Molecular Psychiatry (2011) 16, 800–808

www.nature.com/mp

ORIGINAL ARTICLE

Genome-wide association study of conduct disorder symptomatology

DM Dick[1], F Aliev[1], RF Krueger[2], A Edwards[1], A Agrawal[2], M Lynskey[2], P Lin[2], M Schuckit[3], V Hesselbrock[4], J Nurnberger Jr[5], L Almasy[6], B Porjesz[7], HJ Edenberg[5], K Bucholz[2], J Kramer[8], S Kuperman[8] and L Bierut[2], as part of the SAGE and GENEVA Consortia

[1]*Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA;* [2]*Washington University, St Louis, MO, USA;* [3]*University of California, San Diego, CA, USA;* [4]*University of Connecticut Health Center, Farmington, CT, USA;* [5]*Indiana University School of Medicine, Indianapolis, IN, USA;* [6]*Southwest Foundation, San Antonio, TX, USA;* [7]*State University of New York (SUNY) Downstate Medical Center, Brooklyn, NY, USA and* [8]*University of Iowa College of Medicine, Iowa City, IA, USA*

Conduct disorder (CD) is one of the most prevalent childhood psychiatric conditions, and is associated with a number of serious concomitant and future problems. CD symptomatology is known to have a considerable genetic component, with heritability estimates in the range of 50%. Despite this, there is a relative paucity of studies aimed at identifying genes involved in the susceptibility to CD. In this study, we report results from a genome-wide association study of CD symptoms. CD symptoms were retrospectively reported by a psychiatric interview among a sample of cases and controls, in which cases met the criteria for alcohol dependence. Our primary phenotype was the natural log transformation of the number of CD symptoms that were endorsed, with data available for 3963 individuals who were genotyped on the Illumina Human 1M beadchip array. Secondary analyses are presented for case versus control status, in which caseness was established as endorsing three or more CD symptoms ($N=872$ with CD and $N=3091$ without CD). We find four markers that meet the criteria for genome-wide significance ($P<5\times10^{-8}$) with the CD symptom count, two of which are located in the gene *C1QTNF7* (C1q and tumor necrosis factor-related protein 7). There were six additional SNPs in the gene that yielded converging evidence of association. These data provide the first evidence of a specific gene that is associated with CD symptomatology. None of the top signals resided in traditional candidate genes, underscoring the importance of a genome-wide approach for identifying novel variants involved in this serious childhood disorder.
Molecular Psychiatry (2011) **16,** 800–808; doi:10.1038/mp.2010.73; published online 29 June 2010

Keywords: conduct disorder; association analysis; genetics; GWAS

Introduction

Childhood conduct disorder (CD) involves a persistent pattern of rule-breaking and aggressive behaviors, including bullying other children, stealing, vandalizing and skipping school. CD is one of the most prevalent childhood disorders. Although rates vary according to the population under study, approximately 6–16% of males, and 2–9% of females, are diagnosable with CD.[1] CD is associated with serious problems in home and school functioning, and is a strong risk factor for concurrent and future alcohol and other substance problems; several studies of adolescents who have been diagnosed with alcohol use disorders have concluded that among childhood behavioral disorders, CD has the strongest association with alcohol problems.[2–5] CD also shows considerable evidence for genetic influence. Retrospective reports of CD have estimated heritability ranging from 40 to 70%.[6,7] Prospective reports of CD symptoms in children also show considerable evidence of genetic influence, with heritability estimates on the order of 40–50% in both boys and girls.[8] A review of >100 quantitative genetic studies of antisocial behavior (measured using various methods) converged on a heritability estimate of 50%,[9] and a recent multi-informant study of childhood antisocial behavior showed evidence of even stronger genetic influence.[10] Furthermore, twin studies indicate that genetic influences on CD show considerable overlap with alcohol and other substance dependence.[11–14]

Despite strong evidence for a considerable genetic component to CD, there have been relatively few studies aimed at identifying genetic variation that

Correspondence: Assistant Professor DM Dick, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Department of Psychiatry, Richmond, VA, USA.
E-mail: ddick@vcu.edu
Received 9 December 2009; revised 23 April 2010; accepted 5 May 2010; published online 29 June 2010

contributes to the risk for this disorder. Two genome-wide linkage scans have been conducted on retrospectively reported CD, using data from the Collaborative Study on the Genetics of Alcoholism (COGA)[15] and the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD).[16] CD symptoms have also been analyzed in the context of linkage scans that focused on externalizing psychopathology more generally.[17–19] A number of chromosomal regions have been implicated by these studies, although the reported LOD scores have generally been modest, and none has yet led to the subsequent identification of an associated gene in the region. In addition, there are a small number of candidate gene studies of CD, with largely negative or inconsistent results. Association has been reported between the short allele of 5HTTLPR and CD with aggressive symptoms among adolescents ascertained through a substance abuse treatment program.[20] However, another study failed to detect an association between this polymorphism and externalizing behavior in a sample of high-risk adoptees, although secondary analyses suggested that the short allele increased risk for externalizing behaviors in conjunction with a genetic diathesis for alcohol dependence.[21] There has also been a negative report of association between the *DAT1* gene and CD in adolescents.[22] A number of other studies have examined CD symptomatology in the context of ADHD, also with mixed results. One recent study reported an association between the catechol-O-methyltransferase (*COMT*) valine/methionine polymorphism and CD symptoms in ADHD cases,[23] whereas another study failed to replicate the association with CD symptoms in an independent sample of ADHD cases, although some evidence for association was detected with a subset of the aggressive CD symptoms.[24] In addition, CD traits among ADHD probands have been evaluated in the GAIN-ADHD sample, with no markers meeting the criteria for genome-wide significance across any of the three CD problem traits that were examined.[25]

To date, no gene has yet been identified which is reliably associated with this serious and prevalent disorder. Furthermore, the gene identification efforts aimed at elucidating the underlying genetic susceptibility to CD have paled in comparison with, e.g., the considerable and extensive effort surrounding the identification of genes involved in another common childhood disorder, ADHD.[26] In this study, we report results from a genome-wide association study (GWAS) of retrospectively reported CD symptomatology.

Materials and methods

Sample

Data for this study come from the Study of Addiction: Genes and Environment (SAGE),[27] which was one of the eight phase 1 studies in the Gene Environment Association (GENEVA) consortium.[28] Cases and controls for the SAGE sample were drawn from three contributing projects: the Collaborative Study on the Genetics of Alcoholism (COGA),[29] the Collaborative Study on the Genetics of Nicotine Dependence (COGEND)[30] and the Family Study of Cocaine Dependence (FSCD). Although cases in these studies were ascertained for alcoholism, nicotine dependence (based on an FTND score of $\geqslant 4$ in current smokers, controls being smokers) and cocaine dependence, respectively, cases for SAGE were uniformly defined as those meeting DSM-IV criteria for alcohol dependence ($N = 1899$). Controls ($N = 1946$) were unrelated individuals who were largely past the highest risk period for developing alcohol dependence, and who reported drinking alcohol but not meeting criteria for alcohol dependence at any time during their life. An additional 143 subjects who met criteria for dependence on illicit drugs but not for alcohol dependence were also genotyped and included in this study.

Measure

CD was assessed in all three studies that comprise SAGE using versions of the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA).[31,32] Our primary phenotype was the natural log transformation of the number of CD symptoms [$\ln(1+x)$] that were endorsed (raw symptom counts ranged from 0 to 13). There were 17 individuals missing CD information, yielding a total N of 3963 for analyses. Table 1 shows the distribution of the raw CD symptom counts, split by sex, before transformation. We log transformed the data due to the skew apparent in the symptom counts, and to minimize the impact of outliers at the upper end of the distribution. As expected, males were more likely to endorse CD symptoms than were females; therefore, gender was used as a covariate in all genetic analyses. In addition, we analyzed a dichotomous CD case status variable, because this phenotype and accompanying odds ratios may also be of interest. Case status was defined by the endorsement of $\geqslant 3$ of the 15 DSM-IV CD

Table 1 Distribution of individuals by CD symptom count

CD symptom count	*Males*	*Females*	*Total*
0	565	1166	1731
1	366	465	831
2	290	239	529
3	194	124	318
4	126	68	194
5	115	35	150
6	63	17	80
7	38	19	57
8	21	4	25
9	25	4	29
10	9	0	9
11	3	2	5
12	2	0	2
13	2	1	3
Total	1819	2144	3963

Abbreviation: CD, conduct disorder.

criteria under Criterion A, without assessing the requirement of clustering of different symptoms within a 12-month time period. Furthermore, we did not use Criterions B and C (significant clinical impairment and absence of Antisocial Personality Disorder) in defining cases. We refer to this phenotype as CD case status throughout this paper, although we note that full diagnostic criteria were not applied. Controls were defined as those who endorsed fewer than three symptoms for DSM-IV CD. A total of 872 subjects met these criteria for CD case status, and there were 3091 controls. The mean age of the CD cases was 37.32 years (s.d. = 8.86; range 18–74 years). The mean age of the CD controls was 39.63 years (s.d. = 9.13; range 18–77 years).

Genotyping

DNA samples were genotyped on the Illumina Human 1M beadchip (Illumina) by the Center for Inherited Diseases Research (CIDR) at Johns Hopkins University. A thorough data-cleaning procedure was applied, including using HapMap controls, detection of gender and chromosomal anomalies, hidden relatedness, population structure, missing call rates, batch effects, Mendelian error detection, duplication error detection and Hardy–Weinberg equilibrium.[33] A total of 948 658 SNPs passed data cleaning procedures.

Population stratification

The software package EIGENSTRAT/EIGENSOFT[34] was used to calculate principal components reflecting continuous variation in allele frequencies representing ancestral differences in subjects. A description of the principal components extraction process may be found in related publications.[27,33] Briefly, two principal components, the first distinguishing African-American participants from European-American participants and the second distinguishing Hispanic and non-Hispanic subjects, were identified and used to control for effects of population stratification. The self-reported ethnicity of the sample analyzed in this study is 2698 European Americans, 1257 African Americans and 8 individuals of other ethnicity.

Genomewide association analyses

GWAS was conducted using linear and logistic regressions in PLINK,[35] for CD symptom counts and case status, respectively. Genotypes for each of the 917 694 autosomal SNPs were coded log additively (0, 1, 2 copies of the minor allele). Covariates representing sex, age (defined using quartiles as three dummy measures representing younger than 34 years, 35–39 years, 40–44 years, with older than 45 years as the reference group, as in other SAGE publications[27]), study site, and the two aforementioned principal component scores were included as covariates.

Results

Table 2 lists all SNPs that yielded *P*-values $<10^{-5}$ with either phenotype. The results are ordered by significance of *P*-values for the primary phenotype of natural log-transformed CD symptom count, although the SNPs listed under the horizontal rule in the table are included for $P<10^{-5}$ with the CD case status. The left portion of the table reports results for the natural log-transformed CD symptom count, and the corresponding results for CD case status are included in the right portion of the table. All SNPs that met criteria for $P<10^{-5}$ for one of the phenotypes also showed a corresponding *P*-value <0.01 with the other phenotype, indicating general consistency across the results for both the quantitative CD symptom count variable and CD case status. Figure 1 shows a q–q plot for the observed versus expected distribution of *P*-values for the primary CD symptom count analysis. Four SNPs reached genome-wide significance with CD symptom count, based on $P<5\times10^{-8}$. Two of these SNPs (rs16891867 and rs1861046) were in the gene *C1QTNF7* (C1q and tumor necrosis factor-related protein 7). These SNPs were in high LD ($r^2=0.97$). There were 40 SNPs tested across this gene in the GWAS panel; 6 additional SNPs in the gene yielded $P<0.05$. The average r^2 between these SNPs and rs16891867 and rs1861046 was 0.24 (range 0.09–0.76), indicating that they provide some independent, converging evidence of association in the gene. The other top two SNPs meeting genome-wide significance were intergenic and located on chromosomes 11 and 13.

As SAGE is an ethnically diverse sample, as described in the 'Methods' section, our primary analyses used covariates to account for ethnicity. However, for the top SNPs listed in Table 2, we conducted secondary analyses for the primary CD symptom count variable on the European American and African American subsets of the sample to test whether the evidence for association was evident in both groups. These results are presented in Table 3. Most SNPs yielded evidence for association in both the European-American and African-American samples, with the exception of four SNPs (rs11838918, rs8179116, rs13398848 and rs2720508), in which the evidence for association was largely limited to the African-American subgroup ($P>0.20$ in the European American sample). One additional SNP, rs10776612, provided stronger evidence of association in the African-American subgroup, with $P=0.14$ in the European-American subgroup. In all cases, the direction of effect was consistent in both samples, including for four SNPs in which the minor allele differed between the groups (indicated in Table 3).

Finally, we know that there is considerable overlap between CD and alcohol dependence, and that twin studies suggest shared genetic liability across these disorders, as described in the 'Introduction' section.[12] The extensive comorbidity between the disorders, both in our sample (83% of the CD cases also met

Table 2 SNPs yielding $P<1\times10^{-5}$ with either CD symptom count or CD case status

						CD symptom count		*CD case status*			
Chr	*SNP*	*BP*	*Allele*	*MAF*	*Gene*	*BETA*	P-*value*	OR	*L95*	*U95*	P-*value*
4	rs16891867	15 006 462	G/A	0.097	C1QTNF7	0.145	**3.275E-09**	1.624	1.357	1.943	1.185E-07
11	rs7950811	92 290 650	A/C	0.071		0.157	**1.295E-08**	1.647	1.350	2.009	8.601E-07
13	rs11838918	78 308 575	C/T	0.022		0.282	**1.332E-08**	2.262	1.607	3.184	2.880E-06
4	rs1861046	15 007 004	A/G	0.095	C1QTNF7	0.140	**1.432E-08**	1.650	1.377	1.976	5.795E-08
6	rs7762160	165 935 421	C/T	0.365	PDE10A	0.071	1.128E-06	1.266	1.126	1.424	8.313E-05
4	rs4698107	15 023 110	A/G	0.098	C1QTNF7	0.118	1.204E-06	1.527	1.275	1.830	4.296E-06
12	rs8179116	107 549 556	A/G	0.021	SELPLG	0.229	3.321E-06	1.669	1.182	2.358	3.646E-03
5	rs2122554	165 889 664	A/C	0.162		−0.101	3.410E-06	0.739	0.626	0.874	3.912E-04
14	rs1256531	64 817 512	G/A	0.234		0.088	4.479E-06	1.260	1.088	1.459	1.989E-03
5	rs1450624	165 888 868	C/T	0.161		−0.099	5.124E-06	0.749	0.633	0.885	7.073E-04
2	rs16831128	132 860 862	G/A	0.189		0.083	5.458E-06	1.449	1.255	1.674	4.465E-07
2	rs6750486	35 380 422	T/C	0.126		0.103	5.583E-06	1.317	1.112	1.560	1.435E-03
20	rs6031252	42 007 236	A/C	0.142	TOX2	−0.112	5.745E-06	0.765	0.635	0.921	4.686E-03
16	rs3136202	13 945 675	A/G	0.429	ERCC4	−0.067	6.199E-06	0.782	0.693	0.882	6.513E-05
12	rs12302829	5 010 745	G/A	0.008		−0.351	7.639E-06	0.423	0.219	0.820	1.077E-02
1	rs4434872	152 040 900	T/C	0.263	LOC343052	0.076	7.683E-06	1.326	1.163	1.513	2.508E-05
17	rs4792394	13 625 642	C/A	0.469		−0.062	9.126E-06	0.856	0.763	0.959	7.209E-03
16	rs3136166	13 939 594	G/T	0.448	ERCC4	−0.066	9.171E-06	0.799	0.708	0.901	2.475E-04
11	rs7129870	38 492 315	T/C	0.007		0.355	9.916E-06	2.513	1.434	4.406	1.292E-03
2	rs13398848	84 098 679	G/A	0.074		0.120	1.472E-05	1.594	1.304	1.947	5.163E-06
4	rs17007017	142 362 179	G/A	0.315	ZNF330	−0.066	2.271E-05	0.747	0.658	0.849	7.313E-06
4	rs1861050	15 091 458	T/C	0.063	KIAA1345	0.118	4.742E-05	1.626	1.314	2.013	7.787E-06
10	rs2184898	119 408 094	A/G	0.327		0.059	7.790E-05	1.332	1.181	1.502	2.973E-06
4	rs1550057	142 336 874	C/T	0.243		−0.063	1.229E-04	0.717	0.623	0.825	3.557E-06
8	rs2720508	17 374 188	C/T	0.093		−0.101	1.861E-04	0.623	0.507	0.766	7.409E-06
4	rs17350440	142 380 929	C/A	0.260		−0.059	2.116E-04	0.726	0.634	0.830	3.152E-06
2	rs7581919	15 582 207	C/T	0.041	NAG	0.115	1.264E-03	1.854	1.414	2.431	8.049E-06
10	rs2419006	110 178 904	C/T	0.169		0.053	5.104E-03	1.389	1.201	1.607	9.642E-06
10	rs10776612	49 405 569	T/C	0.481	ARHGAP22	−0.039	6.308E-03	0.752	0.669	0.846	1.874E-06

Abbreviations: CD, conduct disorder; MAF, minor allele frequency; OR, odds ratio; SNP, single-nucleotide polymorphism.
SNPs with $P<5\times10^{-8}$ shown in bold.

criteria for alcohol dependence) and in the general population, makes it unrealistic to completely tease apart genetic effects on the two disorders. However, we did want to determine whether our results for CD were driven largely by association with alcohol dependence. The correlation between the natural log transformations of CD symptom count and alcohol dependence symptom count was 0.45 ($P<0.001$) in our sample. We reran the top SNPs from Table 2 including the log-transformed alcohol dependence symptom count as a covariate. Results are listed in Table 4. As expected, the *P*-values dropped in magnitude, though all were still significant.

Discussion

This paper reports results from one of the first (GWASs) of childhood CD symptomatology. To our knowledge, only one previous GWAS of CD traits has been conducted, and that study examined CD traits only among individuals with ADHD[24,25]). We find four markers that pass the threshold for genome-wide significance, and another 25 SNPs that show association with CD symptoms or CD case status at

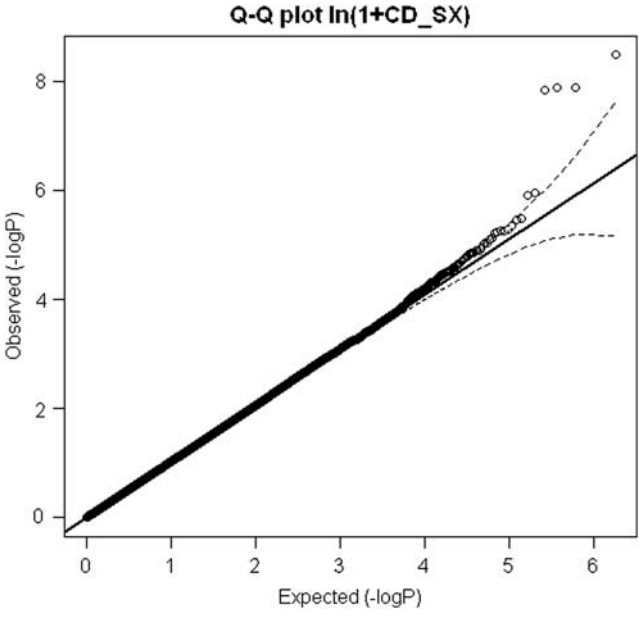

Figure 1 Q–Q plot for the conduct disorder symptom count variable showing a significant deviation of findings from what would be expected by chance.

Table 3 Results from analyses of CD symptom count, split by race for all SNPs yielding $P<10^{-5}$ (from Table 2)

		Results for European Americans				*Results for African Americans*			
SNP	P-*value full sample*	*MA*	*MAF*	*BETA*	P-*value*	*MA*	*MAF*	*BETA*	P-*value*
rs16891867	3.28E-09	G	0.041	0.142	4.38E-04	G	0.218	0.148	4.35E-06
rs7950811	1.30E-08	A	0.048	0.149	8.72E-05	A	0.121	0.163	7.76E-05
rs11838918	1.33E-08	C	0.001	−0.314	3.62E-01	C	0.066	0.278	2.67E-07
rs1861046	1.43E-08	A	0.041	0.142	4.38E-04	A	0.212	0.140	1.71E-05
rs7762160	1.13E-06	C	0.408	0.056	7.72E-04	C	0.273	0.112	1.46E-04
rs4698107	1.20E-06	A	0.042	0.137	6.02E-04	A	0.217	0.108	8.08E-04
rs8179116	3.32E-06	A	<0.001	0.405	3.38E-01	A	0.065	0.218	4.51E-05
rs2122554	3.41E-06	A	0.038	−0.142	8.15E-04	A	0.428	−0.086	1.24E-03
rs1256531	4.48E-06	G	0.098	0.107	8.96E-05	A	0.461	−0.070	1.09E-02
rs1450624	5.12E-06	C	0.038	−0.135	1.47E-03	C	0.428	−0.086	1.33E-03
rs16831128	5.46E-06	G	0.213	0.077	1.32E-04	G	0.135	0.106	8.06E-03
rs6750486	5.58E-06	T	0.047	0.070	6.67E-02	T	0.297	0.119	6.63E-05
rs6031252	5.75E-06	A	0.019	−0.196	1.11E-03	A	0.409	−0.092	1.26E-03
rs3136202	6.20E-06	A	0.338	−0.070	5.79E-05	G	0.376	0.068	1.52E-02
rs12302829	7.64E-06	G	<0.001	−1.044	8.06E-02	G	0.026	−0.347	4.89E-05
rs4434872	7.68E-06	T	0.169	0.048	2.71E-02	T	0.465	0.101	2.12E-04
rs4792394	9.13E-06	C	0.477	−0.049	2.63E-03	C	0.451	−0.073	7.30E-03
rs3136166	9.17E-06	G	0.344	−0.065	1.34E-04	T	0.325	0.071	1.23E-02
rs7129870	9.92E-06	T	<0.001	0.918	2.97E-02	T	0.022	0.338	1.28E-04
rs13398848	1.47E-05	G	0.024	−0.008	8.84E-01	G	0.181	0.168	1.13E-06
rs17007017	2.27E-05	G	0.246	−0.064	7.61E-04	G	0.465	−0.066	1.44E-02
rs1861050	4.74E-05	T	0.041	0.160	8.17E-05	T	0.112	0.080	6.01E-02
rs2184898	7.79E-05	A	0.382	0.046	4.88E-03	A	0.208	0.097	2.62E-03
rs1550057	1.23E-04	C	0.263	−0.047	1.13E-02	C	0.199	−0.094	4.73E-03
rs2720508	1.86E-04	C	0.005	−0.010	9.24E-01	C	0.282	−0.102	6.67E-04
rs17350440	2.12E-04	C	0.243	−0.062	1.04E-03	C	0.297	−0.053	7.36E-02
rs7581919	1.26E-03	C	0.054	0.094	8.86E-03	C	0.013	0.265	2.57E-02
rs2419006	5.10E-03	C	0.153	0.048	3.44E-02	C	0.204	0.063	6.02E-02
rs10776612	6.31E-03	T	0.429	−0.025	1.40E-01	C	0.405	0.074	6.40E-03

Abbreviations: CD, conduct disorder; MA, minor allele; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.

$P<5\times10^{-5}$. Two of the genomewide-significant SNPs were in the gene *C1QTNF7* (C1q and tumor necrosis factor-related protein 7). Very little is known about this gene. Gene expression information available from the BioGPS online database (http://biogps.gnf.org/#goto=welcome)[36] indicates that *C1QTNF7* is expressed at comparable levels across tissues. It remains unclear at this time how *C1QTNF7* may be functionally involved in CD. However, the involvement of multiple other SNPs, not in complete LD with the most highly associated SNPs in the gene, bolsters confidence that this gene is associated with CD symptoms.

The other two SNPs meeting genome-wide significance were intergenic and located on chromosomes 11 and 13. The SNP rs7950811, on chromosome 11q14.3, is located at the edge of the gene *LOC642791*, which is similar to elongation factor 1-α; *EF-1-*α. The other SNP, rs11838918, on 13q31.1 is ~13 kb from the nearest gene *RP11–600P1.1/ LOC647298*, a heat-shock 60-kDa protein 1 pseudogene. None of the SNPs significant at $P<10^{-5}$ in our GWAS represents an obvious candidate for involvement in CD. We used the online databases ToppGene[37] and DAVID[38] to assess whether our list of genes containing significant markers is functionally enriched for gene ontology categories, and whether these genes have been previously implicated in human phenotypes. Using an FDR cutoff of $P<0.05$, the only enriched gene ontology category among genes implicated in the current report is the molecular function 'magnesium ion binding', which applies to two genes (*ERCC4* and *PDE10A*). We further investigated potential commonalities among significant genes using the BioGPS online database. Six of the genes (*ARHGAP22*, *ERCC4*, *NAG*, *SELPLG*, *TOX2* and *ZNF330*) are highly expressed in the blood; *NAG* is also expressed in the testis germ cell, and *TOX2* is highly expressed in the thymus and lung. *LOC343052* exhibits high expression levels in the ciliary ganglion, thalamus and trachea, whereas *PDE10A* is most highly expressed in the caudate nucleus.

None of the top SNPs (or associated genes) identified in our study (at $P<5\times10^{-5}$) overlapped with any of the top 54 markers reported by Anney and colleagues to be associated with CD traits in ADHD probands at $P<10^{-5}$. Furthermore, combined examination of the list of genes from our sample and the

Table 4 Comparison of *P*-values with and without AD symptoms included as a covariate for the CD symptom count

SNP	*Original* P-*value CD symptoms*	P-*value CD symptoms with AD symptoms as covariate*
rs16891867	3.275E-09	1.605E-07
rs7950811	1.295E-08	2.410E-07
rs11838918	1.332E-08	5.668E-08
rs1861046	1.432E-08	5.154E-07
rs7762160	1.128E-06	2.548E-04
rs4698107	1.204E-06	6.187E-05
rs8179116	3.321E-06	2.352E-06
rs2122554	3.410E-06	4.632E-04
rs1256531	4.479E-06	2.063E-04
rs1450624	5.124E-06	6.873E-04
rs16831128	5.458E-06	2.654E-06
rs6750486	5.583E-06	3.868E-07
rs6031252	5.745E-06	1.243E-06
rs3136202	6.199E-06	1.302E-05
rs12302829	7.639E-06	4.303E-06
rs4434872	7.683E-06	4.811E-04
rs4792394	9.126E-06	2.683E-05
rs3136166	9.171E-06	3.019E-05
rs7129870	9.916E-06	7.262E-06
rs13398848	1.472E-05	1.441E-05
rs17007017	2.271E-05	9.237E-05
rs1861050	4.742E-05	7.721E-04
rs2184898	7.790E-05	3.296E-04
rs1550057	1.229E-04	1.016E-03
rs2720508	1.861E-04	2.127E-04
rs17350440	2.116E-04	6.474E-03
rs7581919	1.264E-03	4.489E-04
rs2419006	5.104E-03	1.084E-02
rs10776612	6.308E-03	1.376E-02

Abbreviations: AD, alcohol dependence; CD, conduct disorder; SNP, single-nucleotide polymorphism.

Anney sample, using the databases indicated above, did not indicate that the findings were enriched for any gene ontology category. However, three genes from the combined list (*PDX1*, *ATP8B1* and *ERCC4*) have been implicated in pancreas abnormalities; these results are significant (FDR = 0.000047) and might be relevant given the role of the pancreas in the endocrine system. However, the largely nonoverlapping findings from the studies likely reflects differences in the sample ascertainment. Anney and colleagues studied CD traits among individuals with ADHD. Furthermore, the average age of that sample was 10.88 years (s.d. = 2.8). Our sample was ascertained through substance-dependent probands, although phenotypic information was available on the controls, in addition to the cases, which enabled us to study CD symptomatology among individuals with and without alcohol dependence. Furthermore, CD symptoms were assessed retrospectively in our sample, among adults who had passed through the period of risk for childhood CD. Accordingly, the different sample characteristics may have yielded populations with different underlying etiological factors.

We conducted several secondary analyses to examine the robustness of the detected effects. Although log transformation of the CD symptom scores has the effect of reducing the impact of outliers at the upper end of the distribution, we also conducted analyses removing the 16 males and 11 females at the upper end of the symptom count distribution (Table 1), with results largely unchanged (data not shown). In addition, results with the binary CD case status variable were consistent with the primary analysis of log-transformed CD symptom counts. Furthermore, most of the top SNPs implicated in the primary analysis showed consistent effects in the European and African-American subgroups. Accordingly, these analyses bolster our confidence in the robustness of the findings reported in this study.

We examined whether any of the most highly associated SNPs (from Table 2) were located in the linkage regions implicated in the two previous genome-wide scans focused on CD symptomatology (among samples originally ascertained for alcohol dependence, as in this report), in the COGA and IASPSAD samples. Peaks in the IASPSAD sample for CD symptoms were reported with markers located on chromosomes 1, 2, 7, 8, 10 and 14,[39] and in COGA on chromosomes 1, 2, 3, 12 and 19 for CD diagnoses and symptom counts.[40] One of the associated SNPs from the CD analyses reported in this study, rs13398848, was located in a region on chromosome 2 with converging evidence of linkage across the two samples. The SNP is located ~2 MB from the peak marker in the COGA scan (D2S1331), and <8 MB from the peak marker from the IASPSAD scan (D2S2116). The SNP is intergenic; however, the closest gene downstream is *CTNNA2* (Catenin, -2), located ~4 MB away. *CTNNA2* is an interesting candidate, as it is considered to be involved in stability of synaptic contacts,[41] and disruption of the gene in mice has been associated with fear conditioning and prepulse inhibition of the startle response.[42] Dysregulation of inhibitory responses is considered to be related to the development of conduct problems and antisocial behavior (as well as drug dependence).[43] Two additional SNPs on chromosome 10 showing evidence of association in the GWAS, rs2419006 and rs2184898, were located 1 and 4 MB, respectively, from markers yielding significant evidence of linkage (D10S597 and D10S1679) in the IASPSAD sample. These SNPs were also intergenic. None of the other top SNPs from Table 2 were located within 10 MB of peak markers from the linkage scans.

The results from this study should be interpreted in the context of several limitations. The sample analyzed in this study (SAGE) consisted of cases and controls originally ascertained for alcohol dependence; accordingly, rates of CD were higher in our sample (22% meeting case status for CD) than in general population samples; for example, the National Epidemiological Survey of Alcohol and Related Conditions (NESARC) has reported rates of CD of ~5% among a general population-based sample.[44]

However, NESARC also observes a strong relationship between CD and alcohol dependence;[44] accordingly, it is not surprising that we would find elevated rates of CD among a sample ascertained for alcohol dependence. We would also expect the converse to be true: that a sample ascertained based on CD would show high rates of alcohol dependence; accordingly, it would be difficult, and perhaps artificial, to study one disorder outside the context of the other. Our secondary analyses using alcohol dependence symptom counts as a covariate suggested that the results observed in this study for CD were not driven solely through association with alcohol dependence. Similarly, there was no overlap in the most highly associated SNPs reported with alcohol dependence (*P*-value $\leqslant$1E-06) in the SAGE sample,[27] and the most highly associated SNPs with CD, reported in this study. Another limitation is that we did not have sufficient information on the sample to make full CD diagnoses according to DSM-IV; accordingly, we defined case status solely on the basis of the symptom threshold for DSM-IV diagnoses, and without taking into account clustering; this practice has been applied in previous studies.[45] Another limitation is that we used retrospective reports of CD. The age range of the sample was quite broad (18–77); accordingly, some participants were reporting on adolescent behavior at a much later time in life. To ensure that the results were not unduly influenced by this, we reran the top hits from Table 2 in the younger half of the sample, as defined by a median split on the age variable ($\leqslant$39 years). The magnitude of the *P*-values was not as significant, as would be expected with a reduction in sample size of $\sim$50%; however, nearly all *P*-values were still on the order of 10^{-3} or lower, and, importantly, the top hits in *C1QTNF7* were still significant at $P<10^{-7}$. Accordingly, we feel confident that the results are not driven by poor retrospective reports among older participants.

There is a literature suggesting that there may be etiologically different subtypes of CD symptomatology, with distinctions between aggressive and rule-breaking forms of antisocial behavior.[46] We do not find evidence for two different factors in this sample; rather, all symptoms load onto a single factor (data not shown). Nonetheless, we did create sum symptom scores for the symptoms that loaded onto the aggressive and rule-breaking factors as reported in previous studies.[47] We reran the top hits reported in Table 2 for the aggressive and rule-breaking symptom counts, respectively. There was association across both symptom dimensions. This is perhaps not surprising as the symptoms all loaded onto a single factor in this sample. Thus, we find no evidence that the findings reported in this study are differentially associated with aggressive and rule-breaking dimensions of CD.

In summary, we report results from the first GWAS of CD symptomatology not occurring solely in the context of ADHD. We find four markers that meet criteria for genome-wide significance, and several more with highly significant ($P<10^{-5}$) evidence of association. The current literature on genes involved in CD is extremely limited, despite the prevalence and long-term serious consequences associated with the disorder. None of our top hits reside in genes whose functions are well characterized. This is one of the strengths of GWAS—the ability to identify novel genes that force us to expand our theories surrounding the underlying biological underpinnings of disease etiology. Replication of our findings will be key. It is our hope that these results will drive additional studies aimed at elucidating the genetic basis for CD.

Conflict of interest

LJ Bierut is an inventor on the patent 'Markers for Addiction' (US 20070258898) covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction. Dr Bierut served as a consultant for Pfizer Inc. in 2008. No other authors reported any conflicts of interest to disclose.

Acknowledgments

Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative [GEI] (U01 HG004422). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract 'High throughput genotyping for studying the genetic contributions to human disease' (HHSN268200782096C). This project also was supported by a 2007 NARSAD Young Investigator Award to Danielle M. Dick.

References

1 Association AP. *Diagnostic and Statistical Manual of Mental Disorders*. 4th edn Association AP: Washington, DC, 1994.
2 Kuperman S, Schlosser SS, Kramer JR, Bucholz KK, Hesselbrock V, Reich T *et al*. Developmental sequence from disruptive behavior diagnosis to adolescent alcohol dependence. *Am J Psychiatry* 2001; **158**: 2022–2026.

3 Kuperman S, Schlosser SS, Kramer JR, Bucholz KK, Hesselbrock V, Reich T *et al.* Risk domains associated with an adolescent alcohol dependence diagnosis. *Addiction* 2001; **96**: 629–636.
4 Moss HB, Lynch KG. Comorbid disruptive behavior disorder symptoms and their relationship to adolescent alcohol use disorders. *Drug Alcohol Dependence* 2001; **64**: 75–83.
5 Molina BSG, Bukstein OG, Lynch KG. Attention-deficit/hyperactivity disorder and conduct disorder symptomatology in adolescents with alcohol use disorder. *Psychol Addict Behav* 2002; **16**: 161–164.
6 Slutske WS, Heath AC, Dinwiddie SH, Madden PAF, Bucholz KK, Dunne MP *et al.* Modeling genetic and environmental influences in the etiology of conduct disorder: a study of 2682 adult twin pairs. *J Abnormal Psychol* 1997; **106**: 266–279.
7 Goldstein RB, Prescott C, Kendler KS. Genetic and environmental factors in conduct problems and adult antisocial behavior among adult female twins. *J Nerv Ment Disord* 2001; **189**: 201–209.
8 Rose RJ, Dick DM, Viken RJ, Pulkkinen L, Nurnberger Jr JI, Kaprio J. Genetic and environmental effects on conduct disorder, alcohol dependence symptoms, and their covariation at age 14. *Alcoholism Clin Exp Res* 2004; **28**: 1541–1548.
9 Moffitt TE. Genetic and environmental influences on antisocial behaviors: evidence from behavioral-genetic research. *Advances in Genetics* 2005; **55**: 41–104.
10 Baker LA, Jacobson KC, Raine A, Lozano DI, Bezdjian S. Genetic and environmental bases of childhood antisocial behavior: a multi-informant twin study. *J Abnormal Psychol* 2007; **116**: 219–235.
11 Kendler KS, Prescott C, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatry* 2003; **60**: 929–937.
12 Slutske WS, Heath AC, Dinwiddle SH, Madden PAF, Bucholz KK, Dunne MP *et al*. Common genetic risk factors for conduct disorder and alcohol dependence. *J Abnormal Psychol* 1998; **107**: 363–374.
13 Krueger RF, Hicks BM, Patrick CJ, Carlson SR, Iacono WG, McGue M. Etiologic connections among substance dependence, antisocial behavior, and personality: modeling the externalizing spectrum. *J Abnormal Psychol* 2002; **111**: 411–424.
14 Young SE, Stallings MC, Corley RP, Krauter KS, Hewitt JK. Genetic and environmental influences on behavioral disinhibition. *Am J Med Genet* 2000; **96**: 684–695.
15 Dick DM, Li TK, Edenberg HJ, Hesselbrock V, Kramer JR, Foroud T. A genome-wide screen for genes influencing conduct disorder. *Mol Psychiatry* 2003; **9**: 81–86.
16 Kendler KS, Kuo PH, Todd WB, Kalsi G, Neale MC, Sullivan PF *et al.* A joint genomewide linkage analysis of symptoms of alcohol dependence and conduct disorder. *Alcoholism Clin Exp Res* 2006; **30**: 1972–1977.
17 Dick DM, Aliev F, Wang JC, Grucza RA, Schuckit M, Kuperman S *et al.* Using dimensional models of externalizing psychopathology to aid in gene identification. *Arch Gen Psychiatry* 2008; **65**: 310–318.
18 Stallings MC, Corley RP, Dennehey B, Hewitt JK, Krauter KS, Lessem JM *et al.* A genome-wide search for quantitative trait loci that influence antisocial drug dependence in adolescence. *Arch Gen Psychiatry* 2005; **62**: 1042–1051.
19 Ehlers CL, Gilder DA, Slutske WS, Lind PA, Wilhelmsen KC. Externalizing disorders in American Indians: comorbidity and a genome wide linkage analysis. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 690–698.
20 Sakai JT, Young SE, Stallings MC, Timberlake D, Smolen A, Stetler GL *et al.* Case-control and within-family tests for an association between conduct disorder and 5HTTLPR. *Am J Med Genet B Neuropsychiatr Genet* 2006; **141B**: 825–832.
21 Cadoret R, Langbehn D, Caspers K, Troughton E, Yucuis R, Sandhu H *et al.* Associations of the serotonin transporter polymorphism with aggressivity, attention deficit, and conduct disorder in an adoptee population. *Comprehens Psychiatry* 2003; **44**: 88–101.
22 Schulz-Heik RJ, Maentz SK, Rhee SH, Gelhorn HL, Young SE, Timberlake DS *et al.* Case-control and within-family tests for an association between conduct disorder and DAT1. *Psychiatr Genet* 2008; **18**: 17–24.
23 Caspi A, Langley K, Milne B, Moffitt TE, O'Donovan M, Owen MJ *et al.* A replicated molecular genetic basis for subtyping antisocial behavior in children with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* 2008; **65**: 203–210.
24 Monuteaux MC, Biederman J, Doyle AE, Mick E, Faraone SV. Genetic risk for conduct disorder symptom subtypes in an ADHD sample: specificity to aggressive symptoms. *J Am Acad Child Adolesc Psychiatry* 2009; **48**: 757–764.
25 Anney RJ, Lasky-Su J, O'Dushlaine C, Kenny E, Neale BM, Mulligan A *et al.* 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* 2008; **147B**: 1369–1378.
26 Franke B, Neale BM, Faraone SV. Genome-wide association studies in ADHD. *Hum Genet* 2009; **126**: 13–50.
27 Bierut L, Agrawal A, Bucholz K, Doheny KF, Laurie CC, Pugh E *et al.* A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci USA* 2010; **107**: 5082–5087.
28 Cornelis MC, Agrawal A, Cole JC. The Gene, Environment Association Studies Consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genet Epidemiol* 2010; **34**: 364–372.
29 Begleiter H, Reich T, Hesselbrock V, Porjesz B, Li TK, Schuckit M *et al.* The Collaborative Study on the Genetics of Alcoholism. *Alcohol Health Res World* 1995; **19**: 228–236.
30 Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF *et al.* Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet* 2007; **16**: 24–35.
31 Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger Jr JI *et al.* A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J Studies on Alcohol* 1994; **55**: 149–158.
32 Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V. A validity study of the SSAGA–A comparison with the SCAN. *Addiction* 1999; **94**: 1361–1370.
33 Laurie CC, Doheny KF, Mirel DB. Quality control and quality assurance in genotypic data for genome-wide association studies. Under review 2009.
34 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.
35 Purcell S, Neale BM, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 2007; **81**: 559–575.
36 Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 2004; **101**: 6062–6067.
37 Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucl Acids Res* 2009.
38 Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* 2009; **4**: 44–57.
39 Kendler KS, Kuo PH, Webb BT, Kalsi G, Neale MC, Sullivan PF *et al.* A joint genomewide linkage analysis of symptoms of alcohol dependence and conduct disorder. *Alcohol Clin Exp Res* 2006; **30**: 1972–1977.
40 Dick DM, Li TK, Edenberg HJ, Hesselbrock V, Kramer JR, Foroud T. A genome-wide screen for genes influencing conduct disorder. *Mol Psychiatry* 2004; **9**: 81–86.
41 Abe K, Chisaka O, Van RF, Takeichi M. Stability of dendritic spines and synaptic contacts is controlled by alpha N-catenin. *Nat Neurosci* 2004; **7**: 357–363.
42 Park C, Falls W, Finger JH, Longo-Guess CM, Ackerman SL. Deletion in Catna2, encoding alpha N-catenin, causes cerebellar and hippocampal lamination defects and impaired startle modulation. *Nat Genet* 2002; **31**: 279–284.
43 Iacono WG, Carlson SR, Taylor J, Elkins IJ, McGue M. Behavioral disinhibition and the development of substance-use disorders: findings from the Minnesota Twin Family Study. *Dev Psychopathol* 1999; **11**: 869–900.

44 Compton W, Conway KP, Stinson FS, Colliver JD, Grant BF. Prevalence, correlates, and comorbidity of DSM-IV antisocial personality syndromes and alcohol and specific drug use disorders in the United States: results from the national epidemiological survey on alcohol and related conditions. *J Clin Psychiatry* 2005; **66**: 6770685.
45 Dick DM, Bierut L, Hinrichs AL, Fox L, Bucholz KK, Kramer JR *et al*. The role of GABRA2 in risk for conduct disorder and alcohol and drug dependence across developmental stages. *Behav Genet* 2006; **36**: 577–590.
46 Burt SA. Are there meaningful etiological differences within antisocial behavior? Results of a meta-analysis. *Clin Psychol Rev* 2009; **29**: 163–178.
47 Tackett JL, Krueger RF, Iacono WG, McGue M. Symptom-based subfactors of DSM-defined conduct disorder: evidence for etiologic distinctions. *J Abnorm Psychol* 2005; **114**: 483–487.