Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs

Scott F. Saccone1,*,†, Anthony L. Hinrichs1,‡, Nancy L. Saccone1, Gary A. Chase3, Karel Konvicka4, Pamela A.F. Madden1, Naomi Breslau5, Eric O. Johnson6, Dorothy Hatsukami7, Ovide Pomerleau8, Gary E. Swan9, Alison M. Goate1,2, Joni Rutter10, Sarah Bertelsen1, Louis Fox1, Douglas Fugman11, Nicholas G. Martin12, Grant W. Montgomery12, Jen C. Wang1, Dennis G. Ballinger4, John P. Rice1,2 and Laura Jean Bierut1

1Department of Psychiatry and 2Department of Genetics, Box 8134, Washington University School of Medicine, 660 South Euclid Avenue, St Louis, MO 63110, USA, 3Department of Health Evaluation Sciences, Penn State College of Medicine, Hershey, PA 17033, USA, 4Perlegen Sciences, Mountain View, CA 94043, USA, 5Department of Epidemiology, Michigan State University, East Lansing, MI 48824, USA, 6Research Triangle Institute International, Research Triangle Park, NC 27709, USA, 7Department of Psychiatry, University of Minnesota, Minneapolis, MN 55454, USA, 8Department of Psychiatry, University of Michigan, Ann Arbor, MI 48109, USA, 9Center for Health Sciences, SRI International, Menlo Park, CA 94025, USA, 10National Institute on Drug Abuse, Bethesda, MD 20892, USA, 11Rutgers University Cell and DNA Repository, Rutgers University, Piscataway, NJ 08854, USA and 12Queensland Institute of Medical Research, Queensland 4029, Australia

Received August 11, 2006; Revised September 26, 2006; Accepted November 10, 2006

Nicotine dependence is one of the world’s leading causes of preventable death. To discover genetic variants that influence risk for nicotine dependence, we targeted over 300 candidate genes and analyzed 3713 single nucleotide polymorphisms (SNPs) in 1050 cases and 879 controls. The Fagerström test for nicotine dependence (FTND) was used to assess dependence, in which cases were required to have an FTND of 4 or more. The control criterion was strict: control subjects must have smoked at least 100 cigarettes in their lifetimes and had an FTND of 0 during the heaviest period of smoking. After correcting for multiple testing by controlling the false discovery rate, several cholinergic nicotinic receptor genes dominated the top signals. The strongest association was from an SNP representing CHRNA3, the β3 nicotinic receptor subunit gene (P = 9.4 × 10⁻³). Biologically, the most compelling evidence for a risk variant came from a non-synonymous SNP in the α5 nicotinic receptor subunit gene CHRNA5 (P = 6.4 × 10⁻⁴). This SNP exhibited evidence of a recessive mode of inheritance, resulting in individuals having a 2-fold increase in risk of developing nicotine dependence once exposed to cigarette smoking. Other genes among the top signals were KCNJ6 and GABRA4. This study represents one of the most powerful and extensive studies of nicotine dependence to date and has found novel risk loci that require confirmation by replication studies.

© The Author 2006. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org
INTRODUCTION

The World Health Organization estimates that if current trends continue, the annual number of deaths from tobacco-related diseases will double from five million in the year 2000 to 10 million in 2020 (1,2). Nicotine, a naturally occurring alkaloid found in tobacco, mimics acetylcholine, and nicotine’s ability to bind to nicotinic cholinergic receptors (nAChRs) underlies the molecular basis of nicotine dependence [susceptibility to tobacco addiction (MIM 188890)]. Chronic nicotine exposure produces long-lasting behavioral and physiological changes that include increased synaptic strength, altered gene expression and nAChR up-regulation (3). Although nAChRs are expressed throughout the central nervous system, the addictive effects of nicotine are thought to be mediated through mesocorticolimbic dopamine (DA) pathways (4). It is believed that the interplay among glutamate, dopamine and gamma-aminobutyric acid (GABA) systems is critical for the reinforcing effects of nicotine (3,5). Cigarettes are the predominant form of tobacco used worldwide (6), and genetic factors are important to the etiology of nicotine dependence, with estimates of the heritability ranging from 44 to 60% (7).

Efforts to identify susceptibility loci influencing cigarette smoking behavior through association studies have used a candidate gene approach with both case-control and family-based designs. Several candidate genes that may influence smoking have been studied, including nicotinic receptors (8–10), nicotine metabolizing genes (11–13), dopamine system receptors (14–17), GABA receptors (18) and other neurotransmitters and receptors (19–21). There appears to be very little concordance among linkage findings and association studies (19–21). It is believed that the interplay among glutamate, dopamine and GABA receptors, which are known to be part of the biological pathways involved in dependence. This was done in conjunction with a GWAS conducted in the companion paper by Bierut et al. (23) which was conducted in parallel with our study and used the same case-control sample.

Our approach was to target an extensive set of candidate genes for single nucleotide polymorphism (SNP) genotyping to detect variants associated with nicotine dependence using a case-control design. We targeted over 300 genes for genotyping, with a design that allowed for approximately 4000 SNPs. These included the gene families encoding nicotinic receptors, dopaminergic receptors and GABA receptors, which are known to be part of the biological pathways involved in dependence. This was done in conjunction with a GWAS conducted in the companion paper by Bierut et al. (23). Both studies used a large sample of cases and controls of European descent. The 1050 nicotine dependent cases were contrasted with a unique control sample of 879 individuals who are non-dependent smokers. The size of the sample and strict control criteria should provide ample power to detect variants influencing nicotine dependence, but the depth of the coverage of known candidate genes is ambitious and requires delicate handling to deal with the complex issue of multiple testing. We used the false discovery rate (FDR) to limit the effects of multiple testing (24,25) and to report on the top FDR-controlled list of associations.

RESULTS

Our list of candidate genes initially numbered 448 and was divided into categories ‘A’ and ‘B’. All 55 category A genes were targeted for SNP genotyping, but because it was beyond our resources to target all of the remaining 393 category B genes, these were prioritized for SNP genotyping according to the results of the pooled genotyping in the parallel GWAS (23). Table 1 shows a summary of the results of the pooled genotyping in the candidate genes. Out of the 393 category B genes considered for SNP selection, 296 were targeted for individual genotyping in our candidate gene study. These were chosen using the lowest corrected minimum P-values, as defined in Eq. (1), where the cutoff was approximately $P < 0.95$. We individually genotyped 4309 SNPs in these candidate genes, and after quality control filtering, 3713 SNPs were tested for association. There were 515 SNPs tested for 52 category A genes and 3198 SNPs tested for 296 category B genes.

In the individual genotyping for the candidate genes, the 10 smallest P-values from our primary association analysis ranged from $9.36 \times 10^{-5}$ to $1.22 \times 10^{-3}$. There were 39 SNPs with an FDR <40%, indicating the presence of about 24 true signals (Tables 2 and 3; Fig. 1). These top 39 signals were dominated by nicotinic receptor genes (Figs 2 and 3). The top five FDR values corresponded to the genes CHRNA3, CHRNA7 and CHRNA5 and ranged from 0.056 to 0.166. Our best evidence was that four of these five signals were from genuine associations and were not due to random effects. The permutation FDR estimates were roughly the same as the FDR, differing by not more than 0.02, with a minimum permutation FDR of 0.07 at the SNP rs6474413. After selecting a single SNP from each linkage disequilibrium (LD) bin, three of these 39 SNPs showed significant evidence of a non-multiplicative model (Table 4) and several SNPs were found to have a significant gender by genotype interaction (Table 5; also see Supplementary Material, Table S1 for a list of all SNPs from Table 2 showing gender by genotype P-values and gender-specific odds ratios).

The β3 nicotinic receptor subunit gene CHRNA3, located on chromosome 8, accounted for the two strongest signals from our analysis: rs6474413 and rs10958726 (Fig. 2A). These two SNPs effectively contributed to a single signal because they were in very high LD with an $r^2$ correlation $>0.99$. They are in the putative 5′ promoter region: the SNP rs6474413 is within 2 kb of the first 5′ promoter and the SNP rs10958726 is...
SNPs from category A genes were weighted 10-fold more heavily than those from category B genes when estimating FDR. The signals are sorted by the primary two-degree-of-freedom $P$-value of adding the genotype term and the genotype by gender interaction term to the base model in the logistic regression. SNPs with function ‘FP’ are within the footprint of the gene, defined for display purposes as ±10 kb of the transcribed region. Those labeled LD BIN are outside of the footprint and were selected for genotyping for being in LD with SNPs near an exon. Genes in parentheses are having no copies, the odds ratios for having one copy and two copies of Allele A of rs16969968 has a frequency of 38% in cases and 14% in controls, and the odds ratio for having two copies is 5.6 times greater. The most interesting signal appears to be the non-synonymous coding SNP in exon 5 of CHRNA5, the α5 nicotinic receptor subunit gene. This SNP was in very strong LD with rs1051730, a synonymous coding SNP in CHRNA3, with an $r^2$ correlation $\geq 0.99$. The most interesting signal appears to be the non-synonymous SNP rs16969968 in CHRNA5. As discussed earlier, it is completely correlated with an SNP in the CHRNA3 gene (Fig. 2B). Allele A of rs16969968 has a frequency of 38% in cases and 32% in controls. There is convincing evidence for a recessive mode of inheritance for this SNP (Table 4). Compared to having no copies, the odds ratios for having one copy and two copies of this SNP is 20 times greater.

### Table 2. Top associations with nicotine dependence where the weighted FDR is <0.4%

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Function</th>
<th>Category</th>
<th>Chr</th>
<th>Pos (bp)</th>
<th>LD Bin ID</th>
<th>Min ($r^2$)</th>
<th>Risk allele</th>
<th>Primary $P$-value</th>
<th>Rank</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs445597</td>
<td>CHRNB3</td>
<td>FP B</td>
<td>A 21</td>
<td>16</td>
<td>100,267,028</td>
<td>–</td>
<td>–</td>
<td>G (0.74/0.71)</td>
<td>2.95E-03</td>
<td>104</td>
<td>0.334</td>
</tr>
<tr>
<td>rs445600</td>
<td>CHRNB3</td>
<td>FP B</td>
<td>A 21</td>
<td>16</td>
<td>100,267,028</td>
<td>–</td>
<td>–</td>
<td>C (0.74/0.71)</td>
<td>4.49E-03</td>
<td>104</td>
<td>0.334</td>
</tr>
<tr>
<td>rs445602</td>
<td>CHRNB3</td>
<td>FP B</td>
<td>A 21</td>
<td>16</td>
<td>100,267,028</td>
<td>–</td>
<td>–</td>
<td>C (0.74/0.71)</td>
<td>4.49E-03</td>
<td>104</td>
<td>0.334</td>
</tr>
<tr>
<td>rs445603</td>
<td>CHRNB3</td>
<td>FP B</td>
<td>A 21</td>
<td>16</td>
<td>100,267,028</td>
<td>–</td>
<td>–</td>
<td>G (0.74/0.71)</td>
<td>2.95E-03</td>
<td>104</td>
<td>0.334</td>
</tr>
<tr>
<td>rs445604</td>
<td>CHRNB3</td>
<td>FP B</td>
<td>A 21</td>
<td>16</td>
<td>100,267,028</td>
<td>–</td>
<td>–</td>
<td>C (0.74/0.71)</td>
<td>4.49E-03</td>
<td>104</td>
<td>0.334</td>
</tr>
</tbody>
</table>

SNPs from category A genes were weighted 10-fold more heavily than those from category B genes when estimating FDR. The signals are sorted by the primary two-degree-of-freedom $P$-value of adding the genotype term and the genotype by gender interaction term to the base model in the logistic regression. SNPs with function ‘FP’ are within the footprint of the gene, defined for display purposes as ±10 kb of the transcribed region. Those labeled LD BIN are outside of the footprint and were selected for genotyping for being in LD with SNPs near an exon. Genes in parentheses are having no copies, the odds ratios for having one copy and two copies of Allele A of rs16969968 has a frequency of 38% in cases and 14% in controls, and the odds ratio for having two copies is 5.6 times greater. The most interesting signal appears to be the non-synonymous coding SNP in exon 5 of CHRNA5, the α5 nicotinic receptor subunit gene. This SNP was in very strong LD with rs1051730, a synonymous coding SNP in CHRNA3, with an $r^2$ correlation $\geq 0.99$. The most interesting signal appears to be the non-synonymous SNP rs16969968 in CHRNA5. As discussed earlier, it is completely correlated with an SNP in the CHRNA3 gene (Fig. 2B). Allele A of rs16969968 has a frequency of 38% in cases and 32% in controls. There is convincing evidence for a recessive mode of inheritance for this SNP (Table 4). Compared to having no copies, the odds ratios for having one copy and two copies of this SNP is 20 times greater.
Table 3. Details of all category A genes and any category B genes with SNPs among our top signals (i.e. SNPs that appear in Table 2)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr</th>
<th>5' position (Mb)</th>
<th>Size (kb)</th>
<th>Strand</th>
<th>SNPs tested</th>
<th>SNPs tested per kb</th>
<th>SNPs in top signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADRBK2</td>
<td>22</td>
<td>24.286</td>
<td>159</td>
<td>+</td>
<td>5</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>ANKK1</td>
<td>11</td>
<td>112.764</td>
<td>12.6</td>
<td>+</td>
<td>23</td>
<td>1.8</td>
<td>0</td>
</tr>
<tr>
<td>ARRB2</td>
<td>17</td>
<td>4.561</td>
<td>11.0</td>
<td>+</td>
<td>3</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>BDNF</td>
<td>11</td>
<td>27.700</td>
<td>66.8</td>
<td>-</td>
<td>10</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>CCK</td>
<td>3</td>
<td>42.281</td>
<td>6.9</td>
<td>-</td>
<td>13</td>
<td>1.9</td>
<td>0</td>
</tr>
<tr>
<td>CHRNA1</td>
<td>2</td>
<td>175.455</td>
<td>16.6</td>
<td>-</td>
<td>3</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>CHRNA10</td>
<td>11</td>
<td>3.649</td>
<td>5.8</td>
<td>-</td>
<td>3</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>CHRNA2</td>
<td>8</td>
<td>27.393</td>
<td>18.5</td>
<td>-</td>
<td>17</td>
<td>0.9</td>
<td>0</td>
</tr>
<tr>
<td>CHRNA3</td>
<td>15</td>
<td>76.700</td>
<td>25.7</td>
<td>-</td>
<td>18</td>
<td>0.7</td>
<td>5</td>
</tr>
<tr>
<td>CHRNA4</td>
<td>20</td>
<td>61.463</td>
<td>16.7</td>
<td>-</td>
<td>8</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>CHRNA5</td>
<td>15</td>
<td>76.645</td>
<td>28.6</td>
<td>+</td>
<td>18</td>
<td>0.6</td>
<td>6</td>
</tr>
<tr>
<td>CHRNA6</td>
<td>17</td>
<td>4.561</td>
<td>11.0</td>
<td>+</td>
<td>3</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>CHRNA7</td>
<td>15</td>
<td>30.11</td>
<td>138.5</td>
<td>+</td>
<td>13</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>CHRNA9</td>
<td>8</td>
<td>40.178</td>
<td>19.5</td>
<td>+</td>
<td>11</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td>CHRNA10</td>
<td>15</td>
<td>76.721</td>
<td>23.0</td>
<td>+</td>
<td>10</td>
<td>0.4</td>
<td>2</td>
</tr>
<tr>
<td>CHRNB1</td>
<td>17</td>
<td>7.289</td>
<td>12.5</td>
<td>+</td>
<td>10</td>
<td>0.8</td>
<td>0</td>
</tr>
<tr>
<td>CHRNB2</td>
<td>17</td>
<td>151.353</td>
<td>8.8</td>
<td>+</td>
<td>4</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>CHRNB3</td>
<td>17</td>
<td>3.649</td>
<td>5.8</td>
<td>-</td>
<td>3</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td>CHRNB4</td>
<td>17</td>
<td>42.783</td>
<td>6.0</td>
<td>+</td>
<td>6</td>
<td>1.0</td>
<td>4</td>
</tr>
<tr>
<td>CHRNK1</td>
<td>6</td>
<td>88.912</td>
<td>5.5</td>
<td>-</td>
<td>9</td>
<td>1.6</td>
<td>0</td>
</tr>
<tr>
<td>COMT</td>
<td>22</td>
<td>18.304</td>
<td>27.2</td>
<td>+</td>
<td>13</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>CYP2A6</td>
<td>19</td>
<td>46.048</td>
<td>6.9</td>
<td>-</td>
<td>3</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>CYP2B6</td>
<td>19</td>
<td>46.189</td>
<td>27.1</td>
<td>+</td>
<td>14</td>
<td>0.5</td>
<td>2</td>
</tr>
<tr>
<td>DBH</td>
<td>9</td>
<td>133.531</td>
<td>23.0</td>
<td>+</td>
<td>10</td>
<td>0.4</td>
<td>2</td>
</tr>
<tr>
<td>DDC</td>
<td>7</td>
<td>50.386</td>
<td>85.7</td>
<td>-</td>
<td>30</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>DRD1</td>
<td>5</td>
<td>174.804</td>
<td>3.1</td>
<td>-</td>
<td>4</td>
<td>1.3</td>
<td>0</td>
</tr>
<tr>
<td>DRD2</td>
<td>11</td>
<td>112.851</td>
<td>65.6</td>
<td>-</td>
<td>29</td>
<td>0.4</td>
<td>2</td>
</tr>
<tr>
<td>DRD3</td>
<td>3</td>
<td>115.38</td>
<td>50.2</td>
<td>-</td>
<td>8</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>DRD5</td>
<td>4</td>
<td>9.460</td>
<td>2.0</td>
<td>+</td>
<td>4</td>
<td>2.0</td>
<td>0</td>
</tr>
<tr>
<td>FAAH</td>
<td>1</td>
<td>46.572</td>
<td>19.5</td>
<td>+</td>
<td>5</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>FMO1</td>
<td>1</td>
<td>167.949</td>
<td>37.5</td>
<td>+</td>
<td>14</td>
<td>0.4</td>
<td>4</td>
</tr>
<tr>
<td>FMO3</td>
<td>1</td>
<td>167.792</td>
<td>26.9</td>
<td>+</td>
<td>23</td>
<td>0.9</td>
<td>0</td>
</tr>
<tr>
<td>GABRB2</td>
<td>5</td>
<td>160.908</td>
<td>254.3</td>
<td>-</td>
<td>14</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>GPR51</td>
<td>9</td>
<td>98.551</td>
<td>421.1</td>
<td>-</td>
<td>29</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>HTR1A</td>
<td>5</td>
<td>63.293</td>
<td>1.3</td>
<td>+</td>
<td>5</td>
<td>3.9</td>
<td>0</td>
</tr>
<tr>
<td>HTR2A</td>
<td>13</td>
<td>46.368</td>
<td>62.7</td>
<td>-</td>
<td>20</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>HTR5A</td>
<td>7</td>
<td>154.3</td>
<td>13.6</td>
<td>+</td>
<td>13</td>
<td>1.0</td>
<td>3</td>
</tr>
<tr>
<td>MAOA</td>
<td>23</td>
<td>43.272</td>
<td>90.7</td>
<td>+</td>
<td>5</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>MAOB</td>
<td>23</td>
<td>43.498</td>
<td>115.8</td>
<td>-</td>
<td>10</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>NPY</td>
<td>7</td>
<td>24.097</td>
<td>7.7</td>
<td>+</td>
<td>22</td>
<td>2.9</td>
<td>0</td>
</tr>
<tr>
<td>OPRD1</td>
<td>1</td>
<td>28.959</td>
<td>51.6</td>
<td>+</td>
<td>1</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>OPRK1</td>
<td>8</td>
<td>54.327</td>
<td>22.2</td>
<td>-</td>
<td>8</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>OPRM1</td>
<td>6</td>
<td>154.43</td>
<td>80.1</td>
<td>+</td>
<td>12</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>PDYN</td>
<td>20</td>
<td>1.923</td>
<td>15.3</td>
<td>-</td>
<td>11</td>
<td>0.7</td>
<td>1</td>
</tr>
<tr>
<td>PENK</td>
<td>8</td>
<td>57.521</td>
<td>5.1</td>
<td>-</td>
<td>6</td>
<td>1.2</td>
<td>0</td>
</tr>
<tr>
<td>POMC</td>
<td>2</td>
<td>25.303</td>
<td>7.7</td>
<td>-</td>
<td>2</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>SLC6A3</td>
<td>17</td>
<td>25.587</td>
<td>37.6</td>
<td>-</td>
<td>8</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>SLC6A4</td>
<td>7</td>
<td>25.178</td>
<td>9.1</td>
<td>+</td>
<td>3</td>
<td>0.2</td>
<td>1</td>
</tr>
<tr>
<td>TH</td>
<td>11</td>
<td>2.150</td>
<td>7.9</td>
<td>-</td>
<td>6</td>
<td>0.8</td>
<td>0</td>
</tr>
<tr>
<td>TPH1</td>
<td>11</td>
<td>18.019</td>
<td>19.8</td>
<td>-</td>
<td>14</td>
<td>0.7</td>
<td>0</td>
</tr>
</tbody>
</table>

Category B

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr</th>
<th>5' position (Mb)</th>
<th>Size (kb)</th>
<th>Strand</th>
<th>SNPs tested</th>
<th>SNPs tested per kb</th>
<th>SNPs in top signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVPR1A</td>
<td>12</td>
<td>61.83</td>
<td>6.4</td>
<td>-</td>
<td>15</td>
<td>2.4</td>
<td>1</td>
</tr>
<tr>
<td>CLTCL1</td>
<td>22</td>
<td>17.654</td>
<td>112.2</td>
<td>-</td>
<td>15</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>DAO</td>
<td>12</td>
<td>107.776</td>
<td>20.8</td>
<td>+</td>
<td>7</td>
<td>0.3</td>
<td>1</td>
</tr>
<tr>
<td>FMO4</td>
<td>1</td>
<td>168.015</td>
<td>27.7</td>
<td>+</td>
<td>12</td>
<td>0.4</td>
<td>4</td>
</tr>
<tr>
<td>GABRA4</td>
<td>4</td>
<td>46.837</td>
<td>74.7</td>
<td>-</td>
<td>29</td>
<td>0.4</td>
<td>2</td>
</tr>
<tr>
<td>GRM2</td>
<td>3</td>
<td>51.718</td>
<td>9.1</td>
<td>+</td>
<td>3</td>
<td>0.2</td>
<td>1</td>
</tr>
<tr>
<td>KCNJ6</td>
<td>21</td>
<td>38.211</td>
<td>291.9</td>
<td>-</td>
<td>18</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>PIP5K2A</td>
<td>10</td>
<td>23.043</td>
<td>177.7</td>
<td>-</td>
<td>15</td>
<td>0.1</td>
<td>1</td>
</tr>
</tbody>
</table>

The column ‘SNPs tested’ refers to the number of SNPs tested for association and the column ‘SNPs in top signals’ refers to the SNPs that appear in Table 2. Some SNPs represent multiple genes, particularly when two genes are near each other; hence there is overlap between genes for the SNPs represented by these two columns. Genes with SNPs in our top signals are shown in boldface.
Results of the candidate gene association analysis. The $P$-values from the primary analysis are plotted for each chromosome below an ideogram using the $-\log_{10}(P)$ transformation. The bottom axis is $P = 1$ and the top axis is $P = 10^{-3}$. Category A genes are shown below the plots in red and category B genes are shown in cyan below the category A genes. Regions on chromosomes 8 and 15, which are shown in more detail in Figure 2, are highlighted in red.
copies of the A allele were 1.1 (95% CI 0.9–1.4) and 1.9 (95% CI 1.4–2.6), respectively. That is, compared with individuals with other genotypes, individuals with the AA genotype were nearly twice as likely to have symptoms of nicotine dependence.

**DISCUSSION**

Nicotine addiction from tobacco smoking is responsible for over three million deaths annually, making it the leading cause of preventable mortality in the world (1). In the USA in 2003, 21.6% of adults were smokers, where 24% of men and 19% of women were smokers (26). Previous association studies have been limited to narrowly focussed candidate gene studies. Our candidate gene study was more extensive, genotyping 3713 SNPs for 348 candidates in 1050 nicotine-dependent cases and 879 non-dependent smokers, where our control group definition was particularly strict.

Our top FDR-controlled findings were dominated by nicotinic receptor genes. Our positive association findings for the \( \alpha_5 \) and \( \beta_3 \) nicotinic receptor subunits are novel. To date, most human genetic and biological studies of the nicotinic receptors and nicotine dependence have focussed on the \( \alpha_4 \) and \( \beta_2 \) subunits because they co-occur in high-affinity receptors and are widely expressed in the brain (27). However, mouse studies have demonstrated that of the \( \alpha_4 \beta_2 \) containing receptors that mediate dopamine release, a substantial proportion contain \( \alpha_5 \) as well (28). This is consistent with our evidence for an important role of \( \alpha_5 \) in nicotine dependence susceptibility. Furthermore, in a brain \( \alpha_4 \beta_2 \) receptor, an \( \alpha_5 \) or \( \beta_3 \) subunit can take the fifth position in the pentamer, corresponding to \( \beta_1 \) of muscle. Although neither \( \alpha_5 \) nor \( \beta_3 \) is thought to participate in forming binding sites, they are able to affect channel properties and influence agonist potency because they participate in the conformational changes associated with activation and desensitization (27).

The most compelling biological evidence of a risk factor for nicotine dependence is from the non-synonymous SNP rs16969968 in CHRNA5. This SNP causes a change in amino acid 398 from asparagine (encoded by the G allele) to aspartic acid (encoded by A, the risk allele), which results in a change in the charge of the amino acid in the second intracellular loop of the \( \alpha_5 \) subunit (29). The risk allele appeared to act in a recessive mode, in which individuals who were homozygous for the A allele are at a 2-fold risk to develop nicotine dependence. Although the \( \alpha_5 \) subunit has not been studied extensively and there are no reports of known functional effects of this polymorphism, it is striking that a non-synonymous charge-altering polymorphism in the corresponding intracellular loop of the \( \alpha_4 \) nAChR subunit has been shown to alter nAChR function.
in mice in response to nicotine exposure (30–33). This variant is common in the populations of European descent (allele frequency of A allele ~42%), but uncommon in populations of Asian or African descent (<5%, data from International HapMap project, http://www.hapmap.org).

Also among the top 39 FDR-controlled signals were the genes KCNJ6 (also known as GIRK2) and GABRA4. These were the only other genes besides nicotinic receptors with SNPs that had P-values less than 0.001. KCNJ6 belongs to the inwardly rectifying potassium channel (GIRK) family of...
genes. GIRK provides a common link between numerous neurotransmitter receptors and the regulation of synaptic transmission (34). GABA is the major inhibitory neurotransmitter in the mammalian central nervous system and is critical for the reinforcing effects of nicotine (3,5). We found significant evidence that the risk due to genotype is much stronger in men than in women (Table 5), where the male odds ratio was 2.2 (95% CI 1.4–3.3).

Previously reported findings in other nicotinic receptors were not among our most significant findings. In prior studies of CHRNA4, nominal association with nicotine dependence measures was reported for the SNPs rs2236196 and rs3787137 in African-American families and rs2273504 and rs1044396 in European-Americans, but only rs2236196 in African-Americans remained after multiple testing correction (9). Also in CHRNA4, rs1044396 and rs1044397 were associated with both Fagerström test for nicotine dependence (FTND) score and qualitative nicotine dependence in a family-based sample of Asian male smokers (8). In our sample of European descent, we tested 11 SNPs for CHRNA4 including the above-mentioned SNPs except rs2273504, which did not pass our stringent quality control standards. The lowest primary P-value across all 11 SNPs was 0.026 for rs2236196 (study-wide rank = 132); this particular result may be considered a single test given the specific prior finding for this SNP, thus providing modest evidence for replication. The remaining four previously reported SNPs that we analyzed showed P-values greater than 0.8. Contrasts in these results are possibly due in part to the different ethnicities of the respective samples.

A recent study of smoking initiation and severity of nicotine dependence in Israeli women (10) analyzed 39 SNPs in 11 nicotinic receptor subunit genes. Their single SNP analyses also did not detect association with SNPs in a4, including rs2236196, rs1044396 and rs1044397, although finding nominal significance in the a7, a9, b2 and b3 subunits. Their study did not include the same SNPs in the b3 subunit and a5–a3–b4 cluster comprising our four strongest associations in nicotinic receptor genes; they did analyze our fifth ranking nicotinic receptor SNP, rs1051730, and found a suggestive P-value of 0.08 when comparing ‘high’ nicotine-
Table 5. Gender-specific odds ratios and 95% confidence intervals for SNPs in Table 2

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Primary P-value</th>
<th>Rank</th>
<th>Gender * genotype P-value</th>
<th>Male odds ratio</th>
<th>Female odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10508649</td>
<td>PIP5K2A</td>
<td>1.02E-03</td>
<td>8</td>
<td>1.09E-02</td>
<td>9.7 (2.1–44.2)</td>
<td>1.0 (0.3–3.1)</td>
</tr>
<tr>
<td>rs17041074</td>
<td>DAO</td>
<td>1.12E-03</td>
<td>9</td>
<td>3.70E-04</td>
<td>0.8 (0.6–1.0)</td>
<td>1.3 (1.1–1.6)</td>
</tr>
<tr>
<td>rs3762967</td>
<td>GABRA4</td>
<td>1.22E-03</td>
<td>10</td>
<td>3.43E-02</td>
<td>2.2 (1.4–3.3)</td>
<td>1.2 (0.9–1.6)</td>
</tr>
<tr>
<td>rs6772197</td>
<td>DOC5</td>
<td>1.66E-03</td>
<td>12</td>
<td>6.35E-04</td>
<td>1.6 (1.2–2.2)</td>
<td>0.9 (0.7–1.1)</td>
</tr>
<tr>
<td>rs3021529</td>
<td>AVPR1A</td>
<td>1.73E-03</td>
<td>13</td>
<td>8.96E-04</td>
<td>0.8 (0.5–1.0)</td>
<td>1.5 (1.1–1.9)</td>
</tr>
<tr>
<td>rs6320</td>
<td>HTR5A</td>
<td>6.50E-03</td>
<td>39</td>
<td>1.61E-03</td>
<td>0.7 (0.6–1.0)</td>
<td>1.2 (1.0–1.5)</td>
</tr>
<tr>
<td>rs4802100</td>
<td>CYP2A7</td>
<td>6.76E-03</td>
<td>41</td>
<td>2.82E-02</td>
<td>0.9 (0.6–1.4)</td>
<td>1.6 (1.2–2.1)</td>
</tr>
<tr>
<td>rs4245150</td>
<td>DRD2</td>
<td>1.08E-02</td>
<td>61</td>
<td>2.79E-03</td>
<td>0.8 (0.6–1.0)</td>
<td>1.2 (1.0–1.4)</td>
</tr>
<tr>
<td>rs1657273</td>
<td>HTR5A</td>
<td>1.11E-02</td>
<td>64</td>
<td>3.06E-03</td>
<td>0.8 (0.6–1.0)</td>
<td>1.2 (1.0–1.5)</td>
</tr>
<tr>
<td>rs6045733</td>
<td>PDYN</td>
<td>1.55E-02</td>
<td>84</td>
<td>4.25E-03</td>
<td>1.3 (1.1–1.7)</td>
<td>0.9 (0.7–1.0)</td>
</tr>
</tbody>
</table>

Only SNPs where the gender by genotype interaction was significant ($P < 0.05$) are shown, and the SNP with the most significant primary $P$-value was selected from each LD bin. The odds ratios are based on the coefficient of the genotype term $G$ in Eq. (2) and represent the increase in risk for every unit increase in $G$; i.e. the risk follows a log-linear model (Tables 8 and 9).

dependent subjects with ‘low’ nicotine-dependent subjects in a much smaller sample than ours.

Our study was unable to corroborate reported association findings of Beuten et al. (18) for the β2 subunit of the GABA<sub>B</sub> receptor GABBR2 (also known as GABAB<sub>2</sub>, GABAB2 and GPR51). We genotyped 32 SNPs in GABBR2 including five SNPs reported by Beuten et al. (18), three of which were the most significant in European-Americans by at least one test in that study. The primary $P$-value in our study was greater than 0.07 for all 32 SNPs and greater than 0.3 for the five previously reported SNPs.

Similarly, we do not find evidence for nominal association in our primary test of the 31 SNPs which we genotyped for the DDC gene, which includes an SNP previously reported significant in European-Americans (35). And of the 11 SNPs covering the gene BDNF, three (rs60265, rs2030324 and rs7934165) were previously reported as associated in European-American males (21); these three were not significant in our sample (primary $P = 0.86$, 0.088 and 0.12, respectively), and the lowest primary $P$-value among the remaining eight SNPs was 0.02, which does not survive correction for the six LD bins covering the gene. Note that our primary test uses a log-additive model, whereas previous reports sometimes found their strongest results under other models (e.g. recessive and dominant); however, for these previously reported associations, our tests for departure from the log-additive model did not find evidence for improvement under alternative modes of inheritance.

Our primary association analysis was a two-degree-of-freedom test of the significance of adding genotype and genotype by gender interaction terms to the base predictors sex and site. This approach helps to ensure that we detect associations that are significantly influenced by gender. The disadvantage is that the extra degree of freedom makes associations with significant gender interaction appears to be less significant overall.

Because our controls were highly selected and could even be considered ‘protected’ against susceptibility to nicotine dependence, interpretation of our results must consider the possibility that an association signal from our study may actually represent protective rather than risk effects. We used the allele more frequent in cases for reporting these data as a convention to facilitate comparison of the odds ratios among SNPs; this should not be viewed as a conclusion of how a particular variant influences the risk for nicotine dependence. The precise determination of the mechanism by which a variant alters risk can only come from functional studies.

We performed additional tests for association using only the individuals from the US sample to determine whether our primary conclusions still hold in this subset of 797 cases and 813 controls (the Australian sample alone is too small to test for association, with only 253 cases and 66 controls). We used the same logistic regression method as for the entire sample except for the omission of the term ‘site’. The Spearman rank-order correlation of the $P$-values between the two tests for association was 0.87. Supplementary Material, Table S2 shows the results of the US-only analysis for the 39 SNPs from our list of top associations (Table 2), with the original ordering and FDR filtering, side by side with results from the US sample. Supplementary Material, Table S3 describes the result of completely starting over and using only the US sample to order by $P$-value, filter by FDR <40% and compute LD bins. In this case, 30 of 39 (77%) SNPs in our original set of top signals (Table 2) appeared in the list of top signals in the US-only analysis (Supplementary Material, Table S3), which includes the genes CHRNA5 and CHRNA3, the top genes from our initial analysis. Hence, although there were some changes in the order of the results, the primary conclusion of association with the nicotinic receptors CHRNA3 and CHRNA5 remains valid when the analysis is performed on the US subsample.

As a companion to the candidate gene study, a GWAS was carried out in parallel (23). Approximately 2.4 million SNPs were genotyped across the human genome in a two-stage design that began with pooled genotyping in a portion of the sample and followed with individual genotyping of the entire sample for the top 40 000 signals. The 21st strongest signal from the GWAS was due to an SNP 3 kb upstream of the first S′ promoter of CHRNA3, the gene with the strongest signal from our candidate gene study. This signal came from the SNP rs13277254 (genotyped only for the GWAS and not for our candidate gene study) and had a $P$-value of 6.52 × 10<sup>-5</sup>. This convergence from two different study designs provides further support that the signals in this gene are not random effects.
In conclusion, we have identified several genetic variants as being associated with nicotine dependence in candidate genes, the majority of which are nicotinic receptor genes. One of the SNPs implicated has a number of biologically relevant consequences, making it a particularly plausible candidate for influencing smoking behavior. These variants should be considered potential sources of genetic risk. Additional research is required to establish replication and possibly its role in the pharmacogenetics of response to nicotine dosing as well as to treatments for nicotine dependence.

MATERIALS AND METHODS

Subjects

All subjects (Table 6) were selected from two ongoing studies. The Collaborative Genetic Study of Nicotine Dependence (US) recruited subjects from three urban areas in the USA and the Nicotine Addiction Genetics (Australian) study collected subjects of European ancestry from Australia. Both studies used community-based recruitment and equivalent assessments were performed. Subjects who were identified as being smokers, using the criteria that they had smoked 100 or more cigarettes in their lifetimes, were queried in more detail using the FTND questionnaire. The US samples were enrolled at sites in St Louis, Detroit and Minneapolis, where a telephone screening of community-based subjects was used to determine whether subjects met criteria for case (current FTND ≥ 4) or control status. The study participants for the Australian sample were enrolled at the Queensland Institute of Medical Research in Australia, where families were identified from two cohorts of the Australian twin panel, which included spouses of the older of these two cohorts, for a total of approximately 12 500 families with information about smoking. The ancestry of the Australian samples is predominantly Anglo-Celtic and Northern European. The Institutional Review Boards approved both studies, and all subjects provided informed consent to participate. Blood samples were collected from each subject for DNA analysis and submitted, together with electronic phenotypic information about smoking. The ancestry of the Australian samples is predominantly Anglo-Celtic and Northern European.

Case subjects were required to score 4 or more on the FTND (36) during the heaviest period of cigarette smoking (the largest possible score is 10). This is a common criterion for defining nicotine dependence. Control subjects must have smoked 100 or more cigarettes in their lifetimes, yet never exhibited symptoms of nicotine dependence: they were smokers who scored 0 on the FTND during the heaviest period of smoking. By selecting controls that had a significant history of smoking, the genetic effects that are specific to nicotine dependence can be examined. Additional data from the Australian twin panel support this designation of a control status (23). In the US study, using the sample of 15 086 subjects who were determined to be smokers (smoked 100 or more cigarettes in their lifetimes) during the screening process, the prevalence of ‘nicotine dependence’ (FTND ≥ 4) was 46.4% and the prevalence of ‘smoking without nicotine dependence’ (FTND = 0) was 20.1%.

Candidate gene selection

The criteria for the selection of the candidate genes were based on known biology, correlations between nicotine dependence and other phenotypes and previous reports on the genetics of nicotine dependence and related traits. Genes were nominated by an expert committee of investigators from the NIDA Genetics Consortium (http://zork.wustl.edu/nida), with expertise in the study of nicotine and other substance dependence. These included classic genes that respond to nicotine, such as the nicotinic receptors, and other genes involved in the addictive process.

In total, 448 genes were considered for SNP genotyping. The genes were divided into two categories: A and B. Category A genes, which included the nicotinic and dopaminergic receptors, were considered to have a higher prior probability of association and were guaranteed to be targeted for genotyping. As our study design allowed for individual genotyping of approximately 4000 SNPs, the category B genes were too numerous to receive adequate SNP coverage once the A genes had been sufficiently covered. We therefore prioritized the category B genes using the results of the pooled genotyping from the companion GWAS study (23). Genes exhibiting the most evidence for association with nicotine dependence were prioritized for coverage. Some genes are larger than others and, therefore, may receive more SNPs. These genes may therefore appear more significant because of the increased number of tests performed. Hence, we corrected for multiple testing as follows. For a given candidate gene on the B list, if \( P_{\text{min}} \) is the minimum \( P \)-value found in the pooled genotyping of stage I of the GWAS for all the SNPs genotyped in the gene and \( N \) is the number of SNPs tested, then we computed the corrected minimum \( P \)-value.
by guest
Downloaded from https://academic.oup.com/hmg/article-abstract/16/1/36/2355934

Pcorr using the formula

\[ P_{corr} = 1 - (1 - P_{min})^{(N+1)/2} \]  

(1)

As roughly 50% of the SNPs in any chromosomal region are in high LD (37), we used \((N + 1)/2\) as the exponent. The category B genes were then ranked by these corrected minimum \(P\)-values and SNPs were selected from the top of the ranked list until our resources were exhausted.

### SNP selection

We chose all SNPs within exons, regardless of the allele frequency, and all SNPs within \(\pm 2\) kb of annotated gene promoters where the European-American minor allele frequency was at least 4%. We then chose tag SNPs for all European-American LD bins (38) crossing the exons of the candidate genes, with two SNPs for each bin with three or more SNPs. SNPs meeting these criteria were chosen first from those selected for individual genotyping in the companion pooled study (23) and then to cover the physical regions as uniformly as possible if there was choice available for the other SNPs. In addition, we included specific SNPs that have been reported in the literature as being associated with nicotine dependence (8,9,18,34).

### Pooled genotyping

See the companion paper by Bierut et al. (23) for a description of the pooled genotyping.

### Individual genotyping

For individual genotyping, we designed custom high-density oligonucleotide arrays to interrogate SNPs selected from candidate genes, as well as quality control SNPs. Each SNP was interrogated by 24 25mer oligonucleotide probes synthesized on a glass substrate. The 24 features comprise four sets of six features interrogating the neighborhoods of SNP reference and alternate alleles on forward and reference strands. Each allele and strand is represented by five offsets: \(-2, \,-1, 0, 1\) and 2, indicating the position of the SNP within the 25mer, with 0 being at the 13th base. At offset 0, a quartet was tiled, which includes the perfect match to reference and alternate SNP alleles and the two remaining nucleotides as mismatch probes. When possible, the mismatch features were selected as purine nucleotide substitution for a purine perfect match nucleotide and as a pyrimidine nucleotide substitution for a pyrimidine perfect match nucleotide. Thus, each strand and allele tiling consisted of six features comprising five perfect match probes and one mismatch.

### Individual genotype cleaning

Individual genotypes were cleaned using a supervised prediction algorithm for the genotyping quality, compiled from 15 input metrics that describe the quality of the SNP and the genotype. The genotyping quality metric correlates with a probability of having a discordant call between the Perlegen platform and outside genotyping platforms (i.e. non-Perlegen HapMap project genotypes). A system of 10 bootstrap aggregated regression trees was trained using an independent data set of concordance data between Perlegen genotypes and HapMap project genotypes. The trained predictor was then used to predict the genotyping quality for each of the genotypes in this data set (see Supplementary Material for more information regarding cleaning).

### Population stratification analysis

In order to avoid false positives due to population stratification, we performed an analysis using the STRUCTURE software (39). This program identifies subpopulations of individuals who are genetically similar through a Markov chain Monte Carlo sampling procedure using markers selected across the genome. Genotype data for 289 high performance SNPs were analyzed across all 1929 samples. This analysis revealed no evidence for population admixture.

### Genetic association analysis

An ANOVA analysis testing the predictive power of various phenotypes indicated that gender and site (USA or Australia) were the most informative and that age and other demographic variables did not account for significant additional trait variance (Table 7). Our primary method of analysis was based on a logistic regression: if \(P\) is the probability of being a case, then our linear logistic model has the form

\[ \log\left(\frac{P}{1-P}\right) = \alpha + \beta_1 g + \beta_2 s + \beta_3 G + \beta_4 gG \]  

(2)

where \(\alpha\) is the intercept, \(g\) the gender coded 0 or 1 for males or females, respectively, and \(s\) the site coded as 0 or 1 for USA or Australia, respectively. The variable \(G\) represents genotype and is coded as the number of copies of the risk allele, defined as the allele more common in cases than in controls. It follows from Eq. (2) that the risk due to genotype is being modeled using a log-linear (i.e. multiplicative) scale rather than an additive scale. Maximum likelihood estimates for the coefficients and confidence intervals for odds ratios were computed using the SAS software package (40).

The predictors of our base model were gender and site. We then tested whether the addition of genotype and gender by genotype interaction to the base model significantly increased the predictive power and used the resulting two-degree-of-freedom \(\chi^2\) statistic to rank the SNPs by the corresponding \(P\)-values. Table 8 shows the formulas for the odds ratios in terms of the coefficients.

Following these primary analyses, we further analyzed the top ranked SNPs for significant evidence of dominant or recessive mode of inheritance. This was done using a logistic regression of the form

\[ \log\left(\frac{P}{1-P}\right) = \alpha + \beta_1 g + \beta_2 s + \beta_3 G + \beta_4 gH \]  

(3)

where \(H\) is 1 for heterozygotes and 0 otherwise. When \(H\) is significant, the interpretation is that the genetic effect deviates signifi-
inclusion in our list of FDR-filtered top signals. For each category B genes must have stronger association signals for SNPs a moderate 10-fold more heavily. Therefore, the prior probability of association, we followed the recommendations of Roeder et al. (43) and weighted category A gene SNPs a moderate 10-fold more heavily. Therefore, the category B genes must have stronger association signals for inclusion in our list of FDR-filtered top signals. For each

<table>
<thead>
<tr>
<th>Model</th>
<th>ANOVA evaluated covariate</th>
<th>(\chi^2) (1 df)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Gender</td>
<td>40.0</td>
<td>(4.2 \times 10^{-10})</td>
</tr>
<tr>
<td>Gender + age</td>
<td>Age</td>
<td>10.3</td>
<td>(1.3 \times 10^{-03})</td>
</tr>
<tr>
<td>Gender + site</td>
<td>Site</td>
<td>100.4</td>
<td>(1.2 \times 10^{-23})</td>
</tr>
<tr>
<td>Gender + site + age</td>
<td>Age</td>
<td>0.25</td>
<td>0.62</td>
</tr>
<tr>
<td>Gender + site + gender*site</td>
<td>Gender*site</td>
<td>0.84</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Logistic regression, modeling the probability of being a case, was performed for the indicated covariates. The \(\chi^2\) statistic is from the formula \(-2(\Delta \log \mathcal{L})\), where \(\Delta \log \mathcal{L}\) is the change in likelihood in the logistic regression. The variable site has two levels: USA and Australia.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>g</th>
<th>G</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0</td>
<td>0</td>
<td>(-)</td>
</tr>
<tr>
<td>AA</td>
<td>0</td>
<td>1</td>
<td>(e^{b_1})</td>
</tr>
<tr>
<td>aa</td>
<td>0</td>
<td>2</td>
<td>(e^{b_2})</td>
</tr>
<tr>
<td>AA</td>
<td>1</td>
<td>0</td>
<td>(-)</td>
</tr>
<tr>
<td>aa</td>
<td>1</td>
<td>1</td>
<td>(e^{b_3}) (e^{b_4})</td>
</tr>
<tr>
<td>aa</td>
<td>1</td>
<td>2</td>
<td>(e^{b_2}) (e^{b_3})</td>
</tr>
</tbody>
</table>

The allele a is the risk allele, the allele more common in cases than in controls. The variable G is defined as the number of copies of the risk allele, and g is 0 or 1 for male or female, respectively. The last column shows the expression for the gender-specific odds ratio for a given genotype compared with the AA genotype, which follows directly from the logistic regression model in Eq. (2).

## Linkage disequilibrium

We estimated \(r^2\) correlation separately in cases and controls for all pairs of SNPs within 1 Mb windows using an EM algorithm as implemented in the computer program Haplovie (version 3.2, http://www.broad.mit.edu/mpg/haplovie) (41). Our final measure of LD is the minimum \(r^2\) from the two samples. Following the algorithm in Hinds et al. (38) and Carlson et al. (42), SNPs were grouped into bins, where every bin contains at least one ‘tag SNP’ satisfying \(\min(r^2) \geq 0.8\) with every SNP in the bin. The group of association signals from such an LD bin can be viewed essentially as a single signal.

## Correcting for multiple testing

To account for multiple testing, we estimated the FDR (24,25) to control the proportion of false positives among our reported signals. As category A genes were considered to have a higher prior probability of association, we followed the recommendations of Roeder et al. (43) and weighted category A gene SNPs a moderate 10-fold more heavily. Therefore, the category B genes must have stronger association signals for inclusion in our list of FDR-filtered top signals. For each

\[
P_w = \left\{ \begin{array}{ll}
wP & \text{category A genes} \\
10wP & \text{category B genes} \\
\end{array} \right.
\]

The odds ratios follow directly from Eq. (3). Note that for a dominant model, the two odds ratios are equal, and for a recessive model, the odds ratio for \(aa\) is 1.

\[P\text{-value}, \text{ we computed a weighted } P\text{-value } P_w \text{ using the formula}\]

where \(w\) was defined so that the average of the weights is 1 (this depends on the number of SNPs selected for A and B genes). For every weighted \(P\text{-value } P_w\), we computed a \(q\text{-value } q_w\) that has the property that the FDR is no greater than \(q_w\) among all SNPs with \(q_w < q_w\) (25,44). This was done using the computer program QVALUE (version 1.1, http://faculty.washington.edu/jstorey/qvalue) (45). Our estimates of the FDR are based on the \(q\)-values.

This method of estimating the FDR does not take into account LD. Therefore, as an additional measure to correct for multiple testing and to assess statistical significance, we estimated the FDR using permutations and \(P\text{-values weighted for A and B genes, which preserves the LD structure. This was done by performing 1000 random permutations of the case–control status and testing the permuted data for association. The significance of a } P\text{-value from the original data was assessed by counting the number of times a more significant weighted } P\text{-value occurs in the random permutations, where the weights were the same as those used for the FDR estimates.}\]

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

## ACKNOWLEDGEMENTS

The authors wish to acknowledge the contributions of advisors to this project. The NIDA Genetics Consortium and two NCId committees were vital to the success of the research. The Candidate Gene Committee helped review and finalize the list of candidate genes to be genotyped with individual SNP genotyping. In addition to the authors, committee members included Andrew Bergen, Joseph Cubells, Ken Krauter, Mary Jeanne Kreek, Sharon Murphy, Huijun Ring, Ming Tsuang and Kirk Wilhelmsen. The Data Analysis Committee helped oversee analyses for the candidate gene and genome-wide association studies and investigated methodological issues in association analyses. Further, the committee assisted in data management and data sharing functions. In addition to the authors, committee members included Andrew Bergen, Gerald Dunn, Mary Jeanne Kreek, Huijun Ring, Lei Yu and Hongyu Zhao. At Perlegen
Sciences, we would like to acknowledge the work of Laura Stuve, Curtis Kautzer, the genotyping laboratory, Laura Kami- 
gaki, the sample group and John Blanchard, Geoff Nilsen and 
the bioinformatics and data quality groups for excellent techni-
cal and infrastructural support for this work performed under 
NIDA Contract HHSN271200477741C. Figures 1 and 2 were 
generated with the Generic Genome Browser (version 1.64, 
http://www.gmod.org) (46). In memory of Theodore Reich, 
founding Principal Investigator of COGEND, we are indebted 
to his leadership in the establishment and nurturing of 
COGEND and acknowledge with great admiration his 
seminal scientific contributions to the field. This work was 
supported by the NIH grants CA89392 from the National 
Cancer Institute, DA12854 and DA015129 from the 
National Institute on Drug Abuse and the contract N01DA-0-7079 
from NIDA.

Data access: Phenotypes and genotypes are available 
through the NIDA Genetics consortium (http://nidagenetics. 
org) to the scientific community at time of publication.

Conflict of Interest statement. D.G.B. and K.K. are employed 
by Perlegen Sciences, Inc. With the exception of D.G.B. 
and K.K., none of the authors or their immediate families 
are currently involved with, or have been involved with, 
any companies, trade associations, unions, litigants or 
other groups with a direct financial interest in the subject 
matter or materials discussed in this manuscript in the past 
5 years.

REFERENCES


Global Tobacco Surveillance System (GTSS) collaborative group. 
Patterns of global tobacco use in young people and implications for 

acetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madras, B.K., 
von Zastrow, M. (eds), Pharmacology of two classes of striatal presynaptic 
icetylcholine receptors and nicotine dependence. In Madra...


