Original investigation

Genome-Wide Meta-Analyses of FTND and TTFC Phenotypes

Jingchun Chen MD, PhD1, Anu Loukola PhD2,3, Nathan A. Gillespie PhD4, Roseann Peterson PhD4, Peilin Jia PhD5, Brien Riley PhD4, Hermine Maes PhD4, Daniella M. Dick PhD6, Kenneth S. Kendler MD4, M. Imad Damaj PhD7, Michael F. Miles MD, PhD7, Zhongming Zhao PhD5, Ming D. Li PhD8, Jacqueline M. Vink PhD8,10, Camelia C. Minica PhD8,11,12, Gonneke Willemsen PhD9,11,12, Dorret I. Boomsma PhD9,11,12, Beenish Qaiser PhD2,3, Pamela A. F. Madden MD13, Tellervo Korhonen PhD2,14, Pekka Jousilahti PhD15, Jenni Hälfors PhD3, Joel Gelernter MD16, Henry R. Kranzler MD17, Richard Sherva PhD18, Lindsay Farrer PhD18, Brion Maher PhD19, Michael Vanyukov PhD20, Michelle Taylor PhD21, Jenifer J. Ware PhD21, Marcus R. Munafò PhD21, Sharon M. Lutz PhD22, John E. Hokanson PhD22, Fangyi Gu PhD23, Maria T. Landi MD, PhD23, Neil E. Caporaso PhD23, Dana B. Hancock PhD24, Nathan C. Gaddis PhD25, Timothy B. Baker PhD26, Laura J. Bierut MD13, Eric O. Johnson PhD24,27, Meghan Chenoweth PhD28, Caryn Lerman PhD29, Rachel Tyndale PhD28, Jaakko Kaprio MD, PhD2,3, Xiangning Chen PhD1,4,30

1Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV; 2Department of Public Health, University of Helsinki, Helsinki, FI-00014, Finland; 3Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 4Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA; 5School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX; 6Department of Psychology, Virginia Commonwealth University, Richmond, VA; 7Department of Pharmacology and Toxicology, Virginia Commonwealth University, Richmond, VA; 8State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China; 9Netherlands Twin Register, Department of Biological Psychology, VU University, the Netherlands; 10Behavioural Science Institute, Radboud University, Nijmegen, the Netherlands; 11Neuroscience Campus Amsterdam, the Netherlands; 12EMGO+ Institute for Health and Care Research, VU Medical Center, Amsterdam, the Netherlands; 13Department of Psychiatry, Washington University, St. Louis, MO; 14Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Finland; 15National Institute for Health and Welfare, Helsinki, Finland; 16Department of Psychiatry, Yale University, New Haven, CT; 17Department of Psychiatry, University of Pennsylvania, Philadelphia, PA; 18Section of Biomedical Genetics, Department of Medicine, Boston University School of Medicine, Boston, MA; 19Department of Mental Health, Johns Hopkins University, Baltimore, MD; 20Department of Pharmaceutical Sciences, University of Pittsburgh, Pittsburgh, PA; 21MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, Bristol, BS8 2BN, UK; 22Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO; 23Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, United States Department of Health and Human Services, Bethesda, MD; 24Behavioral Health and Criminal Justice Division, RTI International, Research Triangle Park, NC; 25Research Computing Division, RTI
Abstract

Introduction: FTND (Fagerström test for nicotine dependence) and TTFC (time to smoke first cigarette in the morning) are common measures of nicotine dependence (ND). However, genome-wide meta-analysis for these phenotypes has not been reported.

Methods: Genome-wide meta-analyses for FTND ($N = 19,431$) and TTFC ($N = 18,567$) phenotypes were conducted for adult smokers of European ancestry from 14 independent cohorts.

Results: We found that $SORBS2$ on 4q35 ($p = 4.05 \times 10^{-9}$), $BG182718$ on 11q22 ($p = 1.02 \times 10^{-4}$), and $AA333164$ on 14q21 ($p = 4.11 \times 10^{-9}$) were associated with TTFC phenotype. We attempted replication of leading candidates with independent samples (FTND, $N = 7010$ and TTFC, $N = 10,061$), however, due to limited power of the replication samples, the replication of these new loci did not reach significance. In gene-based analyses, $COPB2$ was found associated with FTND phenotype, and $TFC2PL1$, $RELN$, and $INO80C$ were associated with TTFC phenotype. In pathway and network analyses, we found that the interconnected interactions among the endocytosis, regulation of actin cytoskeleton, axon guidance, MAPK signaling, and chemokine signaling pathways were involved in ND.

Conclusions: Our analyses identified several promising candidates for both FTND and TTFC phenotypes, and further verification of these candidates was necessary. Candidates supported by both FTND and TTFC ($CHRNA4$, $THSD7B$, $RBFOX1$, and $ZNF804A$) were associated with addiction to alcohol, cocaine, and heroin, and were associated with autism and schizophrenia. We also identified novel pathways involved in cigarette smoking. The pathway interactions highlighted the importance of receptor recycling and internalization in ND.

Implications: Understanding the genetic architecture of cigarette smoking and ND is critical to develop effective prevention and treatment. Our study identified novel candidates and biological pathways involved in FTND and TTFC phenotypes, and this will facilitate further investigation of these candidates and pathways.

Introduction

Cigarette smoking imposes a high health and financial toll on the smokers as well as society at large. Regular cigarette smoking often leads to nicotine dependence (ND). The tendency to develop ND is influenced by both genetic predisposition and environmental factors. In recent years, genetic studies of ND have made significant progress, exemplified by the identification of the $CHRNA5-CHRNA3-CHRNB4$, $CHRNA3-CHRNB3$, $CHRNB6$, $CHRNA4$, and $CYP2A6$ loci. However, variants identified in these genes explain only a small proportion of heritability. For example, cigarettes smoked per day (CPD), a common measure used in ND studies, explained about 5.6% of the heritability. Many more risk genes remain unidentified.

Cigarette smoking is a complex behavior, and different measures are used to assess ND in both clinical and research settings. The quantity of consumption, typically assessed by self-reported CPD, is a measure used in many studies, including those that identified the $CHRNA5-CHRNA3-CHRNB4$, $CHRNA3-CHRNB6$, and $CYP2A6$ loci. The Fagerström Test for Nicotine Dependence (FTND) is another commonly used measure in genetic studies. CPD is one of the six questions included in the FTND and accounts up to three points in the 10-point FTND scale. Another FTND question, “How soon after you wake up do you smoke your first cigarette?” or time to smoke first cigarette in the morning (TTFC), is a strong predictor of difficulty to quit smoking and smoking relapse, and lung cancer. Although both FTND and TTFC are important measures of ND, comparing with CPD, only a few genome-wide association studies (GWASs) have been conducted using these phenotypes. $CHRNA4$ and a few intergenic loci were found to be associated with FTND, but no locus for TTFC was identified.

Here we report results from our GWAS meta-analyses on FTND ($N = 19,431$) and TTFC ($N = 18,567$). Our motivation to analyze TTFC separately is to understand what genetic factors contributing to the difficulty of quitting and relapse. Another benefit is that we can compare the results between FTND and TTFC phenotypes and identify convergent candidates. A lesson learned from comprehensive studies of the $CHRNA5-CHRNA3-CHRNB4$ locus is that a true signal can be detected with many correlated phenotypes. Hence, a candidate with corroborative support from both FTND and TTFC would be more credible.
Methods

Datasets
We assembled 14 independent cohorts (FTND, N = 19 431; TTFC, N = 18 567) to examine the association between FTND/TTFC phenotypes and genome-wide genetic variants. A description of the cohorts and the summary of descriptive statistics of the cohorts are provided in Supplementary Material and Supplementary Table S1, respectively.

Genotype Quality Control and Imputation
Genotype quality controls were conducted separately by individual groups for the discovery data sets ISIB, ALSPAC, AUTW, CEDAR, FTC1, FTC2, NTR, and GCD. For discovery data sets obtained from dbGaP (http://www.ncbi.nlm.nih.gov/gap) (SAGE, SC, MGS, COPDGene1, CIDR370v1, CIDR370v3, CIDRc3, CIDRc4, and EAGLE), we used the genotypes provided by the original investigators who conducted quality control procedures following the dbGaP standards. Genotype imputations were also conducted separately by each group using MaCH17,18 or IMPUTE219,20 with the 1000 Genomes reference haplotypes (EUR panel, March 2012 release), using the default settings of the programs. Similarly, genotype quality control for the replication data sets (VTSABD, COPDGene2, 54S, PNAT2, and EAGLE [used for TTFC only]) was conducted by each group separately.

Inclusion Criteria and Phenotypes
Regular smokers, defined as those who smoked daily for at least 1 month or those who smoked at least 100 cigarettes lifetime, of European ancestry were included. FTND scores (0–10) and TTFC scores (0–3) were obtained from self-reported FTND questionnaire, ascertaining the smoking behaviors during the heaviest smoking period of the smokers. Both FTND and TTFC scores were treated as quantitative traits without transformation.

Association and Meta-Analyses
GWAS analyses were performed separately for each data set by individual groups using the PLINK program.21 Assuming a linear mixed effect model, FTND and TTFC were treated as continuous outcomes and genotypes as predictors, whereas sex, age, and the first 10 principal components were included as covariates. Summary statistics from each data set were combined by GWAMA23 program using inverse variance-weighted meta-analysis approach with fixed effects. Because the individual samples were analyzed using the same model, the summary statistics were used directly without further normalization.

For the meta-analyses, only bi-allelic single nucleotide polymorphisms (SNPs) with minor allele frequency ≥1% and with high imputation quality (INFO value ≥ 0.4 from IMPUTE2, or r² ≥ 0.4 from MaCH) were included. p-Values from the meta-analyses were corrected for genomic-control. p-Values below 5 × 10⁻⁸ were considered as genome-wide significant, whereas p-values below 5 × 10⁻⁵ were considered as suggestive association and this threshold was used for selection of loci for convergence analysis.

SNP Heritability Analyses
To estimate the SNP heritability with genome-wide data, we used the linkage disequilibrium (LD)-adjusted kinship algorithm.22,23 Specifically, we used the control subjects from two large studies23,24 as reference to select LD-adjusted SNP predictors and used the meta-analysis results from the FTND and TTFC to estimate the weights for these SNPs. The heritability was then estimated with these LD-tagged SNPs. The number of SNPs used for FTND heritability estimate was 2 857 113 and that for TTFC was 2 981 471.

Replication Analyses
Six independent data sets of European descent (N = 7010 for FTND and N = 10 061 for TTFC) were used for a replication study of selected SNPs. We conducted replication study for all loci meeting these four criteria: (1) having five or more SNPs with p-value ≤ 5 × 10⁻⁸; (2) at least one SNP with p-value ≤ 5 × 10⁻⁵ having a minor allele frequency ≥5%; (3) minor allele frequency variation at the locus > 5% between SNPs with p-value ≤ 5 × 10⁻⁴; and (4) at least four data sets contributed to the signal at the locus. For each locus, we selected the SNP with smallest p-value for replication testing. The selection of these criteria was based on lessons learned from recent GWASs where true loci have multiple associated SNPs with different frequencies and many loci have multiple independent signals.14,27 The requirement of 5% minor allele frequency was intended to minimize the influence of potential outlier SNPs, and the inclusion of variation of minor allele frequency was to maximize the likelihood that the locus could harbor more than one association signal. p-Values below .0031 (.05/16) were considered as significant replication.

Gene-Based Association Analyses
We performed gene-based association analyses using the results from the GWAS meta-analyses. Specifically, we used the Knowledge-based mining system for Genome-wide Genetic studies (KGG) software,20 which uses an extended Simes test to integrate functional information and association evidence to combine the SNP p-values within a gene to obtain an overall p-value for each gene. In these analyses, we filtered out SNPs found in less than four data sets. All analyses were conducted using KGG default settings. We used the Benjamini–Hochberg method30 to correct for multiple testing, and considered FDR q-values below 0.05 as statistically significant.

Pathway and Network Analyses
We conducted pathway enrichment analysis for genes with at least one marker with p ≤ 5 × 10⁻⁸ from GWAS meta-analyses of either FTND or TTFC. If a marker was within a gene region, it was assigned to the gene; otherwise, it was mapped to its most proximate gene using the 30-kb flanking regions (both 5’ and 3’ sides). Genes identified using SNPs associated with FTND and TTFC were merged for the pathway enrichment analyses. We used the hypergeometric test implemented in the tool WebGestalt (2014 update)31,32 and the canonical pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. We required each pathway to have at least three genes from our gene list and no more than 300 genes from the reference genome. The p-values from hypergeometric tests were further adjusted by the Benjamini–Hochberg method.30 Only pathways with adjusted p-values < .05 were considered statistically significantly enriched.

We further examined how enriched pathways interacted with each other in function and regulation. Specifically, we applied the Characteristic Sub-Pathway Network (CSPN) algorithm15 to search for significantly interacting pathway pairs.15 CSPN was designed to prioritize pathway pairs with significant interaction of molecules from each pathway pair in the human protein–protein interaction (PPI) network (details in ref. 33). We used the human PPI data from the Protein Interaction Network Analysis (PINA) platform34 as the reference network in this pathway cross-talk analysis. Our working
PPI network included a total of 11,318 nodes (protein-coding genes) and 67,936 interactions. We restricted the analysis specifically to the aforementioned merged gene set and their enriched pathways. When running CSPN, a mode “OR” was selected, i.e., we considered all PPIs formed by the supplied genes as well as their one-step extension. In the final step, we selected the significant pathway interaction pairs based on permutation p-values ≤ 0.05.

Results

FTND and TTFC GWAS Meta-Analyses

In the GWAS meta-analyses of the discovery sample, about 19,000 subjects from 14 independent cohorts were included. For FTND (N = 19,431), only the CHRNA5-CHRNA3-CHRNB4 locus reached genome-wide significance (Table 1). The smallest p-value was observed at rs16969968 (β = −0.21, SE = 0.02, p = 3.96 × 10−19), which is a functional variant in CHRNA5 known to influence smoking behavior, in particular the quantity of cigarettes smoked. Several additional loci (CIB4 on 2p23, BG182718 on 11q12, and DSC3 on 18q12) were promising (Table 1).

<p>| Table 1. The most significant loci identified in the FTND and TTFC GWAS meta-analyses in the discovery sample |</p>
<table>
<thead>
<tr>
<th>Chr</th>
<th>BP</th>
<th>SNP</th>
<th>EA</th>
<th>NEA</th>
<th>EAF</th>
<th>β</th>
<th>SE</th>
<th>Z</th>
<th>p</th>
<th>N</th>
<th>Effect*</th>
<th>GENE</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTND</td>
<td>2</td>
<td>26,860,822</td>
<td>rs17005545</td>
<td>T</td>
<td>C</td>
<td>0.94</td>
<td>−0.23</td>
<td>0.05</td>
<td>−4.97</td>
<td>6.99E-07</td>
<td>18621</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>97,713,579</td>
<td>rs117029742</td>
<td>T</td>
<td>C</td>
<td>0.98</td>
<td>−0.64</td>
<td>0.12</td>
<td>−5.32</td>
<td>1.04E-07</td>
<td>12977</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>78,882,925</td>
<td>rs16969968</td>
<td>G</td>
<td>A</td>
<td>0.64</td>
<td>−0.21</td>
<td>0.02</td>
<td>−8.83</td>
<td>3.96E-19</td>
<td>17860</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>28,355,510</td>
<td>rs28510557</td>
<td>C</td>
<td>T</td>
<td>0.60</td>
<td>−0.10</td>
<td>0.02</td>
<td>−4.85</td>
<td>1.23E-06</td>
<td>18673</td>
<td>--------</td>
</tr>
<tr>
<td>TTFC</td>
<td>1</td>
<td>40,464,894</td>
<td>rs34022242</td>
<td>G</td>
<td>A</td>
<td>0.94</td>
<td>−0.13</td>
<td>0.02</td>
<td>−5.16</td>
<td>2.48E-07</td>
<td>14286</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>103,715,385</td>
<td>rs223441</td>
<td>C</td>
<td>T</td>
<td>0.51</td>
<td>0.05</td>
<td>0.01</td>
<td>4.58</td>
<td>4.66E-06</td>
<td>18533</td>
<td>++????++?????</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>186,514,078</td>
<td>rs28567706</td>
<td>C</td>
<td>G</td>
<td>0.99</td>
<td>0.49</td>
<td>0.09</td>
<td>5.49</td>
<td>4.05E-08</td>
<td>6533</td>
<td>+++????????????</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>42,527,386</td>
<td>rs11785369</td>
<td>A</td>
<td>C</td>
<td>0.89</td>
<td>0.09</td>
<td>0.02</td>
<td>5.35</td>
<td>8.83E-08</td>
<td>18634</td>
<td>++++++++?+++</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>21,673,589</td>
<td>rs18404282</td>
<td>C</td>
<td>T</td>
<td>0.98</td>
<td>−0.31</td>
<td>0.06</td>
<td>−5.43</td>
<td>5.70E-08</td>
<td>9005</td>
<td>+---??---??++</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>97,713,579</td>
<td>rs117029742</td>
<td>T</td>
<td>C</td>
<td>0.98</td>
<td>−0.30</td>
<td>0.05</td>
<td>−5.73</td>
<td>1.02E-08</td>
<td>13061</td>
<td>+---??---??++</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>44,608,436</td>
<td>rs10133756</td>
<td>C</td>
<td>G</td>
<td>0.98</td>
<td>−0.31</td>
<td>0.05</td>
<td>−5.88</td>
<td>4.11E-09</td>
<td>13283</td>
<td>+???-?+++---</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>78,882,925</td>
<td>rs16969968</td>
<td>G</td>
<td>A</td>
<td>0.66</td>
<td>−0.06</td>
<td>0.01</td>
<td>−5.82</td>
<td>6.21E-09</td>
<td>17834</td>
<td>+++??+++??</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>51,940,280</td>
<td>rs2877510</td>
<td>C</td>
<td>T</td>
<td>0.17</td>
<td>0.06</td>
<td>0.01</td>
<td>4.52</td>
<td>3.66E-06</td>
<td>18647</td>
<td>++++++++????</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>15,901,883</td>
<td>rs16969968</td>
<td>G</td>
<td>A</td>
<td>0.66</td>
<td>−0.06</td>
<td>0.01</td>
<td>−4.55</td>
<td>5.33E-06</td>
<td>18644</td>
<td>+++???-?++??</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>40,358,846</td>
<td>rs132786</td>
<td>G</td>
<td>A</td>
<td>0.19</td>
<td>−0.07</td>
<td>0.01</td>
<td>−4.61</td>
<td>4.07E-06</td>
<td>14605</td>
<td>+????-?+++??</td>
</tr>
</tbody>
</table>

EA, effect allele; NEA, non-effect allele; EAF, effect allele frequency; BP, base-pair position according to GRCh37/hg19; FTND, Fagerström Test for Nicotine Dependence; TTFC, time to first cigarette. *, minus and plus signs refer to the direction of effect in each independent dataset (SAGE, SC, MGS, COPDGene1, ISIB, CIDR370v1, CIDR370v3, CIDRC3, CIDRC4, EAGLE, ALSPAC, AUTW, VTSABD, CEDAR, FTC1, FTC2, NTR, GCD), whereas “?” refers to missing data. Genome-wide significant signals are highlighted in bold.

Figure 1. Network interaction based on markers with p values ≤ 5 × 10−5.
For FTND, four loci were selected for replication in several independent samples (Table 2). Of the four loci tested, only rs16969698 at the CHRNA5-CHRNB3-CHRNA6 locus was successfully replicated ($p = 3.61 \times 10^{-10}$). Rs17005545 in the novel locus CIB4 had a $p$-value of $2.68 \times 10^{-7}$ in the combined samples. For TTFC, rs16969698 in CHRNA5 and 11 other loci were selected for replication. Rs16969698 in CHRNA5 ($p = 2.84 \times 10^{-13}$) was successfully replicated and rs11785369 near the CHRNA3 gene reached genome-wide significance in the combined samples (Table 2).

We conducted SNP heritability analyses for both FTND and TTFC using the LD-adjusted kinship algorithm. The SNP heritability for FTND was estimated at 0.45, and that for TTFC was 0.193.

Convergent Loci Between FTND and TTFC

We compared the results obtained for FTND and TTFC to identify genes showing convergent association signals. We then selected genes/loci showing suggestive association ($p \leq 5 \times 10^{-5}$)
in both FTND and TTFC meta-analyses. A total of 15 genes/loci with convergent association signals were found (Table 3). The CHRNA5-CHRNA3-CHRNB4 locus was the only one reaching genome-wide significance for both FTND and TTFC. Other nicotinic receptors, CHRN3-CHRNA6 and CHRNA4, also showed convergent association signals. Other highlighted genes included long noncoding RNAs (DA409732 and BG182718) and a RNA-binding gene (RBFOX1), microtubule and actin regulation genes (KIF2B and VAV2), a cAMP/cGMP modulating gene (PDE2A), and an apolipoprotein gene (APOL3). For some genes (THSD7B, PDE2A, KIF2B, and APOL3), the same SNPs showed suggestive association for both phenotypes; for other genes (ZNF804A, VAV2, RBFOX1, and CHRNA4), the association signals for the two phenotypes were from different SNPs.

Gene-Based Meta-Analyses

We conducted gene-based analyses using the KGG program. In these analyses, in addition to the CHRNA5 locus and nearby genes, a few novel genes were identified (Table 4). COPB2 and CHRNA4 reached significance for association with FTND, and TTFCP2L1, RELN, and INO80C were significant for the TTFC phenotype. Table 4 also lists other candidates from gene-based analyses (i.e., genes with q-value ≤ 0.1).

### Pathway and Network Analyses

There were 647 SNPs with p-value ≤ 5 × 10^{-5} in the FTND meta-analysis and they were mapped to 134 known genes. There were 936 markers with p-value ≤ 5 × 10^{-5} in the TTFC meta-analysis and they were mapped to 145 genes. Altogether 15 genes were shared between the two phenotypes, yielding a total of 265 unique genes for pathway analyses. In the enrichment analyses using canonical pathways defined in KEGG database, a total of 26 pathways were found to be overrepresented among these genes (Table 5). The neuroactive ligand–receptor interaction pathway was the most significant pathways defined in KEGG database, a total of 26 pathways were found to be overrepresented among these genes (Table 5). The neuroactive ligand–receptor interaction pathway was the most significant pathway, including nicotinic receptors (CHRNA5, CHRNA3, and CHRNA4), glutamate receptor (GABRG3), gamma-aminobutyric acid receptor (GABRG3), glutamate receptor (GMR5), and somatostatin receptors (STTR1 and STTR4). In addition to several pathways known to be involved in ND (cell adhesion molecules, MAPK signaling, tight junction, and axon guidance), our analyses suggested that endocytosis, lysosome, chemokine signaling, and regulation of actin cytoskeleton pathways are also involved in ND.

We further tested if and how these pathways interacted with each other. Multiple immune related pathways were identified by two highlighted genes, HLA-DMB and HLA-DRB5; to avoid the extensive network of HLA-related immune responses, we excluded these two genes in our pathway crosstalk analyses. Our analyses revealed interacting networks among the endocytosis, regulation of...
Conclusion and Discussion

We conducted GWAS meta-analyses for the FTND and TTFC phenotypes in about 19,000 regular smokers of European ancestry from 14 independent cohorts. We confirmed the known association of the functional variant D398N (rs16969968) in the CHRNA5-CHRNA3-CHRNB4 locus for both FTND and TTFC phenotypes. Although the association between FTND and CHRNA5-CHRNA3-CHRNB4 locus had been reported before, ours is the first to report its association with TTFC. GWAS meta-analysis of FTND identified one potential novel locus (CIB4 on 2p23) with suggestive association in the discovery samples and a trend in the replication samples. CIB4 encodes a calcium binding protein that interacts with integrin. It may be involved in the integrin signaling. Our GWAS meta-analyses of TTFC, the largest for this phenotype, identified three novel loci (SORBS2 on 4q35, BG182718 on 11q22, and AA333164 on 14q21) in the discovery samples. SORBS2 encodes an adaptor protein involved in the regulation of actin cytoskeleton and is recently suggested to be involved in intellectual disability. The 11q22 signal peaks at rs117029742 located 25 kb from the 3′ end of BG182718 (also referred to as RP11-379J13.2). BG182718 is a long noncoding RNA with unknown function. In the interval of 1.5 MB centered at rs117029742, there are no other known genes. Interestingly, ClinVar database (http://www.ncbi.nlm.nih.gov/clinvar/) reports multiple large copy number variants in this region and these variants are reported to be associated with developmental disabilities. AA333164 (also referred to as RP11-305B6.3) on 14q21 is another long noncoding RNA with unknown function; it also overlaps with known copy number variants associated with global developmental disabilities.

<table>
<thead>
<tr>
<th>Table 5. Summary of pathway enrichment analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathway</td>
</tr>
<tr>
<td>Neuroactive ligand-receptor interaction</td>
</tr>
<tr>
<td>Endocytosis</td>
</tr>
<tr>
<td>Intestinal immune network for IgA production</td>
</tr>
<tr>
<td>Lysosome</td>
</tr>
<tr>
<td>Protein export</td>
</tr>
<tr>
<td>Cell adhesion molecules (CAMs)</td>
</tr>
<tr>
<td>Arrhythmogenic right ventricular cardiomyopathy (ARVC)</td>
</tr>
<tr>
<td>Dilated cardiomyopathy</td>
</tr>
<tr>
<td>Purine metabolism</td>
</tr>
<tr>
<td>Huntington’s disease</td>
</tr>
<tr>
<td>MAPK signaling pathway</td>
</tr>
<tr>
<td>Melanogenesis</td>
</tr>
<tr>
<td>Allograft rejection</td>
</tr>
<tr>
<td>Type 1 diabetes mellitus</td>
</tr>
<tr>
<td>Ashema</td>
</tr>
<tr>
<td>Graft-versus-host disease</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>Chemokine signaling pathway</td>
</tr>
<tr>
<td>Regulation of actin cytoskeleton</td>
</tr>
<tr>
<td>Autoimmune thyroid disease</td>
</tr>
<tr>
<td>Hedgehog signaling pathway</td>
</tr>
<tr>
<td>Basal cell carcinoma</td>
</tr>
<tr>
<td>Tight junction</td>
</tr>
<tr>
<td>Inositol phosphate metabolism</td>
</tr>
<tr>
<td>Axon guidance</td>
</tr>
<tr>
<td>Staphylococcus aureus infection</td>
</tr>
</tbody>
</table>

Statistically significant adjusted p-values are highlighted in bold.
It is not clear how TTFC relates to intellectual disability or developmental delay, or whether there are other functions that account for the association. Independent replication of the three novel loci for TTFC was not evident, presumably due to the lack of power of the replication samples. Statistically significant evidence for independent replication of the observed SNP associations may require even larger sample sizes.

TTFC is a commonly used measure in studies of smoking behavior and ND. It consists of six questions, with a sum score ranging from 0 to 10. Many studies have used TTFC scores to define ND, with varying thresholds (e.g., TTFC ≤ 3 as low ND and > 6 as high ND, but also TTFC ≥ 4 as dependent and TTFC = 0 as unaffected and those in-between as uncertain), whereas others have treated TTFC as a quantitative trait. Hancock and colleagues conducted a GWAS using categorized TTFC (low-smoking and mild ND [TTFC = 0 to 3]; moderate ND [TTFC = 4 to 6]; and severe ND [TTFC = 7 to 10]) and discovered that CHRNA4 was associated with TTFC at genome-wide significance across GWAS discovery and replication samples. Loukola and colleagues used TTFC as a binary trait (TTFC ≥ 4 as affected) in a study of 1,114 Finns but detected no genome-wide significant associations. Gelernter and colleagues conducted a GWAS with FTND as a quantitative trait in 7,084 cases for which cigarette smoking is a significant risk factor, such as cocaine addiction (RBFOX1), alcohol dependence (THSD7B), schizophrenia (ZNF804A) and heroin addiction (ZNF804A), rheumatoid arthritis (PDE2A), and prostate cancer (APOL3) (Table 3). Furthermore, RBFOX1 was highlighted in a recent study on the genetic relationship between schizophrenia and ND. It remains to be elucidated whether these genes are independent risk factors for these diseases among nonsmokers, or is the genetic association arising from mediation or moderation by smoking. For example, the CHRNAS-CCHRNAS-CHRNB4 locus reached genome-wide significance. When the discovery and replication samples were combined, the signal amplified. The signal observed in Gelernter et al.'s study, rs13225753, was not significant in our analyses (p = .7188), presumably due to heterogeneity at the locus (Cochran's Q = 36.86).

Compared with FTND, the TTFC phenotype is more related to the ability to quitting and relapse. Loukola and colleagues conducted a GWAS of TTFC in 1,114 Finnish subjects, but detected no genome-wide significant signals. In our discovery sample, which included the Caucasian subjects from the study of Gelernter and colleagues, only the CHRNA5-CHRNAS-CHRNB4 locus reached genome-wide significance. In our twin studies, 53, but it was close to the estimate (p = 0.193) for TTFC. Our convergent analyses identified several intriguing genes in addition to those previously highlighted for ND (CHRNAS-CCHRNAS-CHRNB4, CHRNA3, and CHRNA4). These included genes previously associated with diseases for which cigarette smoking is a significant risk factor, such as alcohol dependence (RBFOX1), alcohol dependence (THSD7B), schizophrenia (ZNF804A) and heroin addiction (ZNF804A), rheumatoid arthritis (PDE2A), and prostate cancer (APOL3). Our results further reassured that for a true signal, such as the CHRNA5-CHRNAS-CHRNB4 locus, convergent signals are seen with multiple-related phenotypes.

We also noticed the difference in SNP heritability estimates between the FTND and TTFC phenotypes. The SNP heritability estimate of 0.645 for FTND was close to the heritability estimated from twin studies for ND. The estimate of 0.193 for TTFC seemed low when compared with twin studies, but it was close to the estimate of 0.154 from a study with SNP measure. The implication of this difference was not immediately clear. As SNP heritability estimates were influenced by the power of the GWASs, it was likely that our TTFC meta-analysis did not have the power to have a good estimate of heritability.

Gene-based analyses highlighted several other genes besides the CHRNA5-CHRNAS-CHRNB4 and CHRNA4 loci, including a gene involved in pluripotency demethylation and development (TTFCP2L1), a gene associated with lung cancer (COP2B1), a gene implicated in schizophrenia (RELN), and a gene involved in chromatin remodeling (INO80C) (Table 4). These were novel genes first reported to be associated with smoking-related phenotypes. It would be interesting to see whether these genes could be replicated in future studies. Our pathway analyses revealed two statistically significantly enriched pathways. Although the involvement of neuroactive ligand-receptor interaction pathway in ND was expected since nicotinic receptors and other surface receptors belong to this pathway, the identification of endocytosis pathway in ND was novel and interesting. Over the years, there has been accumulating evidence that the recycling and internalization of surface receptors, such as nicotinic receptors, NMDA receptors, etc. It is not clear how TTFC relates to intellectual disability or developmental delay, or whether there are other functions that account for the association. Independent replication of the three novel loci for TTFC was not evident, presumably due to the lack of power of the replication samples. Statistically significant evidence for independent replication of the observed SNP associations may require even larger sample sizes.

TTFC is a commonly used measure in studies of smoking behavior and ND. It consists of six questions, with a sum score ranging from 0 to 10. Many studies have used TTFC scores to define ND, with varying thresholds (e.g., TTFC ≤ 3 as low ND and > 6 as high ND, but also TTFC ≥ 4 as dependent and TTFC = 0 as unaffected and those in-between as uncertain), whereas others have treated TTFC as a quantitative trait. Hancock and colleagues conducted a GWAS using categorized TTFC (low-smoking and mild ND [TTFC = 0 to 3]; moderate ND [TTFC = 4 to 6]; and severe ND [TTFC = 7 to 10]) and discovered that CHRNA4 was associated with TTFC at genome-wide significance across GWAS discovery and replication samples. Loukola and colleagues used TTFC as a binary trait (TTFC ≥ 4 as affected) in a study of 1,114 Finns but detected no genome-wide significant associations. Gelernter and colleagues conducted a GWAS with FTND as a quantitative trait in 7,084 cases for which cigarette smoking is a significant risk factor, such as cocaine addiction (RBFOX1), alcohol dependence (THSD7B), schizophrenia (ZNF804A) and heroin addiction (ZNF804A), rheumatoid arthritis (PDE2A), and prostate cancer (APOL3). Our results further reassured that for a true signal, such as the CHRNA5-CHRNAS-CHRNB4 locus, convergent signals are seen with multiple-related phenotypes.

We also noticed the difference in SNP heritability estimates between the FTND and TTFC phenotypes. The SNP heritability estimate of 0.645 for FTND was close to the heritability estimated from twin studies for ND. The estimate of 0.193 for TTFC seemed low when compared with twin studies, but it was close to the estimate of 0.154 from a study with SNP measure. The implication of this difference was not immediately clear. As SNP heritability estimates were influenced by the power of the GWASs, it was likely that our TTFC meta-analysis did not have the power to have a good estimate of heritability.

Gene-based analyses highlighted several other genes besides the CHRNA5-CHRNAS-CHRNB4 and CHRNA4 loci, including a gene involved in pluripotency demethylation and development (TTFCP2L1), a gene associated with lung cancer (COP2B1), a gene implicated in schizophrenia (RELN), and a gene involved in chromatin remodeling (INO80C). These were novel genes first reported to be associated with smoking-related phenotypes. It would be interesting to see whether these genes could be replicated in future studies. Our pathway analyses revealed two statistically significantly enriched pathways. Although the involvement of neuroactive ligand-receptor interaction pathway in ND was expected since nicotinic receptors and other surface receptors belong to this pathway, the identification of endocytosis pathway in ND was novel and interesting. Over the years, there has been accumulating evidence that the recycling and internalization of surface receptors, such as nicotinic receptors, NMDA receptors, etc.
glutamate receptors, and opioid receptors are involved in drug addiction. Furthermore, it is known that the recycling and internalization of surface proteins are mediated and regulated by endocytosis, actin cytoskeleton, and lysosome functions. Our pathway analyses explicitly identified these pathways in ND (Table 5). Our network analyses indicated that these pathways were interacting with each other, forming an interconnected network. These findings were consistent with the emerging picture in addiction studies.

In summary, our GWAS meta-analyses of FTND and TTFC identified several promising candidates for both phenotypes. Three novel loci (SORBS2, BG182718, and AA333164) were discovered for TTFC. Although we could not validate these novel loci with statistical significance in our replication sample, further investigation is warranted. With supporting information from both FTND and TTFC phenotypes, we also identified promising candidates for ND, including several genes known for association with other psychiatric disorders. Our pathway analyses highlighted the endocytosis pathway, supporting the importance of recycling and internalization of surface receptors in the development of nicotine addiction. In our network analyses, we discovered a multinode network with several interacting pathways known for involvement in substance abuse and psychiatric disorders. This information provides new insights for our understanding of ND and nicotine withdrawal.

Supplementary Material
Supplementary data are available at Nicotine and Tobacco Research online.

Acknowledgments
This study was supported, in part, by NIH grants DA032246, DA035825. This work used data from multiple dbGaP studies. We acknowledge the investigators involved in the collection, genotyping, and analysis of these data sets. Detailed description of these samples, their funding agencies, and principal investigators are available in Supplementary Material.

Declaration of Interests
LJB is listed as an inventor on Issued U.S. Patent 8,080,371, “Markers for Addiction” covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction. HRK has been a consultant, advisory board member, or CME speaker for Indivior, Lundbeck, and Otsuka and is a member of the American Society of Clinical Psychopharmacology’s Alcohol Clinical Trials Initiative (ACTIVE), which was supported in the last 3 years by AbbVie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, and XenoPort.

References


