Human Molecular Genetics, 2010, Vol. 19, No. 12 **2497–2506**
doi:10.1093/hmg/ddq112
Advance Access published on March 23, 2010

A systematic gene-based screen of chr4q22–q32 identifies association of a novel susceptibility gene, *DKK2*, with the quantitative trait of alcohol dependence symptom counts

Gursharan Kalsi[1,*], Po-Hsiu Kuo[2], Fazil Aliev[1,3], Jeffry Alexander[1], Omari McMichael[1], Diana G. Patterson[4], Dermot Walsh[5], Zhongming Zhao[6], Marc Schuckit[7], John Nurnberger Jr[8], Howard Edenberg[9], John Kramer[10], Vladimir Vladimirov[1], Carol A. Prescott[11], Danielle M. Dick[1], Kenneth S. Kendler[1] and Brien P. Riley[1]

[1]Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, 800 E Leigh Street, Richmond, VA, USA, [2]Department of Public Health, National Taiwan University, Taipei, Taiwan, [3]Department of Statistics, Ankara University, Ankara, Turkey, [4]Shaftesbury Square Hospital, Belfast, UK, [5]Health Research Board, Dublin, Ireland, [6]Department of Biomedical Informatics, Vanderbilt University Medical Center, 2525 West End Avenue, Suite 600, Nashville, TN 37203, USA, [7]School of Medicine, University of California, San Diego, 8950 Villa La Jolla Drive, Suite B-218, San Diego, CA, USA, [8]Department of Psychiatry, Institute of Psychiatric Research and [9]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 791 Union Drive, Indianapolis, IN 46202, USA, [10]Department of Psychiatry, Medical Education Building, University of Iowa School of Medicine, 500 Newton Road, Iowa City, IA 52242-1000, USA and [11]Department of Psychology, SGM 501, University of Southern California, 3620 So. McClintock Avenue, Los Angeles, CA, USA

Received October 5, 2009; Revised March 1, 2010; Accepted March 16, 2010

Studies of alcohol dependence (AD) have consistently found evidence of linkage on chromosome 4q21–q32. A genome-wide linkage scan in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) sample also provided its strongest evidence of linkage on chromosome 4q22–q32 using an index of AD severity based on the count of DSM-IV AD symptoms (ADSX; LOD = 4.59). We conducted a systematic, gene-centric association study using 518 LD-tagging single nucleotide polymorphisms (SNPs) in the 65 known and predicted genes within the 1-LOD interval surrounding the linkage peak. Case-only regression analysis with the quantitative variable of ADSX was performed in the 562 genetically independent cases; nominal support for association was demonstrated by 32 tagging SNPs in 14 genes. We did not observe study-wide significance, but gene-wise correction for multiple testing with the Nyholt procedure yielded empirical evidence of association with two genes, *DKK2* (dickkopf homolog 2) ($P = 0.007$) and *EGF* (epidermal growth factor) ($P = 0.025$) in the IASPSAD sample. Three SNPs in *DKK2* (rs427983; rs419558; rs399087) demonstrated empirical significance. Assessment of possible replication in 847 cases of European descent from a large independent sample, the Collaborative Study of the Genetics of Alcoholism, yielded replication for *DKK2* but not *EGF*. We observed genotypic and phenotypic replication for *DKK2* with the three SNPs yielding significant association with ADSX in the IASPSAD sample. Haplotype-specific expression measurements in post-mortem tissue samples suggested a functional role for *DKK2*. This evidence notwithstanding, replication is needed before confidence can be placed in these findings.

*To whom correspondence should be addressed at: Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Biotech One Suite 110, 800 E Leigh Street, Richmond, VA 23298, USA. Tel: +1 8048283595; Fax: +1 8048281471; Email: gkalsi@vcu.edu

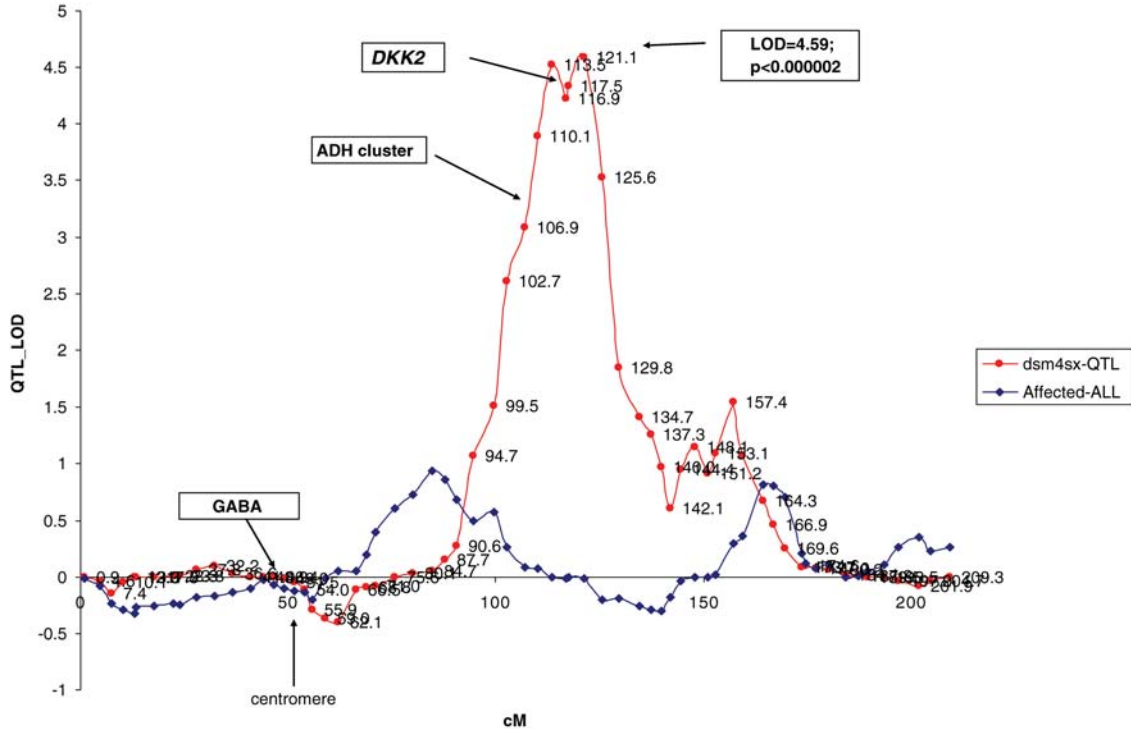


Figure 1. Linkage support interval on chr4q22–q32 includes the *ADH* cluster but is not included in the 1-LOD interval surrounding the peak. The *ADH* cluster is located between the markers D4s2986 and D4s1572 at 106.9 and 110.1 cM, respectively. The map positions (in cM), denoted on this figure, refer to the DeCode map.

INTRODUCTION

Alcohol dependence (AD) is a common, complex, behavioral disorder with both environmental and genetic contributions (1). Epidemiological studies suggest that genetic effects account for 40–60% of the variance in risk for alcoholism (2). Much progress has been made recently in identifying susceptibility loci and candidate genes (3–5). One region that has shown consistent linkage is the susceptibility locus on chromosome 4q21–q32, particularly when the phenotype has indexed quantitative measures, such as severity of alcoholism. The strongest evidence was reported by the Collaborative Study on the Genetics of Alcoholism (COGA) for two severity-related phenotypes: a quantitative measure of the maximum number of drinks consumed in a 24 h period (LOD of 3.5) (6) and a multiple-threshold definition of AD adjusted for age and sex, which also yielded a peak LOD of 3.5 but with different markers (7). We previously reported a genome-wide linkage scan using 1407 sibs and relatives from the 474 linkage-informative families in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) sample, a large, ethnically homogeneous, severely affected sample collected in Ireland and Northern Ireland (8). Non-parametric linkage analyses for AD and the number of DSM-IV AD symptoms endorsed (ADSX) yielded support for a large linkage interval (~55 Mb) on 4q21–q32, with the strongest evidence being a maximum multipoint LOD score of 4.59 ($P = 2.1 \times 10^{-6}$) with ADSX as the phenotype (Fig. 1).

The genomic region on 4q21–q32 is particularly relevant to AD, because several functionally plausible candidate genes are localized to this region, including genes encoding subunits of alcohol dehydrogenase (ADH). The *ADH* genes have the potential for a direct relationship with the physiological aspects of AD (9), and associations between *ADH* genes and AD risk are among the most widely replicated of any gene (3,10–15). Our study in the IASPSAD sample demonstrated strongest support for association with AD and single nucleotide polymorphisms (SNPs) in *ADH5* ($P = 0.004$) and *ADH1B* ($P = 0.005$) (16). In the COGA sample, a comprehensive examination of all seven *ADH* genes supported association of DSM-IV criteria of AD with SNPs located in a broad region across A*DH4* and in the intergenic region between *ADH4* and *ADH5* (13). Additional tests in *ADH1A* and *ADH1B* produced association, not with DSM-IV criteria but with the broader COGA definition of AD DSM-IIIR plus Feighner definition of alcoholism (17). Recently, the crucial relationship between drinking measures and different variants of *ADH* and *ALDH2* (aldehyde dehydrogenase subunit 2) was examined; considering that different variants have differing enzymatic activity, this would be expected to affect drinking behaviors and therefore severity (18). The results substantiated the hypothesis; a non-synonymous SNP (rs1229984) in *ADH1B* was significantly associated with higher level of alcohol consumption, reaction to alcohol and flushing in Europeans. Furthermore, association was demonstrated with quantitative traits that may affect risk, namely *ALDH2* and AD symptom score, and *ADH5* and frequency of alcohol use. One other gene in this region is worth mentioning; *SNCA* (α-synuclein) showed association with craving in the COGA sample (19). In another study, differential expression

of α-synuclein mRNA was observed in blood samples of alcohol-dependent subjects compared with healthy control samples. These differences were correlated with craving and were significantly higher during withdrawal (20).

Here, we conducted a systematic study of the 1-LOD interval surrounding the linkage peak; 65 genes and expressed sequence tags (ESTs) were identified in this interval from HapMap Phase II database. We used LD-tagging SNPs (tSNPs) to test genetically independent cases from the IASPSAD sample and Irish population controls. Prior studies in the COGA sample had tested just two of the genes: *NFKB1* (nuclear factor of kappa light polypeptide gene enhancer in B-cells 1) was associated with AD, particularly in subjects with earlier age of onset (21) and *TACR3* (Tachykinin receptor 3) was associated with comorbid alcohol and cocaine dependence (22).

RESULTS

Results in the IASPSAD sample

Overall, 397 of 518 (76.7%) tSNPs were available for the final analysis of association for AD and ADSX (full results are presented in Supplementary Material, Table S1). Because genotyping success was <100%, we assessed the maximum linkage disequilibrium (LD) captured by submitting successfully genotyped tSNPs to TAGGER (23) and noting the percentage LD captured. In 92% of the genes (60 of 65), our data captured at least 60% of the LD, 70% of the total number of genes captured at least 75% of the LD and 26 (40%) captured 90% or more of the LD within the gene (Supplementary Material, Table S2). Only three genes had <50% of the tSNP data: *PAPSS1* (3′ phosphoadenosine 5′-phosphosulfate synthase 1), *CYP2U1* (cytochrome P450, family 2, subfamily U, polypeptide 1) and *IF* (I factor).

Association results for AD

Single-marker association tests with the dichotomous AD case/control status produced nine genes with at least one tSNP showing nominal $P \leq 0.05$. Estimated values of the odds ratios (ORs) varied from 1.19 to 1.29 (Table 1). The lowest *P*-value ($P = 0.012$) was obtained with rs4441820 in *ARSJ* (arylsulfatase, family member J) and the highest OR (1.29) was produced by rs298991 in *SEC24D* (SEC24-related gene family, member D), ($P = 0.013$). Haplotype tests of association produced nominal significance in the following genes: *DKK2* (dickkopf homolog 2), *PAPSS1*, *COL25A1* (collagen, type XXV, alpha 1), *EGF* (epidermal growth factor), *ANK2* (ankyrin 2), *SYNPO2* (synaptopodin 2) and a cluster including *IF*, *NOLA1* (nucleolar protein family A, member 1), *RRH* (retinal pigment epithelium-derived rhodopsin homolog) and FLJ44691. However, none of the results survived permutation testing.

Association results for ADSX

Consistent with the original linkage evidence, our results with ADSX were stronger than those using the AD diagnostic status (Fig. 2). In case-only regression tests (with age and gender as covariates), 32 tSNPs in 12 genes produced nominal significance (Table 1). *DKK2* demonstrated the strongest evidence, yielding four SNPs with $P < 0.05$ and two SNPs produced $P = 0.0004$ (rs419558 and rs399087) (Table 2). Gene-wise testing showed that *DKK2* and *EGF* survived Bonferroni-corrected significance. *DKK2* produced a gene-wise empirical *P*-value of 0.007; three SNPs (rs427983, rs419558 and rs399087) produced *P*-values smaller than this value, thus could be considered to have survived correction. *EGF* yielded an empirical *P*-value of 0.02; two tSNPs (rs6850557 and rs2237052) survived correction under the null hypothesis (Table 1). In the regression-based haplotype analyses for *DKK2*, a four-marker haplotype ($P = 0.0004$) was associated with increased symptoms counts (Table 3).

Replication of association with *DKK2* and *EGF* in the COGA sample

In the interest of making a parallel comparison, assessment in the COGA data set was limited to SNPs yielding genotypes in the Irish sample. There were several overlapping SNPs, particularly in *DKK2*; where direct overlap was missing, a proxy SNP from the same LD block was selected for the comparison. Using a parallel quantitative phenotype of ADSX, unrelated cases of European descent in the COGA sample produced five SNPs yielding $P < 0.05$ (Table 4). In comparing the four tSNPs yielding $P < 0.05$ in the IASPSAD sample, we found three SNPs (rs427983, rs419558 and rs419764) were available for direct comparison in the COGA sample; all three were also associated with ADSX in the COGA sample. The most significant results in the IASPSAD sample were observed for rs419558 and rs399087 ($P = 0.0004$); in the COGA sample, rs419558 produced $P = 0.021$ and the proxy SNP for rs399087 (rs411143) produced $P = 0.038$. The SNPs rs427983 and rs419764 produced *P*-values of 0.003 and 0.009, respectively, in the IASPSAD sample and 0.021 and 0.004, respectively, in the COGA sample. In both samples, the minor alleles of all four markers were associated with increased symptom counts. Moreover, the risk alleles (increased symptom counts) were the same, with similar allele frequencies (Table 4) and the LD pattern was also similar in both samples (Supplementary Material, Fig. S1a and b). No replication was observed for *EGF* in the COGA sample (Table 4).

Expression results for *DKK2*

We assessed gene expression differences for *DKK2* in RNA from the prefrontal cortex. Allele frequencies of the SNPs making up the haplotype associated with higher ADSX (rs427983, rs419558, rs419764 and rs447372) were similar in both the IASPSAD sample and the post-mortem sample. Our analysis in the control samples ($n = 29$) produced the Kruskal–Wallis H statistic of 7.118; $P = 0.028$, when fold expression means were compared in the groups of different haplotype pairings. Haplotype reconstruction identified $n = 5$ samples carrying the haplotype associated with higher ADSX (Fig. 3); results showed that there was a reduced expression in this group. When the analysis was assessed in the larger sample ($n = 84$), we observed a similar trend (Kruskal–Wallis H statistic 7.403; $P = 0.025$). These differences were not confounded by the effect of potential covariates (age, pH, smoking, etc).

Table 1. Single-marker association results with $P < 0.05$ for AD and ADSX

Gene symbol	SNP ID	Position	AD, uncorrected *P*-values	OR	ADSX, uncorrected *P*-values	ADSX, empirical *P*-values
MANBA	rs223489	104038182			0.029	0.012
DKK2	rs427983	108185605			0.003	**0.007**
DKK2	rs419558	108201696			0.0004	
DKK2	rs419764	108201761			0.009	
DKK2	rs447372	108211600	0.016	1.238		
DKK2	rs399087	108222755			0.0004	
PAPSS1	rs1514730	108888489			0.014	0.012
PAPSS1	rs9569	108892939	0.015	1.269		
PAPSS1	rs2726670	108893989			0.015	
COL25A1	rs2030844	110199642			0.031	0.002
COL25A1	rs1526141	110286663	0.046	1.194		
EGF	rs2881559	111198550			0.047	**0.025**
EGF	rs6850557	111268619			0.015	
EGF	rs2237052	111268629			0.016	
ELOVL6	rs7694475	111434765			0.016	0.004
ELOVL6	rs7681062	111439774			0.021	
PITX2	rs976568	111908325			0.03	0.025
PITX2	rs994978	111910002			0.033	
ALPK1	rs7688190	113687069	0.021	1.232		0.005
ALPK1	rs6812485	113689394			0.039	
NEUROG2	rs701760	113796816			0.026	NA
ANK2	rs313975	114343317			0.015	0.001
ANK2	rs10029516	114368692			0.018	
ANK2	rs592670	114370890			0.033	
ANK2	rs626267	114374582			0.024	
ANK2	rs695002	114375568	0.047	1.187		
ANK2	rs313956	114381453	0.016	1.276	0.025	
ANK2	rs17045600	114391220			0.044	
ANK2	rs1351998	114426531	0.014	1.245		
ANK2	rs10516593	114436416			0.006	
ANK2	rs13134375	114436535			0.042	
CAMK2D	rs17046072	114732417	0.019	1.281		0.004
CAMK2D	rs1880529	114736983			0.035	
CAMK2D	rs13113625	114751348			0.042	
CAMK2D	rs6533690	114784348			0.047	
CAMK2D	rs13130985	114789941			0.026	
CAMK2D	rs4834349	114804795			0.022	
CAMK2D	rs2040744	114990968	0.034	1.203		
ARSJ	rs12645879	115200346	0.027	1.197		0.005
ARSJ	rs4441820	115200729	0.012	1.262		
ARSJ	rs10034059	115220756			0.012	
KIAA1627	rs298998	119981070	0.015	1.265		NA
SEC24D	rs298991	120003603	0.013	1.288		NA
SEC24D	rs4834700	120021831	0.038	1.216		
SYNPO2	rs1490512	120229244			0.027	0.002

Results for ADSX were corrected using the Nyholt procedure which yielded a gene-wise empirical *P*-value. Single-marker *P*-values smaller than this threshold value were considered to have survived correction. Bold values denote $P < 0.05$.

DISCUSSION

Our systematic screen of 65 genes and ESTs in the18.5 Mb region spanning the 1-LOD linkage support interval on chromosome 4q22–q32 identified two novel candidate genes (*DKK2* and *EGF)*, yielding empirically significant evidence of association in the IASPSAD sample for the DSM-IV symptom count severity index ADSX. *DKK2* emerged as the most strongly associated gene in our sample; however, this is a novel finding therefore we were further interested in assessing possible replication in an independent sample. Novel associations in complex disorders often remain unconfirmed due to difficulties in replicating the initial association (24), particularly as first reports of positive findings can suffer from the so-called 'winner's curse', such that the genetic effect can be larger than the true overall effect (25). As discussed in a thorough review of association studies, replication studies may also interrogate different variants, suffer from population stratification or lack power (26). In conducting the replication in the COGA sample, we had the advantage of a larger sample with cases of European ancestry; observed similarities in allele frequencies suggested that population stratification was unlikely to confound the results. A parallel comparison was also possible because either matching SNPs were available or proxies could be used; comparison of the LD pattern showed similarity in both samples. Our results suggested genotypic and phenotypic replication for *DKK2* for AD severity of problematic alcohol use thereby

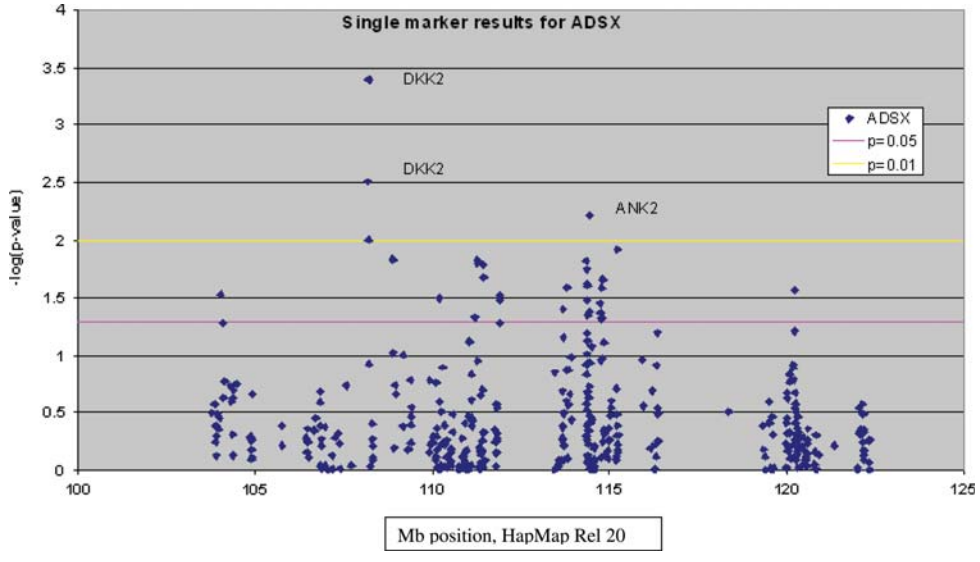


Figure 2. Single-marker results with $-\log P$-values plotted against HapMap base pair location (HapMap Rel 20).

conforming with the definition of genetic replication advocated by Sullivan (27), namely replication with the same SNPs, the same phenotype (*ADSX*) and the same direction of association. Nonetheless, as our original finding did not achieve experiment-wide significance, it is possible that the result was a false positive and confidence should be placed in these results only if they are replicated in independent samples. We note that the distribution of symptoms counts in the two populations tested here does vary; the IASPSAD sample is more severely affected with over 86.7% of probands endorsing six or seven of the seven DSM-IV criteria, whereas in the COGA sample, a similar percentage (86%) endorse four, five, six or seven symptoms (Supplementary Material, Fig. S3). Our data suggest that *DKK2* is associated with variation in symptoms endorsed and that it does not appear to be specific to six or seven symptoms.

DKK2 belongs to a multigene family of Dickkopf proteins (28) involved in the Wnt/β-catenin signaling pathway, a key signaling pathway comprising of a complex network of proteins regulating cell fate, polarity, axon guidance and limb development (29,30). *DKK2* modulates the Wnt/β-catenin pathway; the C-terminal domains of the extracellular *DKK2* bind to low-density lipoprotein receptor-related protein (*LRP6*) at the cell surface, causing activation of the Dishevelled (Dvl) proteins, leading to the downstream inhibition of glycogen synthase kinase 3 (*GSK-3*β) (31). A study investigating the regulation of the transcription factor NF-kappa B1 (NF-κB) reported that phosphorylation by *GSK-3*β leads to negative regulation of basal p65 NF-κB activity (32). Measurements of mRNAs in post-mortem brain samples of chronic alcoholics have revealed the downregulation of NF-κB expression (33). The combination of these findings with the results of our expression data suggests a possible hypothesis of mechanism involving *DKK2*. Our observation of reduced expression with the haplotype associated with higher ADSX suggests that downregulation may relax the control of the downstream molecules in the Wnt signaling pathway. Thus, downregulation of *DKK2* could lead to upregulation of GSK-3β, and accumulation of this key molecule could lead to enhanced downregulation of NF-κB. This hypothesis is consistent with the observed reduction in expression levels in chronic alcoholics, as mentioned above. Recent findings in the COGA sample have shown that variants in *NFKB1* are associated with alcoholism and with early age of onset of AD (21); earlier onset can be a risk factor for a more severe disease (34).

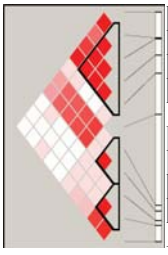
In conclusion, we have identified a functionally plausible candidate, *DKK2*, which may contribute to the pathophysiology of alcoholism. Expression measurements in control post-mortem brain samples suggested that *DKK2* might affect expression unaffected by neuroadaptive changes arising from dependent alcohol consumption. Our association results are strengthened by replication in a large, independent sample, the COGA sample. Although the support for *DKK2* is encouraging, a number of caveats need to be considered. First, we had adopted a gene-centric approach and gene-level correction for multiple testing was applied, however, had we applied study-wide correction, these results would not have survived correction. Secondly, as mentioned above, the finding may be a false positive and despite replication in the COGA sample, it requires further replication in other samples. Thirdly, our study and the subsequent replication were conducted in clinically ascertained samples with a high proportion of severe cases; it is possible that *DKK2* may be more relevant to a particular subgroup of alcoholics. Finally, both samples were of European origin; it remains to be seen whether *DKK2* might have a role in non-European samples. The evidence that *DKK2* has a relationship with other susceptibility genes through a functional pathway provides evidence for a widespread role and opens the door for further examination.

MATERIALS AND METHODS

Subjects

Participants in IASPSAD were recruited in Ireland and Northern Ireland between 1998 and 2002. Complete details of the study design, ascertainment and clinical characteristics

Table 2. *DKK2*: single-marker association results for AD and ADSX



Gene symbol	tSNP	Position (HM Rel 20)	Case–contol	OR	ADSX, uncorrected	ADSX, corrected
DKK2	rs427983	108185605	0.465	1.071	**0.003**	
DKK2	rs419558	108201696	0.297	1.107	**0.0004**	**0.007**
DKK2	rs419764	108201761	0.322	1.097	**0.009**	
DKK2	rs447372	108211600	**0.016**	**1.238**	0.118	
DKK2	rs399087	108222755	0.255	1.119	**0.0004**	
DKK2	rs396196	108247402	0.276	1.102	0.921	
DKK2	rs2156505	108301999	0.164	1.164	0.542	
DKK2	rs2212968	108305412	0.291	1.121	0.396	
DKK2	rs13136418	108311307	0.316	1.099	0.622	
DKK2	rs2850412	108313883	0.807	1.029	0.779	
DKK2	rs419178	108323648	0.796	1.025	0.814	

Assessment in HapMap indicated that there were three haplotype blocks in the gene (shown in the figure and in the table as shaded blocks). The SNPs rs419558 and rs399087 appear to be in close LD with each other in our data. Haplotype analysis was limited to the block (shown with shading) composed of rs427983, rs419558, rs419764 and rs447372.

of this sample are described elsewhere (35). In brief, ascertainment of probands was conducted in community alcoholism treatment facilities and public and private hospitals. Probands were eligible for study inclusion if they met the DSM-IV (American Psychiatric Association, 1994) criteria for current AD and if all four grandparents had been born in Ireland, Northern Ireland, Scotland, Wales or England. After a prospective family was identified through a proband, parents and potentially affected siblings whom the probands provided permission to contact were recruited. Probands, siblings and parents were interviewed by clinically trained research interviewers, most of whom had extensive clinical experience with alcoholism. The assessment included demographic characteristics, lifetime history of AD and other comorbid conditions, alcohol-related traits, personality features and clinical records. The DSM-IV AD diagnosis was assessed in probands and siblings using a modified SSAGA (Semi-Structured Assessment of the Genetics of Alcoholism) interview (version 11) (36) to reduce assessment time. All participants provided informed consent.

A total of 1238 subjects met criteria for DSM-IV AD diagnosis, including 591 probands, 620 affected siblings and 27 additional affected first-degree relatives from 10 complex families. Controls were recruited in Northern Ireland from volunteers donating at the Northern Ireland Blood Transfusion Service ($n = 554$) and in the Republic from the Garda Siochana (the national police force, $n = 38$) and the Forsa Cosanta Aituil (the army reserve, $n = 34$). Controls were screened and their samples excluded if they reported a history of heavy drinking or problem alcohol use; the mode of sample collection did not permit more detailed screening. DNA samples were collected from blood and buccal swabs, and all samples were re-quantified using fluorimetry and tested for genotyping quality prior to use in this study. Samples non-compliant with our stringent quality control (QC) were discarded from the study and replaced with a sample from an affected sibling, where these were available and satisfied our QC criteria. In the present case–control study design, we included 562 independent AD cases that were selected from the IASPSAD families and 569 population-matched controls.

Tag SNP selection and genotyping

The 1-LOD support interval around our maximum linkage peak, defined by the markers D4S1572 and D4S427, represents a physical distance of 17.6 Mb. We extended this to 18.5 Mb to include the complete genomic loci for *NFKB1* and *FLJ23191*, the genes at either end of the interval. The region contained 65 genes and ESTs in the January 2006 build of the human genome (HapMap data Rel20, NCBI B35 assembly, dbSNP b125). We assessed LD-tSNPs in all 65 loci for association with AD and the severity index phenotype ADSX. Tagging SNPs (tSNPs) were selected using TAGGER (23) as implemented in HAPLOVIEW 3.2 (37) using the default criteria of $r^2 = 0.8$ and minor allele frequency (MAF) ≥ 0.2. TAGGER selects a minimum set of tSNPs by performing pairwise and aggressive tagging through evaluation of multi-marker haplotypes; overfitting is avoided by constraining multi-marker predictors to be in strong LD with each other. We limited the sequence included in TAGGER assessment to the gene boundaries defined in the January 2006 build of the human genome (HapMap data Rel 20, NCBI B35 assembly, dbSNP b125). For genes displaying several isoforms, the longest isoform was chosen for tag selection but in order to limit genotyping load and cost, 5′ and 3′ regions of the genes and ESTs were not directly tagged. TAGGER assessment did not identify any tSNPs in *TRAM1L1* (translocation-associated membrane protein 1-like) but we selected one SNP within this gene to ensure full data collection; total list of SNPs for testing was 518 in the 65 genes and ESTs (Supplementary Material, Table S1).

Genotyping and data checking

Multiplex genotyping was conducted on the GenomeLab SNPstream (Beckman Coulter, Fullerton, CA, USA) following manufacturer's protocols, in panels of 12 or 48 SNPs matched for their extension type. SNP sequences were screened for repeats and homology with other genomic sequences prior to using the proprietary Beckman Coulter primer design program, Autoprimer, for constructing the multiplex panels. In instances where SNP sequences were repeat-rich or

Table 3. Haplotype analysis of the block containing the tSNPs rs427983, rs419558, rs419764 and rs447372

Haplotype	Freq	Alt (B)	χ^2	*P*-value
1111	0.587	−0.100	2.754	0.097
2222	0.249	0.239	12.402	0.0004
1122	0.081	−0.169	2.404	0.121
1112	0.038	−0.095	0.396	0.529
2112	0.025	−0.250	1.756	0.185
2111	0.020	0.063	0.091	0.763

The four-marker haplotype demonstrates that the minor alleles of all four markers are associated with ADSX and that the association is with increased symptom counts.

matching extension types were necessary for successful multiplex paneling, proxy SNPs with matching r^2 and MAF criteria were substituted. Tag SNPs failing in the first round of genotyping were re-paneled and tSNPs failing twice on the SNPstream platform and those which could not be paneled were genotyped as monoplex reactions using Taqman Assays-on-Demand (Applied Biosystems, Foster City, CA, USA). A total of 510 tSNPs were suitable for the first round of genotyping using the high-throughput multiplex genotyping technology, followed by another round of multiplex genotyping and 114 Taqman assays. To ensure uniformity and accuracy, all reaction steps were performed using the Eppendorf 5075 automated liquid handling platform.

Stringent evaluation of initial data is important to avoid artifactual effects of genotyping errors; therefore, all genotypes were independently assessed by two raters. Ambiguous calls were discussed and in cases of non-resolution, genotypes were dropped from the analyses. Individual DNA samples with 30% or more missing data across the entire study were also excluded. Individual SNPs were excluded if they failed SNPstream QC parameters or showed deviation from Hardy–Weinberg equilibrium ($P = 0.001$) in control samples. After data cleaning, a total of 397 (76%) tSNPs were analyzed for tests of association.

Quantitative real-time expression measurements

DKK2 expression was analyzed on the StepOne Plus instrument (ABI, Foster City, CA, USA) using quantitative real-time PCR with Taqman expression assay (ABI). Post-mortem brain samples from the Stanley Medical Research Institute (SMRI) collection were used for the gene expression studies. The collection consists of DNA and mRNA samples extracted from dorsolateral prefrontal cortex (Brodmann's area 46) from 35 individuals each with schizophrenia and bipolar affective disorder and 35 controls (38). Each sample was assayed in triplicate, and mean values from the triplicates were used for all analyses. From a pool of three reference genes, *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase), *TBP* (TATA box binding protein) and *HPRT1* (hypoxanthine phosphoribosyltransferase 1), *TBP* was chosen as the endogenous gene for normalizing expression for *DKK2* based on its low coefficient of variation (calculated from the triplicates of each sample) and matching expression range of *DKK2*. Differences in PCR efficiency were measured at five concentrations using log serial dilutions of total RNA for *DKK2* and *TBP*. These were used to generate the calibration curve and the regression coefficient (r^2). The PCR efficiency for *TBP* was 97% and for *DKK2* was 92%. We recognize that the $2^{-\Delta\Delta}C_T$ algorithm, widely used to compare differences in gene expression, is valid only if PCR efficiency in the reference gene and target gene are equal (39). However, despite expression efficiency difference between the two genes, we were able to normalize the data using the $2^{-\Delta\Delta}C_T$ algorithm by using a PCR efficiency correction (40,41). As we were interested in assessing haplotype-specific functional effect in *DKK2*, comparison of expression differences was stratified on the basis of association results. The SMRI samples were genotyped using Taqman SNP genotyping assays (ABI); genotyping was limited to those SNPs yielding significant haplotype association with *DKK2* (rs427983, rs419558, rs419764 and rs447372).

Statistical analyses

Association in the IASPSAD sample. In the first instance, we were interested in examining the binary diagnostic phenotype of AD for which we performed case–control tests for single-marker and haplotype association in HAPLOVIEW v3.2 (37). Calculation of the OR between cases and controls was conducted to index the effect size. Empirical significance of the case–control data was assessed by implementing 5000 permutations in HAPLOVIEW v3.2. However, since the linkage evidence is maximized using the quantitative phenotype of the number of DSM-IV AD symptoms (ADSX), we were additionally interested in testing for association with ADSX. Tests of association were performed in PLINK (http://pngu.mgh.harvard.edu/purcell/plink) (42) using case-only regression procedures with age and sex as covariates. We first assessed how many SNPs would be expected by chance at false discovery rate of 5% and in this case, the number expected by chance would be 20; the number of SNPs with $P < 0.05$ in the experimental data is 32. Next, we applied the Nyholt (43) procedure for implementing gene-wise correction for multiple testing for genes yielding two or more markers with $P < 0.05$. Although the SNPs tested in our study are all independently tagging in the HapMap population, it is possible that background LD may exist between the SNPs in our data. The Nyholt method takes into account LD information between pairs of genotyped SNPs to compute the number of independent SNPs and calculates the significance threshold value required to keep the type 1 error rate at 5%. The output produces a *P*-value threshold for the gene-wise correction and SNPs yielding *P*-values less than or equal to this threshold value are considered significant.

For the purpose of conducting haplotype-specific expression measurements in *DKK2*, haplotype association tests were conducted in the IASPSAD sample. LD blocks were first identified using control samples and three blocks were identified (Table 2). For the analysis, we used the block containing the tSNPs yielding association in the single-marker tests. In addition, we included rs427983 in the analysis because of evidence of LD with SNPs in the block. Two SNPs (rs419558 and rs399087) appeared to give same haplotype information in the

Table 4. Replication of association results for *DKK2* in the COGA sample

Gene	SNP id	bp, HapMap Rel20	*P*-values, ADSX IASPSAD	Risk allele	Risk allele freq.	SNP id	bp, HapMap Rel20	*P*-values, ADSX COGA	Risk allele	Risk allele freq.
DKK2	rs427983	108047450	0.003	C	0.29	rs427983	108047450	0.021	C	0.30
DKK2	rs419558	108063541	0.0004	T	0.25	rs419558	108063541	0.021	T	0.25
DKK2	rs419764	108063606	0.009	T	0.33	rs419764	108063606	0.004	T	0.32
DKK2	rs447372	108073445	0.118	A	0.39	*rs433201*	108070724	0.001	A	0.39
DKK2	rs399087	108084500	0.0004	C	0.25	*rs411143*	108093282	0.038	G	0.25
DKK2	rs396196	108109247	0.921	C	0.46	NA		NA		
DKK2	rs2156505	108163844	0.542	G	0.80	*rs4956275*	108165610	0.635	T	0.77
DKK2	rs2212968	108167257	0.396	T	0.80	rs2212968	108167257	0.648	T	0.77
DKK2	rs13136418	108173152	0.622	G	0.50	*rs10032810*	108170686	0.141	T	0.50
DKK2	rs2850412	108175728	0.779	C	0.75	rs2850412	108175728	0.029	C	0.76
DKK2	rs419178	108185493	0.814	T	0.73	rs419178	108185493	0.152	T	0.74
EGF	rs2881559	111060495	0.045	G	0.42	*rs3796944*	111093375	0.792	G	0.42
EGF	rs6850557	111130464	0.015	A	0.36	*rs11568994*	111116984	0.767	A	0.33
EGF	rs2237052	111130474	0.016	G	0.41	*rs2074390*	111129465	0.762	T	0.39
EGF	rs2298999	111131356	0.112	C	0.45	rs2298999	111131356	0.760	C	0.42

The assessment was limited to the tSNPs yielding results in the IASPSAD sample and to the parallel phenotype of ADSX. Light shading denotes $P < 0.05$ and darker shading implies $P < 0.01$. Alleles associated with increased symptom counts represent risk allele. Allele frequencies for the risk alleles are presented

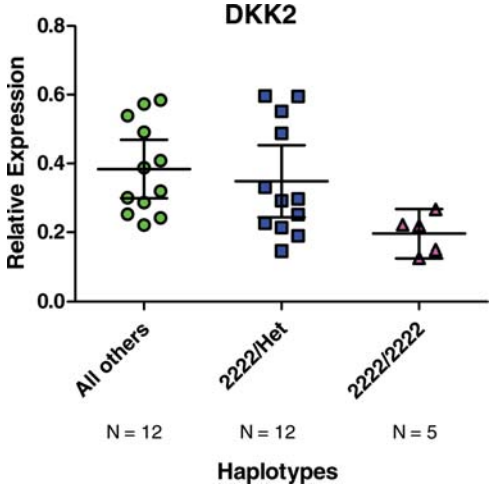


Figure 3. Haplotype-specific expression analysis of *DKK2* shows significantly reduced ($P = 0.028$) RNA levels in five individuals homozygous for the haplotype associated with higher ADSX compared with pairing of homozygotes/heterozygous and all other haplotype combinations.

IASPSAD sample, thus one (rs399087) was dropped from the haplotype analysis.

Replication in an independent sample. Because of the novel findings in our study, we were interested in testing significant results for replication in an independent sample. For this, we used data from the genome-wide association study (GWAS) conducted in the COGA sample, which is a large multi-site project for which families were collected across six centers across the USA: Indiana University, State University of New York Health Science Center, University of Connecticut, University of Iowa, University of California/San Diego and Washington University at St Louis. Individuals had been diagnosed for AD using DSM-IV criteria, among others (APA, 1994) and had also been administered the SSAGA interview (36). For the GWAS, a case–control sample of 1205 cases and 700 control individuals was genotyped on the Illumina Beadstation platform by the Center for Inherited Disease Research (CIDR). From this large admixed sample, we analyzed 847 cases of European descent. Analyses in this sample were limited to the phenotype of ADSX among the cases (parallel to the phenotype in the IASPSAD sample). In the COGA GWAS study, a greater number of SNPs had been tested within *DKK2* ($n = 31$) and *EGF* gene ($n = 39$); the COGA data set has independent and non-independent SNPs as well as rare SNPs. For parallel comparison, tSNPs were identified in the COGA sample using the same release of HapMap database (HM Rel20) used in the Irish sample and the assessment was limited to those SNPs yielding genotypes in the IASPSAD sample (*DKK2*: $n = 11$; *EGF*: $n = 4$). In some instances, the same SNP had been genotyped in both samples; where this was not the case, a proxy SNP was identified in the COGA data set and used for the comparison. The criteria used for making the tagging selection in the COGA data set was also the same as that used in the IASPSAD sample; MAF = 0.2 and $r^2 = 0.8$.

Statistical analyses of expression data. The potential confounding effects of diagnosis, age, gender, pH, post-mortem interval, refrigerator interval, smoking and drug abuse in the analysis of *DKK2* expression were controlled by the analysis of covariance. As the distribution of expression data was not normal, *DKK2* expression values were normalized by raising them by a power of 0.25 to achieve normal distribution. To verify normality of the transformed data, we used the Anderson–Darling EBF test that has been shown to be a powerful statistic for detecting departures from normality even with a relatively small sample size ($n \leq 100$). Samples performing poorly either in the qPCR reactions or the genotyping assays were omitted from further analysis; the cutoff used here was ± 2SD from the group mean. Thus, the number of samples available for the final analysis was reduced from 35 to 29 in the controls, and in the full sample, the total was reduced from 105 to 84.

The haplotype significantly associated with higher ADSX was made up the tSNPs rs427983, rs419558, rs419764 and rs447372; these were genotyped in the SMRI sample and haplotypes were reconstructed using PHASE v2.1 (44,45). PHASE implements a Bayesian statistical method using an iterative approach and in this analysis, we used 100 iterations. The means of the fold expression differences were compared in the individuals carrying two copies of the minor alleles, the heterozygous individuals and individuals carrying all other combinations. For the assessment of haplotype-specific expression differences, we were interested in testing first the controls only. By restricting the analysis thus, we were able to assess expression levels without the confounding effects of changes in the brain due to alcohol. However, due to the small sample size ($n = 29$) available for this analysis, no conclusions could be made about the normality of the data hence the non-parametric Kruskal–Wallis H test was used. Additionally, as the sample size of the controls was small, we extended the assessment to the entire SMRI sample ($n = 84$), for which the possible confounding effect of disease (schizophrenia and bipolar) had been overcome through using disease status as a covariate.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We are grateful to the study participants and their families for their time and effort. F.A. O'Neill, the Northern Ireland Blood Transfusion Service of the British National Health Service and the Irish Garda provided assistance in obtaining control blood samples. Figures 1A and B are reproduced with kind permission from Dr Carol Prescott.

Conflict of Interest statemet. None declared.

FUNDING

This work was funded by US National Institutes of Health grant R01-AA-11408 (K.S.K., PI) with support from the Irish Health Research Board.

REFERENCES

1. Schuckit, M.A. (2000) Genetics of the risk for alcoholism. *Am. J. Addict.*, **9**, 103–112.
2. Kendler, K.S., Neale, M.C., Heath, A.C., Kessler, R.C. and Eaves, L.J. (1994) A twin-family study of alcoholism in women. *Am. J. Psychiatry*, **151**, 707–715.
3. Edenberg, H.J. and Foroud, T. (2006) The genetics of alcoholism: identifying specific genes through family studies. *Addict. Biol.*, **11**, 386–396.
4. Mayfield, R.D., Harris, R.A. and Schuckit, M.A. (2008) Genetic factors influencing alcohol dependence. *Br. J. Pharmacol.*, **154**, 275–287.
5. Oroszi, G. and Goldman, D. (2004) Alcoholism: genes and mechanisms. *Pharmacogenomics*, **5**, 1037–1048.
6. Saccone, N.L., Kwon, J.M., Corbett, J., Goate, A., Rochberg, N., Edenberg, H.J., Foroud, T., Li, T.K., Begleiter, H., Reich, T. *et al.* (2000) A genome screen of maximum number of drinks as an alcoholism phenotype. *Am. J. Med. Genet.*, **96**, 632–637.
7. Corbett, J., Saccone, N.L., Foroud, T., Goate, A., Edenberg, H., Nurnberger, J., Porjesz, B., Begleiter, H., Reich, T. and Rice, J.P. (2005) A sex-adjusted and age-adjusted genome screen for nested alcohol dependence diagnoses. *Psychiatr. Genet.*, **15**, 25–30.
8. Prescott, C.A., Sullivan, P.F., Kuo, P.H., Webb, B.T., Vittum, J., Patterson, D.G., Thiselton, D.L., Myers, J.M., Devitt, M., Halberstadt, L.J. *et al.* (2006) Genomewide linkage study in the Irish affected sib pair study of alcohol dependence: evidence for a susceptibility region for symptoms of alcohol dependence on chromosome 4. *Mol. Psychiatry*, **11**, 603–611.
9. Crabb, D.W., Matsumoto, M., Chang, D. and You, M. (2004) Overview of the role of alcohol dehydrogenase and aldehyde dehydrogenase and their variants in the genesis of alcohol-related pathology. *Proc. Nutr. Soc.*, **63**, 49–63.
10. Birley, A.J., James, M.R., Dickson, P.A., Montgomery, G.W., Heath, A.C., Martin, N.G. and Whitfield, J.B. (2009) ADH single nucleotide polymorphism associations with alcohol metabolism in vivo. *Hum. Mol. Genet.*, **18**, 1533–1542.
11. Chen, Y.C., Lu, R.B., Peng, G.S., Wang, M.F., Wang, H.K., Ko, H.C., Chang, Y.C., Lu, J.J., Li, T.K. and Yin, S.J. (1999) Alcohol metabolism and cardiovascular response in an alcoholic patient homozygous for the ALDH2*2 variant gene allele. *Alcohol. Clin. Exp. Res.*, **23**, 1853–1860.
12. Edenberg, H.J. (2007) The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res. Health*, **30**, 5–13.
13. Edenberg, H.J., Xuei, X., Chen, H.J., Tian, H., Wetherill, L.F., Dick, D.M., Almasy, L., Bierut, L., Bucholz, K.K., Goate, A. *et al.* (2006) Association of alcohol dehydrogenase genes with alcohol dependence: a comprehensive analysis. *Hum. Mol. Genet.*, **15**, 1539–1549.
14. Thomasson, H.R., Edenberg, H.J., Crabb, D.W., Mai, X.L., Jerome, R.E., Li, T.K., Wang, S.P., Lin, Y.T., Lu, R.B. and Yin, S.J. (1991) Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. *Am. J. Hum. Genet.*, **48**, 677–681.
15. Whitfield, J.B. (2002) Alcohol dehydrogenase and alcohol dependence: variation in genotype-associated risk between populations. *Am. J. Hum. Genet.*, **71**, 1247–1250. author reply 1250–1251.
16. Kuo, P.H., Kalsi, G., Prescott, C.A., Hodgkinson, C.A., Goldman, D., van den Oord, E.J., Alexander, J., Jiang, C., Sullivan, P.F., Patterson, D.G. *et al.* (2008) Association of ADH and ALDH genes with alcohol dependence in the Irish Affected Sib Pair Study of alcohol dependence (IASPSAD) sample. *Alcohol. Clin. Exp. Res.*, **32**, 785–795.
17. Feighner, J.P., Robins, E., Guze, S.B., Woodruff, R.A. Jr, Winokur, G. and Munoz, R. (1972) Diagnostic criteria for use in psychiatric research. *Arch. Gen. Psychiatry*, **26**, 57–63.
18. Macgregor, S., Lind, P.A., Bucholz, K.K., Hansell, N.K., Madden, P.A., Richter, M.M., Montgomery, G.W., Martin, N.G., Heath, A.C. and Whitfield, J.B. (2009) Associations of *ADH* and *ALDH2* gene variation with self report alcohol reactions, consumption and dependence: an integrated analysis. *Hum. Mol. Genet.*, **18**, 580–593.
19. Foroud, T., Wetherill, L.F., Liang, T., Dick, D.M., Hesselbrock, V., Kramer, J., Nurnberger, J., Schuckit, M., Carr, L., Porjesz, B. *et al.* (2007) Association of alcohol craving with alpha-synuclein (SNCA). *Alcohol. Clin. Exp. Res.*, **31**, 537–545.
20. Bonsch, D., Reulbach, U., Bayerlein, K., Hillemacher, T., Kornhuber, J. and Bleich, S. (2004) Elevated alpha synuclein mRNA levels are associated with craving in patients with alcoholism. *Biol. Psychiatry*, **56**, 984–986.
21. Edenberg, H.J., Xuei, X., Wetherill, L.F., Bierut, L., Bucholz, K., Dick, D.M., Hesselbrock, V., Kuperman, S., Porjesz, B., Schuckit, M.A. *et al.* (2008) Association of NFKB1, which encodes a subunit of the transcription factor NF-kappaB, with alcohol dependence. *Hum. Mol. Genet.*, **17**, 963–970.
22. Foroud, T., Wetherill, L.F., Kramer, J., Tischfield, J.A., Nurnberger, J.I. Jr, Schuckit, M.A., Xuei, X. and Edenberg, H.J. (2008) The tachykinin receptor 3 is associated with alcohol and cocaine dependence. *Alcohol. Clin. Exp. Res.*, **32**, 1023–1030.
23. de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J. and Altshuler, D. (2005) Efficiency and power in genetic association studies. *Nat. Genet.*, **37**, 1217–1223.
24. Spector, T.D., Ahmadi, K.R. and Valdes, A.M. (2006) When is a replication not a replication? Or how to spot a good genetic association study. *Arthritis Rheum.*, **54**, 1051–1054.
25. Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. and Hirschhorn, J.N. (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.*, **33**, 177–182.

26. Hirschhorn, J.N., Lohmueller, K., Byrne, E. and Hirschhorn, K. (2002) A comprehensive review of genetic association studies. *Genet. Med.*, **4**, 45–61.
27. Sullivan, P.F. (2007) Spurious genetic associations. *Biol. Psychiatry*, **61**, 1121–1126.
28. Krupnik, V.E., Sharp, J.D., Jiang, C., Robison, K., Chickering, T.W., Amaravadi, L., Brown, D.E., Guyot, D., Mays, G., Leiby, K. *et al.* (1999) Functional and structural diversity of the human Dickkopf gene family. *Gene*, **238**, 301–313.
29. Cadigan, K.M. and Nusse, R. (1997) Wnt signaling: a common theme in animal development. *Genes Dev.*, **11**, 3286–3305.
30. Wodarz, A. and Nusse, R. (1998) Mechanisms of Wnt signaling in development. *Annu. Rev. Cell. Dev. Biol.*, **14**, 59–88.
31. Brott, B.K. and Sokol, S.Y. (2002) Regulation of Wnt/LRP signaling by distinct domains of Dickkopf proteins. *Mol. Cell. Biol.*, **22**, 6100–6110.
32. Buss, H., Dorrie, A., Schmitz, M.L., Frank, R., Livingstone, M., Resch, K. and Kracht, M. (2004) Phosphorylation of serine 468 by GSK-3beta negatively regulates basal p65 NF-kappaB activity. *J. Biol. Chem.*, **279**, 49571–49574.
33. Okvist, A., Johansson, S., Kuzmin, A., Bazov, I., Merino-Martinez, R., Ponomarev, I., Mayfield, R.D., Harris, R.A., Sheedy, D., Garrick, T. *et al.* (2007) Neuroadaptations in human chronic alcoholics: dysregulation of the NF-kappaB system. *PLoS One*, **2**, e930.
34. Dick, D.M., Agrawal, A., Wang, J.C., Hinrichs, A., Bertelsen, S., Bucholz, K.K., Schuckit, M., Kramer, J., Nurnberger, J. Jr, Tischfield, J. *et al.* (2007) Alcohol dependence with comorbid drug dependence: genetic and phenotypic associations suggest a more severe form of the disorder with stronger genetic contribution to risk. *Addiction*, **102**, 1131–1139.
35. Prescott, C.A., Sullivan, P.F., Myers, J.M., Patterson, D.G., Devitt, M., Halberstadt, L.J., Walsh, D. and Kendler, K.S. (2005) The Irish Affected Sib Pair Study of Alcohol Dependence: study methodology and validation of diagnosis by interview and family history. *Alcohol. Clin. Exp. Res.*, **29**, 417–429.
36. Bucholz, K.K., Cadoret, R., Cloninger, C.R., Dinwiddie, S.H., Hesselbrock, V.M., Nurnberger, J.I. Jr, Reich, T., Schmidt, I. and Schuckit, M.A. (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**, 149–158.
37. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
38. Torrey, E.F., Webster, M., Knable, M., Johnston, N. and Yolken, R.H. (2000) The stanley foundation brain collection and neuropathology consortium. *Schizophr. Res.*, **44**, 151–155.
39. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
40. Meijerink, J., Mandigers, C., van de Locht, L., Tonnissen, E., Goodsaid, F. and Raemaekers, J. (2001) A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time PCR. *J. Mol. Diagn.*, **3**, 55–61.
41. Tichopad, A., Didier, A. and Pfaffl, M.W. (2004) Inhibition of real-time RT–PCR quantification due to tissue-specific contaminants. *Mol. Cell. Probes*, **18**, 45–50.
42. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
43. Nyholt, D.R. (2004) A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.*, **74**, 765–769.
44. Stephens, M. and Donnelly, P. (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.*, **73**, 1162–1169.
45. Stephens, M., Smith, N.J. and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.*, **68**, 978–989.