Association of *COL25A1* with Comorbid Antisocial Personality Disorder (ASPD) and Substance Dependence

Dawei Li[1], **Hongyu Zhao**[2,3], **Henry R. Kranzler**[4,5], **David Oslin**[4,5], **Raymond F Anton**[6], **Lindsay A Farrer**[7], and **Joel Gelernter**[1,3,8,9,*]

[1]Department of Psychiatry, School of Medicine, Yale University, New Haven, Connecticut

[2]Department of Epidemiology and Public Health, School of Medicine, Yale University, New Haven, Connecticut

[3]Department of Genetics, School of Medicine, Yale University, New Haven, Connecticut

[4]Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania

[5]VISN 4 MIRECC, Philadelphia VAMC, Philadelphia, Pennsylvania

[6]Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston, South Carolina

[7]Departments of Medicine (Biomedical Genetics), Neurology, Ophthalmology, Genetics & Genomics, Biostatistics, and Epidemiology, Boston University Schools of Medicine and Public Health, Boston, Massachusetts

[8]VA Connecticut Healthcare Center, West Haven, Connecticut

[9]Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut

Abstract

Background—Antisocial Personality Disorder (ASPD) is a psychiatric disorder characterized by a long-term pattern of manipulating, exploiting or violating the rights of others.

Methods—Subjects ascertained for genetic studies of substance dependence (SD) and diagnosed with ASPD and comorbid SD were included in a two-stage genetic association study. In the discovery stage, 627 single nucleotide polymorphisms (SNPs) located in 179 candidate genes for addiction were analyzed in a case-control cohort and family-based cohort. The significant findings were replicated in an independent case-control cohort.

Results—One SNP, rs13134663, in the collagen XXV alpha 1 gene (*COL25A1*), was significantly associated with ASPD in both African-Americans (AAs) and European Americans (EAs) (smallest P values were 0.0002 and 0.0004, respectively). There was also evidence of association with the same SNP in independent samples of AA and EA cases and controls (P = 0.035 and 0.033, respectively). Analysis of the combined set of case-control subjects yielded an allelic P value of 9×10^{-6} with odds ratio (95% confidence interval) of 1.3 (1.16, 1.47) (smallest P = 1×10^{-7}; Bonferroni threshold P = 0.00012).

*To whom correspondence should be addressed: Joel Gelernter, Department of Psychiatry, Yale University School of Medicine, VA CT Healthcare Center, 950 Campbell Avenue, 116A2, West Haven, CT 06516. joel.gelernter@yale.edu.

Conclusions—The *COL25A1* gene, located at chromosome 4q25, encodes the collagen-like Alzheimer amyloid plaque component precursor, a type II transmembrane protein specifically expressed in neurons; it co-localizes with Aβ in senile plaques in Alzheimer disease brains. This SNP maps to the transcription factor binding site and is conserved in 17 vertebrates, including mice and rats. Our findings suggest that *COL25A1* may be associated with ASPD, especially in the context of SD.

Keywords

Antisocial personality disorder (ASPD); Substance Dependence; Addiction; Association; Collagen XXV alpha 1 (*COL25A1*); Amyloid Plaque

Introduction

Antisocial personality disorder (ASPD) is a "pervasive pattern of manipulating, exploiting, or violating the rights of others that begins in childhood or early adolescence and continues into adulthood"(1). Many of the behaviors characteristic of ASPD result in interventions by the criminal justice system(2). A longitudinal study of a cohort of abusers of several classes of drugs identified ASPD as a predictor of criminal behavior(3). Although the causes of ASPD are largely unknown, a strong hereditary component has been revealed by family, twin and adoption studies(4). Genetic epidemiological studies of antisocial personality and behavior indicated that as much as 56% of the variance in these phenotypes can be explained through genetic factors (5). People with an antisocial or alcoholic parent are at increased risk of ASPD, and people with ASPD have higher rates of alcohol dependence and more alcohol-related problems than people without ASPD(6). The prevalence of ASPD and other adult antisocial behaviors among individuals with alcohol use disorders ranges from 20–33%(7, 8). The combination of alcoholism and ASPD may lead to greater frontal brain deficits than the sum of each(9). Adolescent boys with antisocial conduct disorder and substance use disorder have decreased brain neuronal activity during risky decision-making and reward and increased activity during loss(10).

Genetic association studies show that the dopamine and serotonin transporter protein genes may be associated with ASPD in alcoholics(11). Such studies provide evidence for genetic influences on both antisocial behavior and substance dependence and substantial genetic overlap in these behaviors(12). A functional polymorphism in the *MAOA* gene can moderate the effect of maltreatment; and maltreated children with a genotype conferring high levels of MAOA expression were less likely to develop antisocial problems(13). A few other gene polymorphisms (i.e, *COMT* val/met variant(14), 5-HTTLPR(15), and 5-HTTVNTR(16)) and gene interactions (i.e., *ALDH2**1*1 (glu504ly) and *ANKK1* rs1800497(8)) have also been reported to be associated with ASPD(5). However, our understanding of the genetic contributions to ASPD remains limited. Further evaluation of susceptibility genes for ASPD is warranted(15). In the present study, we identified a single nucleotide polymorphism (SNP) in the collagen XXV alpha 1 gene (*COL25A1*) as being significantly associated with ASPD in African American (AA) and European American (EA) case-control samples and in two independent family samples from the same populations. This association was replicated in independent case-control samples of AAs and EAs. The *COL25A1* gene encodes a type II transmembrane protein, and is specifically expressed in neurons that colocalize with amyloid β (Aβ) in senile plaques in brains from persons with Alzheimer disease, and as such might have implications for brain development as well as brain degeneration.

Methods

Subjects

A total of 4,063 case-control and family-based (including affected sibling pairs) subjects were genotyped in the discovery stage and an additional 4,737 subjects in the replication stage (supplementary Tables 1a and 1b). The subjects were recruited at the Yale University School of Medicine (APT Foundation, New Haven, Connecticut), the University of Connecticut Health Center (Farmington), the Medical University of South Carolina (Charleston), the University of Pennsylvania School of Medicine (Philadelphia), and McLean Hospital (Harvard Medical School, Belmont, Massachusetts) (3,838, 3,334, 770, 555, and 303 subjects, respectively). Families were ascertained through sibling pairs meeting Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) criteria(17) for cocaine or opioid dependence (or both). All subjects were interviewed using an e-version of the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA), as described previously(18, 19). Subjects gave informed consent, as approved by the institutional review board at each clinical site. After a complete description of the study, written informed consent was obtained. Certificates of confidentiality for the work were obtained from both the National Institute on Drug Abuse (NIDA) and the National Institute on Alcohol Abuse and Alcoholism (NIAAA).

Genotyping

DNA was extracted from immortalized cell lines or directly from blood or saliva. The discovery subjects were genotyped using 1,536 SNPs of the NIAAA Addiction Array by Illumina GoldenGate Assay (Illumina, San Diego, California)(20). The replication subjects were genotyped using TaqMan Real-Time PCR method probes (Applied Biosystems, Carlsbad, California) at the significant SNP from discovery analyses. For quality control, 8% of the TaqMan-genotyped subjects and >8% of the discovery subjects were re-genotyped. All the discrepant genotypes (<1%) were excluded from the analyses.

Statistical Analyses

To measure the significance level of the genotype missing rate, a chi-square test was performed on the missing rates of cases and controls for each SNP. The exact Hardy-Weinberg Equilibrium (HWE) test, which is more accurate for rare genotypes than the standard asymptotic test, was carried out, and the unrelated subjects were used for the calculations(21). A revised principal component analysis (PCA)(22) was applied to calculate the principal components of the genotypes of each individual (both case-control and family samples) and to detect population stratification using 137 ancestry informative markers (AIMs) that were included in the microarray. The identification of the outliers is shown in the supplementary method. To identify the admixed individuals, the model-based clustering method implemented in Structure(23, 24) was applied to infer population structure and calculate the European and African ancestry membership of the samples, using 115 unlinked markers among the 137 AIMs where the pairwise distance was larger than 10 kilobases. The quality control flowchart is shown in supplementary Figure 1.

In the discovery stage, for unrelated subjects (case-control samples and probands of family samples), the allelic analysis was carried out as the primary study between ASPD patients and ASPD-free controls using a two-by-two chi-square association test, followed by the additional genotypic analyses and dominant/recessive models. The initial AA-EA cross-validation threshold of the discovery association tests was set at P 0.001 in the allelic or genotypic analysis. To identify whether sex, age, comorbid substance dependence (SD) status, and/or population stratification had a significant influence on the findings, logistic regression was applied to confirm the findings from the chi-square tests. For example, the

first 3 principal components from the discovery genotypes were considered as covariates in the regressions, correcting for population stratification. The resulting corrections minimized spurious associations while maximizing the power to detect true associations. The permutation test implemented in PLINK(25) was used to account for multiple testing. We also compared the findings resulting from the true phenotypes and those from permutated phenotypes (null hypothesis), and this procedure was repeated eight times. In addition, to explore the significant associations, a unified association approach(22) was applied to the combined unrelated and family samples for the family-based association tests correcting for stratification by using the principal components of all the samples.

The combined analyses of the AA and EA case-control samples were performed via genotype pooling and meta-analysis. For the meta-analyses, when heterogeneity was found, the random effects model, which was more conservative and yielded wider confidence intervals (CIs) than the fixed effects model, was adopted; otherwise, both the fixed and random effects models were considered appropriate(26). The number of independent tests was calculated based on the number of uncorrelated SNPs, measured by the pair-wise LD (r^2 values) between the adjacent SNPs(25). Of the 627 SNPs analyzed, there were 407 independent tests ($r^2 < 0.3$), which was used to establish the Bonferroni significance threshold for multiple testing correction in the discovery stage ($P < 0.05/407 = 0.00012$). The threshold was set at $P < 0.05$ for the replication of the identified SNP.

To measure whether the significant associations were due to a technical genotyping artifact that was specific to those SNPs, the flanking SNPs that were in strong linkage disequilibrium (LD) with each significant SNP were identified, and the haplotypes (proxies) formed from those flanking SNPs (excluding the reference SNP) were analyzed for association using the EM algorithm(27). For SNPs with no haplotypes identified in our SNP panel, the raw phased genotypes from the 1000 Genomes Project were used to calculate and build a LD block for each population using Haploview(28) and SNAP(29).

The case-control association analysis procedures that were described above were also applied for the replication and combined samples. Three sub-group/comorbidity analyses were carried out to explore whether the association finding of ASPD was driven by the comorbid phenotypes (e.g., the comorbid SD and conduct disorders). The sub-group/comorbidity analyses included 1) those between the ASPD/SD patients and ASPD-free/SD controls, 2) those between the ASPD/SD patients and ASPD-free controls, and 3) those between the SD/ASPD patients and SD-free/ASPD controls in both the discovery and replication samples as well as in the combined samples. Figure 1 shows the diagram of the analyses.

Results

Quality Control

In the discovery stage, 4,063 subjects were genotyped at 1,536 SNP loci. The genotypes of the samples with Mendelian inconsistency errors (MIEs) were deleted (314 genotypes or < 0.01%). After the subjects with a missing rate > 20% and SNPs with a missing rate > 5% were removed, the rates were recalculated, and the subjects with a missing rate > 5% were excluded. 16 SNPs and 5 SNPs (3 common SNPs) significantly deviating from HWE were excluded for the AA and EA samples, respectively (the threshold was $P < 0.001$ in controls). 3,547 subjects were identified as EA or AA based on subjects' self-identification (and initial classification of EA-Hispanics and AA-Hispanics). Using the revised PCA, seventy-five individuals were identified as outliers for removal, and 1,714 AA and 1,758 EA individuals remained. The overall plots of the first three principal components are shown in Figure 2. Supplementary Figure 2 shows the three-dimensional plots before and after the 75

individuals were removed from the AA and EA samples. The European and African ancestry memberships of the subjects are shown in Figure 3 using the model-based clustering method. Based on the histogram of the inferred ancestry memberships, the individuals with European ancestry membership > 90% were identified as EA and the individuals with African membership > 76% as AA, resulting in 1,548 AA and 1,584 EA subjects. After excluding the subjects with unknown phenotypes for ASPD, 1,486 AA and 1,235 EA subjects remained. Only autosomal SNPs with minor allele frequency (MAF) > 0.1 in both cases and controls were analyzed in the association study, 627 SNPs in 179 genes. Supplementary Table 1a shows the samples and SNPs that were excluded at each quality control step in the discovery stage, and Table 1 shows the final sample size of the discovery and replication subjects.

Association Analyses in the Discovery Samples

In the discovery stage, seven and two SNPs in the AA and EA populations, respectively, were associated with ASPD in the unrelated subjects with a statistical significance level of P 0.001 and were also significant ($P < 0.05$) in the combined unrelated and family samples (supplementary Table 2). The *COL25A1* SNP rs13134663 was the only SNP that was significant in both the AA and EA samples (e.g., the allelic P values were 0.0028 and 0.0004, respectively), and also was the only SNP that achieved the threshold of SNP-level Bonferroni correction ($P < 0.00012$) in the discovery samples. For example, the allelic P values of the combined AA and EA samples were 7×10^{-7} and 5×10^{-6} based on the genotype pooling (Table 2) and meta-analysis (supplementary Table 3), respectively. The genotypic analysis, dominant/recessive models, and permutation tests yielded results similar to those for the allelic analysis (supplementary Table 2), as did the regression models with age, sex, comorbid SD status, and/or principal components from genotypes as covariates (data not shown). For example, the P value was 0.0002 for rs13134663 in the AA samples under the recessive model. The supplementary Table 2 shows the odds ratios (ORs), 95% CIs, and P values of the case-control samples, and the results from the family samples for the SNPs with P 0.001.

Sub-group Analyses of Rs13134663 in the Discovery Samples

For the *COL25A1* SNP rs13134663, the overall OR (95% CI) was 3.65 (1.75, 7.62) with a P value of 0.0002 in the discovery AA population and 1.72 (1.15, 2.57) with a P value of 0.007 in the discovery EA population under the recessive model. When only the subjects of both cases and controls affected with alcohol dependence (AD), cocaine dependence (CD), or marijuana dependence (MjD) were included in the analyses (ASPD/SD patients vs. ASPD free/SD controls), the ORs were 4.22, 2.45, and 6.42 in the AAs, and 1.86, 1.81, and 2.32 in the EAs, respectively (Table 2). Furthermore, when the subjects affected with AD and CD, AD and MjD, CD and MjD, or dependence on all three substances were analyzed, the ORs were larger (e.g., 8.21 and 2.73 in the subjects affected with both AD and MjD in the AA and EA populations under the recessive model, respectively). The supplementary Table 4 shows the allele counts corresponding to Table 2.

Association Analyses of Rs13134663 in the Replication Samples

The same trend described above was also observed in the replication stage. For example, the ORs were 2.78 and 1.19 under the recessive model in the replication AA and EA subjects, respectively, where both cases and controls were diagnosed with dependence on all three substances (ASPD/SD patients vs. ASPD free/SD controls). The replication P values were smaller in AAs than EAs. For example, AA subjects with both CD and MjD showed a P value of 0.0017 (OR = 3.52 (1.54, 8.09)) under the recessive model, and the EA subjects with both AD and MjD had a P value of 0.01 (OR = 2.23 (1.19, 4.19)) under the dominant model. The combined replication AA and EA samples with AD, CD, and MjD had an allelic

P value of 8×10^{-5} with an OR of 1.71 (1.31, 2.24) (ORs = 1.99 (1.34, 2.96) and 1.84 (1.16, 2.91) under the dominant and recessive models, respectively). Table 2 and it's full version (supplementary Table 5) show the summary of the association results in both the discovery and replication samples for the *COL25A1* SNP rs13134663. Supplementary Table 1b shows the samples that were excluded at each quality control step in the replication stage.

Combined Discovery and Replication Analyses of Rs13134663

When the discovery and replication samples were combined, the overall allelic *P* value was 9×10^{-6} with an OR of 1.3 (1.16, 1.47), which was more significant in the case-control subjects comorbid with SD (Table 2). The meta-analysis approach showed similar results, with the overall allelic *P* of 8×10^{-5} (supplementary Table 6). The allelic *P* values were 3×10^{-7} (OR = 1.45 (1.26, 1.67)), 7×10^{-7} (OR = 1.4 (1.22, 1.6)), 8×10^{-6} (OR = 1.5 (1.25, 1.79)), 1×10^{-7} (OR = 1.52 (1.3, 1.77)), 3×10^{-6} (OR = 1.62 (1.32, 1.98)), 5×10^{-6} (OR = 1.56 (1.29, 1.89)), and 2×10^{-6} (OR = 1.68 (1.36, 2.08)) for the samples with AD, CD, MjD, AD and CD, AD and MjD, CD and MjD, and all three substances, respectively (Table 2). When the samples were divided by sex, women (allelic OR = 1.34) showed a larger effect than men (allelic OR = 1.26); however, the Cochran's Q test showed no evidence of significant heterogeneity between the two groups (*P*(Q) > 0.05). The results of sex difference and heterogeneity analyses are shown in supplementary Table 7.

Alternate Subgroup Analyses

By comparison with the analyses where both ASPD patients and controls were stratified by comorbid SD (ASPD/SD patients vs. ASPD-free/SD controls, Table 2), when only the patients had comorbid AD, CD, and MjD dependence (ASPD/SD patients vs. ASPD-free controls; the comorbid SD status was not considered for the controls), the results showed similar evidence of association with the allelic *P* values ≤ 0.035 in both the replication AA and EA populations (ORs = 1.51 (1.03, 2.24) and 1.4 (1.03, 1.91), respectively). However, the combined discovery and replication samples showed slightly stronger evidence of association (*P* = 1×10^{-6} and OR = 1.58 (1.31, 1.9) in the allelic analysis or *P* = 3×10^{-5} and OR = 1.9 (1.4, 2.58) under the recessive model). These results show that there may be minor contribution of SD status to our association finding of ASPD (supplementary Table 8). To answer this question, the following analyses were carried out.

Rs13134663 with Comorbid SD and Conduct Disorders

The same procedure of association tests was also applied between *COL25A1* rs13134663 and each of the comorbid disorders, including the subjects' SD status (AD, CD, and MjD) and conduct disorders; however, there was no evidence of significant association in either AA or EA in the discovery or replication samples using the allelic analysis or dominant/ recessive models (*P* > 0.01). Furthermore, for the observations with 0.05 > *P* > 0.01, the association tests were repeated by controlling for ASPD status (both case and control subjects had ASPD), however, there was no significant evidence observed (*P* > 0.1). All these results suggested that our association findings between *COL25A1* rs13134663 and ASPD are not likely to be significantly driven by the comorbid phenotypes (e.g., the comorbid SD and conduct disorders) (supplementary Tables 9a and 9b).

Permutation Analyses

As another measure to control for false positive results, we repeated the whole discovery analysis pipeline with the permuted data and investigated whether randomness would generate similar findings. As shown above, our study demonstrated that there were 12, 4, and 1 significant SNPs identified in the AA, EA, and both AA and EA samples, respectively, based on the true unrelated samples. However, the permutation analyses

revealed fewer significant findings using the exact same analysis procedures and parameters in each of the permutated datasets (null hypothesis and random phenotypes). The permutation analyses were repeated eight times, and only a range of 1–4, 0–1, and 0 SNPs were identified in the AA, EA, and both AA and EA samples, respectively. Table 3 shows the comparison of the counts of significant hits among the true data and eight permutation datasets. This suggests that we observed an excess of positive results compared to what would have been predicted by chance.

Analyses of Flanking Markers and Haplotypes and LD Structure of Rs13134663

The LD and haplotype analyses showed that the haplotypes of the other six SNPs identified in our AA samples (except rs13134663), as shown in supplementary Table 2, showed evidence of significant associations (available upon request). However, there was no SNP that met the criteria(25) to form a haplotype with rs13134663, based on our SNP panel. Therefore, to understand the LD structure of rs13134663, the *COL25A1* gene, and the flanking regions, the genotype data from the 1000 Genomes Project were used to build plots of LD structure and mapping boundaries by computing all proxy SNPs, and the recombination rate, r^2, and D' values. Strong LD was identified between rs13134663 and other *COL25A1* SNPs. For example, seven SNPs were in complete LD ($r^2 = 1$) and 77 SNPs were in strong LD ($0.8 < r^2 < 1$) based on the European-ancestry samples while only five SNPs were in strong LD ($r^2 = 0.92$) based on the African-ancestry samples. This LD block can be easily distinguished by the recombination hotspot on each side of the block. The block was smaller in the African-ancestry samples than European-ancestry samples, consistent with the earlier origins of African populations. The plots are shown in supplementary Figures 3–4, and the SNPs in complete or strong LD with rs13134663 are shown in supplementary Tables 10.1–10.2.

Discussion

Our findings suggested that *COL25A1* rs13134663 was associated with ASPD, especially in subjects with comorbid SD. The association was cross-validated in the discovery EA and AA samples with the smallest *P* values being 0.0002 and 0.0004, respectively, which were below or close to the gene-level Bonferroni correction threshold for multiple testing (0.05/179 = 0.0003). The combined discovery samples also achieved the SNP-level threshold (0.00012) with the allelic *P* values of 7×10^{-7} and 5×10^{-6} based on the genotype pooling and meta-analysis, respectively. The association was further replicated in the independent EA and AA samples with comorbid SD (*P* = 0.035 and 0.033, respectively), where the threshold was *P* < 0.05 for the one SNP tested. Analysis of the combined discovery and replication subjects yielded the overall allelic *P* values of 9×10^{-6} and 8×10^{-5} based on the genotype pooling and meta-analysis, respectively, which remained significant after the SNP-level Bonferroni correction. The characteristics of the gene or SNP as discussed below support the biological importance of our findings.

Rs13134663 maps to the *COL25A1* (Collagen XXV alpha 1) gene on chromosome 4q25. Collagen XXV alpha 1, also known as collagen-like Alzheimer's amyloid plaque component precursor (CLAC or CLACP), is a type II transmembrane protein specifically expressed in neurons. It colocalizes with Aβ in senile plaque in brains of subjects with Alzheimer disease(30, 31). CLAC is a brain-specific membrane-bound collagen that was originally identified in amyloid preparations in an Alzheimer disease brain. Proteolytic processing releases CLAC, a soluble form of COL25A1 containing the extracellular collagen domains associated with senile plaques in Alzheimer disease brains(32). Forsell et al. reported an association between *COL25A1* SNPs and Alzheimer's disease(33). *COL25A1* was found to trigger and promote Alzheimer's disease-like pathology *in vivo*(30). Rs13134663 maps to a conserved transcription factor binding site and is conserved in 17 vertebrates, including mice

and rats(34), supporting the possible functional importance of this SNP. One possible explanation for the observed associations is that rs13134663 could affect the expression levels of *COL25A1* mRNA.

We found that the association of rs13134663 was stronger in the discovery datasets than the replication datasets. A possible explanation for this is that our two discovery cohorts were more homogeneous based on the analyses of PCA, population structure, and genomic control (e.g., there was no evidence of significant stratification after excluding the samples that might result in ethnic heterogeneity or imbalance of ancestry proportions between cases and controls based on these measures). However, not all of these analyses could be done in the replication datasets, since only one SNP was genotyped in the replication stage. This could also be explained by the "winner's curse", where an initial finding is more significant than what is found in a replication cohort, perhaps due to regression to the mean.

It is notable that multiple strategies were adopted to address the sample ethnicity heterogeneity issue in this study, as described in the quality control section. For example, the outliers were excluded using the automatic outlier-removal approach implemented in EIGENSTRAT(35) and a manual procedure based on the eigenvectors generated by PCA was also applied. By comparison with the automatic approach, our manual procedure to exclude outliers was more stringent and precise in maintaining homogeneity and matching the case and control groups.

By comparison with the overall analyses, the sub-group analyses of case-control subjects with SD (both cases and controls had comorbid SD) showed stronger association. Furthermore, among patients with comorbid AD, CD, and MjD dependence (only the patients had the comorbid phenotypes), a consistent association was observed in both the AA and EA populations, and the effects were larger than those from the overall analyses. Nevertheless, we identified a new genetic susceptibility variant for ASPD, with a greater effect size in people who suffer from comorbid SD.

It may be worthwhile to investigate the expression levels of the gene specifically in brains from the ASPD patients, and explore whether there is a defect related to expression of the *COL25A1* gene that is related to brain development. Many genes required for embryo development are expressed again at later stages of the life cycle(36). A possible model of the *COL25A1* gene is that it may be expressed in childhood, but suppressed and subsequently released and expressed again later in life. Brain eQTL mapping(37) may be useful to corroborate the potential hypothesis, and clarify the functional role of this gene in the pathogenesis of ASPD. Montgomery et al.(38) found four putative eQTL cis SNPs (rs11942576, rs13115905, rs1452688, and rs4368681) at *COL25A1* using RNA-Seq in 60 European HapMap samples (supplementary Figure 5).

In summary, we found seven SNPs significantly associated with ASPD in AAs and two SNPs associated with EAs, after strict quality control in the discovery stage. Evidence of association of the cross-validated *COL25A1* rs13134663 was obtained in independent replication cohorts, particularly in AAs. The overall allelic *P* value was 9×10^{-6} (smallest $P = 1 \times 10^{-7}$) (Bonferroni threshold $P = 0.00012$) with an OR of 1.3 (1.16, 1.47) in the combined AA and EA populations, which was more significant in the subjects with SD. Thus, *COL25A1* rs13134663 may be associated with ASPD, especially when comorbid with AD, CD and/or MjD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition Revision (DSM-IV-R). American Psychiatric Press; Washington, DC: 2000.

2. Hesse M. What should be done with antisocial personality disorder in the new edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V)? BMC Med. 2010; 8:66. [PubMed: 20979622]

3. Fridell M, Hesse M, Jaeger MM, Kuhlhorn E. Antisocial personality disorder as a predictor of criminal behaviour in a longitudinal study of a cohort of abusers of several classes of drugs: relation to type of substance and type of crime. Addict Behav. 2008; 33:799–811. [PubMed: 18258375]

4. van den Bree MB, Svikis DS, Pickens RW. Genetic influences in antisocial personality and drug use disorders. Drug Alcohol Depend. 1998; 49:177–187. [PubMed: 9571383]

5. Ferguson CJ. Genetic contributions to antisocial personality and behavior: a meta-analytic review from an evolutionary perspective. J Soc Psychol. 2010; 150:160–180. [PubMed: 20397592]

6. Moeller FG, Dougherty DM. Antisocial personality disorder, alcohol, and aggression. Alcohol Res Health. 2001; 25:5–11. [PubMed: 11496966]

7. Goldstein RB, Dawson DA, Saha TD, Ruan WJ, Compton WM, Grant BF. Antisocial behavioral syndromes and DSM-IV alcohol use disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Alcohol Clin Exp Res. 2007; 31:814–828. [PubMed: 17391341]

8. Lu RB, Lee JF, Huang SY, Lee SY, Chang YH, Kuo PH, et al. Interaction between ALDH2*1*1 and DRD2/ANKK1 TaqI A1A1 genes may be associated with antisocial personality disorder not co-morbid with alcoholism. Addict Biol. 2010 [Epub ahead of print].

9. Oscar-Berman M, Valmas MM, Sawyer KS, Kirkley SM, Gansler DA, Merritt D, et al. Frontal brain dysfunction in alcoholism with and without antisocial personality disorder. Neuropsychiatr Dis Treat. 2009; 5:309–326. [PubMed: 19557141]

10. Crowley TJ, Dalwani MS, Mikulich-Gilbertson SK, Du YP, Lejuez CW, Raymond KM, et al. Risky decisions and their consequences: neural processing by boys with Antisocial Substance Disorder. PLoS One. 2010; 5:e12835. [PubMed: 20877644]

11. Reese J, Kraschewski A, Anghelescu I, Winterer G, Schmidt LG, Gallinat J, et al. Haplotypes of dopamine and serotonin transporter genes are associated with antisocial personality disorder in alcoholics. Psychiatr Genet. 2010; 20:140–152. [PubMed: 20505557]

12. Corley RP, Zeiger JS, Crowley T, Ehringer MA, Hewitt JK, Hopfer CJ, et al. Association of candidate genes with antisocial drug dependence in adolescents. Drug Alcohol Depend. 2008; 96:90–98. [PubMed: 18384978]

13. Caspi A, McClay J, Moffitt TE, Mill J, Martin J, Craig IW, et al. Role of genotype in the cycle of violence in maltreated children. Science. 2002; 297:851–854. [PubMed: 12161658]

14. Thapar A, Langley K, Fowler T, Rice F, Turic D, Whittinger N, et al. Catechol O-methyltransferase gene variant and birth weight predict early-onset antisocial behavior in children with attention-deficit/hyperactivity disorder. Arch Gen Psychiatry. 2005; 62:1275–1278. [PubMed: 16275815]

15. Douglas K, Chan G, Gelernter J, Arias A, Anton R, Poling J, et al. 5-HTTLPR as a potential moderator of the effects of adverse childhood experiences on risk of antisocial personality disorder. Psychiatr Genet. 2011 [Epub ahead of print].

16. Garcia LF, Aluja A, Fibla J, Cuevas L, Garcia O. Incremental effect for antisocial personality disorder genetic risk combining 5-HTTLPR and 5-HTTVNTR polymorphisms. Psychiatry Res. 2010; 177:161–166. [PubMed: 20363030]

17. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM). Fourth Edition. American Psychiatric Press; Washington, DC: 1994.

18. Pierucci-Lagha A, Gelernter J, Chan G, Arias A, Cubells JF, Farrer L, et al. Reliability of DSM-IV diagnostic criteria using the semi-structured assessment for drug dependence and alcoholism (SSADDA). Drug Alcohol Depend. 2007; 91:85–90. [PubMed: 17590536]

19. Gelernter J, Panhuysen C, Weiss R, Brady K, Hesselbrock V, Rounsaville B, et al. Genomewide linkage scan for cocaine dependence and related traits: significant linkages for a cocaine-related trait and cocaine-induced paranoia. Am J Med Genet B Neuropsychiatr Genet. 2005; 136B:45–52. [PubMed: 15909294]

20. Hodgkinson CA, Yuan Q, Xu K, Shen PH, Heinz E, Lobos EA, et al. Addictions biology: haplotype-based analysis for 130 candidate genes on a single array. Alcohol Alcohol. 2008; 43:505–515. [PubMed: 18477577]

21. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet. 2005; 76:887–893. [PubMed: 15789306]

22. Zhu X, Li S, Cooper RS, Elston RC. A unified association analysis approach for family and unrelated samples correcting for stratification. Am J Hum Genet. 2008; 82:352–365. [PubMed: 18252216]

23. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155:945–959. [PubMed: 10835412]

24. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003; 164:1567–1587. [PubMed: 12930761]

25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]

26. Li D, Sham PC, Owen MJ, He L. Meta-analysis shows significant association between dopamine system genes and attention deficit hyperactivity disorder (ADHD). Hum Mol Genet. 2006; 15:2276–2284. [PubMed: 16774975]

27. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol. 1995; 12:921–927. [PubMed: 7476138]

28. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263–265. [PubMed: 15297300]

29. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008; 24:2938–2939. [PubMed: 18974171]

30. Tong Y, Xu Y, Scearce-Levie K, Ptacek LJ, Fu YH. COL25A1 triggers and promotes Alzheimer's disease-like pathology in vivo. Neurogenetics. 2010; 11:41–52. [PubMed: 19548013]

31. Hashimoto T, Wakabayashi T, Watanabe A, Kowa H, Hosoda R, Nakamura A, et al. CLAC: a novel Alzheimer amyloid plaque component derived from a transmembrane precursor, CLAC-P/collagen type XXV. EMBO J. 2002; 21:1524–1534. [PubMed: 11927537]

32. Osada Y, Hashimoto T, Nishimura A, Matsuo Y, Wakabayashi T, Iwatsubo T. CLAC binds to amyloid beta peptides through the positively charged amino acid cluster within the collagenous domain 1 and inhibits formation of amyloid fibrils. J Biol Chem. 2005; 280:8596–8605. [PubMed: 15615705]

33. Forsell C, Bjork BF, Lilius L, Axelman K, Fabre SF, Fratiglioni L, et al. Genetic association to the amyloid plaque associated protein gene COL25A1 in Alzheimer's disease. Neurobiol Aging. 2010; 31:409–415. [PubMed: 18501477]

34. Washington University in St. Louis SoM. SNPseek - a web based software tool for accessing alleles, frequencies, and other information about SNPs. 2011

35. 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. [PubMed: 16862161]

36. Franzmann L, Patton DA, Meinke DW. In vitro morphogenesis of arrested embryos from lethal mutants of Arabidopsis thafiana. Theor Appl Genet. 1989; 77:609–616.

37. Liu C. Brain expression quantitative trait locus mapping informs genetic studies of psychiatric diseases. Neurosci Bull. 2011; 27:123–133. [PubMed: 21441974]

38. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. Nature. 2011; 464:773–777. [PubMed: 20220756]
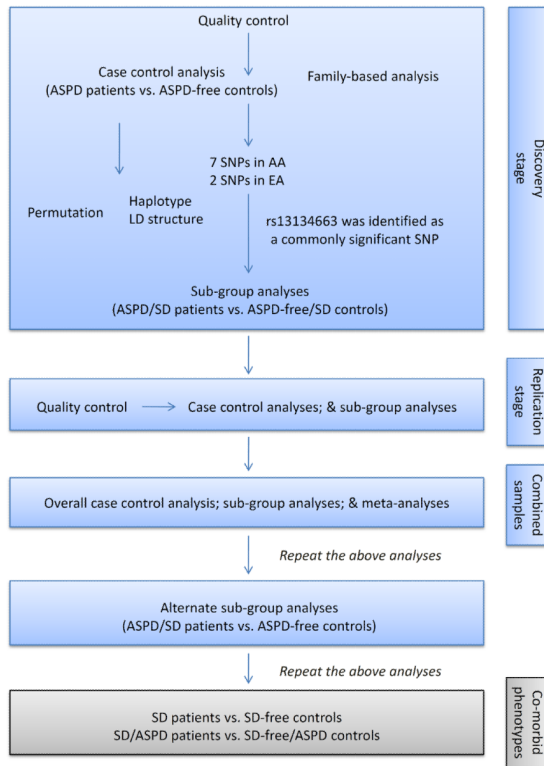
Figure 1.
Diagram of the analyses.
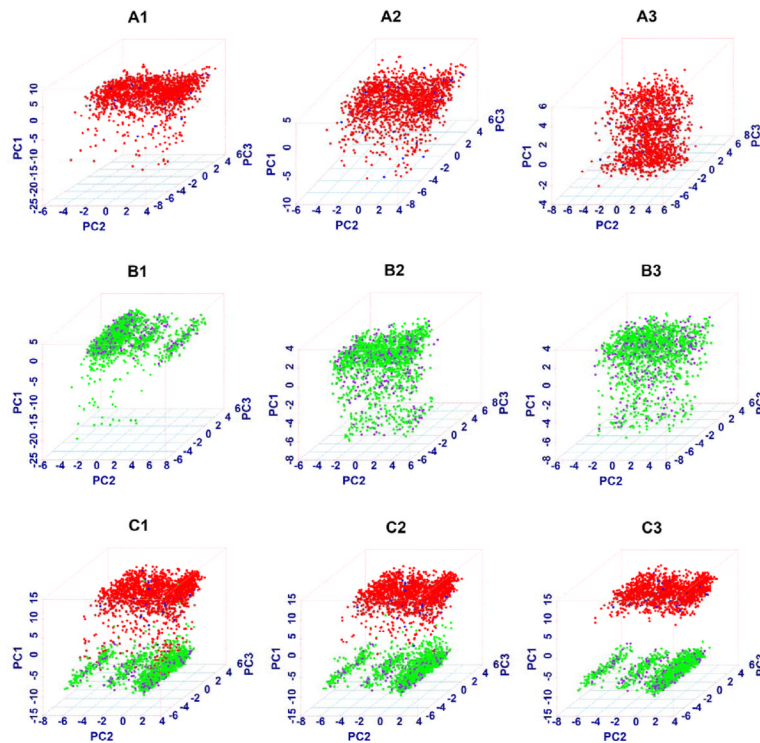
Figure 2.
Three-dimension plots of first three principal components of the genotypes. Figures a1, a2, and a3 represent all the AA samples, those excluding 75 outliers, and those further excluding 340 "in-between" individuals, respectively. Figures b1, b2, and b3 represent the corresponding results for the EA samples. Figures c1, c2, and c3 show the corresponding results when the AA and EA samples are combined. Red: AA patients; Blue: AA controls; Green: EA patients; and Purple: EA controls. After the quality control and filtering, no sample between the AA and EA groups was identified (c3) and the patients and controls were evenly distributed.
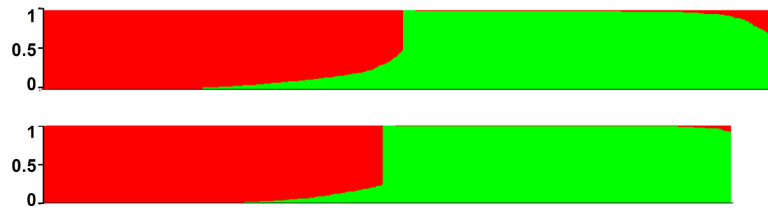
Figure 3.
Ancestry memberships. Upper and lower graphs represent the samples before and after excluding the 340 individuals (i.e., AA membership between 76% and 10%). Red: all of the AA samples (patients and controls) and 60 unrelated HapMap African samples; Green: all of the EA samples and 60 unrelated HapMap European samples. Linkage model(24) was applied and the Hapmap samples were used as ancestry populations.

Table 1

The subjects included in the association study

	AA	Affected (Men/Women)	Unaffected (Men/Women)	EA	Affected (Men/Women)	Unaffected (Men/Women)	Total
Discovery Unrelated	1049	120 (88/32)	929 (422/507)	909	121 (100/21)	788 (431/357)	1958
Discovery Unrelated + Family	1486	166 (118/48)	1320 (594/726)	1235	160 (135/25)	1075 (563/512)	2721
Replication Unrelated	1751	205 (148/57)	1546 (833/713)	1662	212 (54/158)	1450 (766/684)	3413
Total	3237	371 (266/105)	2866 (1427/1439)	2897	372 (189/183)	2525 (1329/1196)	6134

Unrelated = population subjects + probands of family subjects

Table 2

Sub–group and Replication Association Tests for rs13134663

Tests	Discovery AA OR (95% CI)	P	Discovery EA OR (95% CI)	P	Discovery AA + EA OR (95% CI)	P	Replication AA OR (95% CI)	P	Replication EA OR (95% CI)	P	Replication AA + EA OR (95% CI)	P	Combined AA + EA OR (95% CI)	P
All														
Allelic	1.6 (1.17,2.18)	**0.0028**	1.64 (1.24,2.16)	**0.00045**	1.63 (1.34,1.97)	**7.E-07**	1.12 (0.88,1.42)	0.37	1.13 (0.92,1.39)	0.23	1.15 (0.99,1.33)	0.068	1.3 (1.16,1.47)	**9.E-06**
Dom	1.5 (1.02,2.21)	**0.037**	2.21 (1.29,3.78)	**0.0031**	1.7 (1.28,2.25)	**0.00020**	0.98 (0.73,1.32)	0.89	1.23 (0.86,1.74)	0.25	1.11 (0.9,1.36)	0.34	1.29 (1.09,1.53)	**0.0026**
Rec	3.65 (1.75,7.62)	**0.00025**	1.72 (1.15,2.57)	**0.0073**	1.98 (1.43,2.75)	**3.E-05**	2 (1.18,3.39)	**0.0084**	1.14 (0.83,1.57)	0.42	1.33 (1.02,1.72)	**0.035**	1.54 (1.26,1.89)	**3.E-05**
AD														
Allelic	1.52 (1.05,2.21)	**0.028**	1.75 (1.25,2.44)	**0.0010**	1.61 (1.28,2.02)	**5.E-05**	1.28 (0.95,1.71)	0.10	1.26 (0.98,1.62)	0.076	1.36 (1.13,1.63)	**0.00085**	1.45 (1.26,1.67)	**3.E-07**
Dom	1.41 (0.9,2.23)	0.13	2.47 (1.26,4.82)	**0.0067**	1.63 (1.18,2.27)	**0.0033**	1.16 (0.81,1.67)	0.41	1.41 (0.9,2.21)	0.14	1.36 (1.05,1.76)	**0.018**	1.46 (1.2,1.79)	**0.00022**
Rec	4.22 (1.53,11.63)	**0.0026**	1.86 (1.15,2.99)	**0.010**	2.01 (1.36,2.98)	**0.00040**	2.17 (1.14,4.16)	**0.017**	1.32 (0.9,1.95)	0.16	1.62 (1.18,2.23)	**0.0028**	1.76 (1.38,2.26)	**6.E-06**
CD														
Allelic	1.35 (0.95,1.92)	0.089	1.64 (1.19,2.28)	**0.0027**	1.54 (1.23,1.92)	**0.00014**	1.16 (0.89,1.51)	0.28	1.23 (0.97,1.55)	0.087	1.33 (1.12,1.57)	**0.00086**	1.4 (1.22,1.6)	**7.E-07**
Dom	1.28 (0.83,1.96)	0.26	2.03 (1.11,3.73)	**0.020**	1.53 (1.12,2.09)	**0.0079**	1.03 (0.74,1.43)	0.88	1.4 (0.93,2.11)	0.11	1.31 (1.04,1.65)	**0.023**	1.38 (1.15,1.67)	**0.00063**
Rec	2.45 (1.05,5.73)	**0.032**	1.81 (1.13,2.91)	**0.013**	1.96 (1.33,2.88)	**0.00052**	2.08 (1.15,3.75)	**0.013**	1.24 (0.86,1.77)	0.24	1.59 (1.19,2.13)	**0.0018**	1.71 (1.36,2.16)	**5.E-06**
MjD														
Allelic	1.58 (0.97,2.56)	0.063	2.08 (1.33,3.25)	**0.0011**	1.73 (1.28,2.33)	**0.00031**	1.33 (0.92,1.92)	0.12	1.19 (0.88,1.62)	0.26	1.38 (1.11,1.73)	**0.0044**	1.5 (1.25,1.79)	**8.E-06**
Dom	1.42 (0.79,2.55)	0.24	3.58 (1.35,9.49)	**0.0070**	1.72 (1.12,2.65)	**0.013**	1.15 (0.73,1.83)	0.54	1.67 (0.97,2.85)	0.061	1.5 (1.09,2.07)	**0.013**	1.58 (1.22,2.04)	**0.00047**
Rec	6.42 (1.49,27.71)	**0.0047**	2.32 (1.23,4.37)	**0.0083**	2.42 (1.44,4.09)	**0.00071**	2.51 (1.18,5.34)	**0.014**	1 (0.63,1.59)	0.99	1.44 (0.98,2.11)	0.063	1.72 (1.27,2.34)	**0.00050**
AD + CD														
Allelic	1.5 (1.2,2.3)	**0.047**	1.87 (1.27,2.74)	**0.0013**	1.64 (1.27,2.12)	**0.00013**	1.36 (0.99,1.87)	0.057	1.25 (0.95,1.64)	0.12	1.45 (1.19,1.76)	**0.00021**	1.52 (1.3,1.77)	**1.E-07**
Dom	1.36 (0.84,2.21)	0.21	2.44 (1.15,5.16)	**0.017**	1.58 (1.1,2.26)	**0.012**	1.23 (0.83,1.82)	0.30	1.45 (0.88,2.38)	0.15	1.48 (1.12,1.96)	**0.0057**	1.52 (1.22,1.89)	**0.00020**
Rec	4.14 (1.46,11.75)	**0.0041**	2.16 (1.25,3.75)	**0.0052**	2.3 (1.47,3.58)	**0.00018**	2.7 (1.31,5.56)	**0.0052**	1.27 (0.84,1.94)	0.26	1.73 (1.23,2.45)	**0.0017**	1.92 (1.46,2.52)	**2.E-06**
AD + MjD														
Allelic	1.54 (0.92,2.59)	0.10	2.19 (1.29,3.7)	**0.0031**	1.66 (1.19,2.32)	**0.0029**	1.32 (0.87,1.98)	0.19	1.43 (1.01,2.02)	**0.044**	1.59 (1.23,2.04)	**0.00031**	1.62 (1.32,1.98)	**3.E-06**
Dom	1.31 (0.7,2.47)	0.40	3.05 (1.01,9.23)	**0.040**	1.49 (0.93,2.39)	0.099	1.21 (0.72,2.03)	0.48	2.23 (1.19,4.19)	**0.011**	1.81 (1.25,2.61)	**0.0015**	1.68 (1.26,2.25)	**0.00039**
Rec	8.21 (1.55,43.67)	**0.0038**	2.73 (1.31,5.69)	**0.0063**	2.65 (1.47,4.78)	**0.00085**	2.06 (0.88,4.83)	0.091	1.21 (0.72,2.03)	0.48	1.66 (1.08,2.55)	**0.020**	1.96 (1.39,2.77)	**0.00012**
CD + MjD														
Allelic	1.68 (1.02,2.79)	**0.042**	1.96 (1.21,3.19)	**0.0062**	1.65 (1.19,2.28)	**0.0024**	1.55 (1.06,2.28)	**0.025**	1.19 (0.86,1.65)	0.29	1.52 (1.19,1.92)	**0.00060**	1.56 (1.29,1.89)	**5.E-06**

	Discovery AA		Discovery EA		Discovery AA + EA		Replication AA		Replication EA		Replication AA + EA		Combined AA + EA	
Dom		1.55 (0.83,2.88)	0.16	2.81 (1.04,7.61)	0.036	1.61 (1.02,2.56)	0.042	1.35 (0.83,2.2)	0.23	1.57 (0.89,2.78)	0.12	1.65 (1.17,2.33)	0.0043	1.64 (1.24,2.16)	0.00045
Rec	0.0095	5.7 (1.32,24.68)	0.016	2.35 (1.16,4.77)	0.0031	2.35 (1.32,4.19)	0.0017	3.52 (1.54,8.09)	0.87	1.04 (0.64,1.71)	0.014	1.66 (1.1,2.5)	0.00022	1.86 (1.33,2.6)	
AD + CD + MjD															
Allelic	0.048	1.74 (1,3.03)	0.014	2 (1.14,3.5)	0.0076	1.63 (1.14,2.33)	0.070	1.5 (0.97,2.31)	0.074	1.4 (0.97,2.03)	8.E-05	1.71 (1.31,2.24)	2.E-06	1.68 (1.36,2.08)	
Dom	0.21	1.54 (0.78,3.02)	0.11	2.44 (0.79,7.56)	0.13	1.48 (0.89,2.46)	0.26	1.37 (0.79,2.4)	0.021	2.19 (1.11,4.32)	0.00061	1.99 (1.34,2.96)	0.00026	1.78 (1.3,2.44)	
Rec	0.0083	7.16 (1.34,38.24)	0.016	2.62 (1.18,5.81)	0.016	2.53 (1.34,4.77)	0.0035	2.78 (1.08,7.12)	0.54	1.19 (0.68,2.07)	0.0087	1.84 (1.16,2.91)	0.00013	2.05 (1.41,2.97)	

All: all the samples without considering the status of substance dependence (SD); AD: alcohol dependence; CD: cocaine dependence; MjD: marijuana dependence. Some of the upper 95% CI values of ORs are large due to the small sample size in the sub-groups of comorbid substance(s). The full version is shown in Supplementary Table 5.

Table 3

Comparison of the counts of significant hits among the real data and permuted datasets

	Real	Perm 1	Perm 2	Perm 3	Perm 4	Perm 5	Perm 6	Perm 7	Perm 8
AA									
	12	1	4	4	2	3	3	4	4
EA									
	4	0	0	1	1	0	0	0	1
Common in AA and EA									
	1	0	0	0	0	0	0	0	0

Perm represents permutation.