable when the fidelity of DNA synthesis is essential. There is a

reciprocal relationship of $Mg^{2+}$ ions and dNTPs, therefore a change in the ion concentration usually requires similar modification to the amount of dNTPs (Vance and Othmane, 1998). The other reagents were added accordingly by our laboratory protocols and other references (Vance and Othmane, 1998; Starchan and Read, 2004). Optimization of PCR assays by considering these factors will help in the downstream steps for an accurate SNP calls on any SNP genotyping assay.

PCR-RFLP SNP genotyping assay is based on a single base pair change that creates a cutting site for a specific restriction enzyme (Vance and Othmane, 1998). Our PCR products were digested with their specific enzymes and ran on an ethidium bromide treated agarose gels (Appendix D for specific protocols). Although this method of genotyping is straightforward and does not require the purchase of an expensive equipment to detect the SNP alleles, it has its own limitations. We had encountered difficulties in interpreting the banding patterns of our PCR-RFLP gels due to the incomplete digestion by the restriction enzymes, which thus may give a false genotype read. Due to the known inheritance pattern of our family trios and comparing our SNP genotype data on the HapMap database we were able to confirm the genotypes of our individual samples. We were also limited with the tag SNPs selection because there were no restriction enzymes available for several of the SNPs on each gene and there were no enzymes available for the SNPs that were genotyped on the HapMap database for *DRD4* gene. For a larger sample study a high-throughput efficient genotyping method such as SNuPe assay is a better approach as it would allow us to type more SNPs and thus help in

selecting a denser 'tag' SNPs to better capture LD patterns on these candidate genes. However, optimization of methodologies is critical in order to generate reliable genotypes. For our single base extension assay (SNuPe) the SNP is identified by using a third oligonucleotide primer to the 3' end of the SNP of interest (added to the PCR product). This primer recognizes the SNP of interest and labeled with fluorochromes for detection. SNPs were then scored by capillary gel electrophoresis on a MegaBace 500 sequencer machine (Protocols followed in suggestion by Amersham Biosciences). Our SNuPe genotype results were not robust and of low quality as probably resulted from the low recovery of products during the final spin of the clean-up step (see Appendix); and even with alternative clean-up method did not improve the quality of the genotypes. Although this method is more efficient than PCR-RFLP, it is more expensive and like with any emerging new technologies sometimes available technical support is scarce. Sobrino et al reviewed other SNP genotyping discrimination assays (such as allele specific hybridization, primer extension methods) and detection methodologies (fluorescence arrays, mass spectrometry and luminescence) that could be used in addition to the two that were explored in this study (2005). The Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight (MALDI-TOF) mass spectrometry has become one of the better tools for SNP genotyping (Tost and Gut, 2005; Pusch et al., 2002; Bray et al., 2001). Mass spectrometry detection of single-base extension, such as Sequenom MassARRAY platform (Gabriel et al., 2009) could possibly be an alternative for the sequencer-based approaches such as SNuPe for lower-throughput assays with fewer loci. In addition, High-throughput genotyping platforms based on Affymetrix GeneChip or

Illumina's Beadchip are becoming the standard for high-throughput assays of genome-wide marker sets (Ragoussis, 2009). The optimization and the advancement of methodologies are essential considerations when selecting a genotyping method for a genetic association study. Efficient genotyping is as important as the marker selection because unreliable genotype calls may lead to spurious results in association studies.

### *Linkage Disequilibrium Patterns*

In part to test whether sparse sampling of SNPs contributes to inconsistent detection of potentially causative genetic variants among human populations (CEU, YRI and CHB+JPT), we used three or fewer SNPs per gene to evaluate the LD ($r^2$) patterns on the two genes. Using ≤ 3 SNPs per gene did not permit this evaluation, but our reasons also included costs and technical limitations (i.e. the unavailability of restriction enzymes for recognizing many sites), as mentioned above. We found that some SNPs show poor association with the assayed SNPs (Tables 9, 10 and 11) on the both of the genes, but more predominantly observed in the *SLC6A3/DAT1* gene (either using 2 or 3 tag SNPs, Table 11). In the *COMT* gene, the majorities of the captured SNPs had the strongest LD (r2) with the tag SNPs (0.71-0.8) in the CEU population and lower in the YRI (0-0.1) and the CHB+JPT (0.31-0.4) populations. In the *DAT1/SLC6A3* gene (using three tag SNPs), the majority of the SNPs (15) had a low $r^2$ (0-0.1), but there were also as many SNPs (14) that showed strong LD (0.91-1) with the tag SNPs in the CEU population; the YRI population captured SNPs showed significantly lower $r^2$ values (0-0.1). The use of only

two tag SNPs in the *DAT1/SCL6A3* gene showed an average low $r^2$ of the captured SNPs on the three populations (CEU=0.11-0.1; YRI=0-0.1; CHB+JPT=0-0.02). Furthermore, we observed that $r^2$ can be close to 1 only when MAFs are similar, though SNPs with similar MAFs could have very low $r^2$. Our results could suggest that sparse sampling of tag SNPs in these genes (*COMT* and *SLC6A3*) does not capture the other SNPs well and that a denser set of tag SNPs is needed.

The common-disease common-variant (CDCV) hypothesis suggests that causative SNPs are common. Thus, under this hypothesis, common tag SNPs should capture the disease risk associated with causative SNPs most efficiently. We hypothesized that potentially causative uncommon SNPs are captured more poorly by common tag SNPs, thus making the efficacy of association studies as usually designed dependent on the validity of the CDCV hypothesis. We evaluated this (using our sparse tag SNP selection) by doing a regression of SNP-tag set $r^2$ values on MAF for each gene, with the expectation that the regression coefficient will be significantly positive. On the *COMT* gene, the three populations showed similar trend of regression (with positive slopes), but only the CEU population showed significance ($p$=0.048). This is partly because we selected the tag SNPs accordingly to the MAF on the CEU and therefore LDs were better captured on this population (Figure 14). Consequently, on the *SLC6A3/DAT1* gene (using 3 tag SNPs), the CEU and YRI population showed similar trends of regression with positive slopes and of significance ($p$=0.006, $p$=0.001 respectively; Figure 15). On the other hand, no significant changes on the regression using 2 tag SNPs on the CEU and

YRI populations in this gene. The CHB+JPT population showed significance using 2 tag SNPs but with a negative slope (Figure 16). The regression results suggest that there is a trend toward poorer capturing of rare SNPs, which would mean lesser detection of AD/HD associations if rare SNPs were causative. The non-significant regression results on the SLC6A3/DAT1 (2 tag SNPs) gene may be due to the relatively low frequency of the two tag SNPs especially in the CHB+JPT populations and causing the slope to be negative. Our results could also suggest that 3 tag SNPs is better than 2 tag SNPs in capturing LD on these genes particularly the *SLC6A3/DAT1* gene. However, our method does not represent a true haplotype analysis because we selected the tag SNP having the greatest LD with each captured SNP on these genes and used one reference population to select our tag SNPs. Thus, in support with our results on the previous paragraph, a denser SNP and/or haplotype-based tag SNPs is suggested to further test our hypothesis.

The LD ($r^2$) comparison of each identified SNPs between the tag SNP sets in the CEU population with that in the other two populations predicted that the tag SNPs selected are less useful in the other populations. Our results showed on average that the LD ($r^2$) values between the tag SNPs and captured SNPs for *COMT* and *SLC6A3/DAT1* genes were lower in the YRI and CHB+JPT populations compared to the CEU population (Tables 9, 10, and 11) and their LD distributions amongst populations were variable (Figures 11,12, and 13). Overall, the Wilcoxon Sign Test (one-tailed) showed significant differences in the mean $r^2$ values between CEU and the other two populations (lower) on these two genes. We expected these results because the tag SNPs were chosen according

to the MAF patterns on the CEU population to capture LD on these genes; and the tag SNPs may not do as well in the other populations for which they were not evaluated ahead of time. The significant reductions of the $r^2$ in the other two populations relative to the CEU would support that associations on these genes varies on between populations and suggest that prior assessment of tag SNPs using the HapMap is a essential step in the design of genetic association studies. This will help tailor the tag SNP for a specific population sample as LD patterns may vary among the human populations.

It is estimated that there are about 7 million of SNPs with MAF of at least 5% across the entire human population (Kruglyak and Nickerson, 2001). They are known to contribute to the population diversity and phenotypic differences between individuals and their predispositions to diseases. Linkage disequilibrium in human populations has been influenced by factors affecting human evolution (Goldstein and Weale, 2001). Studies by Kruglyak et al (1999) concluded that LD would not extend beyond 3kb, but populations that have undergone severe genetic bottlenecks and recent admixture have resulted in extended regions of LD (Wilson and Goldstein, 2000). LD also deteriorates rapidly as the distance between markers increases in population of unrelated individuals (Orr and Chanock, 2008). Currently, there are no explanations on the regulation on the size of haplotype block boundaries, but the size and distribution of haplotype blocks are variable between populations (The International HapMap Consortium, 2005; Reich et al., 2001; Conrad et al., 2006). As a general assumption, the average block size in African populations is smaller than for the other populations studied (European Caucasians or

East Asians (Orr and Chanock, 2008); this is also evident with the LD pattern results on the *COMT* gene on these populations (Figure 17).

The considerations above suggest that population-specific markers may be needed for genetic association studies, and our results ($r^2$ and MAF regression analyses) support this idea. Ideally for genetic association studies, an investigator can select one of the four populations that is representative or demographically similar to their population samples and test the transferability of the (HapMap population) SNPs (tag SNPs) to their population sample. Mueller et al surveyed the linkage disequilibrium patterns and tagSNP transferability on their European sample population (randomly chosen with no specific diagnosis) and used the CEU population as their reference to select tag SNPs. They scanned four genomic regions (749 kb total) containing candidate genes for complex traits. In the two regions analyzed, the LD patterns of sets of tag SNPs that were selected from CEU population performed well or transferable to their sample population (2005). On the other hand, a study comparing the performance of HapMap SNP data with their sample population (Shanghai Chinese) found that tag SNPs on the CHB population on the HapMap data had better correlation with their sample group compared to the other populations on the HapMap database looking at chromosome 1q21-q25 (Hu et al, 2008). Both of these studies used the performance of HapMap-derived tag SNPs in their sample population study using a denser set of SNP tags. These studies (Mueller et al, 2005; Hu et al, 2008) further support that demographics can affect LD patterns and provide suggestions for ways on how the HapMap database could be

utilized. In addition, our study also supports that LD is affected by demography and it is important to consider when selecting populations for future genetic studies of AD/HD on these candidate genes. Consequently, the three populations on the HapMap database do not represent all the human populations and perhaps it could extend its population group as reference for future association studies on different populations.

### *General Suggestions and Recommendations*

In this current study, we have addressed some of the essential issues (molecular assays and human variation patterns) that may explain the inconsistent results of AD/HD association studies in the current literature and may assist in the design of any genetic association studies based on linkage disequilibrium. The International HapMap Project is continually generating genome-wide and densely spaced sequence variation data in different human populations (The International HapMap Consortium, 2005). This type of data will help in the design of LD-based genetic association studies and it will promote multi-locus LD measures to assess the variability of background correlation across the genome (Mueller, 2004). The prior assessment of markers, particularly SNPs, using this data and other software such as Haploview will better inform an investigator on the selection of markers for their particular study. We recommend the selection of a denser tag SNPs (more than 3) and/or using 2-3 tag SNPs per haplotype blocks rather than what we used in this study, selecting the best SNPs (2-3 tag SNPs) to cover the gene region. Currently, genetic association studies on the candidate genes and genome-wide scans

involving the monoamine system (dopamine, serotonin and norepinephrine) on AD/HD are using high-density tag SNP and hapolotype-based LD SNP selection however, their results are still inconclusive (Guan et al., 2009; Elia et al., 2009; Kollins et al, 2008; Ouelle-Morin et al, 2007). Recently, Elia et al (2009) had identified common inherited copy variations (CNVs) in their AD/HD probands relative to their controls in genes associated with neurodevelopmental pathways and disorder. This is the first study in AD/HD that showed a potential role of CNVs and also suggested other potential candidate genes such as *PTPRD* (protein tyrosine phosphatase receptor type D) and *GRM5* (Glutamate receptor 5) for further study.

The Illumina Infinium (Illumina, Inc., San Diego, CA) genotyping platform seems to be the choice for SNP genotyping assays with large sample populations (Guan et al., 2009; Kollins et al, 2008). A thorough testing of new SNP genotyping technologies is essential as false genotyping calls can results to spurious association results. We have suggested several technologies that are currently available (SNP genotyping section) and we recommend selecting a technology that has a reliable technical support. Our analysis of $r^2$ between the tag SNPs and captured SNPs using the best individual SNPs on each of the genes may not be the best approach to capture LD; rather a true haplotype-based approach might be better in capturing the LD patterns on these genes and also extending the tag SNPs for more coverage will be helpful. Furthermore, there are also other methods for estimating and testing of LD. In this project we have used a Pairwise LD method using the Haploview program by selecting a

sparse set of tag SNPs based on the best MAF and did not select tag SNPs based on haplotype blocks. Since SNPs can arise on independent haplotypic backgrounds and that many common haplotypes exist at a given locus (Orr and Channock, 2008), a SNP haplotype-based method is a complementary alternative of selecting SNPs (Stram, 2004). The use of haplotype-tagging SNP allows a researcher to examine multiple independent SNPs across the gene rather than focusing on a specific repetitive SNP on a gene. This provides a more thorough examination on how variation in specific genes may be associated with the specific phenotypes (Kollins et al., 2008). The LD patterns are very important in for the design and interpretation of association studies (Mueller, 2004) and there are a variety of LD measures available online in addition to the Haploview that can be used for estimating and testing for LD (Genetix, http://www.univ-montp2.fr/%7Egenetix/genetix/genetix.htm; DNaSP, http://www.ub.es/dnasp/; The R Project for Statistical Computing, http://www.r-project.org/ are few of the websites available). In addition, other programs such as WCLUSTAG are being developed by combining functional and linkage disequilibrium information in the selection of tag SNPs in order to prioritize the SNPs being genotyped (Sham et al., 2007).

The common-disease common-variant hypothesis states that the genetic variation underlying susceptibility to common heritable disease/trait existed within the founding population of contemporary human population and that these genetic variants have escaped selective pressure and have small to modest effect on the disease/trait (Doris, 2002; Chen et al., 2006; Collins et al., 1999; Risch and Merikangas, 1996;

Lander, 1996). Understanding the forces that drive and shape human genetic variations are important considerations in LD-based genetic association studies. As we have shown in this study, LD varies between populations; it is affected by evolutionary forces such as bottlenecks and varies based on the populations' demographics (Goldstein and Weale, 2001; Kruglyak et al, 1999; Wilson et al., 2000 Orr and Chanock, 2008; Mueller et al, 2005; Hu et al, 2008). LD patterns are sufficiently consistent to allow efficient representation of common variation with the use of tag SNPs. However, association studies based on CDCV hypothesis has been criticized for not fully explain the genetic component of many diseases (such as the inconsistent results on AD/HD). Perhaps we have reached a plateau where there are no more common variants to discover or no more that are worth discovering and we need to turn our focus on rare variants (Goldstein, 2009; Iyengar and Elston, 2007). The variability of the LD patterns observed in our results (i.e. MAF and $r^2$ regressions) deviate from the CDCV hypothesis and may suggest that different strategy for selecting tag SNPs may be needed to further study the association of these genes with AD/HD.

## REFERENCES


Acosta MT, Acros-Burgos M, Muenke M. (2003). Attention Deficit/Hyperactivity Disorder (ADHD): Complex Phenotype, Simple Genotype? *Genetics in Medicine* 6: 1-15.

Altepeter TS and Korger JN. (1999). Disruptive Behavior: Oppositional Defiant Disorder and Conduct Disorder. In S. Netherton, D. Holmes & CE Walker (Eds). *Child and Adolescent Psychological Diosrder: A Comprehensive Texbook* (pp 118-138). New York: Oxford University Press.

American Psychiatric Association. (1994*). Diagnostic and statistical manual of mental disorders DSM-IV (4th Edition).* Washington, D.C.: Author.

Anastopoulos A. (1999). Attention-Deficit/Hyperactivity Disorder. In S. Netherton, D. Holmes & C.E. Walker (Eds). *Child and Adolescent Psychological Disorder: A Comprehensive Textbook* (pp. 98–117). New York: Oxford University Press.

Anastopoulos A and Shelton T. (2001). *Assessing Attention-Deficit Hyperactivity Disorder*. New York: Kluwer Academic/Plenum Publishers.

Angold A, Costello EJ, Erkanli A. (1999). Comorbidity. *Journal of Child Psychology* 40: 57-87.

Bakker PW, Yelensky R, P'er I, Gabriel S, Daly MJ, Altshuler D. (2005). Efficiency and power in genetic association studies. *Nature Genetics* 37(11): 1217-1223.

Barnes MR. (2006). Navigating the HapMap. *Briefings in Bioinformatics* 7(3): 211-224.

Biederman J. (2005). Attention-Deficit/Hyperactivity Disorder: A Selective Overview. *Biological Psychiatry* 57: 1215-1476.

Biederman J, Newcorn J, Sprich S. (1991). Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety, and other disorders. *American Journal of Psychiatry* 148: 564-577.

Biederman J. and Spencer T. (1999). Attention-Deficit/Hyperactivity Disorder (AD/HD) as a Noradrenergic Disorder. *Biological Psychiatry* 46: 1234-1242.

Bobb AJ, Castellanos FX, Addington AM, Rapport JL. (2004). Molecular genetic studies of ADHD: 1991-2004. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. (Early View, Articles online in advance print) Published Online: 2000; accessed via UNCG Library.

Bray MS, Boerwinkle E, Doris PA. (2001). High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: practice, problems and promise. *Human Mutation* 17(4): 296-304.

Burns GL and Walsh JA. (2002). The Influence of ADHD-Hyperactivity/Impulsivity Symptoms on the Development of Oppositional Defiant Disorder Symptoms in a 2-Year Longitudinal Study. *Journal of Abnormal Child Psychology* 30: 245-256.

Cardon LR and Bell JI. (2001). Association Study Designs for Complex Diseases. *Nature Reviews in Genetics* 2: 91-99.

Chang FM, Kidd JR, Livak KJ, Pakstis AJ, Kidd KK. (1996). The world-wide distribution of allele frequencies at human dopamine D4 receptor locus. *Human Genetics* 98: 91-101.

Chen GK, Jorgenson E, Witte JS. (2007). An empirical evaluation of the common disease-common variant hypothesis. *BMC Proceedings*. Retrieved from http://www.biomedcentral.com/1753-6561/SI/S5.

Chen J, Lipska BK, Halim N, Ma QD, Matsumoto M, Melhem S, Kolachana BS, Hyde TM, Herman MM, Apud J, Egan MF, Kleinman JE, Weinberger DR. (2004). Functional analysis of genetic variation in catechol-o-methyltransferase (COMT): Effects on mRNA, protein, and enzyme activity in postmortem human brain. *American Journal of Human Genetics* 75: 807-821.

Chen X and Sullivan PF. (2003). Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics Journal* 3(2): 77-96.

Chen Y, Lin C, Sabatti C. (2006). Volume measures for linkage disequilibrium. *BMC Genetics*. Retrieved from http//www.biomedcentral.com/1471-2156/7/54.

Chu Y, Su X, Huang Q, Zhang X (2009). Patterns of DNA sequence variation at candidate gene loci in black poplar (*Populus nigra* L.) as revealed by single nucleotide polymorphisms. *Genetica*. Retriieved from http://www.springerlink.com/content/m80p546281706374/

Collins A, Lonjou C, Morton NE. (1999). Genetic epidemiology of single-nucleotide polymorphisms. *Proc Natl Acad Sci USA* (96): 15173-15177.

Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, Pritchard JK. (2006). A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nature Genetics* 38: 1251-1260.

Cooper, P and Bilton, K (Eds.). (1999). *ADHD: Research, Practice, and Opinion*. London, Whurr Publishing. Retrieved: April, 2004 from UNCG Library database on the World Wide Web: www.netlibrary.com.

DiMaio S, Grizenko N, Joober R. (2003). Dopamine genes and atteintion-deficit hyperactivity disorder: a review. *Journal of Psychiatry and Neuroscience* 28 (1), 27-38.

Ding YC, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK, Flodman P, Spence MA, Schuck S, Swanson JM, Zhang YP, Moyzis RK. (2002). Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proceedings of the National Academy of Science of the USA* 99(1): 309-314.

Doris PA. (2002). Hypertension genetics, single nucleotide polymorphisms, and the common disease: common variant hypothesis. *Hypertension* 39(2): 323-331.

Elia J, Capasso M, Zaheer Z, Lantieri F, Ambrosini P, Berrettini W, Devoto M, Hakonarson H. (2009). Candidate gene analysis in an on-going genome-wide association study of attention-deficit hyperactivity disorder: suggestive association signals in ADRA1A. *Psychiatric Genetics* 19(3): 134-141.

Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT, D'arcy M, deBerardinis R, Frackelton E, Kim C, Lantieri F, Muganga BM, Wang L, Takeda T, Rappaport EF, Grant SFA, Berrettini W, Devoto M, Shaikh TH, Hakonarson H, White PS. (2009). Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Molecular Psychiatry* advance online publication 23 June 2009: 1-10

Faraone SV and Biederman J. (1998). Neurobiology of Attention-Deficit Hyperactivity Disorder. *Biological Psychiatry* 44: 951-958.

Faraone SV, Doyle AE, Mick E, Biederman J. (2001). Meta-analysis of the association between the 7-repeat allele of the dopamine D (4) gene and attention deficit hyperactivity disorder. *American Journal of Psychiatry* 158: 1052-1057.

Faraone SV, Perlis RH, Doyle AE, Smoller JW, Goralnick JJ, Homgren MA, Sklar P. (2005). Molecular Genetics of Attention-Deficit/Hyperactivity Disorder.

*Biological Psychiatry* 57: 1313-1323.

Fisher SE, Franks C, McCracken JT, McGough JJ, Marlow AJ, MacPhie IL, Newbury DF, Crawford LR, Plamer CGS, Woodward AJ, Del'Homme M, Cantwell DP, Nelson, SF, Monaco AP, Smalley SL. (2002). A Genomewide Scan for Loci Involved in Attention-Deficit/Hyperactivity Disorder. *American Journal of Human Genetics* 70: 1183-1196.

Frazer KA, Murray SS, Schork NJ, Topol EJ. (2009). Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics* 10: 241-251.

Gabriel S, Ziaugra L, Tabbaa D (2009). SNP Genotyping using Sequenom MassARRAY iPlex platform. *Current Protocols in Human Genetics* 60: 2.12.1-2.12.18.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Lui-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. (2002). The structure of haplotype blocks in the human genome. *Science* 296: 2225-2229.

Genro JP, Polanczyk GV, Zeni C, Oliveira AS, Roman T, Rohde LA, Hutz MH. (2008). A common haplotype at the dopamine transporter gene 5' region is associated with attention-deficit/hyperactivity disorder. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)* 147B: 1568-1575.

Giros B, Jaber M, Jone SR, Wightman RM, Caron MG. (1996). Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. *Nature* 379: 606-612.

Goldstein DB. (2009). Common Genetic Variation and Human Traits. *New England Journal of Medicine* 360(17): 1696-98.

Goldstein DB and Weale ME. (2001). Population genomics: Linkage disequilibrium holds the key. *Current Biology* 11: R576-R579.

Gonzalez-Neira A, Ke X, Lao O, Calafell F, Navarro A, Comas D, Cann H, Bupstead S, Ghori J, Hunt S, Deloukas P, Dunham I, Cardon LR, Bertranpetit J. (2006). The portability of tagSNPs across populations: a worldwide survey. *Genome Research* (16): 323-330.

Grady DL, Chi HC, Ding YC, Smith M, Wang E, Schuck S, Flodman P, Spence MA, Swanson JM and Moyzis RK. (2003). High prevalence of rare dopamine receptor D4 alleles in children diagnosed with attention-deficit hyperactivity disorder. *Molecular Psychiatry* 8: 536-545.

Guan L, Wang B, Chen Y, Yang L, Li J, Qian Q, Wang Z, Faraone SV, Wang Y. (2009). A high-density single nucleotide polymorphism screen of 23 candidate genes in attention deficit hyperactivity disorder: suggesting multiple susceptibility genes among Chinese Han population. *Molecular Psychiatry* 14: 546-554.

Haiman CA and Stram DO. (2008). Utilizing HapMap and Tagging SNPs. In R.J. Trent (Ed.), *Methods in Molecular Medicine Volume 141: Clinical Bioinformtics* (37-69). Retrieved from http://books.google.com.

Halleland H, Lundervold AJ, Halmoy A, Haavik J, Johansson S. (2008). Association between catechol o-methyltransferase (COMT) haplotypes and severity of hyperactivity symptoms in adults. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)* 150B: 403-410.

Heiser P, Friedel S, Dempfle A, Konrad K, Smidt J, Grabarkiewicz J, Herpertz-Dahlmann B, Remschmidt H, Hebebrand J. (2004). Molecular genetic aspects of attention-deficit/hyperactivity disorder. *Neuroscience and Behavioral Reviews* 28: 625-641.

Hirschhorn JN. (2005). Genetic Approaches to Studying Common Diseases and Complex Traits. *Pediatric Research* 57: 74R-77R.

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. (2002). A comprehensive review of genetic association studies. *Genetics in Medicine* 4(2): 45-61.

Hu C, Jia W, Zhang W, Wang C, Zhang R, Wang J, Ma X, The International Type 2 Diabetes 1q Consortium, Xiang K. (2008). *BMC Genetics* 9: 1-8.

Iyengar SK and Elston RC. (2007). The genetic basis of complex traits: rare variants or "common gene, common disease"?. *Methods in Molecular Biology* 376: 71-84.

Jensen PS, Martin D, Cantwell DP. (1997). Comorbidity in ADHD: implication for research, practice, and DSM-V. *Journal Am. Acad. Child Adolesc. Psychiatry* 36: 1065-1079.

Jiang S, Xin R, Lin S, Quian Y, Tang G, Wang D, Wu X. (2001). Linkage studies between attention-deficit hyperactivity disorder and the monoamine oxidase genes. *American Journal of Medical Genetics* 105: 783-788.

Kereszturi E, Tarnok Z, Bognar E, Lakatos K, Farkas L, Gadoros J, Sasvari-Szekely M, Nemoda Z. (2008). Catechol-O-Methyltransferase Val158Met Polymorphisms is associated with methylphenidate response in ADHD Children. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)* 147B: 1431-1435

Kollins SH, Anastopoulos AD, Lachiewicz AM, FitzGerald D, Morrissey-Kane E, Garret ME, Keats SL, Ashley-Koch AE. (2008). SNPs in dopamine D2 receptor gene (DRD2) and norepiephrine transporter gene (NET) are associated with continuous performance task (CPT) phenotypes in ADHD children and their families.

Kruglyak L. (1999). Prospects for whole-genome linkage disequilibrium mapping common disease genes. *Nature Genetics* 22: 139-144.

Kruglyak L and Nickerson DA. (2001). Variation is the spice of life. *Nature Genetics* 27(3): 234-236.

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

Lander ES. (1996). The new genomics: global views of biology. *Science* (274): 536-539.

Leung PW, Luk SL, Ho Tp, Taylor E, Mak FL, Bacon-Shone J. (1996). The diagnosis and prevalence of hyperactivity in Chinese schoolboys. *British Journal of Psychiatry* 168(4): 486-496.

Lilienfeld SO, Waldman ID, Israel AC. (1994). A critical examination of the use of the term and concept of comorbidity in psychopathology research. *Clinical Psychology: Science & Practice* 1: 71-83.

Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. (2003). Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics* 33: 177-182.

Lopez RE. (1965). Hyperactivity in twins. *Canadian Psychiatric Association Journal* 10(5): 421-426.

Millichap GJ. (1999). *Attention Deficit Hyperactivity and Learning Disorders Questions & Answers*. PNB Publishers. Chicago.

Monuteaux MC, Sidman LJ, Faraone, SV, Makris N, Spencer T, Valera E, Brown A, Bush G, Doyle AE, Huges S, Helliesen M, Mick E. (2008). A Preliminary Study of Dopamine D4 Receptor Genotype and Structural Brain Alterations in Adults With AD/HD. *American Journal of Medical Genetics Part B(Neuropsychiatric Genetics)* 147B: 1436-1441.

Mueller JC. (2004). Linkage disequilibrium for different scales and applications. *Briefings in Bioinformatics* 5: 355-364.

Mueller JC, Lohmussaar E, Magi R, Remm M, Bettecken T, Lichtner P, Biskup S, Illig T, Pfeufer A, Luedemann J, Schreiber S, Pramstaller P, Pichler I, Romeo G, Gaddi A, Testa A, Wichmann HE, Metspalu A, Meitinger T. (2005). Linkage Disequilibrium Patterns and tagSNP transferability among European Populations. *American Journal of Human Genetics* 76: 387-398.

Ogdie MN, Fisher SE, Yang M, Ishii J, Franks C, Loo SK, Cantor RM, McCracken JT, McGough JJ, Smalley SL, Nelson SF. (2004). Attention Deficit Hyperactivity Disorder: Fine Mapping Supports Linkage to 5p13, 6q12, 16p13, and 17p11. *American Journal of Human Genetics* 75: 661-668.

Ogdie MN, Macphie IL, Minassian SL, Yang M, Fisher SE, Francks C, Cantor RM, McCracken JT, McGough JJ, Nelson SF, Monaco AP, Smalley SL. (2003). A Genomewide Scan for Attention Deficit/Hyperactivity Disorder in an Extended Sample: Suggest Linkage on 17p11. *American Journal of Human Genetics* 72, 1268-1279.

Orr N and Chanock S. (2008). Common Genetic Variation and Human Disease. *Advances in Genetics* (62): 1-32.

Ouellet-Morin I, Wigg KG, Feng Y, Dionne G, Robaey, P, Brendgen M, Vitaro F, Simard L, Schachar R, Tremblay RE, Perusse D, Boivin M, Barr CL. (2008). *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)* 147B: 1442-1449.

Pritchard JK and Przeworski M. (2001). Linkage Disequilibrium in humans: models and data. *American Journal of Human Genetics* 69: 1-14.

Pusch W, Wurmbach JH, Thiele H, Kostrzewa M. (2002). MALDI-TOF mass spectrometry-based SNP genotyping. *Pharmacogenomics* 3(4): 537-548.

Qian QJ, Liu J, Wang YF, Yang Li, Guan LL, Faraone SV. (2009). Attention Deficit Hyperactivity Disorder comorbid oppositional defiant disorder and its predominately inattentive type: evidence for an association with COMT but not MAOA in a Chinese sample. *Behavioral and Bran Functions* 5: 8.

Ragoussis J. (2009). Genotyping Technologies for Genetic Research. *The Annual Review of Genomics and Human Genetics* 10: 5.1-5.7.

Reich DE, Cargill M, Bolk M, Irelan J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian F, Ward R, Lander ES. (2001). Linkage disequilibrium in the human genome. *Nature* 411: 199-204.

Reich DE and Lander ES. (2001). On the allelic spectrum of human disease. *Trends in Genetics* 17: 502-510.

Risch N. (1990). Linkage strategies for genetically complex traits. I. Multilocus models. *American Journal of Human Genetics* 46: 222-228.

Risch N, Merikangas K. (1996). The future of genetic studies of complex human diseases. *Science* 273: 1516-1517.

Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, Stoneking M, Batzer MA, Barbujani G. (2002). Patterns of human diversity within and among continents, inferred from biallelic DNA polymorphis. *Genome Research* 12(4): 602-612.

Sham PC, Ao SI, Kwan JSH, Kao P, Cheung F, Fong PY, Ng MK. (2007). Combining functional and linkage disequilibrium information in the selection of tag SNPs. *Bioinformatics* 23(1): 129-131.

Shastry BS. (2004). Molecular genetics of Attention deficit hyperactivity disorder (ADHD): an update. *Neurochemistry International* 44, 469-474.

Smalley SL, Kustanovich V, Minassian SL, Stone JL, Ogdie MN, McGough JJ, McCracken JT, MacPhie IL, Fracks C, Fisher SE, Cantor RM, Monaco AP, Nelson SF. (2002). Genetic Linkage of Attention-Deficit/Hyperactivity Disorder on Chromosome 16p13, in a Region Implicated in Autism. *American Journal of Human Genetics* 71: 959-963.

Sobrino B, Brion M, Carracedo A. (2005). SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International* 154: 181-194.

Strachan T and Read AP. (2004). *Human Molecular Genetics* (3rd ed.). London and New York: Garland Publishing.

Stram DO. (2004). Tag SNP Selection for Association Studies. *Genetic Epidemiology* 27: 365-374.

Sullivan PF. (2007). Spurious Genetic Association. *Biological Psychiatry* 61: 1121-1126.

Swanson JM, Sunohara GA, Kennedy JL, Regino R, Finberg E, Wigal T, Lerner M, Williams L, Lahoste, GJ, Wigal S. (1998). Association of the dopamine receptor D4 (DRD4) gene with a refined phenotype of attention deficit hyperactivity disorder (ADHD): a family based approach. *Molecular Psychiatry* 3: 38-41.

Syvanen AC. (2001). Accessing Genetic Variation: Genotyping Single Nucleotide Polymorphisms. *Nature* 2: 930-942.

The International HapMap Consortium. (2003). The International HapMap Project. *Nature* 426: 789-796.

The International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature* 437: 1299-1320.

Tost J and Gut IG. (2005). Genotyping single nucleotide polymorphisms by MALDI mass spectrometry in clinical applications. *Clinical Biochemistry* 38(4): 335-50.

Vance JM and Othmane KB. (1998). Methods of Genotyping. In J.L. Haines and M.A. Pericak-Vance (Eds.), *Approaches to Gene Mapping in Complex Human Diseases* (213-26). New York: Wiley-Liss.

Wang E, Ding YC, Flodman P, Kidd JR, Kidd KK, Grady DL, Ryder OA, Spence MA, Swanson JM, Moyzis RK. (2004). The Genetic Architecture of Selection at the Human Dopamine Receptor D4 (*DRD4*) Gene Locus. *American Journal of Human Genetics* 74: 931-944.

Wilson JF and Goldstein DB. (2000). Consistent long-range linkage disequilibrium genetrated by admixture in a Bantu-Semitic hybrid population. *American Journal of Human Genetics* 67: 926-935.

# APPENDIX A

## SNuPe GENOTYPING PROTCOLS I

**I. PCR Protocol for SNuPe**

| Reagent | Final Concentration | Volume (µl) |
|---|---|---|
| 10x PCR buffer | 1x | 2.5 |
| 2 mM dNTPs | 0.2 mM each | 2.5 |
| 50 pmol Forward | 3 pmol | 1.5 |
| 50 pmol Reverse | 3 pmol | 1.5 |
| 25 mM $MgCl_2$ | 2.5 mmol | 2.5 |
| Taq Polymerase (250 units) | - | 0.25 |

Use 25 nanograms of DNA template and add 13.75 µl of deionized $H_20$ to make 25µl total volume

mM=millimolar; pmol=picomolar

**II. 50 Enhanced PCR Program**

| Step | Temperature | Time |
|---|---|---|
| 1 | 95 °C | 5 mins |
| 2 | 95 °C | 45 secs |
| **3** | **68°C** | **5 mins** |
| 4 | 72°C | 1 min |
| 5 | 95 °C | 45 secs |
| **6** | **66 °C** | **5 min** |
| 7 | 72°C | 1 min |
| 8 | 95°C | 45 secs |
| **9** | **64°C** | **5 min** |
| 10 | 72°C | 1 min |
| 11 | 95°C | 45 secs |
| **12** | **62°C** | **5 min** |
| 13 | 72°C | 1 min |
| 14 | 95°C | 45 secs |
| **15** | **60°C** | **5 min** |
| 16 | 72°C | 1 min |
| 17 | 95°C | 45 secs |
| **18** | **58°C** | **2 min** |
| 19 | 72°C | 1 min |
| 20 | 95°C | 45 secs |
| **21** | **56°C** | **2 min** |
| 22 | 72°C | 1 min |
| 23 | 95°C | 45 secs |
| **24** | **54 °C** | **2 min** |
| 25 | 72°C | 1 min |
| 26 | 95°C | 45 secs |
| **27** | **52°C** | **2 min** |
| 28 | 72°C | 1 min |
| 29 | 95°C | 45 secs |
| **30** | **50°C** | **2 min** |
| 31 | 72°C | 1 min |
| **32** | **Go Back Step 29 - 24 times** | |
| 33 | 72°C | 5 min |
| 34 | 4°C | ∞ |

## APPENDIX B

## SNuPe GENOTYPING PROTOCOLS II

**III. After PCR: SNuPe Genotyping (Amersham Biosciences Protocol)**

**Step 1: ExoSap**

5µl of the PCR product

2µl of ExoSap

Thermocycler (5): ExoSap program

37°C for 15min

80°C for 15 min

**Step 2: SNuPe**

**On a 96-well plate prepare sample reactions and control reactions (A,C,T,G)

*Control reactions

4µl – SNuPe premix

1µl – one control primer

1ul – M13 control template

4µl – Deionized $H_2O$

*Sample reactions

4µl – SnuPe premix

1µl – SNP primer

1µl – DNA template (ExoSaped PCR products)

4µl – $dH_2O$

**Step 3: Thermocycler: SNuPe program**

96°C – 10s

50°C – 5s

60°C – 10s

Repeat for 25 cycles

## APPENDIX C

## SNuPe GENOTYPING PROTOCOLS III

**IV. Alcohol Clean Up Method**

2µl Ammonium acetate
60µl 95% Ethanol
10 µl H20

1. These were added to the 10µl SNuPe product
2. Centrifuge palte 2500x g for 30 minutes
3. Wash with 100µl 70% Ethanol, discard Ethanol and let it air dry
4. Added 10µl of the Megabace loading buffer to each sample.
5. Take 5µl of the sample with Megabace loading buffer
   Add 5µl of MB loading buffer and Injection Marker mixture
6. Spin to remove air bubbles and run on the Megabace machine

**V. Making Sephadex plate**

1. Measure sephadex powder using the plate loading device provided.
2. Pour it on the multiscreen filtration plate. (Millipore unifilter 800 plate)
3. Add 250µl ultrapure water on each well
   Allow to expand for at least 3-5 hours
4. Spin at 2400 rpm for 2 minutes
5. Add 100µl of water using the repeater pipette, leave for 5 minutes, spin for 2 minutes at 2000 rpm
6. Repeat step 5; and store plates covered with plastic film and aluminum foil in the 4ºC refrigerator
   It is best to store plates overnight before use.

**Steps before loading SNuPe products on the Sephadex Plate**

1. Add 600µl of water to each well; let it cool and swell for at least 3-4 hours
2. Spin for 4 minutes at 2000rpm
3. Add10µl of water and spin for 5 minutes at 2000rpm
4. Repeat step 3 until you get equal amounts of liquid
5. Then add your samples.

# APPENDIX D

## PCR-RLFP GENOTYPING PROTOCOLS

**I. PCR Protocol for PCR–RFLP**

| Reagent | Final Concentration | Volume (µl) |
|---|---|---|
| 10x PCR buffer | 1x | 2.5 |
| 2mM dNTPs | 0.2 mM each | 2 |
| 10 pmol Forward | 0.6 pmol | 1.5 |
| 10 pmol Reverse | 0.6 pmol | 1.5 |
| 25mM $MgCl_2$ | 2.5mmol | 2.5 |
| Taq Polymerase (250 units) | - | 0.5 |
| Use 25 nanograms of DNA template and add 14 µl of deionized $H_20$ to make 25µl total volume | | |
| For PCR-RFLP: Take 5µl of the PCR product, 1µl of the enzyme buffer, 1µl of the enzyme and add the appropriate amount of water for a 10µl total reaction. Incubate according to the specified time and temperature for each enzyme. | | |

II. PCR Thermal Cycling (From Dr. Remington's Lab)

***A. Intinial Denaturation***

94 °C 2 mins.

***B. 7 Cycles***

94 °C 20 secs.
62 °C 1 min
72 °C 1 min

***C. 28 Cycles***

94 °C 20 secs.
55 °C 1 min.
72 °C 1 min.

***D. Final Extension***

72 °C 7 mins.
4 °C hold

**NOTES**

This is good for most of the primers, except for the one with the low Tm (53 degrees)

Lengthen the Touchdown Portion

Section B: Instead of 7 cycles used 10 Cycles

Section C: Use 52 °C instead of 55 °C

Most of the PCR products are <1000 bp, cut the extension times down to 1:00 or 0:30secs