

# Linkage scan of nicotine dependence in the University of California, San Francisco (UCSF) Family Alcoholism Study

I. R. Gizer<sup>1,2\*</sup>, C. L. Ehlers<sup>3</sup>, C. Vieten<sup>4</sup>, K. L. Seaton-Smith<sup>1</sup>, H. S. Feiler<sup>5</sup>, J. V. Lee<sup>6</sup>, S. K. Segall<sup>1</sup>, D. A. Gilder<sup>3</sup> and K. C. Wilhelmson<sup>1,2,7</sup>

<sup>1</sup> Department of Genetics and the <sup>2</sup> Bowles Center for Alcohol Studies, University of North Carolina, Chapel Hill, NC, USA

<sup>3</sup> Department of Molecular and Integrative Neurosciences, The Scripps Research Institute, La Jolla, CA, USA

<sup>4</sup> California Pacific Medical Center, San Francisco, CA, USA

<sup>5</sup> Department of Laboratory Medicine, University of California, San Francisco, CA, USA

<sup>6</sup> School of Molecular and Cell Biology, University of Illinois, Urbana–Champaign, IL, USA

<sup>7</sup> Department of Neurology, Carolina Genome Center, University of North Carolina, Chapel Hill, NC, USA

**Background.** Nicotine dependence has been shown to represent a heritable condition, and several research groups have performed linkage analysis to identify genomic regions influencing this disorder though only a limited number of the findings have been replicated.

**Method.** In the present study, a genome-wide linkage scan for nicotine dependence was conducted in a community sample of 950 probands and 1204 relatives recruited through the University of California, San Francisco (UCSF) Family Alcoholism Study. A modified version of the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) with additional questions that probe nicotine use was used to derive DSM-IV nicotine dependence diagnoses.

**Results.** A locus on chromosome 2q31.1 at 184 centiMorgans nearest to marker D2S2188 yielded a logarithm (base 10) of odds (LOD) score of 3.54 (point-wise empirical  $p=0.000012$ ). Additional peaks of interest were identified on chromosomes 2q13, 4p15.33-31, 11q25 and 12p11.23-21. Follow-up analyses were conducted examining the contributions of individual nicotine dependence symptoms to the chromosome 2q31.1 linkage peak as well as examining the relationship of this chromosomal region to alcohol dependence.

**Conclusions.** The present report suggests that chromosome 2q31.1 confers risk to the development of nicotine dependence and that this region influences a broad range of nicotine dependence symptoms rather than a specific facet of the disorder. Further, the results show that this region is not linked to alcohol dependence in this population, and thus may influence nicotine dependence specifically.

Received 10 April 2009; Revised 29 April 2010; Accepted 9 May 2010; First published online 1 July 2010

**Key words:** Genetics, genome scan, linkage, nicotine dependence, smoking.

## Introduction

Tobacco-related illnesses have been estimated to claim 435 000 lives annually, thus constituting a serious and costly public health issue (Mokdad *et al.* 2004). Twin studies suggest a substantial genetic component to the development of nicotine dependence (e.g. Carmelli *et al.* 1990; Kendler *et al.* 1999; True *et al.* 1999), and a recent meta-analysis reported heritability estimates for smoking initiation ranging from 37 to 55% and

persistent smoking ranging from 46 to 59% (Li *et al.* 2003).

Given such evidence, several genome-wide linkage analyses have been conducted for a variety of tobacco-related phenotypes [e.g. DSM-IV nicotine dependence (APA, 1994), smoking quantity, and the Fagerström Nicotine Dependence Scale (Heatherton *et al.* 1991)] to identify susceptibility loci related to nicotine dependence (Li, 2008). The earliest studies were conducted using an extended family design in the Collaborative Studies on the Genetics of Alcoholism (COGA) sample (Duggirala *et al.* 1999) and a sib-pair design in samples collected in Christchurch, New Zealand and Richmond, VA, USA (Straub *et al.* 1999). Following these initial studies, more than 20 genome-wide linkage scans of tobacco-related phenotypes have been

\* Address for correspondence: I. R. Gizer, Ph.D., University of North Carolina at Chapel Hill, 120 Mason Farm Road, Room 5015 Genetic Medicine Building CB 7264, Chapel Hill, NC 27599-7264, USA.

(Email: igizer@unc.edu)

published, and several susceptibility regions have been identified across studies, including regions on chromosomes 9, 10, 11 and 17 (Li, 2008).

Association studies have begun to identify specific genes that underlie these linkage regions. For example, three independent studies have reported evidence of linkage between nicotine-related phenotypes and a region of chromosome 17 [20–65 centiMorgans (cM)] (Duggirala *et al.* 1999; Wang *et al.* 2005; Li *et al.* 2008). A recent candidate gene study suggested polymorphisms in the nicotinic acetylcholine  $\beta_1$  receptor subtype gene (*CHRN1*) may be responsible for this linkage signal (Lou *et al.* 2006). Additionally, studies conducted using the COGA sample have reported evidence of linkage between nicotine-related phenotypes and chromosome 2 at 85 cM (Bergen *et al.* 1999; Bierut *et al.* 2004). This region includes the neurexin 1 gene (*NRXN1*), which contains polymorphisms associated with nicotine dependence in a recent genome-wide association study (GWAS) (Bierut *et al.* 2007) as well as recent candidate gene studies (Nussbaum *et al.* 2008; Sherva *et al.* 2008). These studies demonstrate the progress being made in understanding the genetic contributions to tobacco-related phenotypes and how linkage analysis has aided in the identification of susceptibility loci for such traits.

In addition to identifying susceptibility loci for nicotine dependence specifically, several studies have tried to identify loci that contribute to substance dependence more generally by searching for genomic regions that confer risk to both nicotine and alcohol dependence. Such studies are justified by twin studies indicating a shared genetic etiology between these two disorders (Swan *et al.* 1997; True *et al.* 1999), and several have reported positive findings. For example, data from the COGA sample suggest that the susceptibility locus on chromosome 2 containing *NRXN1* contributed to both nicotine and alcohol dependence (Bergen *et al.* 1999). Further, a study of the Mission Indian population found that the chromosome 4 region containing the alcohol dehydrogenase gene cluster contributed to increased risk for both disorders (Ehlers & Wilhelmsen, 2006). In addition, a Finnish twin sample was used to identify loci on chromosomes 7 and 11 (Loukola *et al.* 2008) and sibling pairs collected in Ireland were used to identify loci on chromosomes 7 and 18 (Sullivan *et al.* 2008) conferring risk for both disorders. Such studies provide important insights into how different chromosomal regions confer risk to substance dependence whether it is towards a specific substance or towards a more general tendency toward addictive behavior.

The current study conducted a genome-wide linkage scan for nicotine dependence in the University of California, San Francisco (UCSF) Family Alcoholism

Study to support and extend previous findings. Linkage peaks were followed-up by analysing each of the 14 nicotine dependence symptoms assessed by the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA; Bucholz *et al.* 1994) to identify those symptoms responsible for the reported linkage signals. A further aim was to determine whether the linked genomic regions contributed to nicotine dependence specifically, or whether they might confer increased risk to addiction more generally by showing evidence of linkage to both alcohol and nicotine dependence. Thus, supplementary genome-wide linkage scans of nicotine dependence were conducted utilizing alcohol dependence diagnoses alternatively as a covariate and as an additional predictor in a bivariate analysis.

## Method

### Participants

The present study utilized participants from the UCSF Family Alcoholism Study (Seaton *et al.* 2004; Vieten *et al.* 2004), which consists of 2524 participants from 890 families (average size = 2.83 members). The UCSF study was a nationwide study on the genetics of alcoholism and other substance dependence designed to recruit a large number of small family pedigrees enriched for alcohol dependence. Proband was sampled from the community and invited to participate if they met screening criteria for alcohol dependence at some point in their lifetime and had at least one sibling or both parents available to participate. Proband was excluded if they reported serious drug addictions (defined as use of stimulants, cocaine, or opiates daily for >3 months or weekly for >6 months), any history of intravenous substance use, a current or past diagnosis of schizophrenia, bipolar disorder, or other psychiatric illness involving psychotic symptoms (those with depressive and anxiety disorders were accepted), a life-threatening illness, or an inability to speak and read English. Relatives of qualifying probands were invited by mail to participate.

The UCSF Family Alcoholism Study sample consisted of 1548 women and 976 men with a mean age of 48.5 (s.d. = 13.4) years. The mean educational level of the sample was 14.3 (s.d. = 2.9) years, and the mean annual income was US\$ 54672 (s.d. = 53421) (median, US\$ 45000). The racial distribution was 92% Caucasian, 3% each African-American and Hispanic, and 1% each Native American and other. No attempt was made to exclude or over-sample minorities. A total of 365 participants (15%) were diagnosed with nicotine dependence only, 464 (18%) were diagnosed

with alcohol dependence only, and 880 (35%) were diagnosed with both disorders.

An unselected general population sample of 147 individuals was recruited to assess phenotype base rates. Letters were sent to residents of the same geographical areas as the family samples, requesting participation in a study on 'health behaviors and characteristics' to avoid a sample biased toward participation in a study on alcoholism. No inclusion or exclusion criteria were applied aside from the ability to respond to the telephone interview and complete the questionnaires. Within this population, 36 participants (24%) were diagnosed only with nicotine dependence, 14 (10%) were diagnosed only with alcohol dependence, and 11 (7%) were diagnosed with both disorders. The recruitment details of all participants have been previously published (Seaton *et al.* 2004; Vieten *et al.* 2004).

All participants were administered a modified version of the SSAGA (Bucholz *et al.* 1994), an interview developed by COGA, which was used to diagnose DSM-IV alcohol and other substance dependence and abuse and to collect demographic, medical, psychiatric, alcohol, nicotine, and other drug-use history. This modified version of the SSAGA included questions assessing each of the DSM-IV substance dependence symptoms as applied to tobacco use as well as a question assessing whether these symptoms occurred within a 12-month period to allow for the assignment of DSM-IV diagnoses of nicotine dependence. Participants that did not meet criteria for DSM-IV nicotine dependence included those that had never smoked as well as those that had smoked but did not meet full diagnostic criteria. Only sections of the SSAGA assessing DSM-IV alcohol and substance misuse diagnoses were administered due to time constraints.

#### *Deoxyribonucleic acid (DNA) collection and genotyping*

The DNA extraction procedure and genotyping protocol have been previously described (Wilhelmsen *et al.* 2003). Briefly, DNA was isolated from whole blood using the Puregene kit (Gentra Systems, USA), and genotypes for a panel of microsatellite polymorphisms were generated using fluorescently labeled polymerase chain reaction primers (HD5, version 2.0; Applied Biosystems, USA). The HD5 panel set consisted of 811 markers with an average marker-to-marker distance of 4.6 cM (maximum, 14 cM) and an average heterozygosity of greater than 77%. The sizes of marker amplicons were determined (blinded to pedigree structure and subject characteristics) from the electropherogram using the Genotyper software package (Applied Biosystems, USA). All electropherograms

were visually inspected and exported from Genotyper in base pair sizes relative to the standard measured to one hundredth of a base pair. Allele frequencies observed in the founders were used for all analyses. The sex-averaged marker map order obtained from the manufacturer was used and verified with the family data from the current sample.

Prior to analysis, a number of quality-control steps were undertaken to ensure the accuracy of the genotype calls. First, the pedigree relationship statistical test (McPeck & Sun, 2000) software was used to identify sample and pedigree structure errors based on the genotypes for autosomal markers. For those individuals in which a probable error was detected, DNA was re-isolated from a stored frozen blood specimen and the genotyping was repeated. If the error could not be resolved, the problematic genotype was subsequently treated as missing. A total of 15 families were identified with pedigree structure errors. Of these, five were resolved following re-genotyping. Second, Mendelian errors were identified using the program PEDCHECK (O'Connell & Weeks, 1998). When isolated Mendelian errors were observed, the genotypes for the entire family were excluded for the marker yielding the error. Markers exhibiting a high rate of Mendelian errors across families were excluded from subsequent analysis. PEDCHECK identified 3104 Mendelian errors resulting in 7714 lost genotypes and the exclusion of one marker. Third, the error-checking algorithm implemented in Merlin (Abecasis *et al.* 2002) was used to assess the probability that each genotype was correctly called. A total of 1867 genotypes with probabilities of less than 0.025 of being correct were removed from further analysis. Following these quality-control procedures, a total of 1 269 708 genotypes were accepted for analysis with a success rate of 99.6%.

#### *Analysis*

Both genotype and phenotype data were available for 1647 individuals, and phenotype but not genotype data were available for 875 individuals. Participants not genotyped included relatives of genotyped probands that did not provide DNA samples and probands and their relatives not belonging to a genetically informative pedigree. Notably, rates of alcohol and nicotine dependence remained unchanged when participants not genotyped were removed from the sample (data not shown). Seven hundred and thirteen families that contained sibling, half-sibling, avuncular or cousin pairs were considered genetically informative for linkage analysis. These families ranged in size from three to 20 subjects [average 4.63 (S.D. = 2.13)]. The data include: 1085 sibling, 40 half-sibling,

17 grandparent–grandchild, 238 avuncular and 32 cousin relative pairs. An additional 177 families contained only a single individual with phenotype data. These individuals were included within some analyses to the extent that they contribute information about trait means and variance and the impact of covariates.

Initial analyses were conducted to determine the potential of the DSM-IV nicotine dependence diagnosis and each of its 14 dichotomously scored symptoms as assessed by the SSAGA to exhibit evidence of linkage. Using a variance components approach, SOLAR (Almasy & Blangero, 1998) estimates  $h^2$  by partitioning the trait relative-pair covariance into additive genetic and environmental contributions while correcting for participants' age and sex. The probability that  $h^2$  was greater than zero was determined using a Student's  $t$  test, which was used to evaluate the potential of the nicotine dependence diagnosis and its constituent symptoms to detect linkage.

The variance components approach to linkage incorporates information from affected and unaffected participants by estimating linkage across the entire pedigree (Almasy & Blangero, 1998), and has been shown to possess comparable statistical power compared with relative-pair-based methods [i.e. Kong & Cox (1997) statistic] for both quantitative (Amos *et al.* 1997) and dichotomous phenotypes (Duggirala *et al.* 1997). Thus, the variance components approach implemented in SOLAR was used to calculate multipoint logarithm (base 10) of odds (LOD) scores across the genome at 1-cM intervals for DSM-IV nicotine dependence. Peaks that achieved a LOD score  $>1.0$  were reported as regions of interest as suggested by Lander & Kruglyak (1995).

Dichotomous phenotypes were modeled as latent normally distributed variables with a threshold above which an individual was considered 'affected'. It is notable that when applied to dichotomous phenotypes, variance components linkage analysis can yield inflated LOD scores if distributional assumptions are violated (Duggirala *et al.* 1997). To protect against this potential bias, allele-sharing probabilities for a simulated locus under the null hypothesis of no linkage were generated across 250 000 trials (Duggirala *et al.* 2001), an approach shown to yield appropriate type I error rates (Jung *et al.* 2006). The distribution of simulated LOD scores was used to calculate point-wise estimates of significance according to the formula  $p = (r + 1) / (n + 1)$ , where  $r$  = the number of simulated values greater than the observed value and  $n$  = the number of simulations (North *et al.* 2002). As a further protection, relative-pair analyses utilizing the Kong & Cox (1997) statistic as implemented in Merlin (Abecasis *et al.* 2002) were conducted for each reported linkage peak to test whether evidence for linkage was

robust across analytic methods. All  $p$  values were interpreted using guidelines suggested by Lander & Kruglyak (1995).

SOLAR also allows for the simultaneous analysis of multiple traits as well as the inclusion of multiple covariates. In the bivariate case, SOLAR estimates the proportion of variance in each phenotype that can be explained by a genetic locus as well as estimating the proportion of the genetic correlation between the phenotypes that can be explained by this locus. To ensure that reported linkage peaks were not the result of a small subset of pedigrees, homogeneity tests were performed using the SOLAR  $H_{LOD}$  function (Goring, 2002). This test contrasts a null model in which families belong to a single distribution exhibiting genetic linkage to the tested locus against an alternative model in which families belong to one of two distributions, only one of which shows evidence of genetic linkage to the tested locus.

For the linkage analysis of the 14 SSAGA nicotine dependence symptoms, the support intervals for reported linkage peaks were defined as the region surrounding a linkage peak yielding a LOD score that was greater than the maximum LOD – 1 in each direction. Only these support intervals were subjected to linkage analysis to reduce the number of statistical tests. Further, LOD scores were only interpreted as a measure of magnitude and were not used to determine significance given the exploratory and descriptive nature of these analyses. Because the current sample was selected for alcohol dependence, which is highly correlated with tobacco use (Miller & Gold, 1998), prevalence rates for nicotine dependence and its respective symptoms were estimated in the unselected control sample and included in the tested models to correct for ascertainment bias when calculating  $h^2$  and linkage. Finally, allele frequencies used to calculate allele-sharing probabilities between relative pairs were estimated using study participants from varying ethnic backgrounds. This approach can lead to biased allele frequency and sharing probabilities amongst minority participants. Thus, analyses were conducted for the full sample as well as restricting the sample to only Caucasian participants. To eliminate the potential for the described bias, results for the Caucasian-only sample are reported, though notably only small changes were observed in LOD scores across analyses ( $\Delta LOD < 0.05$  for the variance components analysis and  $\Delta LOD < 0.4$  for the relative-pair analysis).

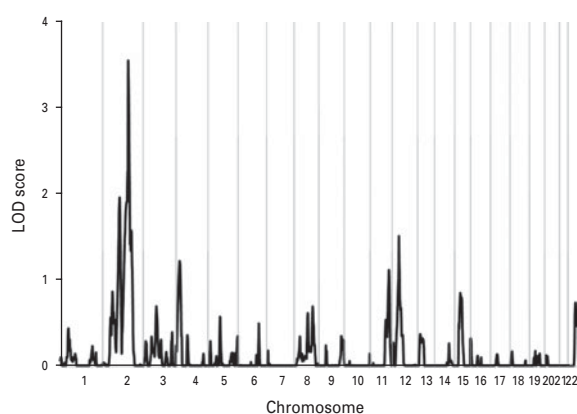
## Results

The  $h^2$  estimate for the nicotine dependence diagnosis was significantly greater than 0 ( $p < 0.001$ ), as were the  $h^2$  estimates for 13 of the 14 nicotine dependence

**Table 1.** Chromosomal regions with evidence of linkage to DSM-IV nicotine dependence

Chromosomal region	Variance components			Kong and Cox: relative pairs		Previous evidence for linkage to nicotine dependence
	cM	Nearest markers	LOD ( <i>p</i> )	cM	LOD ( <i>p</i> )	
2q13	123	D2S160	1.95 (0.00144)	110	1.21 (0.00929)	Straub <i>et al.</i> (1999); Sullivan <i>et al.</i> (2004)
2q31.1	184	D2S2188	3.54 (0.000012)	164	2.98 (0.00012)	Straub <i>et al.</i> (1999); Sullivan <i>et al.</i> (2004); Loukola <i>et al.</i> (2008)
4p15.33-31	27	D4S403/D4S419	1.22 (0.00973)	16	0.60 (0.04891)	Duggirala <i>et al.</i> (1999); Li <i>et al.</i> (2008)
11q25	135	D11S4126	1.11 (0.01292)	132	0.37 (0.09511)	Goode <i>et al.</i> (2003); Morley <i>et al.</i> (2006)
12p11.23-21	50	D12S1640/D12S345	1.51 (0.00452)	56	0.87 (0.02272)	Li <i>et al.</i> (2008); Sullivan <i>et al.</i> (2008)

cM, centiMorgans; LOD, logarithm (base 10) of odds.



**Fig. 1.** Multipoint linkage analysis for nicotine dependence for the entire genome. Chromosome numbers are represented on the *x* axis, and logarithm (base 10) of odds (LOD) scores are represented on the *y* axis. Results for each chromosome are aligned end to end with the p-terminus on the left. Vertical lines indicate the boundaries between chromosomes.

symptoms ( $p < 0.05$ ). The remaining symptom, 'continued smoking despite physical illness', approached significance ( $p = 0.08$ ).

Variance components linkage analysis of the nicotine dependence diagnosis yielded evidence for linkage on chromosome 2 at 184 cM nearest to marker D2S2188 with an associated LOD score of 3.54 (Fig. 1), which achieved genome-wide significance (point-wise empirical  $p = 0.000012$ ) (see Table 1). The parallel relative-pair analysis also revealed a linkage peak in this region at 164 cM nearest to marker D2S151 that failed to reach genome-wide significance (all pairs: LOD = 2.98, point-wise empirical  $p = 0.00012$ ), but, nonetheless, suggested the evidence for linkage to this region was robust across analytic methods. The shift in location of the linkage peaks likely resulted from differences in LOD score calculations across methods

given that the same identity-by-descent (IBD) estimates were used for both sets of analyses.

Notably, there was no evidence of heterogeneity in LOD scores across pedigrees, as each family had an estimated  $\alpha$ , interpreted as the probability of a given family belonging to a single population yielding evidence for linkage, that was  $> 0.99$  for each peak as estimated by the SOLAR<sub>HLOD</sub> function (Goring, 2002). Linkage analysis of the 14 nicotine dependence symptoms assessed by the SSAGA across the support interval of this peak (181–192 cM) showed that the symptom 'thought a lot about quitting or cutting down' yielded the strongest evidence for linkage (LOD = 2.66). The symptoms 'tobacco had less effect with continued regular use' (LOD = 1.74), 'found self smoking more than intended' (LOD = 1.45), and 'continued to smoke despite tobacco-induced emotional problems' (LOD = 1.05) also appeared to contribute to the observed linkage peak (for complete results, see Table 2).

Four additional loci yielded notable evidence of linkage (Table 1). The first locus represented a second peak located on chromosome 2 at 123 cM nearest to marker D2S160 (LOD = 1.95, point-wise empirical  $p = 0.00144$ ) with a support interval that extended from 109 to 132 cM. The second locus was found on chromosome 4 at 27 cM nearest to markers D4S403 and D4S419 (LOD = 1.22, point-wise empirical  $p = 0.00973$ ) with a support interval that extended from 7 to 43 cM. The third peak was found on chromosome 11 at 135 cM nearest to marker D11S4126 (LOD = 1.11, point-wise empirical  $p = 0.01292$ ) with a support interval that spanned from 106 to 146 cM, and the fourth locus was found on chromosome 12 at 50 cM nearest to markers D12S1640 and D12S345 (LOD = 1.51, point-wise empirical  $p = 0.00452$ ) with a support interval that extended from 37 to 67 cM. Among these peaks, the parallel relative-pair analyses only found evidence of linkage to the peak on chromosome 2 at 123 cM,

**Table 2.** LOD scores for symptoms at linkage peak on chromosome 2 at 184 cM

Trait	Position (cM)	LOD score
Thought a lot about quitting	191	2.66
Effects of tobacco decreased over time	184	1.74
Smoked more than intended	191	1.45
Smoked despite emotional problems	191	1.05
Smoked despite having serious illness	190	0.96
Needed to smoke more in a day	184	0.67
Had multiple withdrawal symptoms	184	0.64
Unable to stop or cut down on smoking	189	0.38
Continued smoking to avoid withdrawal	182	0.30
Needed to smoke more to feel comfortable	185	0.29
Continued smoking despite physical problem	185	0.07
Chain smoker or smoking took a lot of time	191	0.05
Ran out of cigarettes sooner than expected	188	0.03
Gave up activities due to tobacco use	184	0.00

LOD, Logarithm (base 10) of odds; cM, centiMorgans.

though this evidence was fairly weak (LOD=1.21, point-wise empirical  $p=0.00929$ ).

Linkage results for the nicotine dependence diagnosis were then compared with the results from a previous linkage scan of alcohol dependence. None of the regions reported in the previous scan yielded evidence of linkage to nicotine dependence. Additionally, when nicotine dependence was included as a covariate in a linkage analysis of alcohol dependence, LOD scores for the alcohol dependence diagnosis approached 0 in those regions that showed linkage to nicotine dependence in the present report (Table 3). These results provide evidence suggesting that the regions reported here confer risk towards nicotine but not alcohol dependence.

This conclusion was supported by subsequent analyses in which the alcohol dependence diagnosis was included in the linkage analysis of nicotine dependence first as a covariate and then as a second phenotype in a bivariate analysis. For the reported peak on chromosome 2 at 184 cM, a slight increase in the initial LOD score was observed when the alcohol dependence diagnosis was added as a covariate (from 3.54 to 3.57), whereas a drop in the LOD score was observed for the bivariate analysis (from 3.54 to 2.17) that probably occurred because of corrections for the inclusion of additional parameters. This result suggests that this region confers risk towards nicotine but not alcohol dependence (see Fig. 2). Similar

conclusions were made regarding the remaining peaks of interest (for complete results, see Table 3).

## Discussion

The primary aim of the current study was to conduct a genome-wide linkage scan of nicotine dependence in the UCSF Family Alcoholism Study sample to support and extend the findings of previous studies. A linkage peak was observed on chromosome 2 at 184 cM that achieved genome-wide significance when a variance components approach was used based on criteria described by Lander & Kruglyak (1995). This region continued to yield strong evidence for linkage when the Kong & Cox (1997) statistic was used to test for linkage, but failed to reach genome-wide significance. This divergence in the results limits claims of genome-wide significance, though the consistency in LOD scores across analytic methods provides strong evidence of linkage to this region. Additional peaks of interest were found on a second region of chromosome 2 at 123 cM, chromosome 4 at 27 cM, chromosome 11 at 135 cM, and chromosome 12 at 50 cM, though these should be interpreted tentatively given the weaker evidence for linkage.

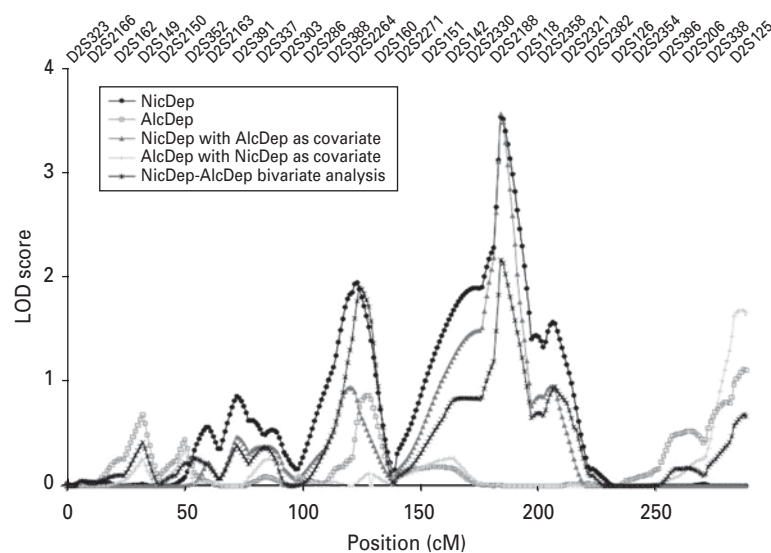
The linkage region on chromosome 2 at 184 cM has been previously identified as harboring a susceptibility locus for nicotine dependence (Straub *et al.* 1999; Loukola *et al.* 2008). The first study reported a linkage signal approximately 35 megabases centromeric, and the second study reported a linkage signal approximately 32 megabases telomeric of the peak reported in this study. In addition, a locus on the short arm of chromosome 2 at 85 cM, which contains *NRXN1*, has been previously linked to both nicotine and alcohol dependence (Yang *et al.* 2005; Bierut *et al.* 2007; Nussbaum *et al.* 2008), but this locus is approximately 100 megabases from the locus identified in the present study.

There are potential candidate genes within the support interval of the locus reported here. For example, the nicotinic acetylcholine  $\alpha 1$  gene (*CHRNA1*) is located near the center of the reported linkage peak. Although originally thought to be found only in muscle tissue, recent gene expression studies have found this gene to be expressed in brain as well (Su *et al.* 2004), suggesting a potential role in nicotine addiction. Evidence for an association between this gene and smoking behavior has been previously reported (Faraone *et al.* 2004), but negative findings have also been described (Sherva *et al.* 2008). In addition, single nucleotide polymorphisms (SNPs) located in the nearby growth factor receptor-bound protein 14 (*GRB14*) and grancalcin (*GCA*) genes were recently associated with nicotine dependence in a GWAS (Vink

**Table 3.** LOD scores for linkage analysis of nicotine and alcohol dependence phenotypes

Chromosome	cM	Nicotine dependence		Alcohol dependence		
		Univariate	Alcohol dependence as covariate	Univariate	Nicotine dependence as covariate	Bivariate: nicotine dependence and alcohol dependence
2	123	1.95	0.82	0.61	0.02	1.77
2	184	3.54	3.57	0.00	0.00	2.17
4	27	1.22	0.87	0.16	0.11	0.49
11	135	1.11	0.50	0.00	0.00	0.57
12	50	1.51	1.87	0.00	0.00	1.12

LOD, Logarithm (base 10) of odds; cM, centiMorgans.



**Fig. 2.** Linkage analysis for the nicotine and alcohol dependence phenotypes on chromosome 2. LOD, logarithm (base 10) of odds; cM, centiMorgans; NicDep, nicotine dependence; AlcDep, alcohol dependence.

*et al.* 2009). Potential mechanisms through which one or both of these genes might confer risk to nicotine dependence are not clear. *GRB14* is thought to be involved in insulin receptor signaling and may influence signaling pathways that regulate growth and metabolism (Carre *et al.* 2008). *GCA* may be involved in the migration and adherence of neutrophils (Jia *et al.* 2000). Thus, further studies are necessary to determine whether a causal variant is located in *GRB14* or *GCA* or whether the associated SNPs are in linkage disequilibrium with a causal variant located in a nearby gene such as *CHRNA1*.

Follow-up analyses of the chromosome 2 linkage peak showed that SSAGA symptoms encompassing a broad range of DSM-IV nicotine dependence symptom clusters, including evidence of tolerance, inability to quit smoking, escalating pattern of use, and persistent use despite negative health consequences, provided

modest contributions to the linkage signal. This suggests that the chromosome 2 locus at 184 cM confers risk for nicotine dependence in general rather than a specific facet of this disorder. As further evidence of this conclusion, it is notable that the linkage analysis of the nicotine dependence diagnosis yielded a higher LOD score than any of the individual nicotine dependence symptoms (LOD=3.54 *v.* maximum LOD=2.66). These results are consistent with a previous study demonstrating that a single genetic factor can explain a predominant proportion of the common variation between DSM-IV nicotine dependence symptoms (Lessov *et al.* 2004).

A further aim of this study was to determine whether the reported genomic regions contributed to nicotine dependence specifically or to addiction more generally by showing evidence of linkage to both alcohol and nicotine dependence. The former conclusion

was supported, as no overlaps between linkage signals reported in the present study were observed with those reported in a previous linkage study of alcohol dependence to chromosomes 1 at 11 cM, 2 at 287 cM, 8 at 163 cM, and 18 at 48 cM (I. R. Gizer *et al.* unpublished observations). Further, supplementary genome-wide linkage scans of nicotine dependence utilizing alcohol dependence diagnoses alternatively as a covariate and as an additional predictor in a bivariate analysis showed little evidence of linkage between alcohol dependence and the regions reported herein. This suggests that the susceptibility loci identified in the present study are specifically involved in the etiology of nicotine dependence and are unrelated to alcohol dependence.

This result was somewhat surprising given the strong correlations between drinking and smoking behaviors (Miller & Gold, 1998). Twin studies suggest that common genetic influences are partially responsible for the observed correlation (Swan *et al.* 1997; True *et al.* 1999), though disorder-specific genetic influences have been identified as well (Kendler *et al.* 2007; Volk *et al.* 2007). Nonetheless, previous family-based samples selected for alcohol dependence have identified genetic loci that confer risk to alcohol- and tobacco-use phenotypes. For example, loci on chromosomes 2 (Bierut *et al.* 2004), 4 (Ehlers & Wilhelmsen, 2006), 7 (Loukola *et al.* 2008; Sullivan *et al.* 2008) and 18 (Sullivan *et al.* 2008) have been shown to contribute jointly to alcohol and nicotine dependence. The lack of such findings in the present study suggests that unique genetic influences contributed to nicotine and alcohol dependence in the UCSF Family Alcoholism Study.

The reported findings have important implications for molecular genetic studies of nicotine dependence, but there are limitations that should be noted. For example, the UCSF Family sample was originally selected for alcohol dependence. Thus, it is not clear how the reported findings will generalize to populations without this bias, though a previous study reporting evidence of linkage to the chromosome 2 region used a sample selected for nicotine rather than alcohol dependence (Loukola *et al.* 2008). Additionally, categorizing 'never smokers' and 'not nicotine-dependent' participants as unaffected individuals may have influenced study results (Munafò *et al.* 2004). Given that unique genetic influences contribute to the initiation and persistent use of tobacco (Heath & Madden, 1995), combining these participants into a single unaffected category likely limited our ability to detect these unique genetic influences. Nonetheless, there is substantial overlap in the genetic influences contributing to these stages of tobacco use (Sullivan & Kendler, 1999), providing

justification for this approach. The statistical power of the present study is another possible limitation. Linkage studies lack sufficient statistical power for identifying loci with small effects, and this may explain the lack of support for loci such as chromosomes 9q and 10q that have been previously linked to nicotine dependence.

In summary, the current study adds to the literature by supporting evidence of genetic linkage of chromosome 2q to nicotine dependence. This study extends this finding by showing that this region confers risk to the full nicotine dependence diagnosis rather than a specific facet of the disorder. Finally, the present study suggests that this locus, as well as the additional loci identified, confers risk to nicotine but not alcohol dependence, thus providing evidence that this genomic region may harbor a gene specifically involved in the physiological effects and/or metabolism of nicotine.

### Acknowledgements

The present study was supported by funds from the State of California for medical research on alcohol and substance abuse through the UCSF. Additional support was provided by the Ernest Gallo Clinic and Research Center (to K.C.W.), AA10201 and U54 RR0250204 (to C.L.E.) and T32 AA007573 (to I.R.G.).

### Declaration of Interest

None.

### References

- Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002). Merlin – rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics* **30**, 97–101.
- Almasy L, Blangero J (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics* **62**, 1198–1211.
- Amos CI, Krushkal J, Thiel TJ, Young A, Zhu DK, Boerwinkle E, de Andrade M (1997). Comparison of model-free linkage mapping strategies for the study of a complex trait. *Genetic Epidemiology* **14**, 743–748.
- APA (1994). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV*. American Psychiatric Association: Washington.
- Bergen AW, Korczak JF, Weissbecker KA, Goldstein AM (1999). A genome-wide search for loci contributing to smoking and alcoholism. *Genetic Epidemiology* **17** (Suppl.), S55–S60.
- Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, Swan GE, Rutter J, Bertelsen S, Fox L, Fugman D, Goate AM, Hinrichs AL, Konvicka K, Martin NG, Montgomery GW, Saccone NL, Saccone SF, Wang JC, Chase GA, Rice JP, Ballinger DG



- (2007). Novel genes identified in a high-density genome wide association study for nicotine dependence. *Human Molecular Genetics* **16**, 24–35.
- Bierut LJ, Rice JP, Goate A, Hinrichs AL, Saccone NL, Foroud T, Edenberg HJ, Cloninger CR, Begleiter H, Conneally PM, Crowe RR, Hesselbrock V, Li TK, Nurnberger Jr. JI, Porjesz B, Schuckit MA, Reich T** (2004). A genomic scan for habitual smoking in families of alcoholics: common and specific genetic factors in substance dependence. *American Journal of Medical Genetics* **124A**, 19–27.
- Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger Jr. JI, Reich T, Schmidt I, Schuckit MA** (1994). A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *Journal of Studies on Alcohol* **55**, 149–158.
- Carmelli D, Swan GE, Robinette D, Fabsitz RR** (1990). Heritability of substance use in the NAS-NRC Twin Registry. *Acta Geneticae Medicae et Gemellologiae (Roma)* **39**, 91–98.
- Carre N, Cauzac M, Girard J, Burnol AF** (2008). Dual effect of the adapter growth factor receptor-bound protein 14 (grb14) on insulin action in primary hepatocytes. *Endocrinology* **149**, 3109–3117.
- Duggirala R, Almasy L, Blangero J** (1999). Smoking behavior is under the influence of a major quantitative trait locus on human chromosome 5q. *Genetic Epidemiology* **17** (Suppl.), S139–S144.
- Duggirala R, Blangero J, Almasy L, Arya R, Dyer TD, Williams KL, Leach RJ, O'Connell P, Stern MP** (2001). A major locus for fasting insulin concentrations and insulin resistance on chromosome 6q with strong pleiotropic effects on obesity-related phenotypes in nondiabetic Mexican Americans. *American Journal of Human Genetics* **68**, 1149–1164.
- Duggirala R, Williams JT, Williams-Blangero S, Blangero J** (1997). A variance component approach to dichotomous trait linkage analysis using a threshold model. *Genetic Epidemiology* **14**, 987–992.
- Ehlers CL, Wilhelmsen KC** (2006). Genomic screen for loci associated with tobacco usage in Mission Indians. *BMC Medical Genetics* **7**, 1–9.
- Faraone SV, Su J, Taylor L, Wilcox M, Van Eerdewegh P, Tsuang MT** (2004). A novel permutation testing method implicates sixteen nicotinic acetylcholine receptor genes as risk factors for smoking in schizophrenia families. *Human Heredity* **57**, 59–68.
- Goode EL, Badzioch MD, Kim H, Gagnon F, Rozek LS, Edwards KL, Jarvik GP; Framingham Heart Study** (2003). Multiple genome-wide analyses of smoking behavior in the Framingham Heart Study. *BMC Genetics* **4** (Suppl.), S102.
- Goring HH** (2002). Hlod (Homo v. 0.2) documentation. *SOLAR Online User's Manual* ([http://solar.sfbgenetics.org/doc/95.appendix\\_5.txt](http://solar.sfbgenetics.org/doc/95.appendix_5.txt)).
- Heath AC, Madden PA** (1995). Genetic influences on smoking behavior. In *Behavior Genetic Approaches in Behavioral Medicine* (ed. J. R. Turner, L. R. Cardon and J. K. Hewitt), pp. 45–66. New York: Plenum.
- Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO** (1991). The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. *British Journal of Addiction* **86**, 1119–1127.
- Jia J, Han Q, Borregaard N, Lollike K, Cygler M** (2000). Crystal structure of human grancalcin, a member of the penta-EF-hand protein family. *Journal of Molecular Biology* **300**, 1271–1281.
- Jung J, Weeks DE, Feingold E** (2006). Gene-dropping vs. empirical variance estimation for allele-sharing linkage statistics. *Genetic Epidemiology* **30**, 652–665.
- Kendler KS, Myers J, Prescott CA** (2007). Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence. *Archives of General Psychiatry* **64**, 1313–1320.
- Kendler KS, Neale MC, Sullivan P, Corey LA, Gardner CO, Prescott CA** (1999). A population-based twin study in women of smoking initiation and nicotine dependence. *Psychological Medicine* **29**, 299–308.
- Kong A, Cox NJ** (1997). Allele-sharing models: LOD scores and accurate linkage tests. *American Journal of Human Genetics* **61**, 1179–1188.
- Lander E, Kruglyak L** (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**, 241–247.
- Lessov CN, Martin NG, Statham DJ, Todorov AA, Slutske WS, Bucholz KK, Heath AC, Madden PAF** (2004). Defining nicotine dependence for genetic research: evidence from Australian twins. *Psychological Medicine* **34**, 865–879.
- Li MD** (2008). Identifying susceptibility loci for nicotine dependence: 2008 update based on recent genome-wide linkage analyses. *Human Genetics* **123**, 119–131.
- Li MD, Cheng R, Ma JZ, Swan GE** (2003). A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. *Addiction* **98**, 23–31.
- Li MD, Ma JZ, Payne TJ, Lou XY, Zhang D, Dupont RT, Elston RC** (2008). Genome-wide linkage scan for nicotine dependence in European Americans and its converging results with African Americans in the Mid-South Tobacco Family sample. *Molecular Psychiatry* **13**, 407–416.
- Lou XY, Ma JZ, Payne TJ, Beuten J, Crew KM, Li MD** (2006). Gene-based analysis suggests association of the nicotinic acetylcholine receptor  $\beta 1$  subunit (CHRN1) and M1 muscarinic acetylcholine receptor (CHRM1) with vulnerability for nicotine dependence. *Human Genetics* **120**, 381–389.
- Loukola A, Broms U, Maunu H, Widen E, Heikkila K, Siivola M, Salo A, Pergadia ML, Nyman E, Sammalisto S, Perola M, Agrawal A, Heath AC, Martin NG, Madden PA, Pelttonen L, Kaprio J** (2008). Linkage of nicotine dependence and smoking behavior on 10q, 7q and 11p in twins with homogeneous genetic background. *Pharmacogenomics Journal* **8**, 209–219.
- McPeck MS, Sun L** (2000). Statistical tests for detection of misspecified relationships by use of genome-screen data. *American Journal of Human Genetics* **66**, 1076–1094.

- Miller NS, Gold MS** (1998). Comorbid cigarette and alcohol addiction: epidemiology and treatment. *Journal of Addictive Diseases* **17**, 55–66.
- Mokdad AH, Marks JS, Stroup DF, Gerberding JL** (2004). Actual causes of death in the United States, 2000. *Journal of the American Medical Association* **291**, 1238–1245.
- Morley KI, Medland SE, Ferreira MA, Lynskey MT, Montgomery GW, Heath AC, Madden PA, Martin NG** (2006). A possible smoking susceptibility locus on chromosome 11p12: evidence from sex-limitation linkage analyses in a sample of Australian Twin Families. *Behavior Genetics* **36**, 87–99.
- Munafó M, Clark T, Johnstone E, Murphy M, Walton R** (2004). The genetic basis for smoking behavior: a systematic review and meta-analysis. *Nicotine & Tobacco Research* **6**, 583–597.
- North BV, Curtis D, Sham PC** (2002). A note on the calculation of empirical *P* values from Monte Carlo procedures. *American Journal of Human Genetics* **71**, 439–441.
- Nussbaum J, Xu Q, Payne TJ, Ma JZ, Huang W, Gelernter J, Li MD** (2008). Significant association of the neurexin-1 gene (*NRXN1*) with nicotine dependence in European- and African-American smokers. *Human Molecular Genetics* **17**, 1569–1577.
- O'Connell JR, Weeks DE** (1998). PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *American Journal of Human Genetics* **63**, 259–266.
- Seaton KL, Cornell JL, Wilhelmsen KC, Vieten C** (2004). Effective strategies for recruiting families ascertained through alcoholic probands. *Alcoholism: Clinical and Experimental Research* **28**, 78–84.
- Sherva R, Wilhelmsen K, Pomerleau CS, Chasse SA, Rice JP, Snedecor SM, Bierut LJ, Neuman RJ, Pomerleau OF** (2008). Association of a single nucleotide polymorphism in neuronal acetylcholine receptor subunit  $\alpha 5$  (*CHRNA5*) with smoking status and with 'pleasurable buzz' during early experimentation with smoking. *Addiction* **103**, 1544–1552.
- Straub RE, Sullivan PF, Ma Y, Myakishev MV, Harris-Kerr C, Wormley B, Kadambi B, Sadek H, Silverman MA, Webb BT, Neale MC, Bulik CM, Joyce PR, Kendler KS** (1999). Susceptibility genes for nicotine dependence: a genome scan and followup in an independent sample suggest that regions on chromosomes 2, 4, 10, 16, 17 and 18 merit further study. *Molecular Psychiatry* **4**, 129–144.
- Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB** (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences USA* **101**, 6062–6067.
- Sullivan PF, Kendler KS** (1999). The genetic epidemiology of smoking. *Nicotine & Tobacco Research* **1**, S51–57.
- Sullivan PF, Kuo PH, Webb BT, Neale MC, Vittum J, Furberg H, Walsh D, Patterson DG, Riley B, Prescott CA, Kendler KS** (2008). Genomewide linkage survey of nicotine dependence phenotypes. *Drug and Alcohol Dependence* **93**, 210–216.
- Sullivan PF, Neale BM, van den Oord E, Miles MF, Neale MC, Bulik CM, Joyce PR, Straub RE, Kendler KS** (2004). Candidate genes for nicotine dependence via linkage, epistasis, and bioinformatics. *American Journal of Medical Genetics* **126B**, 23–36.
- Swan GE, Carmelli D, Cardon LR** (1997). Heavy consumption of cigarettes, alcohol and coffee in male twins. *Journal of Studies on Alcohol* **58**, 182–190.
- True WR, Xian H, Scherrer JF, Madden PA, Bucholz KK, Heath AC, Eisen SA, Lyons MJ, Goldberg J, Tsuang M** (1999). Common genetic vulnerability for nicotine and alcohol dependence in men. *Archives of General Psychiatry* **56**, 655–661.
- Vieten C, Seaton KL, Feiler HS, Wilhelmsen KC** (2004). The University of California, San Francisco Family Alcoholism Study. I. Design, methods, and demographics. *Alcoholism: Clinical and Experimental Research* **28**, 1509–1516.
- Vink JM, Smit AB, de Geus EJ, Sullivan P, Willemsen G, Hottenga JJ, Smit JH, Hoogendijk WJ, Zitman FG, Peltonen L, Kaprio J, Pedersen NL, Magnusson PK, Spector TD, Kyvik KO, Morley KI, Heath AC, Martin NG, Westendorp RG, Slagboom PE, Tiemeier H, Hofman A, Uitterlinden AG, Aulchenko YS, Amin N, van Duijn C, Penninx BW, Boomsma DI** (2009). Genome-wide association study of smoking initiation and current smoking. *American Journal of Human Genetics* **84**, 367–379.
- Volk HE, Scherrer JF, Bucholz KK, Todorov A, Heath AC, Jacob T, True WR** (2007). Evidence for specificity of transmission of alcohol and nicotine dependence in an offspring of twins design. *Drug and Alcohol Dependence* **87**, 225–232.
- Wang D, Ma JZ, Li MD** (2005). Mapping and verification of susceptibility loci for smoking quantity using permutation linkage analysis. *Pharmacogenomics Journal* **5**, 166–172.
- Wilhelmsen KC, Schuckit M, Smith TL, Lee JV, Segall SK, Feiler HS, Kalmijn J** (2003). The search for genes related to a low-level response to alcohol determined by alcohol challenges. *Alcoholism: Clinical and Experimental Research* **27**, 1041–1047.
- Yang HC, Chang CC, Lin CY, Chen CL, Lin CY, Fann CS** (2005). A genome-wide scanning and fine mapping study of COGA data. *BMC Genetics* **6** (Suppl.), S30.