Molecular Psychiatry (2008) 13, 368–373

www.nature.com/mp

IMMEDIATE COMMUNICATION

α-5/α-3 nicotinic receptor subunit alleles increase risk for heavy smoking

W Berrettini[1,2,3], X Yuan[2,3], F Tozzi[2,3], K Song[2,3], C Francks[2,3], H Chilcoat[4], D Waterworth[2,3], P Muglia[2,3,5] and V Mooser[2,3]

[1]*Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA, USA;* [2]*Clinical Pharmacology and Discovery Medicine, GlaxoSmithKline, Upper Merion, PA, USA;* [3]*Clinical Pharmacology and Discovery Medicine, GlaxoSmithKline, Verona, Italy;* [4]*Worldwide Epidemiology, GlaxoSmithKline, Research Triangle Park, NC, USA and* [5]*University of Toronto, Toronto, ON, Canada*

Twin studies indicate that additive genetic effects explain most of the variance in nicotine dependence (ND), a construct emphasizing habitual heavy smoking despite adverse consequences, tolerance and withdrawal. To detect ND alleles, we assessed cigarettes per day (CPD) regularly smoked, in two European populations via whole genome association techniques. In these ~7500 persons, a common haplotype in the CHRNA3–CHRNA5 nicotinic receptor subunit gene cluster was associated with CPD (nominal $P = 6.9 \times 10^{-5}$). In a third set of European populations ($n = \sim 7500$) which had been genotyped for ~6000 SNPs in ~2000 genes, an allele in the same haplotype was associated with CPD (nominal $P = 2.6 \times 10^{-6}$). These results (in three independent populations of European origin, totaling ~15 000 individuals) suggest that a common haplotype in the CHRNA5/CHRNA3 gene cluster on chromosome 15 contains alleles, which predispose to ND.
Molecular Psychiatry (2008) **13,** 368–373; doi:10.1038/sj.mp.4002154; published online 29 January 2008

Keywords: cigarette smoking; nicotine dependence; genetic association; nicotinic receptor subunit; whole genome association

Introduction

If current trends continue, the annual number of deaths (worldwide) from tobacco-related diseases will double from 5 million in the year 2000 to 10 million in 2020.[1] Abundant data from twin and adoption studies provide evidence for the heritability of habitual cigarette smoking. Sullivan and Kendler[2] reviewed twin data in smoking studies. The twin studies suggest that a majority of risk for nicotine dependence (ND) may be attributable to genetic factors. More recent twin smoking research[3–5] suggests that the heritability of ND is even higher.

Lessov *et al.*[3] analyzed multiple ND-related phenotypes in a large twin study. They concluded that the cigarettes per day (CPD) variable has one of the highest genetic loadings among ND-related phenotypes. This phenotype may be a simple measure of tolerance, since most smokers achieve a level of CPD that would have been toxic when they initiated smoking.

While many candidate gene and linkage studies have been published for ND (for review see Li[6]), many have not been replicated consistently, due to small sample sizes, small effect sizes of single alleles/haplotypes and genetic heterogeneity. A whole genome association[7] and candidate gene study[8] of ~1000 ND cases and ~900 controls of European ancestry was recently reported. In this study, cases were defined as individuals with a score of $\geqslant 4$ on the Fagerstrom Test for Nicotine Dependence (FTND[9]), while controls were individuals who smoked at least 100 cigarettes over a lifetime, but had a score of 0 on the FTND. From this perspective, controls are individuals who were exposed to smoking, but never became dependent or never engaged in heavy smoking. In the Beirut *et al.*[7] study, no SNP was found to be associated at a statistical level which would account for searching the entire genome ($\sim 10^{-7}$), but promising, biologically plausible results were obtained for the CHRNA3-CHRNA5-CHRNB4 nicotinic receptor subunit gene cluster on chromosome 15 (minimal $P = 0.0003$). These results were also highlighted in a candidate gene study of the same population.[8]

The results of a whole genome association study of CPD as a quantitative trait in ~7500 people of European origin are reported here. In addition, a complementary study of CPD in a second population of ~7500 people of European origin is also reported. The results indicate that one or more alleles in the CHRNA3-CHRNA5-CHRNB4 nicotinic receptor subunit gene cluster on chromosome 15 increase risk for ND.

Correspondence: Professor W Berrettini, Department of Psychiatry, University of Pennsylvania School of Medicine, room 2206 125 S. 31st Street, Philadelphia, PA 19104, USA.
E-mail: wadeb@mail.med.upenn.edu
Received 14 December 2007; accepted 17 December 2007; published online 29 January 2008

Table 1 Analysis of CPD and DSM-IV nicotine dependence in the NESARC data

Cutoff	*Sensitivity*	*Specificity*
1	0.97	0.0809
5	0.8855	0.308
10	0.7877	0.4307
15	0.603	0.3936
20	0.5373	0.6499
25	0.1813	0.8933
30	0.1676	0.899
35	0.1013	0.9367
40	0.0984	0.9378
45	0.0282	0.98
50	0.0276	0.9804
55	0.0197	0.9846
60	0.0194	0.9847
70	0.0046	0.9963

Abbreviation: CPD, cigarettes per day.
Sensitivity and specificity for usual quantity smoked by lifetime history of nicotine dependence among smokers ($n = 18\,930$).
Approximately 65% specificity for persons habitually smoking $\geqslant 20$ CPD criteria for DSM-IV ND.
When the CPD value is $\geqslant 25$, specificity improves to 90% for DSM-IV criteria for ND.
The National Epidemiologic Survey on alcohol and related conditions (Grant *et al*., 2003)[11] data were analyzed for the specificity and sensitivity of various cigarettes per day definitions of nicotine dependence by DSM-IV criteria.

Methods

CPD was selected as the single phenotype for analysis, because it is a highly heritable,[3] widely used phenotype in genetic studies of smoking.[7,8,10] To understand the relationship between CPD and DSM-IV ND, we analyzed an epidemiologic dataset, the National Epidemiologic Survey of Alcohol and Related Conditions (NESARC[11]). This was a household survey of 43 000 Americans, in which data were collected on CPD and on the diagnosis of ND. Sensitivity and specificity were examined for various values of CPD and the diagnosis of ND. From this analysis (Table 1), specificity and sensitivity of DSM-IV ND are unacceptably low, even when individuals are smoking nearly a pack daily (20 CPD). The analysis suggested that the specificity for a DSM-IV diagnosis of ND improved to 90% when persons smoking $\geqslant 25$ CPD were considered (Table 1). Thus, for dichotomous (case–control) analysis of genotypes, a case was defined as a person who smoked $\geqslant 25$ CPD. Part of the value of this case definition is that it probably defines by DSM-IV a ND person.

GSK sponsored a cardiovascular disease study of a population-based sample of 6205 adult residents of the city of Lausanne. Briefly, participants in the study were randomly selected from a list of 56,694 individuals aged 35–75 years who were permanent residents of the City of Lausanne. Recruitment took place between April 2003 and March 2006, and the overall participation rate was 41%. Only individuals of European origin (persons for whom the four grandparents were of European origin) were included in the study, in an attempt to limit heterogeneity for genetic studies. Participants completed a health questionnaire and underwent a physical exam. They provided a blood sample for genetic studies and clinical chemistries. The health questionnaire included the question: If you were ever a daily smoker, what is the maximal number of CPD regularly smoked? All participants were duly informed about the sponsorship by GSK and were consented for the use of biological samples and data by GSK and its subsidiaries; the study was approved by the Local Ethics Committee.

Genome-wide SNP genotyping was performed on 6000 Lausanne participants, using the Affymetrix 500 K SNP chip, as recommended by the manufacturer. A total of 366 samples were excluded from this analysis as they either had an efficiency $< 90\%$ or showed gender inconsistencies, so that genetic data from 5634 individuals were included in the present study. Markers were excluded if they were monomorphic (4052), had a call rate $< 95\%$ (157) or were out of Hardy–Weinberg equilibrium (35,417), leaving a total of 460,959 markers for analysis.

GSK also sponsored an unrelated case–control genetic association study of dyslipidemia, nested within the GEMS project, in a population of European-origin individuals recruited from medical clinics. A total of 923 cases, defined as individuals with high triglycerides levels and low HDL-cholesterol levels in plasma, and 924 highly discordant controls with low triglycerides, high HDL-cholesterol levels and an excess in body-weight were recruited in this study. Genotypes from Affymetrix 500 K SNP chips were subjected to quality control measures similar to those for the Lausanne study.

Quantitative analyses of CPD with genotype were performed using gender as a covariate, since more men are regular smokers than women.[12] A 2003 US survey revealed that 24% of men and 19% of women are regular smokers.[12] Individuals who denied ever smoking were excluded from this analysis, as they may have never had sufficient exposure to cigarette smoking to become dependent.[7,8] Quantile–quantile plots of the GEMS and Lausanne populations revealed little deviation from the expected distribution (see Supplementary Figure S1). These data suggest an absence of population stratification across the phenotype of CPD. Data were analyzed for association with one and only one phenotype, CPD, using the computer program, PLINK.[13]

GSK also sponsored the establishment of case–control association samples for about 18 common diseases known as the High-Throughput Disease-specific target Identification Program (HITDIP).[14] Each of these samples consisted of approximately 1000 Cases and 1000 controls collected at multiple sites in North America and Europe. These DNA samples were genotyped at $\sim$6000 SNPs in a panel

Table 2 Genes with the same putative CPD risk alleles in GEMS & Lausanne

SNP	*Chromosome*	*StartPos*	*Gene*	*GEMS p*	*Laus* P	*Pooled* P
RS6495308	15	76694711	CHRNA3	0.008723	0.000601	6.90E−05
RS7804771	7	136783133	DGKI	0.007254	0.001059	9.81E−05
RS2645339	5	178348669	GRM6	0.000916	0.02548	0.000272
RS5522	4	149576925	NR3C2	0.000589	0.001751	1.52E−05
RS5525	4	149575966	NR3C2	0.01019	0.000269	3.78E−05
RS10869409	9	76313209	RORB	0.01129	0.000365	5.53E−05
RS7846903	9	76304931	RORB	0.009396	0.001039	0.000122
RS13293006	9	76326716	RORB	0.02351	0.002327	0.000592
RS7873840	9	76340109	RORB	0.01027	0.01714	0.001698
RS4932598	15	90338849	SLCO3A1	0.0249	0.000639	0.000192
RS4932597	15	90338621	SLCO3A1	0.03262	0.000637	0.000245
RS12439738	15	90336555	SLCO3A1	0.00379	0.01282	0.000531
RS12439765	15	90336606	SLCO3A1	0.004227	0.01374	0.000625

The six genes with the lowest *P* values are listed for the GEMS and Lausanne studies. Only those genes with the same allele nominally significant in both studies were included. *P* values were combined using Fisher's method.

of 1800 'drugable' candidate genes (for additional details see Roses A *et al.*[14]). These HITDIP studies collected a common set of data concerning the medical history of each participant, including a question about smoking habits. The only ND-related phenotype available in these HITDIP studies was the answer to the question: If you ever smoked regularly, what was the number of CPD typically smoked? In some of the HITDIP studies, this may have been interpreted as the maximum number of CPD regularly smoked, as opposed to an average or typical number, perhaps due in part to the different languages in the countries in which the studies were executed. Thus, treating CPD in the HITDIP samples as a quantitative trait could have led to errors. A case–control analysis of the HITDIP data was conducted. A control was defined as anyone who reported CPD always <5 CPD and a case anyone who reported smoking $\geqslant 25$ CPD. Individuals who denied ever smoking a single cigarette were excluded from the analysis. In the dichotomous analysis using PLINK,[13] the definition of a case (a person smoking $\geqslant 25$ CPD) rested on the results of our examination of the NESARC data that established the relationship between DSM-IV diagnosis of ND and the maximal number of CPD regularly smoked (see Table 1).

Consent forms were reviewed for each HITDIP study to determine whether the language in the consent form permitted anonymous analysis of the CPD phenotype. In instances where the consent form was narrowly worded (for example, did not permit analysis of phenotypes unrelated to the primary disease), the data set was not analyzed. In instances where the validity of the CPD variable might be questioned (for example, Alzheimer's disease), the data set was not analyzed.

Results

The quantitative CPD phenotype was analyzed in both the GEMS and Lausanne studies, using gender as a covariate, on a total of $\sim$7600 individuals. In these two studies, no SNP reached a proposed genome-corrected level of significance ($\sim 10^{-7}$). There were a total of 117 genes in which at least one SNP had nominal significance ($P<0.05$) in both studies, with the same allele identified as the risk allele (more common among smokers). In Table 2, the six most significant genes are listed, with all the tested SNPs.

CHRNA3 (the α 3 subunit of the nicotinic receptor) is an obvious candidate gene. CHRNA3 has been associated with ND in a case control analysis of $\sim$1000 ND DNA samples and $\sim$900 control DNA samples,[8] with $P=0.0003$. The two associated CHRNA3 SNPs in Saccone *et al.*[8] were not tested in GEMS/Lausanne, as they are not represented in the Affymatrix 500 K chip.

The HITDIP[14] data were then queried to determine whether CHRNA3 SNPs were in linkage disequilibrium (LD) with CPD. HITDIP analyses are limited because $\sim$2000 genes were studied using $\sim$6000 SNPs. Seven HITDIP studies included were COPD, depression, schizophrenia, migraine, Genecard (a cardiovascular disease study), rheumatoid arthritis and osteoarthritis.[14] These seven HITDIP studies were analyzed in a case (>25 CPD, $n=1740$) versus control (<5 CPD, $n=6200$) mode, using gender as a covariate. Across these HITDIP studies, there was only a single SNP in CHRNA3, which was genotyped, rs1317286, an intronic SNP (see Figure 1). This SNP was strongly associated with CPD ($P=0.0000026$). The other genes listed in Table 2 were not genotyped in the HITDIP study.

As can be seen from Figure 1, there are no SNPS in common for Saccone *et al.*,[8] HITDIP and GEMS/Lausanne. None of these SNPs convey known functional difference for the CHRNA3 gene. However, the associated alleles are on a CHRNA5-CHRNA3 common haplotype (see Figure 2; www.hapmap.org). As indicated by the black triangle (in Figure 2) drawn around the CHRNA3-CHRNA5 SNPs, these ND risk alleles all lie within a single haplotype block, while CHRNB4 SNPs lie in an adjacent haplotype block.

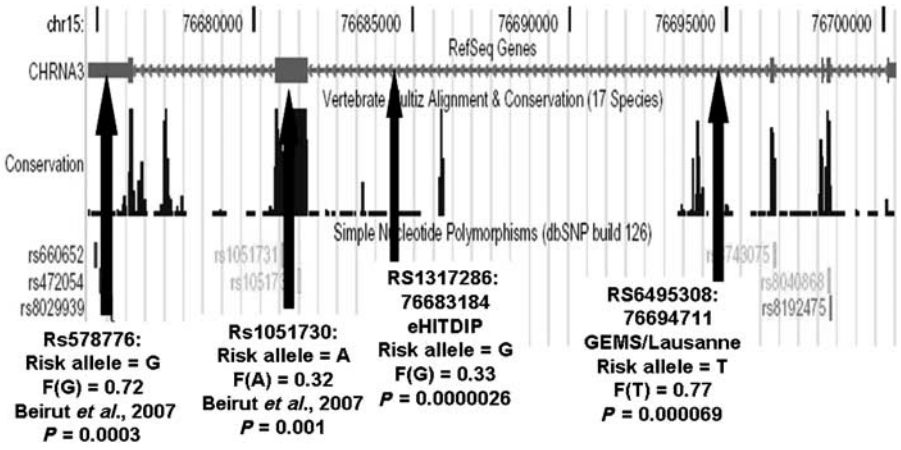

Figure 1 A diagram of the CHRNA3 gene on Chromosome 15 is shown, with base pair number at the top. Exon 1 is on the right side of the figure. Locations of SNPs are indicated by arrows, with base pair given according to www.genome.ucsc.edu. Study of origin is noted, along with the risk allele, risk allele frequency in European-origin individuals and *P* value. Linkage disequilibrium values are taken from www.hapmap.org. F(RA) denotes the frequency of risk allele.

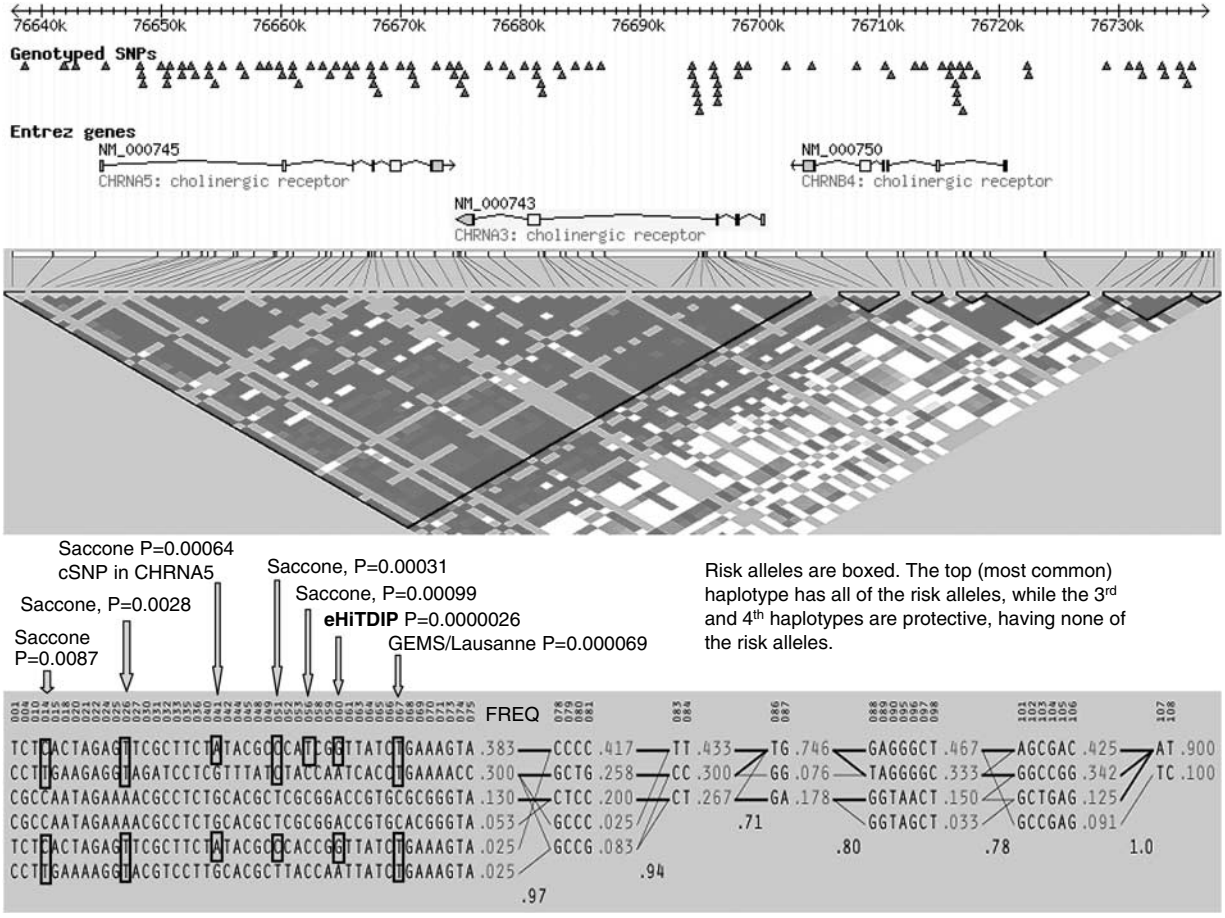

Figure 2 Haplotype Blocks and Linkage Disequilibrium in the CHRNA3 and CHRNA5 Region. Haplotype blocks (indicated by the black triangles) are shown for the CHRNA3/5 region for persons of European origin (www.hapmap.org), with LD values indicated by shading (dark grey = high LD, white = intermediate, light grey = low). Genotyped SNPs used to establish the haplotype blocks are given at the top of the figure. Risk alleles are boxed. The three studies identify the same common haplotype (the first one, with 38% allele frequency) as conveying risk, while the third and fourth haplotypes are protective, in that no risk alleles are present. Figure created using Haploview.

These data are most consistent with the first haplotype conveying risk for ND. However, many risk alleles are found in the second haplotype, while the third and fourth haplotypes are clearly protective, containing no risk alleles. The two remaining haplotypes are uncommon in individuals of European origin. Because the identified risk alleles at these several SNPs lie on the same common haplotype in Europeans, imputation analysis of these data is not likely to reveal one or more causative alleles.[15]

Discussion

Due to extensive LD in this CHRNA3-CHRNA5 region (Figure 2), it is possible that the causative allele(s) may lie within either or both of these genes, which are nicotinic receptor subunit genes. Some data support LD of CHRNA5 SNPs with ND. Saccone *et al*.[8] detected LD for several CHRNA5 SNPs, including rs16969968 ($P=0.0006$), a mis-sense (398 Asp/Asn) SNP which may have functional significance for the CHRNA5 protein. While this mis-sense SNP was not genotyped in the GEMS/Lausanne data, nor in HITDIP, in the Lausanne data, CHRNA5 SNP rs951266 is nominally associated with CPD ($P=0.0006$). This SNP did not pass quality control for GEMS. Thus, the present data do not identify unequivocally the CHRNA3 as a risk gene for ND, to the exclusion of CHRNA5. It will be necessary to conduct additional experiments to clarify the identity and number of ND risk alleles in the CHRNA3-CHRNA5 region. While SNPs in CHRNB4 are less probably involved in ND in these populations, a role for this nicotinic receptor subunit cannot be excluded on the basis of the current data, even though it may lie in an adjacent haplotype block, because there is significant LD between CHRNB4 and the main haplotype block identified as harboring ND risk alleles (see Figure 2).

Both the CHRNA3 and CHRNA5 genes are expressed in human brain areas relevant to addiction, such as the nucleus accumbens, amygdala and entorhinal cortex (see Supplementary Figure S2). Co-ordinated expression of these two genes may occur, as they share some 3′UTR.[16] α-5 subunits are typically found associated with some α-3 and α-4 nicotinic receptor subunits.[17] α-5 and α-3 subunits are obligate accessory subunits which cannot form functional nAChRs by themselves or in combination with only one other type of subunit. α-5 subunits do not participate in the formation of the acetylcholine binding site,[17] but when an α-5 subunit is expressed in an α-3-containing receptor, there are marked changes in Ca^{2+} permeability, desensitization and binding affinities.[18] The regulation of α-5 or α-3 subunit incorporation into a functioning receptor is imperfectly understood.

The effort to identify novel targets for ND through analysis of whole genome association genetic data sets has yielded convincing evidence that alleles of the CHRNA3/5 region of chromosome 15 increases risk for ND (see Figure 2, and Bierut LJ and Saccone SF[7,8]). There is extensive LD across these two genes, which are oriented in opposite directions and share some 3/UTR.[16] All CHRNA3 and CHRNA5 identified risk alleles for several independent populations (GEMS, Lausanne, HITDIP,[7,8]) lie on a single common haplotype in the region. While this haplotype is clearly implicated in risk for ND, the causative allele(s) are not apparent, due to LD across these two genes. The causative allele(s) must be identified through biological studies of the effects of these SNPs.

Acknowledgments

We thank Sreekumar Pillai, for helpful discussions. Allen Roses, Lefkos Midleton, Perry Nisen, Paul Matthews and Emilangelo Ratti are thanked for their support of these genetic studies. We are indebted to the GEMS/Lausanne investigators, including Antero Kesaniemi, Gerard Waeber, Phil Barter, Bob Mahley, Ruth McPherson, Scott Grundy and Tom Bersot. Supplementary information is available at the Molecular Psychiatry website. WHB was supported by NIDA Grant P60 05186 to Charles O'Brien, MD, PhD, by a grant from the Tzedekah Foundation, by the Veterans Administration VISN4 Mental Illness Research, Education and Clinical Center (David Oslin, MD, PI, USA) and by a grant from Philip and Marcia Cohen.

References

1 World Health Organization. *World Health Statistics 2006*. WHO Press: New York, NY, 2006; http://www.who.int/whosis.
2 Sullivan PF, Kendler KS. The genetic epidemiology of smoking. *Nicotine Tob Res* 1999; **1**: S51–S57.
3 Lessov CN, Martin NG, Statham DJ, Todorov AA, Slutske WS, Bucholz KK *et al.* Defining nicotine dependence for genetic research: evidence from Australian twins. *Psychol Med* 2004; **34**: 865–879.
4 Lessov-Schlaggar CN, Pang Z, Swan GE, Guo Q, Wang S, Cao W *et al.* Heritability of cigarette smoking and alcohol use in Chinese male twins: the Qingdao twin registry. *Int J Epidemiol* 2006; **35**: 1278–1285.
5 Maes HH, Sullivan PF, Bulik CM, Neale MC, Prescott CA, Eaves LJ *et al.* A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychol Med* 2004; **34**: 1251–1261.
6 Li MD. The genetics of nicotine dependence. *Curr Psychiatry Rep* 2006; **8**: 158–164.
7 Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF *et al.* Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet* 2007; **16**: 24–35.
8 Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA *et al.* Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet* 2007; **16**: 36–49.
9 Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire. *Br J Addict* 1991; **86**: 1119–1127.
10 Beuten J, Ma JZ, Payne TJ, Dupont RT, Crews KM, Somes G. Single- and multilocus allelic variants within the GABA(B) receptor subunit 2 (GABAB2) gene are significantly associated with nicotine dependence. *Am J Hum Genet* 2005; **76**: 859–864.

11 Grant BF, Kaplan K, Shepard J, Moore T. *Source and Accuracy Statement for Wave 1 of the 2001–2002. National Epidemiological Survey on Alcohol and Related Conditions*. National Institute on Alcohol Abuse and Alcoholism: Bethesda, MD, 2003.
12 CDC. Annual smoking-attributable mortality, years of potential life lost, and productivity losses-United States. *Morbidity & Mortality Weekly Report* 2005; **54**: 625–628.
13 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
14 Roses A, Burns D, Chissoe S, Middleton L, St Jean P. Disease-specific target selection: a critical first step down the right road. *Drug Disc Today* 2005; **10**: 177–186.
15 Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**: 906–913.
16 SoldÃ G, Boi S, Duga S, Fornasari D, Benfante R, Malcovati M *et al*. *In vivo* RNA-RNA duplexes from human alpha3 and alpha5 nicotinic receptor subunit mRNAs. *Gene* 2005; **345**: 155–164.
17 Gotti C, Moretti M, Gaimarri A, Zanardi A, Clementi F, Zoli M. Heterogeneity and complexity of native brain nicotinic receptors. *Biochem Pharmacol* 2007; **74**: 1102–1111.
18 Gerzanich V, Wang F, Kuryatov A, Lindstrom J. A5 subunit alters desensitization, pharmacology, Ca^{++} permeability and Ca^{++} modulation of human neuronal A3 nicotinic receptors. *J Pharmacol Exp Ther* 1998; **266**: 311–320.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)