5-2010

A Family-Based Association Study of Conduct Disorder.

Xueqiu Jian
East Tennessee State University

A Family-Based Association Study of Conduct Disorder

A thesis

presented to

the faculty of the College of Public Health

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Public Health in Epidemiology

---

by

Xueqiu Jian

May 2010

Dr. Kesheng Wang, Chair

Dr. Tiejian Wu

Dr. Joel Hillhouse

Keywords: Conduct Disorder, Genome-Wide Association Study, Family-Based Design, Genes,

Single Nucleotide Polymorphisms

ABSTRACT

A Family-Based Association Study of Conduct Disorder

by

Xueqiu Jian

Conduct disorder is a psychiatric syndrome in childhood and adolescence that is one of the most common childhood disorders with continuously increasing prevalence but uncertain pathogenesis. We performed a genome-wide, family-based association study of CD using P2BAT/FBAT software. The data were gathered from Collaborative Study on the Genetics of Alcoholism (COGA) and International Multi-Center ADHD Genetics Project (IMAGE).

Using COGA data, we identified 20 markers that showed suggestive associations ($p<10^{-3}$) with CD. Nine of them are located in known genes. Two genes, *ADAM10* and *CAMK2A*, which had been reported associated with Alzheimer's disease (AD), bipolar disorder, and depression, were of more concern. Using IMAGE sample, our results were well replicated.

This study identified several CD associated genetic variants, especially 2 novel candidate genes. These findings may serve as a resource for replication in other populations to elucidate the potential role of these genetic variants in CD.

ACKNOWLEDGMENTS

I am greatly indebted to my supervisor Dr. Kesheng Wang for his valuable instructions and suggestions on my thesis as well as his careful reading of the manuscript.

I feel grateful to Dr. Tiejian Wu and Dr. Joel Hillhouse, my advisory committee members, from whom I benefited greatly. My sincere thanks are also given to all the teachers in the College of Public Health at East Tennessee State University who once offered me valuable courses and advice during my study.

Last but not least, I owe much to my family and friends for their valuable suggestions and critiques which are of help and importance in making the thesis a reality.

The Collaborative Study on the Genetics of Alcoholism (COGA) (H. Begleiter, SUNY HSCB, Principal Investigator: T. Reich, Washington University, Co-Principal Investigator) includes nine centers where data collection, analysis, and storage take place. The nine sites and Principal Investigator and Co-Investigators are: Indiana University (T.-K. Li, J. Nurnberger Jr., P.M. Conneally, H.J. Edenberg); University of Iowa (R. Crowe, S. Kuperman); University of California at San Diego (M. Schuckit); University of Connecticut (V. Hesselbrock); State University of New York, Health Sciences Center at Brooklyn (B. Porjesz, H. Begleiter); Washington University in St. Louis (T. Reich, C. R. Cloninger, J. Rice, A. Goate); Rutgers University (J. Tischfield); and Southwest Foundation (L. Almasy). This national collaborative study is supported by NIH grant U10AA08403 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA). The National Institute of General Medical Sciences has provided continuous funding for the Genetic Analysis Workshops (GAW) since 1982, through grant

R01 GM31575 to Jean MacCluer (Southwest Foundation for Biomedical Research). The GAW14 data was kindly provided by Jean MacCluer (Southwest Foundation for Biomedical Research). We acknowledge the contributions of the COGA, supported by NIH Grants U10AA08401 and U10AA08403 (NIAAA) and the contributions of all scientists who have provided genotyping data to the Wave I and/or Wave II--Genetic Analysis Data.

The dataset for replication study were obtained from the GAIN Database found at http://www.ncbi.nlm.nih.gov/projects/gap/ through the dbGAP accession number phs000016.v1.p1. The International Multi-Center ADHD Genetics Project (IMAGE) project is a multi-site, international effort supported by NIH grants R01MH081803 and R01MH62873 to Stephen V. Faraone. Site Principal Investigators are Philip Asherson, Tobias Banaschewski, Jan Buitelaar, Richard P. Ebstein, Stephen V. Faraone, Michael Gill, Ana Miranda, Robert D. Oades, Herbert Roeyers, Aribert Rothenberger, Joseph Sergeant, Edmund Sonuga-Barke, and Hans-Christoph Steinhausen; senior coinvestigators are Ian Craig, Peter McGuffin, Robert Plomin, Pak Sham, Eric Taylor, Iris Manor, Jacques Eisenberg, and Margaret Thompson. Chief Investigators are Evi Bitsakou, Marieke Altink, Wai Chen, Hanna Christiansen, Barbara Franke, Rafaela Marco, UMueller, Fernando Mulas, Lamprini Psychogiou, Nanda Rommelse, Aisling Mulligan, and Henrik Uebel. Other investigators are Cathelijne Buschgens, Frits Boer, Alysa Doyle, Ellen Fliers, Alexander Heise, and Ruud Minderaa. The genotyping of samples was provided through the Genetic Association Information Network (GAIN). Samples and associated phenotype data for The International Multi-Center ADHD Genetics Project (IMAGE) project were provided by Dr Stephen V. Faraone. We thank all the families who kindly participated in this research.

This thesis is part of project “Genetic analysis of alcohol dependence and alcohol-related phenotypes in the COGA sample” approved by IRB, East Tennessee State University.

CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Conduct disorder (CD) is a psychiatric syndrome occurring in childhood and adolescence and is characterized by a longstanding pattern of violations of rules and antisocial behavior (Searight, Rottnek, & Abby, 2001). Symptoms typically include aggression, frequent lying, running away from home overnight, and destruction of property (American Psychiatric Association [APA], 1994). CD is one of the most prevalent childhood disorders that affects 1.5%-3.4% of adolescents (Steiner, 1997) and is more common in males than in females (APA, 1994). Not only does CD associate with social malfunctioning, individuals with CD also show a high comorbidity with attention-deficit/hyperactivity disorder (ADHD) (Biederman, Newcorn, & Sprich, 1991), which is characterized by frequent and severe inattention, distractibility, impulsivity, or hyperactivity (Biederman & Faraone, 2005). Moreover, CD has been proved a robust predictor of both concurrent and future alcohol dependence (Deas & Brown, 2006). The fact that patients with CD may face so many problems leads to the necessity to better understand the pathogenesis of the disease.

Generally, the perception of human diseases can be divided into two levels: individual level and population level. In our field of study, we are not interested in the nosogenesis of CD or its pathophysiological process in certain individuals. Rather, the focus of epidemiologic field is whether the disease has a genetic disposition and if so, the extent to which CD is determined by genetic factors in a population. Twin studies have confirmed that genetic factors play a significant role in CD (Dick et al., 2002; Goldstein, Prescott, & Kendler, 2001; Slutske et al.,

1997), but the location is not clear. Based on this, our study aims to identify novel single nucleotide polymorphisms (SNPs) associated with CD and related known genes. Definitions of several terms are listed in Appendix A.

CHAPTER 2

LITERATURE REVIEW

Genetic Components of Conduct Disorder

A substantial body of epidemiologic evidence indicates that CD has a significant genetic disposition. Twin studies take advantage of the fact that monozygotic (MZ) twins share 100% of their genes, whereas dizygotic (DZ) twins, who are like ordinary sibling pairs genetically, share 50% of their genes identical by descent. Therefore, MZ twins should be more similar than DZ twins on a trait or disorder influenced by genes. Twin studies examine the extent to which observed individual differences are due to differences in genetic, shared environmental, or nonshared environmental influences. Heritability (h^2) is the proportion of observed differences that can be explained by genetic differences. In the Australian Twin Study, more than 70% of the variance in conduct disorder was attributable to genetic factors (Slutske et al., 1997). Other twin and family studies showed CD with modest heritability (about 40%) (Dick et al., 2002; Dick et al., 2004; Goldstein et al., 2001; Hicks, Krueger, Iacono, McGue, & Patrick, 2004; Rhee & Waldman, 2002; Subbarao et al., 2008; True et al., 1999).

Genome-Wide Linkage Studies

Two major approaches have been used to map genetic variants that influence disease risk: linkage analysis and association studies. These can be deployed by using the genome-wide or the candidate gene method. Until recently genome-wide scans were done by linkage only,

whereas association studies usually aimed at candidate genes. This is no longer true, as new genome-wide association (GWA) studies have already proven their usefulness.

Genome-wide linkage analysis involves genotyping of hundreds of polymorphic markers spread uniformly across the whole genome in families with multiple affected relatives (or multiple relatives in whom a trait has been measured). Markers that segregate with the disease (or trait) in relatives more often than expected are used to narrow down the location of the disease gene. The analysis is based upon a logarithmic odds ratio (LOD score) that expresses the likelihood of linkage divided by the likelihood of nonlinkage (Lander & Schork, 1994).

The first genome-wide linkage screen for genes influencing CD was conducted by Dick et al. (2004), which suggested regions on chromosomes 19 (with a maximum LOD score of 2.8 near marker D19S714) and 2 (with a maximum LOD score of 2.4 near marker D2S1331) may contain genes conferring risk to CD. Other candidate regions were also reported (on chromosomes 12 and 3 for CD diagnosis, and chromosomes 1 and 19 for CD symptoms) but have weaker evidence for linkage. Divergently, another genome-wide linkage analysis by Stallings et al. (2005) provided evidence of linkage for CD symptoms to regions on chromosomes 9q34 (near markers D9S1826 and D9S1838) and 17q12 (near marker D17S798). Later, Kendler et al. (2006) reported regions on chromosomes 1q (with a maximal LOD score of 3.16 near marker D1S196) and 14p (with a maximal LOD score of 2.36 near marker D14S261) have the strongest evidence for linkage to CD, followed by regions on chromosomes 2, 7, 8, and 10 with weaker evidence, which moderately provided some support for the findings by Dick et al. (2004).

Candidate Gene Association Studies

While linkage studies have a demonstrated utility in mapping rare, highly penetrant disease genes, results in mapping more moderate loci have been disappointing. The opportunities for mapping low penetrance disease predisposing genes with families alone are diminishing. Risch and Merikangas (1996) argued that case-control association studies present much greater power to detect the locus specific relative risks underlying polygenic disease.

Association studies compare the allele frequency of a polymorphic marker, or a set of markers, in unrelated patients (cases) and healthy controls to identify markers that differ significantly between the two groups. These studies have much greater power to detect the effects of common variants (Risch & Merikangas, 1996).

There are several advantages of association studies compared to linkage studies. First, association studies have increased precision in localizing a disease susceptibility locus. Second, association studies may have increased power compared to linkage studies, particularly for genes with small individual effects (Risch & Merikangas, 1996); this is because the excess sharing of a particular allele over that expected by chance among families (association) is greater than the excess sharing over that expected by chance within families (linkage). Moreover, linkage analysis, traditionally the most reliable of genetic methods when applied to Mendelian traits, has proven to be a much less reliable tool for the study of non-Mendelian diseases with a disappointingly high false-positive rate (Lake, Blacker, & Laird, 2000; Risch, 2000).

Several candidate genes have been tested for their association with CD. *GABRA2*, a gene that has previously been associated with adult AD, was reported significantly associated with

childhood CD symptoms by Dick et al. (2006). But their findings were not supported by a very recent study conducted by Sakai et al. (2010), although they did demonstrate a significant case-control association between a SNP in *GABRA2* (rs279871) and CD. Two studies suggested a strong association between CD and 5HTTLPR, a functional polymorphism in the promoter region of the serotonin transporter gene (*SLC6A4*) (Malmberg, Wargelius, Lichtenstein, Oreland, & Larsson, 2008; Sakai et al., 2006). Inconsistently, in another study using a larger sample from a general population, Sakai et al. (2007) did not find a significant association between 5HTTLPR and conduct problems. In fact, Cadoret et al. (2003) failed to detect a main effect between 5HTTLPR status and CD, although there was some evidence that 5HTTLPR variants play a significant role in CD when interacted with other genetic risk factors. Very recently, Monuteaux, Biederman, Doyle, Mick, and Faraone (2009) reported no significant association between 5HTTLPR functional polymorphism of *SLC6A4* and CD. Another gene, catechol O-methyltransferase gene (*COMT*) was examined by several groups, yielding contradictory results. Thapar et al. (2005) predicted that its homozygous genotype for valine allele, as well as its interaction with low birth weight, is associated with early-onset CD symptoms in ADHD children. Soon afterwards, nonsignificant results were reported by Sengupta et al. (2006). Recently, the positive results were obtained by Caspi et al. (2008) using three independent studies. However, Monuteaux et al. found no significant association between *COMT* Val/Met polymorphism and CD, although they specified that valine-valine homozygosity was associated with CD aggressive symptoms.

Genome-Wide Association Studies

The conventional GWA study approach is a hypothesis-free, systematic search of tagging SNPs across the genome to identify novel associations with common diseases and has emerged as a powerful tool to identify disease-related genes for many common human disorders and other phenotypes (Guessous, Gwinn, & Khoury, 2009; McCarthy et al., 2008; Seng & Seng, 2008; Wellcome Trust Case Control Consortium, 2007).

Recently, the first hypothesis-free genome-wide association analysis of comorbid conduct problems in ADHD was performed by Anney et al. (2008), including CD. However, they did not find any of 1,043,963 autosomal markers that reached genome-wide significance ($p<5\times10^{-7}$), but 54 markers reaching strong genome-wide association signals ($p<10^{-5}$), the top five of which were observed on chromosomes 13, 21, 11, 4, and 12.

To sum, although a large number of linkage and association studies have focused on genetic influence on CD and made great achievements, certainty is lacked and further replication is required before specific loci are confirmed.

CHAPTER 3

METHODS

Datasets

Collaborative Study on the Genetics of Alcoholism (COGA) Data and its Genetic Analysis Workshop 14 (GAW 14) Subset

COGA is a nine-site national collaboration with the goal of identifying and characterizing genes that affect the vulnerability to alcoholism and related phenotypes (Edenberg et al., 2005). Besides alcohol dependence, the dataset contains the disease status for other behavioral problems, including CD, that were diagnosed by using the Semi-Structured Assessment for Genetics of Alcoholism (SSAGA) interview (Bucholz et al., 1994; Hesselbrock, Easton, Bucholz, Schuckit, & Hesselbrock, 1999). In total, 2,282 individuals from 262 families were available for genetic analyses. GAW 14 is a collaborative effort among genetic epidemiologists to evaluate and compare statistical genetic methods. A COGA data subset selected for GAW 14 genotyping includes 1,353 individuals from 143 families and 11,560 SNP markers genotyped by Affymetrix GeneChip Mapping 10K Array and 4,752 SNPs from the Illumina's Linkage III Panel (Edenberg et al., 2005). In order to reduce genetic heterogeneity, we selected 122 Caucasian pedigrees (292 nuclear families with 1,335 individuals) that have a majority of individuals who self-reported to be 'white' (both Hispanic and non-Hispanic). Among the 1,335 individuals, 670 are males and 665 are females. Their ages range from 7 to 93, with a mean of 42 years old and standard deviation of 16.23. Affection status of CD is dichotomous. In our study, we defined 2 as affected, 1 as unaffected, and 0 as unknown. Totally, there are 155

individuals with a CD diagnosis, compared with 957 unaffected and 223 unknown.

International Multi-Center ADHD Genetics Project (IMAGE) Dataset

IMAGE is a project aiming at the detection of genes responsible for the genetic transmission of ADHD in children. The comorbidity of CD is examined in the during the assessment process. Anney et al. (2004) used this dataset to do the first hypothesis-free genome-wide analysis of CD in ADHD patients. We used part of this data to replicate our results for CD. In this study, 205 trios include 205 individuals with CD and their parents.

Study Designs

Family-based association studies, such as the transmission disequilibrium test (TDT), are preferable to case-control studies in allelic association studies when there is population admixture (Spielman, McGinnis, & Ewens, 1993). TDT detects preferential transmission of alleles from heterozygous parents to probands and is robust with respect to population stratification. Therefore, our study applied a genome-wide, family-based association analysis.

Statistical Analysis

Assessment of Hardy-Weinberg Equilibrium (HWE)

Departure from HWE was tested for unaffected founders using PLINK (Purcell et al., 2007).

Family-Based Association Analyses

In this study, family-based association analysis for CD was performed by the P2BAT in R for autosomal SNPs because P2BAT can handle nuclear families as well as extended pedigrees (Hoffmann & Lange, 2006). To deal with X-chromosome SNPs the FBAT empirical variance ("-e") option in FBAT was used to test for association because FBAT divides a large pedigree into small nuclear families and multiple sibs in a family are used (Rabinowitz & Laird, 2000). We also used part of IMAGE data to replicate our results in COGA data. For both two datasets, the additive model was applied. The family-based association test statistics in P2BAT/FBAT software are score tests that do not focus on the strength of association like odds ratio in the general case-control design. The primary aim of genetic association study is to search for the genetic influence of a disease, especially the location of genes affecting CD on the chromosome. We expect that our findings provide some evidence that the genetic component of CD may be probably located in specific loci.

Multiple Testing

For genome-wide statistical significance, we used conservative per-test significance level of $\alpha=5\times10^{-7}$ for genome-wide statistical significance. Meanwhile, a moderate criterion of 'suggestive association' with a cut-off $\alpha=10^{-3}$ was also applied. QVALUE was performed to determine the false discovery rate (FDR), which have been applied to microarray gene expression studies (Storey, 2002).

Fine-Mapping and Haplotype Analysis

Linkage disequilibrium (LD) refers to the nonrandom association of alleles at adjacent loci. LD mapping, which involved searching association between a disease and alleles or haplotypes at mapped marker loci, has been proven a powerful tool for locating disease genes. LD is an association between the genotypes at two or more loci, typically observed as a disease phenotype and marker genotype(s) due to the close physical proximity of the loci, i.e. the disease susceptibility locus to one or more marker loci. When a disease mutation first occurs at a locus, it is on a specific member of two homologous chromosomes and therefore associated with all variants at loci close by on the same chromosome.

Using HAPLOVIEW (Barrett, Fry, Maller, & Daly, 2005), we identified haplotype blocks (within a block, SNPs have strong LD each other) for interesting candidate genes or regions and chose several SNPs within blocks including the associated SNPs in the ADHD data. Using PBAT, we performed haplotyple analysis (using several flanking SNPs instead of single SNP) for SNPs within these candidate genes or regions.

CHAPTER 4

RESULTS

Assessment of Hardy-Weinberg Equilibrium (HWE)

By testing HWE, we removed the SNPs with $p<10^{-4}$ and those with minor allele frequency (MAF) < 0.01. Then, there were 11,120 SNPs left in the Affymetrix and 4,720 SNPs in the Illumina.

Family-Based Association Analyses

Genome-Wide Association Analysis in COGA Sample

For COGA data, no SNPs reached genome-wide significance ($p<5\times10^{-7}$). There were 20 SNPs have suggestive associations with CD ($p<10^{-3}$) 9 of which are located in known genes (Table 1). All 183 SNPs which had a p-value of less than 0.01 are listed in Appendix B (Affymetrix) and Appendix C (Illumina).

Table 1

SNPs Associated with CD Based on HW p-value>10^{-4} and MAF[a]>0.01 Using HAPLOVIEW and p-value<10^{-3} Using P2BAT/FBAT for COGA Data

Marker	Chr	Position[b] (bp)	Known Gene	HWpval	MAF	P
rs272411	19	59803539	*LILRA1*	0.2571	0.31199	0.0000316
rs1380381	7	16780013	*TSPAN13*	0.3082	0.1643	0.0000457

Table 1 (continued)

Marker	Chr	Position[b] (bp)	Known Gene	HWpval	MAF	P
rs1927724	13	98790313	*UBAC2*	0.4868	0.14873	0.0000467
rs1568452	2	57866337	----	0.259	0.37202	0.0000545
rs930983	11	122339624	----	0.3006	0.44052	0.0001426
rs1116327	11	96909790	----	0.4242	0.44598	0.0002086
rs1105009	9	102604800	----	1	0.03382	0.0002302
rs383902	15	56821466	*ADAM10*	0.817	0.31413	0.0003606
rs0725930	21	46014642	----	0.1165	0.10422	0.0006085
rs1366121	5	158098565	*EBF1*	0.3497	0.19126	0.000611
rs1986585	X	117393340	*WDR44*	0.0178	0.154	0.000627
rs953111	1	219929051	----	0.4488	0.35885	0.0007562
rs1883387	22	33536872	----	0.1128	0.28879	0.0007826
rs2262391	4	167341423	----	0.1515	0.19516	0.000785
rs903748	2	240817894	----	0.3013	0.17452	0.0008857
rs805308	2	54048438	*PSME4*	0.212	0.47532	0.000915
rs1381801	3	118723585	----	0.4017	0.38936	0.0009457
rs2053053	5	149589586	*CAMK2A*	0.1941	0.37642	0.0009779
rs720183	10	95141612	*FER1L3*	0.2851	0.14342	0.0009858
rs59232	21	40137420	----	0.6174	0.3512	0.0009934

[a] MAF refers to the minor allele frequency of the SNP.
[b] Position is based on NCBI Genome Build 36.3.

Replication Study in IMAGE Sample

Based on 20 SNPs with p $<10^{-3}$ in the COGA sample, we selected 332 SNPs from IMAGE dataset in order to replicate our results for CD. For the SNPs in COGA data within candidate genes, we chose all the SNPs within those genes while for each SNP in COGA data that are not located in known gene, we chose two flanking SNPs in the IMAGE dataset. In addition, candidate genes identified by Anney et al. (2008) that showed significant association at the level of α=0.01 in our study were also be included. Fourteen of the 332 SNPs were in HWE ($p>10^{-4}$) with MAF>0.01 and significantly ($p<0.05$) associated with CD (Table 2). Among these SNPs, the most significant one was rs4774309 on chromosome 15 (p=0.0016), which is located in the gene *ADAM10*.

Table 2

SNPs Associated with CD Based on HW p-value>10^{-4} and MAF[a]>0.01 Using HAPLOVIEW and p-value<0.05 Using PBAT for IMAGE Sample

Marker	Chr	Position[b]	Known Gene	HWpval	MAF	N[c]	P
rs4774309	15	56722756	*ADAM10*	0.001	0.198	101	0.0016
rs789560	12	68618094	*C12orf28*	0.815	0.13	90	0.003264
rs1152969	12	68599549	*C12orf28*	0.6436	0.036	24	0.004267
rs11187400	10	95119911	*FER1L3*	1	0.014	8	0.004678
rs1345610	5	158149617	*EBF1*	0.12	0.041	27	0.005346
rs11747044	5	158296035	*EBF1*	1	0.028	24	0.009322
rs1152962	12	68608074	*C12orf28*	0.9855	0.379	148	0.015258

Table 2 (continued)

Marker	Chr	Position[b]	Known Gene	HWpval	MAF	N[c]	P
rs1560044	5	158067098	*EBF1*	0.8821	0.344	144	0.015813
rs2241694	5	149582801	*CAMK2A*	0.125	0.082	67	0.018422
rs1240251	12	68510959	*C12orf28*	0.0042	0.152	88	0.024745
rs1674563	12	68529636	*C12orf28*	0.2378	0.499	158	0.032008
rs17056162	5	158109647	*EBF1*	0.8942	0.1	66	0.033895
rs919740	5	149646042	*CAMK2A*	0.0417	0.248	118	0.037881
rs10417589	19	59798475	*LILRA1*	1	0.1	67	0.046624

[a] MAF refers to the minor allele frequency of the SNP.
[b] Position is based on NCBI Genome Build 36.3.
[c] N refers to the number of informative families.

Multiple Testing

Based on the 10,808 p-values for associations with CD in the Affymetrix SNP panel in the COGA sample, the FDR was calculated to handle multiple comparisons. In terms of the q-value, when the p-value cutoff was set to be 0.001, we would expect 59.9% false positive results (FDR=0.599). If the p-value cutoff was set to be 0.0001, we would expect 15.6% false positive results (FDR=0.156).

Fine-Mapping and Haplotype Analysis

We chose SNPs that were in HWE ($p>10^{-4}$) with MAF>0.01 within *ADAM10* and *CAMK2A* from IMAGE sample to do fine-mapping. Among 17 SNPs in *ADAM10*, 13 of them were in five blocks and showed strong linkage disequilibrium (LD) (Figure 1).

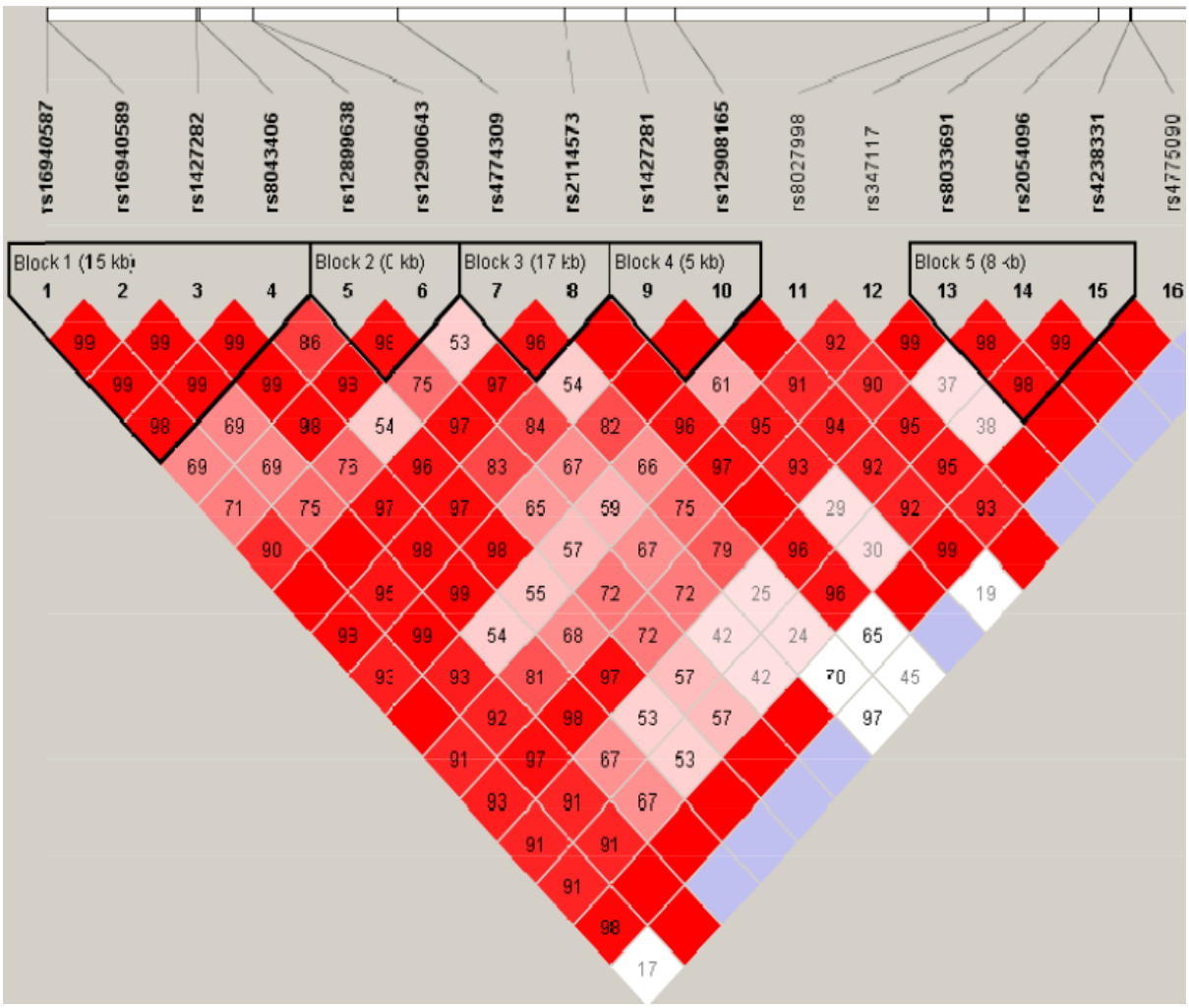

Figure 1. Pairwise LD between 17 SNPs within *ADAM10*, given by the D'.

Among 24 SNPs in *CAMK2A*, 6 of them were in two blocks with strong LD (Figure 2 and Figure 3).

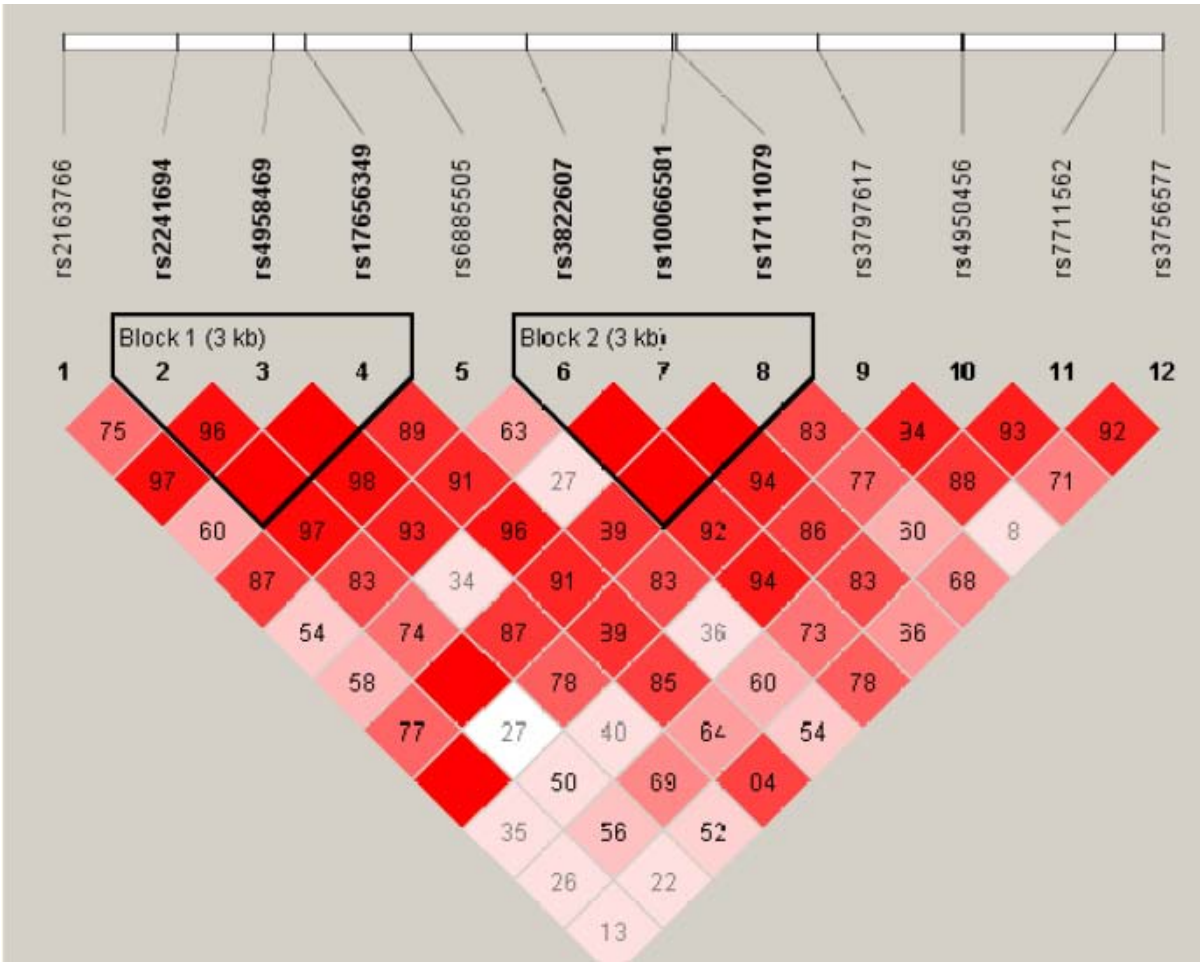

Figure 2. Pairwise LD between first 12 SNPs within *CAMK2A*, given by the D'.

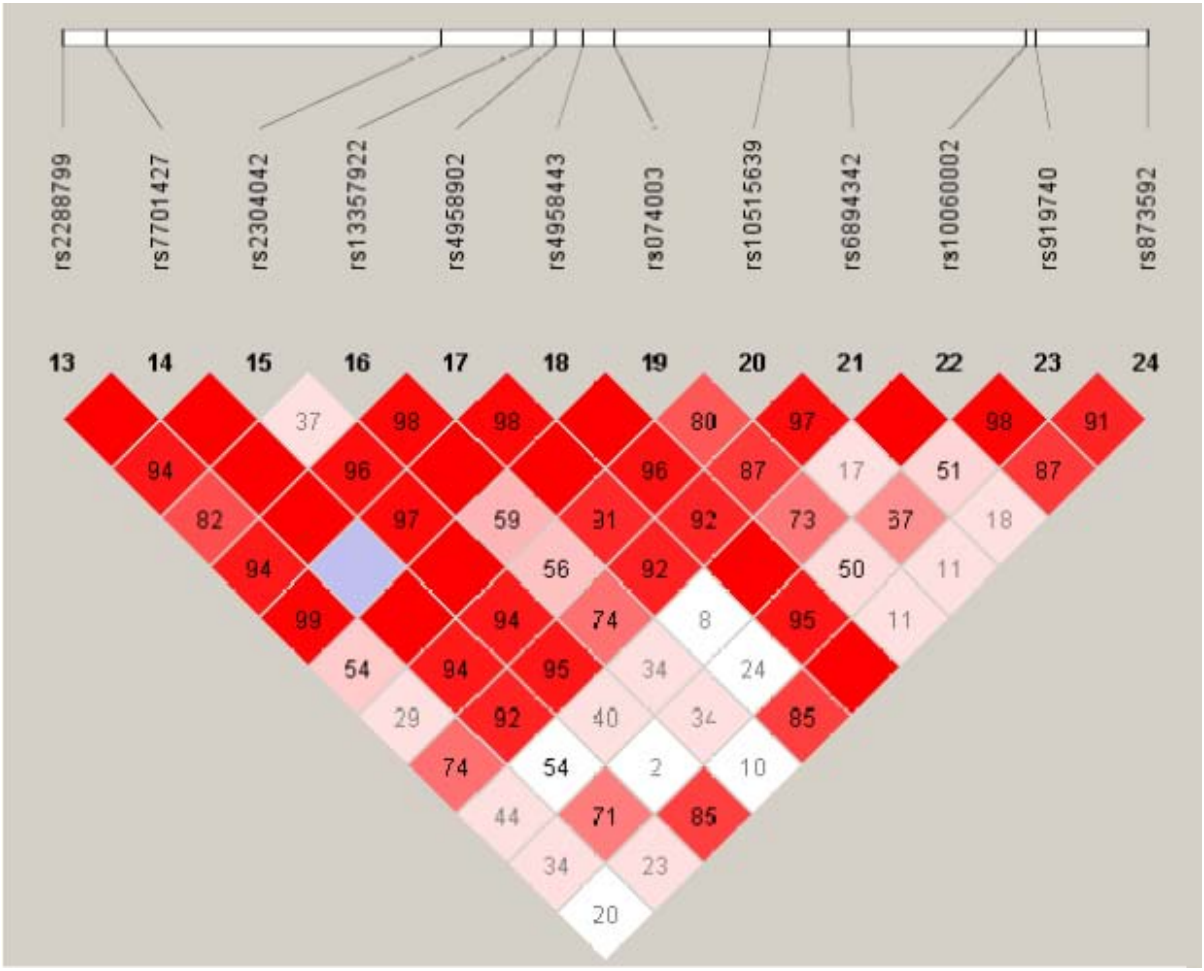

Figure 3. Pairwise LD between the rest 12 SNPs within *CAMK2A*, given by the D'.

We did haplotype analysis for each 2 adjacent SNPs of all 332 SNPs in IMAGE data. Within *ADAM10* and *CAMK2A*, there were 9 and 4 haplotypes significantly ($p<0.05$) associated with CD, respectively (Table 3).

Table 3

Haplotypes Associated with CD Based on p-value<0.05 Using PBAT for IMAGE Sample

Gene	SNPs	Haplotype	Hapfreq[a]	N[b]	p
ADAM10	rs4774309-rs2114573	G-T	0.08608	51	0.00000158

Table 3 (continued)

Gene	SNPs	Haplotype	Hapfreq[a]	N[b]	P
	rs12900643-rs4774309	T-G	0.08072	54	0.0000232
	rs4774309-rs2114573	C-T	0.7832	93	0.000389248
	rs12900643-rs4774309	T-C	0.60795	132	0.001386355
	rs16940587-rs16940589	T-C	0.01226	11	0.014305878
ADAM10	rs8027998-rs347117	G-C	0.02812	19	0.027223353
	rs8033691-rs2054096	T-T	0.15989	88	0.039222871
	rs1427282-rs8043406	A-G	0.06073	42	0.043205003
	rs8043406-rs12899638	A-G	0.05504	38	0.046393843
	rs10068882-rs919740	T-G	0.14629	78	0.00864467
	rs2241694-rs4958469	A-G	0.07962	67	0.024024611
CAMK2A	rs17656349-rs6885505	C-T	0.12141	72	0.03218256
	rs919740-rs873592	T-C	0.24418	116	0.04211522

[a] Hapfreq refers to the frequency of the haplotype.
[b] N refers to the number of informative families.

CHAPTER 5

DISCUSSION

By conducting family-based association studies, we replicated some of the results of previous studies and found some novel genetic variants of CD.

For COGA data, the most significant SNP from Affymetrix is rs272411 on chromosome 19 (p=0.0000316), and the most significant SNP from Illumina is rs1568452 on chromosome 2 (p=0.0000545), which are just the 2 chromosomes that were reported the highest risk to CD by Dick et al. (2004). SNP rs272411 is located in *LILRA1* but no association of this gene with any psychiatric or neurological disorder was reported so far.

Compared with the results of the study conducted by Anney et al. (2008), only one candidate gene (*C12orf28*, rs789560 with p=0.0000072) is supported by our analysis of COGA data. rs0720367 from Affymetrix and rs1240267 from Illumina are both located in *C12orf28*, with p=0.00246849 and 0.002924242, respectively. However, this gene is well supported by our replication using IMAGE data, in which 5 SNPs reached the statistical significance ($p<0.05$). *C12orf28* is an open reading frame on chromosome 12. But the function is still unknown.

Among all SNPs that had a p-value of less than 0.001, 9 are located on known genes. Unfortunately, none of them has been proved associated with CD. However, 2 genes attracted our attention. *ADAM10* on chromosome 15 (rs383902 with p=0.000360549) is a member of *ADAM* family that encodes cell surface proteins with a unique structure possessing both potential adhesion and protease domains. *ADAM10* has been proved as an important role in

amyloid precursor protein proteolysis, which is a key event in the pathogenesis of Alzheimer's disease (AD). Colciaghi et al. (2002) reported a reduced level of *ADAM10* in AD patients' platelets. Contradictorily, Gatta, Albertini, Ravid, and Finazzi (2002) reported a two-fold higher of *ADAM10* mRNA levels in AD patients than controls. The identification of possible association of *ADAM10* with CD may lead to further study of the relationship among *ADAM10*, CD, and AD.

CAMK2A on chromosome 5 (rs2053053 with p=0.000977901) encodes the alpha subunit of Calcium/calmodulin-dependent protein kinase II, which is a ubiquitous serine/threonine protein kinase that is abundant in the brain as a major constituent of the postsynaptic density. Altered expression of *CAMK2A* has been reported associated with bipolar disorder and depression. For example, there is a significant decrease in *CAMK2A* mRNA in bipolar disorder patients (Xing et al., 2002) and a 29% increase in depression patients (Novak, Seeman, & Tallerico, 2006). In order to find out whether there is shared pathogenesis of CD, bipolar disorder, and depression more studies are needed. Although there is no report of association of *CAMK2A* with CD in humans, an animal model has been established. Chen, Rainnie, Greene, and Tonegawa (1994) observed a decreased fear response and an increased aggressive behavior in heterozygous mutant mice deficient for *CAMK2A* that provided some evidence for *CAMK2A* effect on the human psychiatric diseases involving increased risk-taking behaviors, including CD.

Using part of IMAGE data, we replicated our results for most known genes identified in COGA data, including *ADAM10* (rs4774309 with p=0.0016) and *CAMK2A* (rs2241694 with p=0.018422, rs919740 with p=0.037881). When doing two-marker haplotype analysis instead

of single SNP analysis, we found that within gene *ADAM10*, the 4 most significant haplotypes all include SNP rs4774309, which had been identified most significantly associated with CD using in IMAGE sample. Meanwhile, the 2 most significant haplotypes within gene *CAMK2A* just contain 2 SNPs rs919740 and rs2241694 that each showed significant association with CD in our previous single marker analysis. These results provide further evidence that *ADAM10* and *CAMK2A* have association with CD.

Besides the above-discussed 2 genes, there are still other genes that have been replicated by our analysis of IMAGE sample. For example, 4 SNPs were in the gene *EBF1* (early B-cell factor 1) on chromosome 5. This gene has been proved one of the genes that are subject to random monoallelic expression (Gimelbrant, Hutchinson, Thompson, & Chess, 2007). But no psycho-disorders have been reported to be associated with this gene so far.

In our analysis of COGA data, 3 other genes that contain SNPs associated with CD at the level of α=0.01 are of interest to us. *ROBO2* on chromosome 3 (rs876675 with p=0.001593182) encodes a receptor for SLIT2, which functions in axon guidance and cell migration. Besides vesicoureteral reflux type 2, defects in this gene may lead to autism (Anitha et al., 2008) and schizophrenia (Potkin et al., 2009). But the results require confirmation in independent samples.

NTRK2 on chromosome 9 (rs1778970 with p=0.002368143 and rs1838158 with p=0.007260911) encodes the receptor for brain-derived neurotrophic factor. Mutations in this gene are associated with mood disorders, including nicotine dependence (Beuten et al., 2007; Vink et al., 2009), alcohol dependence (Xu et al., 2007), obsessive-compulsive disorder (Alonso et al., 2008), Alzheimer's disease (Chen et al., 2008; Cozza et al., 2008), and bipolar

disorder (Smith et al., 2009). Because the data we used were collected initially for alcohol dependence, the association of this gene with CD may due to confounding.

MME on chromosome 3 (rs1025192 with p=0.007969908) encodes a common acute lymphocytic leukemia antigen that is an important cell surface marker in the diagnosis of human acute lymphocytic leukemia. However, this protein is not restricted to leukemic cells and can be found in many normal tissues. Neurologists have focused on the relationship between this gene and Alzheimer's disease, but the results were inconsistent. Positive results have been reported in Finnish (Helisalmi et al., 2004) and Chinese sample (Shi et al., 2005), while no association was observed in Japanese (Oda et al., 2002), Swedish (Lilius et al., 2003), and a south Chinese population (Fu et al., 2009).

By comparing our study with existed studies, we found some similarities and dissimilarities. First, our results provide support to the findings of Dick et al. (2004) in a linkage analysis. We discovered that the associations of SNPs provided by Affymetrix and Illumina are most significant on chromosome 19 and chromosome 2, respectively, while Dick et al. suggested regions on these two chromosomes confer highest risk of CD. Second, we performed a family-based, genome-wide association analysis, which were the same design as what Anney et al. (2008) did. However, we used different sample, with the focus on the identification of novel genetic variants, not only using a more powerful family-based method in P2BAT, but also dealing with X-chromosome SNPs. In addition, to reduce the genetic heterogeneity, we just used 122 Caucasian pedigrees (292 nuclear families) which have a majority of individuals who self-reported to be 'White' (both Hispanic and non-Hispanic). In our analysis, we found neither significant genome-wide association ($p<5\times10^{-7}$) nor strong genome-wide association

($p<10^{-5}$) of SNPs with CD, probably due to much less amount of SNPs we used (16,312 SNPs), thus limiting the coverage of the genome. Although using a much less stringent criterion (α=0.01), only one candidate gene *C12orf28* identified by Anney et al. (2008) is supported by our analysis. However, we found 2 novel, suggestive genes, *ADAM10* and *CAMK2A*, which was not reported by them or other previous studies. These 2 candidate genes were also proved by replicating family-based analysis and haplotype analysis using part of IMAGE sample.

From a public health point of view, potential application of genetic association studies includes that if specific genes are further confirmed, screening for these genes may be possible in newborns and other populations to detect people of high risk before the disease occurs. Because within each genetic variant that is associated with one complex disease like CD, the allele with relatively higher risk is known, newborn and population screening of these specific loci may differentiate those with high risk alleles from those with low risk alleles, thus narrowing down the range of population who need further follow up and intervention.

One important goal of genetic epidemiology study is to look at gene-environment interaction affecting complex diseases like CD after the main effects of genes have been detected. The present study focused on the identification of the associated genes with CD. However, several studies have shown that environmental factors and their interactions with genetic variants contribute to CD. For example, Caspi et al. (2002) reported an interaction between the level of *MAOA* expression and maltreatment of white male children, suggesting that high levels of *MAOA* may moderate the effect of maltreatment on the development of antisocial behavior. Later, Foley et al. (2004) replicated his findings in white male CD patients that genotypes associated with low *MAOA* increased risk for CD only in the presence of

adverse child environment. Recently, Widom and Brzustowicz (2006) extended their discoveries to show that the *MAOA* gene interacted with child abuse and neglect in both white boys and girls on the development of violent and antisocial behavior. Other environmental factors were also tested. Braun et al. (2008) examined the association of tobacco smoke and environmental lead exposure with CD, suggesting that prenatal tobacco exposure and environmental lead exposure contribute substantially to CD in U.S. children. Very recently, Hay, Pawlby, Waters, Perra, and Sharp (2010) reported that mothers' depression in pregnancy placed their children at a twofold risk for antisocial outcomes, and at a fourfold risk for violent behavior, indicating that mothers' antenatal depression as a predictor of their children's antisocial outcomes. These studies provided sufficient evidence that in searching for the pathogenesis of CD, environmental factors cannot be neglected.

Several limitations of our study merit comment. First, just as Dick et al. (2004) mentioned, the COGA sample was selected through alcoholic probands. In order to yield more accurate support of gene loci contributing to CD, a sample particularly collected for the purpose of CD is needed. Second, the sample size is relatively small and the two SNP panels have limited coverage of the genome. These cause the inability to detect weak to moderate genetic effect (Dick et al., 2004). Third, the genes detected in our study only reached suggestive evidence of association, which need to be replicated in independent samples.

In spite of the above limitations, the present study has several strengths. First, rather than candidate gene studies, which select genes for study based on known or suspected disease mechanisms, we performed GWA studies. By scanning the whole genome, GWA studies have the potential to identify totally novel loci for CD. Second, we used a family-based design

instead of a population based case-control design. This reduces the type 1 error arising from population stratification. Third, by limiting analyses to Caucasian pedigrees, we minimized risk of type 1 error due to genetic heterogeneity.

In summary, this study identified several CD associated genetic variants, especially two genes, *ADAM10* and *CAMK2A* that have been shown in previous studies to affect several other psychiatric or neurological diseases such as Alzheimer's disease, bipolar disorder, and depression. These findings may serve as a resource for replication in other populations and provide a foundation for future investigations. Future studies are warranted to verify the strength of association between the genetic factors and CD and gene-environment interaction on the development of CD.

REFERENCES

Alonso P., Gratacòs M., Menchón J. M., Saiz-Ruiz J., Segalàs C., Baca-García E., et al. (2008). Extensive genotyping of the BDNF and NTRK2 genes define protective haplotypes against obsessive-compulsive disorder. *Biol Psychiatry* 63:619-28.

American Psychiatric Association (1994). Diagnostic and statistical manual of mental disorders. (4th ed.) Washington, DC: American Psychiatric Association.

Anitha A., Nakamura K., Yamada K., Suda S., Thanseem I., Tsujii M., et al. (2008). Genetic analyses of roundabout (ROBO) axon guidance receptors in autism. *Am J Med Genet B Neuropsychiatr Genet* 147B:1019-27.

Anney R. J., Lasky-Su J., O'Dúshláine C., Kenny E., Neale B. M., Mulligan A., et al. (2008). Conduct disorder and ADHD: Evaluation of conduct problems as a categorical and quantitative trait in the international multicentre ADHD genetics study. *Am J Med Genet B Neuropsychiatr Genet* 147B:1369-78.

Barrett J. C., Fry B., Maller J., & Daly M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263-5.

Beuten J., Ma J. Z., Payne T. J., Dupont R. T., Lou X. Y., Crews K. M., et al. (2007). Association of specific haplotypes of neurotrophic tyrosine kinase receptor 2 gene (NTRK2) with vulnerability to nicotine dependence in African-Americans and European-Americans. *Biol Psychiatry* 61:48-55.

Biederman J., & Faraone S. V. (2005). Attention-deficit hyperactivity disorder. *Lancet* 366:237-48.

Biederman J., Newcorn J., & Sprich S. (1991). Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety, and other disorders. *Am J Psychiatry* 148:564-77.

Braun J. M., Froehlich T. E., Daniels J. L., Dietrich K. N., Hornung R., Auinger P., et al. (2008). Association of environmental toxicants and conduct disorder in U.S. children: NHANES 2001-2004. *Environ Health Perspect* 116:956-62.

Bucholz K. K., Cadoret R., Cloninger C. R., Dinwiddie S. H., Hesselbrock V. M., Nurnberger J. I. Jr., et al. (1994). A new, semi-structured psychiatric interview for use in genetic linkage studies: A report on the reliability of the SSAGA. *J Stud Alcohol* 55:149-58.

Cadoret R. J., Langbehn D., Caspers K., Troughton E. P., Yucuis R., Sandhu H. K., et al. (2003). Associations of the serotonin transporter promoter polymorphism with aggressivity, attention deficit, and conduct disorder in an adoptee population. *Compr Psychiatry* 44:88-101.

Caspi A., Langley K., Milne B., Moffitt T. E., O'Donovan M., Owen M. J., et al. (2008). A replicated molecular genetic basis for subtyping antisocial behavior in children with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* 65:203-10.

Caspi A., McClay J., Moffitt T. E., Mill J., Martin J., Craig I. W., et al. (2002). Role of genotype in the cycle of violence in maltreated children. *Science* 297:851-4.

Chen C., Rainnie D. G., Greene R. W., & Tonegawa S. (1994). Abnormal fear response and aggressive behavior in mutant mice deficient for alpha-calcium-calmodulin kinase II. *Science* 266:291-4.

Chen Z., Simmons M. S., Perry R. T., Wiener H. W., Harrell L. E., Go R. C. (2008). Genetic association of neurotrophic tyrosine kinase receptor type 2 (NTRK2) with Alzheimer's disease. *Am J Med Genet B Neuropsychiatr Genet* 147:363-9.

Colciaghi F., Borroni B., Pastorino L., Marcello E., Zimmermann M., Cattabeni F., et al. (2002). [alpha]-Secretase ADAM10 as well as [alpha]APPs is reduced in platelets and CSF of Alzheimer disease patients. *Mol Med* 8:67-74.

Cozza A., Melissari E., Iacopetti P., Mariotti V., Tedde A., Nacmias B., et al. (2008). SNPs in neurotrophin system genes and Alzheimer's disease in an Italian population. *J Alzheimers Dis* 15:61-70.

Deas D., & Brown E. S. (2006). Adolescent substance abuse and psychiatric comorbidities. *J Clin Psychiatry* 67:e02.

Dick D. M., Bierut L., Hinrichs A., Fox L., Bucholz K. K., Kramer J., et al. (2006). The role of GABRA2 in risk for conduct disorder and alcohol and drug dependence across developmental stages. *Behav Genet*. 36:577-90.

Dick D. M., Li T. K., Edenberg H. J., Hesselbrock V., Kramer J., Kuperman S., et al. (2004). A genome-wide screen for genes influencing conduct disorder. *Mol Psychiatry* 9:81-6.

Dick D. M., Viken R. J., Kaprio J., Pulkkinen L., Nurnberger Jr J. I., & Rose R. J. (2002). Adolescent alcohol use and conduct disorder: Age 14 interview data from finnish twin STUDIES. Annual Meeting of the Behavior Genetics Association; Keystone, CO.

Edenberg H. J., Bierut L. J., Boyce P., Cao M., Cawley S., Chiles R., et al. (2005). Description of the data from the Collaborative Study on the Genetics of Alcoholism (COGA) and

single-nucleotide polymorphism genotyping for Genetic Analysis Workshop 14. *BMC Genet 6* Suppl 1:S2.

Foley D. L., Eaves L. J., Wormley B., Silberg J. L., Maes H. H., Kuhn J., et al. (2004). Childhood adversity, monoamine oxidase a genotype, and risk for conduct disorder. *Arch Gen Psychiatry* 61:738-44.

Fu Y., Li A. F., Shi J. J., Tang M. N., Guo Y. B., & Zhao Z. H. (2009). Lack of association of neprilysin gene polymorphisms with Alzheimer's disease in a southern Chinese community. *Int Psychogeriatr* 21:354-8.

Gatta L. B., Albertini A., Ravid R., & Finazzi D. (2002). Levels of beta-secretase BACE and alpha-secretase ADAM10 mRNAs in Alzheimer hippocampus. *Neuroreport* 13:2031-3.

Gimelbrant A., Hutchinson J. N., Thompson B. R., & Chess A. (2007). Widespread monoallelic expression on human autosomes. *Science* 318:1136-40.

Goldstein R. B., Prescott C. A., & Kendler K. S. (2001). Genetic and environemental factors in conduct problems and adult antisocial behavior among adult female twins. *J Nerv Ment Dis* 189:201-9.

Guessous I., Gwinn M., & Khoury M. J. (2009). Genome-wide association studies in pharmacogenomics: Untapped potential for translation. *Genome Med* 1:46.

Hardy J., & Singleton A. (2009). Genomewide association studies and human disease. *N Engl J Med* 360:1759-68.

Hay D. F., Pawlby S., Waters C. S., Perra O., & Sharp D. (2010). Mothers' antenatal depression and their children's antisocial outcomes. *Child Dev* 81:149-65.

Helisalmi S., Hiltunen M., Vepsäläinen S., Iivonen S., Mannermaa A., Lehtovirta M., et al. (2004). Polymorphisms in neprilysin gene affect the risk of Alzheimer's disease in Finnish patients. *J Neurol Neurosurg Psychiatry* 75:1746-8.

Hesselbrock M., Easton C., Bucholz K. K., Schuckit M., & Hesselbrock V. (1999). A validity study of the SSAGA--a comparison with the SCAN. *Addiction* 94:1361-70.

Hicks B. M., Krueger R. F., Iacono W. G., McGue M., & Patrick C. J. (2004). Family transmission and heritability of externalizing disorders: A twin-family study. *Arch Gen Psychiatry* 61:922-8.

Hoffmann T., & Lange C. (2006). P2BAT: A massive parallel implementation of PBAT for genome-wide association studies in R. *Bioinformatics* 22:3103-5.

Kendler K. S., Kuo P. H., Webb B. T., Kalsi G., Neale M. C., Sullivan P. F., et al. (2006). A joint genomewide linkage analysis of symptoms of alcohol dependence and conduct disorder. *Alcohol Clin Exp Res* 30:1972-7.

Lake S. L., Blacker D., & Laird N. M. (2000). Family-based tests of association in the presence of linkage. *Am J Hum Genet* 67:1515-25.

Lander E. S., & Schork N. J. (1994). Genetic dissection of complex traits. *Science* 265:2037-48.

Lilius L., Forsell C., Axelman K., Winblad B., Graff C., & Tjernberg L. (2003). No association between polymorphisms in the neprilysin promoter region and Swedish Alzheimer's disease patients. *Neurosci Lett* 337:111-3.

Malmberg K., Wargelius H. L., Lichtenstein P., Oreland L., & Larsson J. O. (2008). ADHD and Disruptive Behavior scores-associations with MAO-A and 5-HTT genes and with platelet MAO-B activity in adolescents. *BMC Psychiatry* 8:28.

Mayo O. (2008). A century of Hardy-Weinberg equilibrium. *Twin Res Hum Genet* 11:249-56.

McCarthy M. I., Abecasis G. R., Cardon L. R., Goldstein D. B., Little J., Ioannidis J. P., et al. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat Rev Genet* 9:356-69.

Miller G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM* Vol. 38, No. 11: 39-41.

Monuteaux M. C., Biederman J., Doyle A. E., Mick E., & Faraone S. V. (2009). Genetic risk for conduct disorder symptom subtypes in an ADHD sample: Specificity to aggressive symptoms. *J Am Acad Child Adolesc Psychiatry* 48:757-764.

Novak G., Seeman P., & Tallerico T. (2006). Increased expression of calcium/calmodulin-dependent protein kinase IIbeta in frontal cortex in schizophrenia and depression. *Synapse* 59:61-68.

Oda M., Morino H., Maruyama H., Terasawa H., Izumi Y., Torii T., et al. (2002). Dinucleotide repeat polymorphisms in the neprilysin gene are not associated with sporadic Alzheimer's disease. *Neurosci Lett* 320:105-7.

Potkin S. G., Turner J. A., Guffanti G., Lakatos A., Fallon J. H., Nguyen D. D., et al. (2009). A genome-wide association study of schizophrenia using brain activation as a quantitative phenotype. *Schizophr Bull* 35:96-108.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A., Bender D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559-75.

Rabinowitz D., & Laird N. (2000). A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum Hered* 50:211-23.

Rhee S. H., & Waldman I. D. (2002). Genetic and environmental influences on antisocial behavior: A meta-analysis of twin and adoption studies. *Psychol Bull* 128:490-529.

Risch N., & Merikangas K. (1996). The future of genetic studies of complex human diseases. *Science* 273:1516-7.

Risch N. J. (2000). Searching for genetic determinants in the new millennium. *Nature* 405:847-56.

Sakai J. T., Lessem J. M., Haberstick B. C., Hopfer C. J., Smolen A., Ehringer M. A., et al. (2007). Case-control and within-family tests for association between 5HTTLPR and conduct problems in a longitudinal adolescent sample. *Psychiatr Genet* 17:207-14.

Sakai J. T., Stallings M. C., Crowley T. J., Gelhorn H. L., McQueen M. B., Ehringer M. A. (2010). Test of association between GABRA2 (SNP rs279871) and adolescent conduct/alcohol use disorders utilizing a sample of clinic referred youth with serious substance and conduct problems, controls and available first degree relatives. *Drug Alcohol Depend* 106:199-203.

Sakai J. T., Young S. E., Stallings M. C., Timberlake D., Smolen A., Stetler G. L., et al. (2006). Case-control and within-family tests for an association between conduct disorder and 5HTTLPR. *Am J Med Genet B Neuropsychiatr Genet* 141B:825-32.

Searight H. R., Rottnek F., & Abby S. L. (2001). Conduct disorder: Diagnosis and treatment in primary care. *Am Fam Physician* 63:1579-88.

Seng K. C., & Seng C. K. (2008). The success of the genome-wide association approach: A brief story of a long struggle. *Eur J Hum Genet* 16:554-64.

Sengupta S. M., Grizenko N., Schmitz N., Schwartz G., Ben Amor L., Bellingham J., et al. (2006). COMT Val108/158Met gene variant, birth weight, and conduct disorder in children with ADHD. *J Am Acad Child Adolesc Psychiatry* 45:1363-9.

Shi J., Zhang S., Tang M., Ma C., Zhao J., Li T., et al. (2005). Mutation screening and association study of the neprilysin gene in sporadic Alzheimer's disease in Chinese persons. *J Gerontol A Biol Sci Med Sci* 60:301-6.

Slutske W. S., Heath A. C., Dinwiddie S. H., Madden P. A., Bucholz K. K., Dunne M. P., et al. (1997). Modeling genetic and environmental influences in the etiology of conduct disorder: A study of 2,682 adult twin pairs. *J Abnorm Psychol* 106:266-79.

Smith E. N., Bloss C. S., Badner J. A., Barrett T., Belmonte P. L., Berrettini W., et al. (2009). Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry* 14:755-63.

Spielman R. S., McGinnis R. E., & Ewens W. J. (1993). Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506-16.

Stallings M. C., Corley R. P., Dennehey B., Hewitt J. K., Krauter K. S., Lessem J. M., et al. (2005). A genome-wide search for quantitative trait Loci that influence antisocial drug dependence in adolescence. *Arch Gen Psychiatry* 62:1042-51.

Steiner H. (1997). Practice parameters for the assessment and treatment of children and adolescents with conduct disorder. American Academy of Child and Adolescent Psychiatry. *J Am Acad Child Adolesc Psychiatry* 36:122S-39S.

Storey J. D. (2002). A direct approach to false discovery rates. *J R Stat Soc Series B Stat Methodol* 64:479-98.

Subbarao A., Rhee S. H., Young S. E., Ehringer M. A., Corley R. P., & Hewitt J. K. (2008). Common genetic and environmental influences on major depressive disorder and conduct disorder. *J Abnorm Child Psychol* 36:433-44.

Thapar A., Langley K., Fowler T., Rice F., Turic D., Whittinger N., et al. (2005). Catechol O-methyltransferase gene variant and birth weight predict early-onset antisocial behavior in children with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* 62:1275-8.

True W. R., Heath A. C., Scherrer J. F., Xian H., Lin N., Eisen S. A., et al. (1999). Interrelationship of genetic and environmental influences on conduct disorder and alcohol and marijuana dependence symptoms. *Am J Med Genet*. 88:391-7.

Vink J. M., Smit A. B., de Geus E. J., Sullivan P., Willemsen G., Hottenga J. J., et al. (2009). Genome-wide association study of smoking initiation and current smoking. *Am J Hum Genet* 84:367-79.

Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661-78.

Widom C. S., & Brzustowicz L. M. (2006). MAOA and the "cycle of violence:" Childhood abuse and neglect, MAOA genotype, and risk for violent and antisocial behavior. *Biol Psychiatry* 60:684-9.

Xing G., Russell S., Hough C., O'Grady J., Zhang L., Yang S., et al. (2002). Decreased prefrontal CaMKII alpha mRNA in bipolar illness. *Neuroreport* 13:501-5.

Xu K., Anderson T. R., Neyer K. M., Lamparella N., Jenkins G., Zhou Z., et al. (2007). Nucleotide sequence variation within the human tyrosine kinase B neurotrophin receptor gene: Association with antisocial alcohol dependence. *Pharmacogenomics J* 7:368-79.

APPENDIXES

Appendix A. Glossary

Term	Definition
Allele	Either of a pair (or series) of alternative forms of a gene that can occupy the same locus on a particular chromosome and that control the same character (Miller, 1995).
Single Nucleotide Polymorphism (SNP)	Genetic variation in a DNA sequence that occurs when a single nucleotide in a genome is altered (Miller, 1995).
Genome-wide Association Study (GWAS)	Test of the association between markers across the genome and disease (Hardy and Singleton, 2009).
Transmission Disequilibrium Test (TDT)	A family-based association test which considers parents who are heterozygous for an allele associated with disease and evaluates the frequency with which that allele or its alternate is transmitted to affected offspring (Spielman et al., 1993).
Hardy-Weinberg Equilibrium (HWE)	The state of the genotypic frequency of two alleles of one autosomal gene locus after one discrete generation of random mating in an indefinitely large population (Mayo, 2008).
Haplotype	A combination of alleles that are located closely together on the same chromosome and that tend to be inherited together (Miller, 1995).

Appendix B. SNPs Associated with CD Based on the p-value for P2BAT/FBAT (p<0.01) in COGA Data (Affymetrix)

Chr	Marker	Position	Known Gene	HWpval	MAF	P
1	rs0953111	219929051	----	0.4488	0.359	0.0007562
1	rs1343817	111238652	*CD53*	0.5367	0.471	0.00108088
1	rs1926256	69651662	----	1	0.154	0.00417073
1	rs1441834	29945688	----	0.02237	0.237	0.00438844
1	rs0826415	117955500	*FAM46C*	0.8498	0.298	0.00515691
1	rs2095785	99971477	*FRRS1*	0.3636	0.122	0.00568108
1	rs1415260	160351771	*NOS1AP*	0.1799	0.442	0.00912208
2	rs1406418	51609235	----	1	0.015	0.00109905
2	rs2013708	71453862	*ZNF638*	0.08115	0.117	0.0012599
2	rs0956099	53516816	----	0.7489	0.459	0.00130159
2	rs1530272	88196333	----	0.5928	0.388	0.00193553
2	rs2139053	58010043	----	0.7418	0.377	0.00274309
2	rs0951431	117865700	----	0.4791	0.223	0.00403683
2	rs1516238	117865378	----	0.4791	0.223	0.00403683
2	rs0720796	40035896	*LOC100128590*	0.03377	0.464	0.00425925
2	rs1399959	172140552	----	0.0512	0.302	0.00463357
2	rs2078542	148122694	----	0.8616	0.351	0.00471943
2	rs1907544	116966960	----	0.8227	0.315	0.00585939
2	rs1448836	168224982	----	0.8391	0.267	0.00597448

Appendix B (continued)

Chr	Marker	Position	Known Gene	HWpval	MAF	P
2	rs1376586	199759635	----	0.8681	0.459	0.00622715
3	rs1381801	118723585	----	0.4017	0.389	0.00094571
3	rs0952384	116387931	----	0.5991	0.112	0.00137034
3	rs0721779	60604879	----	1	0.199	0.00220199
3	rs1398922	18645164	----	0.03858	0.344	0.00391375
3	rs1554560	154857857	----	0.3955	0.49	0.00530724
3	rs2083948	106828663	----	1	0.145	0.00581919
3	rs1401965	152942197	*AADACL2*	0.241	0.26	0.00691127
3	rs0768324	85222091	----	0.3791	0.075	0.00840476
3	rs1395166	31153785	----	0.3822	0.175	0.00864825
3	rs0951015	24028796	----	0.4924	0.337	0.008885
4	rs2262391	167341423	----	0.1515	0.195	0.000785
4	rs1373475	36607819	----	0.4687	0.16	0.00204416
4	rs0952938	63436594	----	0.4204	0.458	0.0022952
4	rs0725001	125407498	----	0.07951	0.465	0.00458047
4	rs4133280	130406957	----	1	0.122	0.00537809
4	rs1822528	99890067	----	0.6192	0.269	0.00658844
4	rs0939353	25204915	----	0.2713	0.241	0.00685781
4	rs1373053	32879410	----	1	0.028	0.00691539
4	rs1443786	83129627	----	1	0.286	0.00809744

Appendix B (continued)

Chr	Marker	Position	Known Gene	HWpval	MAF	P
4	rs1357032	28572026	----	0.5061	0.408	0.00843092
5	rs1366121	158098565	*EBF1*	0.3497	0.191	0.00061095
5	rs0032524	13282435	----	1	0.182	0.00284293
5	rs701380	79812643	----	1	0.065	0.00487153
5	rs0724865	73826179	----	1	0.113	0.00539712
5	rs1914212	160313280	----	0.1444	0.234	0.00546864
5	rs0324066	40952422	*C7*	0.7306	0.407	0.00621991
5	rs0723852	57466589	----	0.3114	0.071	0.00876685
6	rs0721101	155674536	*TFB1M*	0.653	0.269	0.00100394
6	rs1409098	53990299	*C6orf142*	0.4152	0.421	0.00434218
6	rs0950995	155662744	*TFB1M*	0.7815	0.158	0.00670292
6	rs2327212	132920951	----	1	0.149	0.00724124
6	rs0951318	1263711	----	1	0.14	0.00751216
7	rs1380381	16780013	*TSPAN13*	0.3082	0.164	0.0000457
7	rs1443753	121736554	----	0.5364	0.323	0.0013715
7	rs0719340	133997674	*BPGM*	0.2909	0.394	0.00251761
7	rs0727430	8823441	----	0.04496	0.378	0.00371915
7	rs2392383	35760695	----	0.8657	0.413	0.00404995
8	rs1396977	131791884	----	0.871	0.431	0.0015906
8	rs0344288	63029421	----	0.7424	0.465	0.00207002

Appendix B (continued)

Chr	Marker	Position	Known Gene	HWpval	MAF	P
8	rs2137636	114551461	----	0.7358	0.208	0.00242161
8	rs0344285	63029703	----	0.2564	0.338	0.00259531
8	rs2203657	63034569	----	0.8681	0.465	0.00265834
8	rs0684872	63034323	----	0.8686	0.465	0.00316347
8	rs1381665	76930139	----	1	0.267	0.00350421
8	rs0958374	9235071	----	0.7507	0.496	0.00588367
8	rs1491467	127032094	*LOC100130231*	1	0.049	0.00596978
8	rs0344286	63029664	----	0.6913	0.33	0.00955546
9	rs1105009	102604800	----	1	0.034	0.00023016
9	rs1856203	113532633	*C9orf84*	1	0.218	0.00178912
9	rs2383292	23783644	*ELAVL2*	1	0.01	0.00419286
9	rs1838158	86662829	*NTRK2*	0.3026	0.137	0.00726091
9	rs1419253	19882748	----	0.04635	0.361	0.00802226
9	rs0439269	14481347	----	0.7406	0.124	0.00827884
9	rs3914504	30738943	----	0.4116	0.199	0.00918118
9	rs0063319	74714604	*ALDH1A1*	1	0.446	0.00936777
10	rs0720183	95141612	*FER1L3*	0.2851	0.143	0.00098577
10	rs1408342	102638742	----	0.6776	0.109	0.00231688
10	rs0829108	95395414	*PDE6C*	0.6362	0.231	0.00494065
10	rs0958852	61992624	----	0.565	0.253	0.00587419

Appendix B (continued)

Chr	Marker	Position	Known Gene	HWpval	MAF	P
10	rs0952965	108187118	----	0.009377	0.441	0.00743779
10	rs0604489	95395398	*PDE6C*	0.8256	0.235	0.00853557
10	rs0723009	61992601	----	0.699	0.261	0.00983909
11	rs1116327	96909790	----	0.4242	0.446	0.00020861
11	rs2097160	96909766	----	0.1328	0.401	0.00134155
11	rs0724202	26532254	*TMEM16C*	0.05779	0.295	0.00180448
11	rs1939546	59505456	*LOC255649*	0.6697	0.241	0.00397862
11	rs2170655	18977997	----	0.8739	0.496	0.00622446
11	rs1391610	5384031	----	0.8254	0.416	0.00664355
11	rs0951933	78800124	----	0.3451	0.189	0.0091399
11	rs1404501	63041086	*LGALS12*	0.2651	0.244	0.00948556
12	rs1369035	81928655	*TMTC2*	0.1703	0.14	0.00162996
12	rs0710741	68315603	----	0.4566	0.222	0.00174063
12	rs0720367	68568050	*C12orf28*	0.3065	0.217	0.00246849
12	rs2029692	65279429	*GRIP1*	0.5088	0.154	0.0031165
12	rs1295815	74494077	----	0.2284	0.225	0.00375672
12	rs0726603	44378810	----	0.6046	0.38	0.00608295
12	rs0035426	114054429	----	0.06322	0.392	0.00995411
13	rs1927724	98790313	*UBAC2*	0.4868	0.149	0.0000467
13	rs2149144	105246932	----	0.2132	0.11	0.00326126

Appendix B (continued)

Chr	Marker	Position	Known Gene	HWpval	MAF	P
13	rs1998676	32884219	----	1	0.022	0.00815097
14	rs1950702	32110983	*AKAP6*	0.6571	0.258	0.00306025
14	rs1959619	19490525	----	1	0.325	0.00546271
14	rs1074925	94931266	----	0.5279	0.166	0.00693578
14	rs1959482	70197453	*TTC9*	0.499	0.26	0.00733144
15	rs0717552	84884126	*AGBL1*	0.145	0.454	0.00782124
16	rs2363306	88696275	*LOC728262*	1	0.068	0.00471772
17	rs0953113	22839015	*KSR1*	0.4206	0.446	0.00696989
17	rs1807333	895321	*ABR*	0.3146	0.327	0.00937469
18	rs0959655	48132862	*DCC*	1	0.159	0.00195585
18	rs1073006	1870289	----	0.4678	0.131	0.00238622
19	rs0272411	59803539	*LILRA1*	0.2571	0.312	0.0000316
19	rs3810261	54914726	----	0.346	0.135	0.00874232
20	rs0727662	50708872	----	0.182	0.403	0.00235992
20	rs0715433	50712238	----	0.1013	0.37	0.00315178
20	rs2208970	12325493	----	1	0.336	0.00947568
21	rs0725930	46014642	----	0.1165	0.104	0.00060846
21	rs0059232	40137420	----	0.6174	0.351	0.00099339
21	rs0966179	18808558	----	0.004548	0.166	0.00505555
21	rs0978422	18764448	----	0.3715	0.492	0.00595544

Appendix B (continued)

Chr	Marker	Position	Known Gene	HWpval	MAF	P
21	rs2226672	24923099	----	0.8696	0.488	0.00652364
21	rs1543289	41157388	----	0.05416	0.33	0.00711574
21	rs0724208	33413854	----	1	0.442	0.00846215
22	rs1883387	33536872	----	0.1128	0.289	0.00078255
X	rs1986585	117393340	*WDR44*	0.01778	0.154	0.000627
X	rs1949393	86436628	----	0.226	0.234	0.001073
X	rs0724533	116542720	----	0.562	0.189	0.001435
X	rs0724534	116542916	----	0.2022	0.159	0.002122
X	rs2034391	27198082	----	0.4493	0.363	0.003838
X	rs0742997	142509623	----	0.7614	0.378	0.004138
X	rs2179329	112201908	----	0.8253	0.427	0.005618
X	rs2012646	100945756	----	0.3087	0.492	0.007706
X	rs2366517	136887785	----	1	0.378	0.008602
X	rs0724270	32244571	*DMD*	0.7964	0.311	0.008977

Appendix C. SNPs Associated with CD Based on the p-value for P2BAT/FBAT (p<0.01) in COGA Data (Illumina)

Chr	Marker	Position	Known Gene	HWpval	MAF	P
1	rs636101	120058236	*PHGDH*	0.03237	0.28416	0.004198982
1	rs1557061	56520770	----	0.522	0.40385	0.008725978
1	rs736861	40703590	----	0.6505	0.39006	0.009273092
1	rs1464816	202395477	*REN*	0.6103	0.31091	0.009465632
2	rs1568452	57866337	----	0.259	0.37202	0.0000545
2	rs903748	240817894	----	0.3013	0.17452	0.000885744
2	rs805308	54048438	*PSME4*	0.212	0.47532	0.000914964
2	rs805423	53980545	*PSME4*	0.2984	0.47905	0.001253577
2	rs925229	24087944	*LOC388931*	1	0.47017	0.008256606
2	rs1446596	209993773	----	0.8342	0.47872	0.008381974
3	rs876675	77758328	*ROBO2*	0.6801	0.48616	0.001593182
3	rs531577	139883989	*PIK3CB*	0.6136	0.30425	0.003338441
3	rs13975	126284921	*SLC12A8*	0.2022	0.46283	0.004634245
3	rs1997422	125463298	*KALRN*	0.6767	0.45092	0.005899007
3	rs4796	32498781	*CMTM6*	0.6726	0.48239	0.006523697
3	rs877439	169282596	*GOLIM4*	0.5378	0.49688	0.007811044
3	rs1025192	156310481	*MME*	0.1519	0.37947	0.007969908
3	rs2061719	81336030	----	0.02178	0.34904	0.008160368
3	rs1995137	42106419	*TRAK1*	0.301	0.46211	0.009773593

Appendix C (continued)

Chr	Marker	Position	Known Gene	HWpval	MAF	P
4	rs1968778	21427412	*KCNIP4*	1	0.48212	0.004552297
4	rs1879462	63476530	----	0.5352	0.46425	0.007950792
5	rs2053053	149589586	*CAMK2A*	0.1941	0.37642	0.000977901
5	rs887346	149576185	----	0.1217	0.3821	0.002868128
5	rs949602	157323418	----	0.1339	0.45307	0.007729952
5	rs726847	136100004	----	0.7963	0.28807	0.007938675
5	rs1501656	81822452	----	0.1795	0.35881	0.008460381
6	rs13161	26222681	----	0.06232	0.18871	0.003444944
6	rs612928	144640677	----	1	0.41398	0.007091924
8	rs344278	63031658	----	0.5282	0.4642	0.003427216
9	rs1778970	86585630	*NTRK2*	0.8274	0.35805	0.002368143
9	rs952765	99407127	*C9orf97*	0.484	0.32155	0.003307284
9	rs1336980	128417676	*LMX1B*	0.3635	0.34885	0.008979658
10	rs1887984	110103015	----	0.325	0.35812	0.007722015
11	rs930983	122339624	----	0.3006	0.44052	0.000142593
11	rs160195	87932621	*GRM5*	0.1342	0.40028	0.009626499
12	rs1542707	46921444	----	0.4137	0.22557	0.0016836
12	rs1240267	68570395	*C12orf28*	1	0.23551	0.002924242
13	rs735600	67900291	----	0.522	0.38254	0.008109404
13	rs1856277	108249564	*MYO16*	0.2877	0.39524	0.008370048

Appendix C (continued)

Chr	Marker	Position	Known Gene	HWpval	MAF	P
14	rs221924	70650531	*PCNX*	0.591	0.12117	0.002664231
15	rs383902	56821466	*ADAM10*	0.817	0.31413	0.000360549
15	rs347117	56788249	*ADAM10*	0.809	0.32123	0.002375717
15	rs610877	56890620	*FAM63B*	1	0.32491	0.003242261
15	rs395601	56850488	*FAM63B*	0.2645	0.33428	0.004928419
15	rs387812	56862078	*FAM63B*	0.517	0.40267	0.008158609
17	rs1043127	59144942	*LYK5*	0.4007	0.47327	0.006026961
17	rs727999	24570756	----	0.07223	0.19052	0.008685683
19	rs36633	59338102	*CNOT3*	0.6801	0.43693	0.004842536
21	rs2210267	40685095	*DSCAM*	0.5336	0.43914	0.003792539
X	rs2064034	48708920	*KCND1*	1	0.459	0.001652

VITA

XUEQIU JIAN

PERSONAL DATA:

Date of Birth: March 23, 1984

Place of Birth: Changchun, China

Marital Status: Single

EDUCATION:

Master of Public Health, East Tennessee State University (ETSU), Johnson City, TN, 2010

Concentration: Epidemiology

Bachelor of Medicine (MD equivalent), Sun Yat-sen University (SYSU), Guangzhou, China, 2008

Major: Clinical Medicine

ACADEMIC REWARDS:

Tuition Scholarship and Graduate Assistantship, ETSU, 2008-10

Dr. S. P. Chow Scholarship of Orthopaedics, SYSU, 2007-08

Excellent Student Scholarship, SYSU, 2004-08

Keyes Scholarship for Freshmen, SYSU, 2003-04

MANUSCRIPTS IN PREPARTION:

First Author, *Association of ADAM10 and CAMK2A polymorphisms with conduct disorder in the COGA sample: evidence from family-based studies* (will be submitted to

Psychiatric Genetics), 2010

Fourth Author, *A family-based genome-wide association study of alcohol dependence and related personality traits* (in preparation), 2010

Fourth Author, *Family-based association studies of alcohol dependence and age at onset in the COGA sample* (in preparation), 2010

CONFERENCE ABSTRACTS:

Fourth Author, *Family-based association studies of alcohol dependence and age at onset in the COGA sample*. APHA 137th Annual Meeting, Philadelphia, PA, 2009

Fourth Author, *A family-based genome-wide association study of alcohol dependence and related personality traits*. 6th Annual Primary Care and Prevention Research Day, ETSU, Johnson City, TN, 2009

SOCIAL ACTIVITY EXPERIENCES:

Participant, Tennessee Cancer Coalition Northeast Region Organizational Meeting, 2009

Head, Student Academic Internship Team, SYSU, 2006-08

Volunteer, Free Diagnosis in East Lake Community of Guangzhou, P. R. China, 2005-07

Head, Department of Security, Students' Union, SYSU, 2004-05