Expanding the Genetic Architecture of Nicotine Dependence and its Shared Genetics with Multiple Traits: Findings from the Nicotine Dependence GenOmics (iNDiGO) Consortium

Bryan C. Quach[1], Michael J. Bray[2], Nathan C. Gaddis[3], Mengzhen Liu[4], Teemu Palviainen[5], Camelia C. Minica[7], Stephanie Zellers[4], Richard Sherva[6], Fazil Aliev[8], Michael Nothnagel[9], Kendra A. Young[10], Jesse Marks[1], Hannah Young[4], Yuelong Guo[11,12], Alex Waldrop[3], Nancy Sey[13], Maria T. Landi[14], Daniel W. McNeil[15,16], Lindsay A. Farrer[6,17,18,19,20], Christina A. Markunas[1], Jacqueline Vink[21], Jouke-Jan Hottenga[7],William G. Iacono[4], Henry R. Kranzler[22,23], Nancy L. Saccone[24], Michael C. Neale[25,26], Pamela Madden[2], Marcella Rietschel[27], Mary L. Marazita[28], Matthew McGue[4], Hyejung Won[13], Georg Winterer[29], Richard Grucza[2], Danielle Dick[8,30,31], Joel Gelernter[32,33,34,35], Neil E. Caporaso[14], Timothy B. Baker[36], Dorret I. Boomsma[7], Jaakko Kaprio[5,37], John E. Hokanson[10], Scott Vrieze[4], Laura J. Bierut[2], Eric O. Johnson[1,38], and Dana B. Hancock[1]

1. Center for Omics Discovery and Epidemiology, Behavioral Health Research Division, RTI International, Research Triangle Park, North Carolina, USA;
2. Department of Psychiatry, Washington University, St. Louis, Missouri, USA;
3. Research Computing Division, RTI International, Research Triangle Park, North Carolina, USA;
4. Department of Psychology, University of Minnesota Twin Cities, Minneapolis, Minnesota, USA;
5. Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland;
6. Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, Massachusetts, USA;
7. Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands;
8. Department of Psychology, Virginia Commonwealth University, Richmond, Virginia, USA;
9. Institute of Medical Informatics and Statistics, Christian-Albrechts University, Kiel, Germany;
10. Department of Epidemiology, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA;
11. Center for Genomics in Public Health and Medicine, Biostatistics and Epidemiology Division, RTI International, Research Triangle Park, North Carolina, USA;
12. GeneCentric Therapeutics, Research Triangle Park, North Carolina, USA;
13. Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA;
14. Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, United States department of Health and Human Services, Bethesda, Maryland, USA;
15. Department of Psychology, West Virginia University, Morgantown, West Virginia, USA;

16. Department of Dental Practice and Rural Health, West Virginia University, Morgantown, West Virginia, USA;
17. Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, USA;
18. Department of Ophthalmology, Boston University School of Medicine, Boston, Massachusetts, USA;
19. Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA;
20. Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA;
21. Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands;
22. Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA;
23. VISN 4 MIRECC, Crescenz VA Medical Center, Philadelphia, Pennsylvania, USA;
24. Department of Genetics, Washington University, St. Louis, Missouri, USA;
25. Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia, USA;
26. Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia, USA;
27. Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany;
28. Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA;
29. Experimental & Clinical Research Center, Department of Anesthesiology and Operative Intensive Care Medicine, Charité - University Medicine Berlin, Germany;
30. College Behavioral and Emotional Health Institute, Virginia Commonwealth University, Richmond, Virginia, USA;
31. Department of Human & Molecular Genetics, Virginia Commonwealth University, Richmond, Virginia, USA;
32. Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut, USA;
33. Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA;
34. Department of Neuroscience, Yale University School of Medicine, New Haven, Connecticut, USA;
35. Department of Psychiatry, VA CT Healthcare Center, West Haven, Connecticut, USA;
36. Center for Tobacco Research and Intervention, Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA;
37. Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland;
38. Fellow Program, Behavioral Health Research Division, RTI International, Research Triangle Park, North Carolina, USA;

Abstract

Cigarette smoking is the leading cause of preventable morbidity and mortality. Knowledge is evolving on genetics underlying initiation, regular smoking, nicotine dependence (ND), and cessation. We performed a genome-wide association study using the Fagerström Test for ND (FTND) in 58,000 smokers of European or African ancestry. Five genome-wide significant loci, including two novel loci *MAGI2/GNAI1* (rs2714700) and *TENM2* (rs1862416) were identified, and loci reported for other smoking traits were extended to ND. Using the heaviness of smoking index (HSI) in the UK Biobank (N=33,791), rs2714700 was consistently associated, but rs1862416 was not associated, likely reflecting ND features not captured by the HSI. Both variants were *cis*-eQTLs (rs2714700 for *MAGI2-AS3* in hippocampus, rs1862416 for *TENM2* in lung), and expression of genes spanning ND-associated variants was enriched in cerebellum. SNP-based heritability of ND was 8.6%, and ND was genetically correlated with 13 other smoking traits (r_g=0.40–0.95) and co-morbid diseases. Our results emphasize the FTND as a composite phenotype that expands genetic knowledge of smoking, including loci specific to ND.

Introduction

Cigarette smoking remains the leading cause of preventable death worldwide,[1] despite the well-known adverse health effects. Smoking causes more than 7 million deaths annually from a multitude of diseases including cancer, chronic obstructive pulmonary disease (COPD), and heart disease.[1,2] Cigarette smoking is a multi-stage process consisting of initiation, regular smoking, nicotine dependence (ND), and cessation. Each step has a strong genetic component (for example, twin-based heritability estimates up to 70% for the transition from regular smoking to ND[3,4]), and partial overlaps are expected among the sets of sequence variants correlating with

the different stages,[3] evidenced by findings of the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) with sample sizes up to 1.2 million individuals.[5] GSCAN identified 298 genome-wide significant loci associated with initiation (ever vs. never smoking), age at initiation, cigarettes per day (CPD), and/or cessation (current vs. former smoking); 259 of the loci harbored significant associations with initiation.[5]

In comparison to other stages of smoking, known loci for ND are limited. Only six reproducible, genome-wide significant loci have been identified: *CHRNB3-CHRNA6* (chr8p11), *DBH* (chr9q34), *CHRNA5-CHRNA3-CHRNB4* (chr15q25), *DNMT3B* and *NOL4L* (chr20q11), and *CHRNA4* (chr20q13).[6] A more complete understanding of the genetics underlying ND is needed, as it could help to predict the likelihood of quitting smoking, withdrawal severity, response to treatment, and health-related consequences.[7-10] The Fagerström Test for ND (FTND), also called the Fagerström Test for Cigarette Dependence,[11] provides a composite phenotype that captures multiple behavioral and psychological features of ND.[12] Expanding upon our prior analyses of studies comprising our Nicotine Dependence GenOmics (iNDiGO) Consortium,[13,14] we report findings from the largest GWAS meta-analysis for ND (N=58,000; 46,213 European [EUR] and 11,787 African American [AA] ancestry participants from 23 studies) to identify novel genetic loci associated with ND, assess genetic correlations between ND and other phenotypes and gene expression patterns, and test GSCAN-identified loci[5] for effects on ND.

Results

Cross-ancestry GWAS meta-analysis finds two novel SNP associations with ND

Our cross-ancestry ND GWAS meta-analysis (λ=1.034, **Supplementary Figure 1A**) identified five genome-wide significant loci (**Figure 1**). Associations of the lead SNPs from each

of these five loci are shown in **Table 1**. All genome-wide significant SNP/indel associations from the cross-ancestry meta-analysis are provided in **Supplementary Table 5**.

Three of the genome-wide significant loci have known associations with ND from our prior GWAS and others[6]: chr15q25[13-15] (smallest $P=1.6\times10^{-39}$ for rs16969968, a well-established functional missense [D398N] *CHRNA5* SNP[16]), chr20q13[13] (smallest $P=1.2\times10^{-12}$ for rs151176846, an intronic *CHRNA4* SNP), and chr9q34[14] (smallest $P=1.1\times10^{-8}$ for rs13284520, an intronic *DBH* SNP). The loci spanning nicotinic acetylcholine receptor genes (*CHRNA5-A3-B4* and *CHRNA4*), but no novel loci, were identified at genome-wide significance in the EUR-specific GWAS meta-analysis ($\lambda=1.036$, **Supplementary Figures 1B** and **2A**). No genome-wide significant loci were found in the AA-specific GWAS meta-analysis ($\lambda=1.032$, **Supplementary Figures 1C** and **2B**).

Two genome-wide significant loci from the cross-ancestry meta-analysis represent novel associations with ND. On chr7q21, the most significant SNP ($P=2.3\times10^{-9}$) was rs2714700, a SNP between the *MAGI2* and *GNAI1* genes (**Supplementary Figures 3A–B**). The most significant SNP on chr5q34, rs1862416 ($P=1.5\text{x}10^{-8}$), sits within an intron for *TENM2* (**Supplementary Figures 3C–D**). Both SNPs imputed well: sample size-weighted mean estimated r^2 values were 0.97 for rs2714700 and 0.92 for rs1862416. Further, both SNPs were common, and their associations with ND were observed across EURs and AAs (**Table 1**) and were largely consistent across studies (**Supplementary Figure 4A–B**): rs2714700-T being associated with reduced risk (meta-analysis OR [95% CI]=0.96 [0.94–0.97]) and rs1862416-T being associated with increased risk (meta-analysis OR [95% CI]=1.08 [1.05–1.11]) for severe vs. mild ND. Neither SNP showed evidence for heterogeneity, based on the I^2 index[17], across studies (P=0.83 for rs2714700 and 0.75 for rs1862416). Leave-one-study-out analyses (**Supplementary Table 6**)

revealed some variability in p-values ($P=3.1\times10^{-7}$–7.4×10^{-9} for rs2714700 and $P=5.6\times10^{-9}$–3.9×10^{-6} for rs1862416), likely due to fluctuating statistical power given the significant correlation between N and p-value across iterations: r=-0.65, $P=8.6\times10^{-5}$. However, there was little variation in the effect sizes (range of β values corresponding to the OR for severe vs. mild ND = 0.95–0.96 for rs2714700-T and 1.07–1.08 for rs1862416-T).

We compared the novel ND-associated SNPs with results reported for other smoking traits by GSCAN.[5] Both the *MAGI2/GNAI1* SNP rs2714700 and the *TENM2* SNP rs1862416 were nominally associated at P<0.05 with ever vs. never smoking and rs2714700 with CPD in consistent directions with ND; neither SNP was associated with age at initiation or smoking cessation (**Supplementary Table 7**). Because rs1862416 was located within the boundaries of a genome-wide significant locus for ever smoking (chr5:164,596,435-168,114,971), we used GCTA to assess the independence of association signals via conditional modeling. All 6 lead SNPs in this GSCAN-identified locus were in low LD with rs1862416 (maximum $r^2=0.0047$ [**Supplementary Figure 5**], maximum D'=0.46), and three were nominally associated with ND at P<0.05 (**Supplementary Table 8**). Among our iNDiGO studies, rs1862416 remained associated with ND in models conditioned on each GSCAN lead SNP individually ($P=7.9\times10^{-8}$–1.8×10^{-8}) and with all 6 SNPs taken together ($P=2.2\times10^{-7}$). Rs2714700 was located >1 MB away from any GSCAN-identified locus, so conditional modeling was not necessary. These results suggest that the novel rs2714700 and rs1862416 associations with ND are independent of any GSCAN-identified loci.

For independent testing, we analyzed the two novel SNPs (rs2714700 and rs1862416) for association with HSI in the UK Biobank. Results are shown in **Supplementary Table 9**. The *MAGI2/GNAI1* SNP, rs2714700, was associated with HSI at P=0.014, which surpassed

Bonferroni correction for two SNP tests, and meta-analysis of all discovery studies with UK Biobank (total N=91,791) supported rs2714700-T being associated with milder ND ($P=7.7\times10^{-9}$). The *TENM2* SNP, rs1862416, was not associated with HSI in the UK Biobank (P=0.39).

To determine the factors of ND that drove the novel genome-wide associations, we returned to the iNDiGO studies, tested SNP associations with each specific FTND item, and combined results via cross-ancestry meta-analyses. For rs2714700, we observed the lowest p-values for the two items that comprise the HSI (**Figure 2A**): TTFC ($P=5.3\times10^{-4}$) and CPD ($P=1.1\times10^{-3}$). Rs2714700 was also associated at P<0.05 with difficult in refraining from smoking in forbidden places (P=0.025) and the cigarette most hated to give up (P=0.030). Rs1862416 was associated with TTFC (P=0.018) and two items that are not captured by the HSI: the cigarette most hated to give up (P=0.015) and smoking when ill (P=0.023) (**Figure 2B**).

GWAS findings for other smoking traits extend to ND

We assessed whether genome-wide significant SNPs identified for smoking traits in GSCAN extended to ND using results from the cross-ancestry GWAS meta-analysis. We focused on the 55 genome-wide significant SNPs from 40 loci associated with CPD, given that it displayed the best genetic correlation with ND (**Figure 3**). After applying Bonferroni correction for the 53 SNPs that were available in our meta-analysis ($P<9.4\times10^{-4}$), 17 SNPs had a statistically significant and directionally consistent association with ND (**Table 2**). These SNPs span six loci reported at genome-wide or nominal significance in prior GWAS of ND (*CHRNA5-A3-B4* [chr15], *CHRNA4* [chr20], *DBH* [chr9], *CHRNB3* [chr8], *CYP2A6* [chr19], and *NOL4L* [near *DNMT3B*, chr20])[6] and three loci not reported in prior ND GWAS—*DRD2* (chr11), *C16orf97* (chr16), and *CHRNB2* (chr1).

ND is genetically correlated with 13 other phenotypes

We estimated the heritability explained by common SNPs of ND at h_g^2 (standard error) = 0.086 (0.012), using LDSC[18] and the EUR-specific GWAS meta-analysis results. We also found statistically significant genetic correlations of ND with 13 phenotypes (Bonferroni-corrected P<0.0011; **Figure 3** and **Supplementary Table 3**). Positive correlations indicate that the genetic predisposition to higher ND risk was correlated with genetic risks for other smoking traits (smallest $P=3.1\times10^{-70}$ for higher CPD [r_g=0.95], followed by $P=3.2\times10^{-16}$ for current smoking [r_g=0.54] and $P=3.2\times10^{-16}$ for ever smoking [r_g=0.40]). We repeated LDSC, after removing all chr15q25 variants between 78.5 and 79.5 MB and found only negligible differences in these correlations (r_g=0.94 for higher CPD, r_g=0.51 for current smoking, and r_g=0.42 for ever smoking). Beyond the smoking traits, with all SNPs included, higher ND was genetically correlated with higher risks of alcohol dependence, neuroticism, psychiatric diseases (major depressive disorder and its symptoms and schizophrenia), and smoking-related consequences (lung cancer and coronary artery disease). Among these positively correlated traits, r_g values ranged from 0.16 (schizophrenia) to 0.77 (squamous cell lung cancer). Higher risk of ND was genetically correlated with lower age of smoking initiation (r_g=-0.55) and fewer years of schooling (r_g=-0.33).

Gene expression data implicates target genes for novel ND-associated SNPs and identifies ND heritability enrichment in cerebellum

Rs2714700, an intergenic SNP, is not a significant *cis*-eQTL with any gene-level expression in GTEx (v8), but it was implicated as a *cis*-eQTL for the *MAGI2-AS3* transcript in hippocampus from BrainSeq[19] (N=551; $P=8.5\times10^{-4}$). The protective allele for ND (rs2714700-T) was associated with higher expression of the *MAGI2-AS3* transcript ENST00000414797.5.

Rs1862416 is annotated to enhancer histone marks in brain (specifically, germinal matrix during fetal development and the developed prefrontal cortex, anterior caudate, and cingulate gyrus tissues) and several other tissues in HaploReg.[20] It is also located in the promoter of *CTB-77H17.1*, which is a novel antisense RNA transcript encoded within a *TENM2* intron. In GTEx, rs1862416 was reported as a significant lung-specific *cis*-eQTL SNP *TENM2*. The ND risk-conferring allele (rs1862416-T) was associated with decreased gene-level *TENM2* expression in lung. *CTB-77H17.1* was too lowly expressed across GTEx tissues to test its expression levels by rs1862416.

To assess the enrichment of the EUR-specific ND GWAS meta-analysis results to specific gene expression patterns, we applied LDSC-SEG[21] with reference to 205 tissues/cell types with publicly available gene expression data. We observed statistically significant enrichment in one tissue (cerebellum) at Bonferroni-corrected $P<2.4\times10^{-4}$ (**Supplementary Table 4**), indicating that genes spanning ND-associated SNPs are enriched for specific expression in the cerebellum relative to other tissues/cell types.

Discussion

We expanded current knowledge of ND in this largest GWAS to date, by identifying two novel genome-wide significant loci as well as 3 known loci, extending associations of additional loci implicated for other smoking phenotypes, and detecting significant genetic correlations of ND with 13 other complex phenotypes and with gene expression in cerebellum. The top novel SNPs between *MAGI2* and *GNAI1* (chr7q21) and in *TENM2* (chr5q34) were independent of previously reported GWAS signals for any smoking trait. Three of our genome-wide significant loci were known: (1) *CHRNA5-CHRNA3-CHRNB4* (chr15q25) is irrefutably associated with ND, as driven largely by CPD.[6] (2) Our initial GWAS meta-analysis of 5 studies (now part of the

iNDiGO consortium)[13] identified *CHRNA4* (chr20q13) at genome-wide significance. Subsequent associations were found with heavy vs. never smoking in the UK Biobank[22] and with initiation, CPD, and cessation in GSCAN.[5] (3) *DBH* (chr9q34) was first identified as genome-wide significant for smoking cessation but later associated with ND in our meta-analysis of 15 studies (now part of the iNDiGO consortium)[14] and with CPD and cessation in GSCAN.[5]

The novel ND-associated locus with lead SNP rs2714700 is intergenic between *MAGI2* (membrane associated guanylate kinase, WW and PDZ domain containing 2) and *GNAI1* (G protein subunit alpha i1). We identified rs2714700 at genome-wide significance for its association with ND, which was driven by CPD (unlike rs1862416), TTFC, and other FTND items, indicating that this SNP association may reflect both primary and secondary features of ND. Rs2714700 was also associated with HSI in the independent UK Biobank. The *cis*-eQTL evidence for rs2714700 in the hippocampus suggests that it may influence expression of the long noncoding RNA *MAGI2-AS3* (MAGI2 antisense RNA 3). *MAGI2-AS3* has been mainly studied for its role in the progression of cancer, including glioma in the brain.[23] No genome-wide significant associations have been reported within 1MB of rs2714700 in the GWAS catalog. Our evidence of genome-wide significance for rs2714700 points to a novel locus that has not been associated with smoking or any related trait, and its functional relevance merits further investigation.

We also observed a genome-wide significant association of ND with rs1862416, a lung-specific *cis*-eQTL for *TENM2*. *TENM2* encodes teneurin transmembrane protein 2, a cell surface receptor that plays a fundamental role in neuronal connectivity and synaptogenesis.[24] With rs1862416 residing in the promoter of *CTB-77H17.1*, it could influence this antisense RNA, which in turn could dysregulate its sense transcript, *TENM2*. As an illustrative example, the

autism-associated SNP rs4307059 is annotated to and acts as a promoter region *cis*-eQTL for the antisense RNA *MSNP1AS* (moesin pseudogene 1, antisense) that influences regulation of its sense transcript, *MSN*.[25] However, while rs1862416 is generally indicated for its potential regulatory role (i.e., enhancer and promoter annotations and *cis*-eQTL evidence), its specific effect on either *CTB-77H17.1* or *TENM2* regulation in brain tissue was not evident in currently available data.

Further, independent association testing using HSI in the UK Biobank did not yield statistical significance for rs1862416. This lack of association may be due to rs1862416 influencing components of ND that are not fully captured by the two FTND items that comprise the HSI (TTFC and CPD), as suggested by the specific FTND item association testing among the iNDiGO studies. Rs1862416 was suggestively associated ($P<0.05$) with TTFC, "Which cigarette would you hate most to give up?" (the first one in the morning vs. all others), and "Do/did you smoke if you are so ill that you are in bed most of the day?" (yes/no). These item responses reflect withdrawal symptoms that are indicative of secondary features of ND (smoking based on relief of negative effects or situation), as compared with primary (or core) features of ND that are necessary and sufficient for habit formation (heaviness of smoking [tolerance], automaticity, loss of control, and craving).[26-29] Having the composite ND phenotype may have enhanced our power for discovering *TENM2*, but its detection in the UK Biobank may have been limited by the reliance on the HSI.

Beyond our discovery of rs1862416 with ND, SNPs across the *TENM2* gene have been identified at genome-wide significance, as presented in the GWAS catalog[30], for educational attainment,[31] smoking initiation (ever vs. never smoking),[5,32-34] age of smoking initiation,[5] smoking cessation (current vs. former smoking),[5] alcohol consumption (drinks per week),[5] lung

function,[34,35] height,[34] number of sexual partners,[32] depression,[36,37] risk taking tendency,[32] body mass index,[34] menarche (age at onset)[38], and regular attendance at a religious group[39]. All *TENM2* SNPs in the GWAS catalog have very low r^2 values with our novel SNP, rs1862416 (**Supplementary Figure 5**), and we found that rs1862416 was associated with ND independently from other *TENM2* SNPs implicated in GSCAN. These results suggest that *TENM2* has pleiotropic effects on ND, traits that are genetically correlated with ND, and other traits.

The genetics of smoking behaviors, more broadly, has rapidly evolved with the GSCAN consortium having amassed a very large sample size and identified 298 genome-wide significant loci for smoking traits representing single components: ever vs. never smoking, age of smoking initiation, CPD, and current vs. former smoking.[5] We observed significant genetic correlations of each of these smoking traits with ND, yet despite the nearly complete sharing between ND and CPD specifically, our two novel ND-associated loci were not identified at genome-wide significance by GSCAN (smallest P=0.033 for rs1862416-T; smallest P=0.016 for rs2714700-T), suggesting that these loci are specific to ND. These observations resemble previously reported patterns of genetic correlation between alcohol dependence and alcohol consumption that suggested shared yet distinct genetics underlying specific measures and composite phenotypes.[40] Similarly, the majority of GSCAN-identified loci were trait-specific (191 of the 298 loci), where the other 107 loci were pleiotropic with associations identified for two or more of the smoking traits.[5] In our evaluation of GSCAN-identified loci, we corroborated associations of several previously implicated loci for ND (e.g., nicotine acetylcholine receptors genes *CHRNA5-A3-B4* and *CHRNA4*) and three additional loci (*DRD2*, *C16orf97*, and *CHRNB2*) that have not been reported in prior ND GWAS. Of these loci, *DRD2* is notable as a long-studied addiction

229 candidate gene[4] and its recent identification as genome-wide significant for alcohol use disorder
230 for rs4936277[41], which is correlated (r^2=0.94 in 1000G EUR, 0.82 in 1000G AFR) with
231 rs7125588, the top SNP identified for CPD in GSCAN and associated with ND in iNDiGO; these
232 results support a shared genetic effect of *DRD2* underlying addiction. Notably, rs7125588 is not
233 correlated (r^2=0.04 in 1000G EUR, 0.01 in 1000G AFR) with the *DRD2* variant rs1800497
234 (Taq1A), which is not significantly associated with ND in iNDiGO (P=0.24).
235 Other GSCAN loci were detected for the single component smoking traits but show no
236 evidence for association in our study (**Supplementary Table 10**), suggesting that these loci
237 influence stages of smoking other than ND, or they exert weak effects on ND that we were
238 underpowered to detect. We expect that additional GSCAN-identified loci are associated with
239 ND, but their detection will require a larger sample size. These results demonstrate the utility of
240 studying the genetics of the composite ND phenotype and comparing with GWAS of other
241 smoking traits to tease apart loci that are specific to one stage (i.e., initiation, regular smoking,
242 ND, cessation) vs. loci that influence multiple stages to better understand the full spectrum of
243 smoking behaviors.
244 Beyond the smoking traits, we observed significant genetic correlations between ND and
245 alcohol dependence, years of schooling, neuroticism, comorbid psychiatric traits (major
246 depression and schizophrenia), and smoking-related health consequences (lung cancer and
247 coronary artery disease). Some of these observations corroborate prior findings (for example,
248 alcohol dependence[40] and schizophrenia[42,43] with ND), whereas the other correlations extend to
249 ND prior observations for the single component smoking traits (for example, CPD with years of
250 schooling[5], neuroticism[5], major depression[5], coronary artery disease[5], and lung cancer[44]). The
251 genetic correlation between ND and gene expression in cerebellum is a notable observation

consistent with cerebellum-specific *cis*-eQTL effects observed for the ND-associated *DNMT3B* SNP rs910083[14] and the age of smoking initiation-associated *CHRNA2* SNP rs11780471[44], both of which are also associated with lung cancer. These findings add to the evidence that the cerebellum may be important for ND risk,[45,46] in addition to the other addiction-relevant brain tissues. However, since the cerebellum contains a higher neuronal concentration than other brain tissues,[21,47] future studies are needed to decipher whether the cerebellar gene regulatory effects in the etiology of ND are due to neuronal activity. Additionally, although genetic correlation between ND and another trait suggest shared genetics underlying the phenotypes, multiple mechanisms can produce significant correlations (i.e., unmeasured intermediary phenotypes, correlated risk variants, mediation).[48-50] Identifying the true mechanistic explanation requires further investigations.

The present ND GWAS meta-analysis follows two prior waves of data assembly by the iNDiGO consortium (Ns=17,074[13], 38,602[14], and now 58,000) and is the largest to date for the field. Despite still having substantially smaller sample sizes than the GSCAN GWAS, at each wave, increasing sample size for diverse ancestry groups (EURs and AAs) has illuminated ND-associated loci, some of which are shared with other stages of smoking while others are specific to ND. Our present findings underscore the complexity even within the ND phenotype, as our novel loci displayed patterns of association with specific FTND items that reflect primary or secondary ND features, e.g., the *TENM2* SNP influenced secondary features that are not captured simply by heaviness of smoking. Understanding genetic differences that underlie primary vs. secondary ND may better inform treatment strategies, e.g., changing environmental cues for individuals whose smoking is driven solely by primary ND features vs. treating withdrawal for individuals whose ND is augmented with secondary features.[28] Studying the genetics of ND

alongside other smoking traits (e.g., initiation and cessation) is key to gaining a better understanding of the neurobiological perturbations that influence the trajectory of smoking behaviors and their treatment implications.

Methods

We assembled 58,000 participants from 23 iNDiGO consortium studies with genome-wide single nucleotide polymorphism (SNP) genotypes and FTND phenotype data available for ever smokers to perform ND GWAS meta-analyses. Fifteen of the studies were included from our prior GWAS using their original or updated sample sizes (total N increased from 38,602[14] to 46,098 in the current analysis), while 8 studies were added for the current study (total N=11,902). Participant characteristics are provided in **Supplementary Table 1**, and details of the study designs, genotyping, quality control, 1000 Genomes (1000G) phase 3 imputation (unless otherwise stated), and statistical analyses are provided in the **Supplementary Methods**. Institutional review boards at the respective sites approved the study protocols, and all participants provided written informed consent.

ND GWAS meta-analysis

The FTND is a well-validated, widely used 6-item questionnaire that assesses psychologic dependence on nicotine, with scores ranging from 0 (no dependence) to 10 (highest dependence level).[12,51] As done before,[13,14] we categorized FTND scores as mild (scores 0–3), moderate (scores 4–6), or severe (scores 7–10). FTND data reflected current smoking behaviors at the time of interview (i.e., current FTND) or the period of heaviest smoking among ever smokers (i.e., lifetime FTND). We previously found only small differences in genetic association results due to any measurement variance when using current vs. lifetime FTND.[52] Two studies also included low-intensity smokers, who reported $\leq$10 CPD but had no data available on other

FTND items; these smokers were defined as mildly dependent, given minimal phenotype misclassification when comparing these FTND and CPD thresholds.[14] The other FTND items were always required to define moderately and severely dependent smokers. See **Supplementary Methods** for further details on the ND phenotype data by study.

For each study, genome-wide SNP/indel associations with the 3-level categorical ND outcome were tested within an ancestry group using linear regression. Covariates included age, sex, principal component eigenvectors, and study-specific covariates where warranted. For studies that included relatives, relatedness was accounted for in the regression modeling. See the **Supplementary Methods** for additional study-specific details.

GWAS results were combined using fixed-effect inverse variance-weighted meta-analyses in METAL.[53] Prior to performing meta-analyses, we applied genomic control to results from one study, deCODE, to adjust for inflation due to relatedness among participants (λ=1.12); all other studies had low inflation (λ=0.99–1.04) (**Supplementary Table 1**). We removed SNPs/indels with minor allele frequency (MAF) <1% in the 1000G phase 3 reference panel for the analyzed ancestry group (1000G European or African superpopulations) and SNPs/indels available in only one study. All variant annotations correspond to the National Center for Biotechnology Information (NCBI) build 37. As before[14], the threshold of genome-wide significance was set at $P = 5\times10^{-8}$. Regional association plots of novel genome-wide significant loci were constructed using LocusZoom[54] with references of either 1000G European or African panels to estimate linkage disequilibrium of the lead SNP (based on smallest meta-analysis P-value) and surrounding SNPs. The lead SNP for each novel locus was tested for association with each of the specific FTND items (**Supplementary Methods**).

For any ND-associated SNPs located within the bounds of loci identified by GSCAN (1 MB surrounding the lead SNP),[5] conditional models were analyzed using our GWAS summary statistics and the Genome-wide Complex Trait Analysis (GCTA) tool, adjusted for the lead SNPs in GSCAN.[55,56] To contextualize the magnitude of the observed effect sizes, we calculated odds ratios (ORs) using the β estimate from the single SNP linear regression model ($OR=\exp[2\times\beta_{SNP}]$ for severe vs. mild ND, with OR>1 corresponding to an increased risk of severe ND) and compared these values across studies and ancestries using the Forest Plot Viewer.[57]

Independent testing using heaviness of smoking index in the UK Biobank

Novel, genome-wide significant SNPs from our ND GWAS meta-analysis were tested in the UK Biobank. Although all 6 items of the FTND were not collected in the UK Biobank, two items (CPD and time-to-first-cigarette [TTFC]) were collected among current smokers. These two items together form the heaviness of smoking index (HSI), which is highly correlated with the full-scale FTND (e.g., r=0.7 among nondaily smokers and 0.9 among daily smokers).[58] We derived HSI scores, ranging from 0 (no dependence) to 6 (highest dependence level), and categorized them as follows: mild (scores 0–2), moderate (scores 3–4), and severe (scores 5–6). These HSI categories were highly concordant (89.3%) with our routinely used FTND categories using the COGEND study, which was ascertained specifically for ND (**Supplementary Methods** and **Supplementary Table 2**). The final analysis dataset included 33,791 current smokers (18,063 mildly, 13,395 moderately, and 2,333 severely dependent, as defined by HSI). Our linear regression model included covariates for age, sex, and principal component eigenvectors (**Supplementary Methods**).

Genetic correlations of ND with other complex phenotypes and with gene expression

Summary statistics from the EUR-specific meta-analyses were used as input into linkage disequilibrium (LD) score regression (LDSC)[18] with reference to the 1000G EUR panel to estimate the SNP heritability (h_g^2) of ND and its genetic correlations with 45 other complex phenotypes, including other smoking, drug, and alcohol use and dependence traits, smoking-related health consequences (e.g., cancer, COPD, and coronary heart disease), psychiatric and neurologic disorders, cognitive and educational traits, and brain volume metrics. The full list of phenotypes and GWAS datasets, as obtained from LD Hub[59] or shared by the original study investigators, are provided in **Supplementary Table 3**.

Similarly, EUR-specific GWAS meta-analysis summary statistics were input into stratified LDSC, as applied to specifically expressed genes (LDSC-SEG),[21] with reference to 205 tissues and cell types from two sources—RNA-sequencing data on 53 tissues/cell types in the Genotype-Tissue Expression [GTEx (latest data available on version 7)] resource[60] and array-based data on 152 tissues/cell types made available in Gene Expression Omnibus[61,62] (see full list in **Supplementary Table 4**). Similarly to the initial application of LDSC-SEG,[21] these two sources were selected because their expression data included a wide range of ND-relevant and other tissues and cell types in humans, as opposed to focused information on a particular tissue. LDSC-SEG involved comparing expression of each gene in each tissue/cell type with that in other tissues/cell types, selecting the top 10% of differentially expressed genes, annotating SNPs from the GWAS summary statistics that lie within 100kb windows of the selected genes, and using the stratified LDSC method to estimate the enrichment in SNP heritability for ND for the given gene set compared to the baseline LDSC model with all genes. For each analysis, a Bonferroni correction was applied to assess statistical significance: $P<0.0011$ ($\alpha=0.05/45$ phenotypes) for LDSC and $P<2.4\times10^{-4}$ ($\alpha=0.05/205$ tissues/cell types) for LDSC-SEG.

cis-eQTL assessment of novel ND-associated SNPs

To assess evidence for SNP-gene associations, novel SNPs were queried against GTEx (version 8) *cis*-expression quantitative trait loci (*cis*-eQTL) results derived from SNP genotype and RNA-sequencing data across 44 tissues (N=126–209 for the 13 brain tissues).[60] The GTEx portal (https://gtexportal.org/home/) presents significant single-tissue *cis*-eQTLs, based on a false discovery rate (FDR) <5%.

We also assessed single-tissue *cis*-eQTL evidence from the BrainSeq consortium that includes larger sample sizes with SNP genotype and RNA-sequencing data available in two brain tissues, dorsolateral prefrontal cortex (N=453) and hippocampus (N=447).[19] Of the 551 individuals with data available in at least one brain tissue, 286 were schizophrenia cases; case/control status was included as a covariate for adjustment in the *cis*-eQTL analysis, as described elsewhere.[63] Significant *cis*-eQTLs at FDR <10% are available at http://eqtl.brainseq.org/phase2/eqtl/.

Data Availability

The prior meta-analysis summary statistics[14] are available via dbGaP: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001532.v1.p1.The summary statistics generated from the current study will be included under version 2 of this dbGaP study, or are available upon request to the corresponding author (D.B.H.).

Conflicts of Interest

L.J.B. and the spouse of N.L.S. are listed as inventors on U.S. Patent 8,080,371, 'Markers for Addiction' covering the use of certain SNPs in determining the diagnosis, prognosis and treatment of addiction. Y.G. is an employee of GeneCentric Therapeutics. Although unrelated to this research, H.R.K. has been a consultant or advisory board member for Lundbeck and Indivior

and is a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was supported in the last 3 years by AbbVie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, Arbor and Amygdala Neurosciences. H.R.K. and J.G. are named as inventors on PCT patent application #15/878,640 entitled: "Genotype-guided dosing of opioid agonists," filed January 24, 2018. J.K. consulted for Pfizer in 2012–2015 on ND. All other authors declare no conflict of interest.

Acknowledgements

We are grateful to the many study participants, who made this work possible. The meta-analysis was supported by the National Institute on Drug Abuse grant numbers R01 DA042090 (PI: DBH) and R01 DA036583 (PI: LJB). The authors thank deCODE Genetics / Amgen and its investigators (Gunnar W. Reginsson, Thorgeir E. Thorgeirsson, and Kari Stefansson) for contributing and analyzing their data for inclusion in the meta-analysis. Their work was supported in part by NIDA R01 DA017932 (PI: Kari Stefansson). Acknowledgments for all other ND studies, contributed by the authors and/or made publicly available, are included in the Supplementary Information. This research also leveraged the UK Biobank Resource under Application Number 24603.

References

1. Organization, W.H. WHO report on the global tobacco epidemic, 2017: monitoring tobacco use and prevention policies. (Geneva, 2017).
2. Services, U.S.D.o.H.a.H. *The Health Consequences of Smoking-50 Years of Progress: A Report of the Surgeon General*, (Atlanta (GA), 2014).
3. Sullivan, P.F. & Kendler, K.S. The genetic epidemiology of smoking. *Nicotine Tob Res* **1 Suppl 2**, S51-7; discussion S69-70 (1999).
4. Agrawal, A. *et al.* The genetics of addiction-a translational perspective. *Transl Psychiatry* **2**, e140 (2012).
5. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* **51**, 237-244 (2019).

6. Hancock, D.B., Markunas, C.A., Bierut, L.J. & Johnson, E.O. Human Genetics of Addiction: New Insights and Future Directions. *Curr Psychiatry Rep* **20**, 8 (2018).
7. Baker, T.B. *et al.* Are tobacco dependence and withdrawal related amongst heavy smokers? Relevance to conceptualizations of dependence. *J Abnorm Psychol* **121**, 909-21 (2012).
8. Zelman, D.C., Brandon, T.H., Jorenby, D.E. & Baker, T.B. Measures of affect and nicotine dependence predict differential response to smoking cessation treatments. *J Consult Clin Psychol* **60**, 943-52 (1992).
9. Gu, F. *et al.* Time to smoke first morning cigarette and lung cancer in a case-control study. *J Natl Cancer Inst* **106**, dju118 (2014).
10. Guertin, K.A. *et al.* Time to First Morning Cigarette and Risk of Chronic Obstructive Pulmonary Disease: Smokers in the PLCO Cancer Screening Trial. *PLoS One* **10**, e0125973 (2015).
11. Fagerstrom, K. Determinants of tobacco use and renaming the FTND to the Fagerstrom Test for Cigarette Dependence. *Nicotine Tob Res* **14**, 75-8 (2012).
12. Heatherton, T.F., Kozlowski, L.T., Frecker, R.C. & Fagerstrom, K.O. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict* **86**, 1119-27 (1991).
13. Hancock, D.B. *et al.* Genome-wide meta-analysis reveals common splice site acceptor variant in CHRNA4 associated with nicotine dependence. *Transl Psychiatry* **5**, e651 (2015).
14. Hancock, D.B. *et al.* Genome-wide association study across European and African American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence. *Mol Psychiatry* **23**, 1911-1919 (2018).
15. Thorgeirsson, T.E. *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452**, 638-42 (2008).
16. Bierut, L.J. *et al.* Variants in nicotinic receptors and risk for nicotine dependence. *Am J Psychiatry* **165**, 1163-71 (2008).
17. Huedo-Medina, T.B., Sanchez-Meca, J., Marin-Martinez, F. & Botella, J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods* **11**, 193-206 (2006).
18. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
19. BrainSeq, A.H.B.G.C.E.a.d.l.o. & BrainSeq, A.H.B.G.C. BrainSeq: Neurogenomics to Drive Novel Target Discovery for Neuropsychiatric Disorders. *Neuron* **88**, 1078-1083 (2015).
20. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**, D930-4 (2012).
21. Finucane, H.K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629 (2018).
22. Wain, L.V. *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* **3**, 769-781 (2015).
23. Chen, X.D., Zhu, M.X. & Wang, S.J. Expression of long non-coding RNA MAGI2AS3 in human gliomas and its prognostic significance. *Eur Rev Med Pharmacol Sci* **23**, 3455-3460 (2019).
24. Silva, J.P. *et al.* Latrophilin 1 and its endogenous ligand Lasso/teneurin-2 form a high-affinity transsynaptic receptor pair with signaling capabilities. *Proc Natl Acad Sci U S A* **108**, 12113-8 (2011).
25. Kerin, T. *et al.* A noncoding RNA antisense to moesin at 5p14.1 in autism. *Sci Transl Med* **4**, 128ra40 (2012).
26. Piper, M.E. *et al.* Refining the tobacco dependence phenotype using the Wisconsin Inventory of Smoking Dependence Motives. *J Abnorm Psychol* **117**, 747-61 (2008).

27. Piasecki, T.M., Piper, M.E. & Baker, T.B. Refining the tobacco dependence phenotype using the Wisconsin Inventory of Smoking Dependence Motives: II. Evidence from a laboratory self-administration assay. *J Abnorm Psychol* **119**, 513-23 (2010).
28. Piasecki, T.M., Piper, M.E. & Baker, T.B. Tobacco Dependence: Insights from Investigations of Self-Reported Smoking Motives. *Curr Dir Psychol Sci* **19**, 395-401 (2010).
29. Piasecki, T.M., Piper, M.E., Baker, T.B. & Hunt-Carter, E.E. WISDM primary and secondary dependence motives: associations with self-monitored motives for smoking in two college samples. *Drug Alcohol Depend* **114**, 207-16 (2011).
30. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001-6 (2014).
31. Lee, J.J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* **50**, 1112-1121 (2018).
32. Karlsson Linner, R. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat Genet* **51**, 245-257 (2019).
33. Erzurumluoglu, A.M. *et al.* Meta-analysis of up to 622,409 individuals identifies 40 novel smoking behaviour associated genetic loci. *Mol Psychiatry* (2019).
34. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* **104**, 65-75 (2019).
35. Lutz, S.M. *et al.* A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet* **16**, 138 (2015).
36. Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat Genet* **50**, 920-927 (2018).
37. Wray, N.R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* **50**, 668-681 (2018).
38. Perry, J.R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92-97 (2014).
39. Day, F.R., Ong, K.K. & Perry, J.R.B. Elucidating the genetic basis of social interaction and isolation. *Nat Commun* **9**, 2457 (2018).
40. Walters, R.K. *et al.* Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat Neurosci* **21**, 1656-1669 (2018).
41. Kranzler, H.R. *et al.* Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat Commun* **10**, 1499 (2019).
42. Reginsson, G.W. *et al.* Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. *Addict Biol* (2017).
43. Hartz, S.M. *et al.* Genetic correlation between smoking behaviors and schizophrenia. *Schizophr Res* (2017).
44. McKay, J.D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet* **49**, 1126-1132 (2017).
45. Moulton, E.A., Elman, I., Becerra, L.R., Goldstein, R.Z. & Borsook, D. The cerebellum and addiction: insights gained from neuroimaging research. *Addict Biol* **19**, 317-31 (2014).
46. Miquel, M. *et al.* Have we been ignoring the elephant in the room? Seven arguments for considering the cerebellum as part of addiction circuitry. *Neurosci Biobehav Rev* **60**, 1-11 (2016).
47. Herculano-Houzel, S. & Lent, R. Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain. *J Neurosci* **25**, 2518-21 (2005).
48. Timofeeva, M.N. *et al.* Genetic polymorphisms in 15q25 and 19q13 loci, cotinine levels, and risk of lung cancer in EPIC. *Cancer Epidemiol Biomarkers Prev* **20**, 2250-61 (2011).

49. Martin, J., Taylor, M.J. & Lichtenstein, P. Assessing the evidence for shared genetic risks across psychiatric disorders and traits. *Psychol Med* **48**, 1759-1774 (2018).
50. Hjelmborg, J. *et al.* Lung cancer, genetic predisposition and smoking: the Nordic Twin Study of Cancer. *Thorax* **72**, 1021-1027 (2017).
51. Conway, K.P. *et al.* Data compatibility in the addiction sciences: an examination of measure commonality. *Drug Alcohol Depend* **141**, 153-8 (2014).
52. Glasheen, C. *et al.* Is the Fagerstrom test for nicotine dependence invariant across secular trends in smoking? A question for cross-birth cohort analysis of nicotine dependence. *Drug Alcohol Depend* **185**, 127-132 (2018).
53. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
54. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-7 (2010).
55. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
56. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
57. Boyles, A.L., Harris, S.F., Rooney, A.A. & Thayer, K.A. Forest Plot Viewer: a new graphing tool. *Epidemiology* **22**, 746-7 (2011).
58. DiFranza, J.R. *et al.* What aspect of dependence does the fagerstrom test for nicotine dependence measure? *ISRN Addict* **2013**, 906276 (2013).
59. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279 (2017).
60. Consortium, G.T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).
61. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
62. Fehrmann, R.S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet* **47**, 115-25 (2015).
63. Collado-Torres, L. *et al.* Regional Heterogeneity in Gene Expression, Regulation, and Coherence in the Frontal Cortex and Hippocampus across Development and Schizophrenia. *Neuron* **103**, 203-216 e8 (2019).

Table 1. Lead single nucleotide polymorphism (SNP) associations from the five genome-wide significant loci in the Nicotine Dependence GenOmics (iNDiGO) consortium cross-ancestry meta-analysis for nicotine dependence (ND). Ancestry-specific association results are also presented.

			European ancestry-specific ND meta-analysis (total N = 46,213)			**African American-specific ND meta-analysis (total N = 11,787)**			**Cross-ancestry ND meta-analysis (total N = 58,000)**	
SNP (effect allele)	**Chr:position (NCBI build 37)**	**Gene / closest genes**	**Effect allele freq.[a]**	**β (SE)**	**P**	**Effect allele freq.[a]**	**β (SE)**	**P**	**β (SE)**	**P**
Lead SNPs from novel ND-associated loci										
rs1862416 (T)	5:167,394,595	*TENM2*	0.88	0.037 (0.0074)	5.4×10^{-7}	0.94	0.049 (0.0066)	6.6×10^{-3}	0.039 (0.0068)	1.5×10^{-8}
rs2714700 (T)	7:79,367,667	*MAGI2 / GNAI1*	0.47	-0.022 (0.0045)	1.2×10^{-6}	0.72	-0.026 (0.0094)	5.5×10^{-3}	-0.023 (0.0041)	2.3×10^{-8}
Lead SNPs from known ND-associated loci										

rs13284520 (A)	9:136,502,572	*DBH*	0.83	0.028 (0.0059)	1.7×10^{-6}	0.56	0.029 (0.0092)	1.7×10^{-3}	0.029 (0.0050)	1.1×10^{-8}
rs16969968 (A)	15:78,882,925	*CHRNA5*	0.37	0.061 (0.0047)	4.9×10^{-38}	0.02	0.049 (0.018)	7.1×10^{-3}	0.060 (0.0046)	1.6×10^{-39}
rs151176846 (T)	20:61,997,500	*CHRNA4*	0.92	-0.067 (0.0094)	1.2×10^{-12}	1.00	NA	NA	0.067 (0.0094)	1.2×10^{-12}

Abbreviations: NA, not available (due to monomorphism for rs151176846 among African Americans); NCBI, National Center for Biotechnology Information; SE, standard error.

[a] Frequencies correspond to 1000G European and African superpopulation reference panels.

Table 2. Single nucleotide polymorphisms (SNPs) identified as genome-wide significant for cigarettes per day (CPD) by the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) consortium and associated with nicotine dependence (ND) at $P<9.1\times10^{-4}$ ($\alpha=0.05/55$ tests) in the cross-ancestry meta-analysis by the Nicotine Dependence GenOmics (iNDiGO) consortium. Results are sorted by novelty and then by iNDiGO p-values, and β values correspond to direction of association for the effect alleles.

SNP (effect allele)	Chr:position (NCBI build 37)	Gene / nearest gene(s)	GSCAN consortium meta-analysis for CPD (N=330,721)			iNDiGO consortium meta-analysis for ND (N=58,000)		
			β	SE	P	β	SE	P
SNPs from loci not reported by prior GWAS of ND								
rs7125588 (G)	11:113,436,072	*DRD2* / *TMPRSS5*	-0.014	0.0020	6.5×10^{-12}	-0.016	0.0042	1.8×10^{-4}
rs1592485 (A)	16:52,093,549	*C16orf97*	-0.013	0.0021	1.1×10^{-10}	-0.015	0.0043	4.5×10^{-4}
rs2072659 (G)	1:154,548,521	*CHRNB2*	-0.025	0.0038	2.5×10^{-13}	-0.026	0.0078	8.4×10^{-4}
SNPs from loci reported by prior GWAS of ND								
rs146009840 (T)	15:78,906,177	*CHRNA3*	0.030	0.0036	2.0×10^{-17}	0.060	0.0046	2.6×10^{-39}
rs72740955 (T)	15:78,849,779	*PSMA4* / *CHRNA5*	0.040	0.0033	2.4×10^{-34}	0.058	0.0045	1.5×10^{-38}
rs10519203 (A)	15:78,814,046	*HYKK*	-0.075	0.0021	3.1×10^{-286}	-0.050	0.0042	7.7×10^{-32}

rs8040868 (C)	15:78,911,181	*CHRNA3*	0.022	0.0034	1.8×10^{-10}	0.044	0.0041	7.3×10^{-27}
rs12438181 (A)	15:78,812,098	*HYKK*	-0.023	0.0037	5.0×10^{-10}	-0.039	0.0049	2.6×10^{-15}
rs3743063 (C)	15:79,065,171	*ADAMTS7*	-0.023	0.0035	1.5×10^{-11}	-0.030	0.0042	6.8×10^{-13}
rs28681284 (T)	15:78,908,565	*CHRNA3*	-0.049	0.0030	2.1×10^{-58}	-0.035	0.0051	1.1×10^{-11}
rs2273500 (C)	20:61,986,949	*CHRNA4*	0.031	0.0029	3.5×10^{-26}	0.034	0.0058	4.0×10^{-9}
rs3025383 (C)	9:136,502,369	*DBH*	-0.026	0.0026	9.8×10^{-24}	-0.025	0.0049	1.8×10^{-7}
rs28438420 (T)	15:78,836,288	*PSMA4*	0.020	0.0028	1.3×10^{-12}	0.020	0.0041	7.9×10^{-7}
rs75596189 (T)	9:136,468,701	*FAM163B* / *DBH*	0.035	0.0037	1.8×10^{-20}	0.030	0.0066	8.1×10^{-6}
rs4236926 (G)	8:42,578,059	*CHRNB3*	0.028	0.0024	7.7×10^{-33}	0.021	0.0048	1.6×10^{-5}
rs56113850 (C)	19:41,353,107	*CYP2A6*	0.043	0.0021	4.0×10^{-99}	0.018	0.0042	2.1×10^{-5}
rs1737894 (G)	20:31,054,702	*NOL4L*	0.014	0.0021	9.9×10^{-12}	0.017	0.0043	1.1×10^{-4}

Abbreviations: NCBI, National Center for Biotechnology Information; SE, standard error.

Figure 1. Cross-ancestry nicotine dependence genome-wide association meta-analysis results, comprising 23 iNDiGO studies with total N = 58,000 European and African American ancestry ever smokers. The $-\log_{10}$ meta-analysis p-values of single nucleotide polymorphisms (SNPs; depicted as circles) and insertions/deletions (indels; depicted as triangles) are plotted by chromosomal position. Five loci surpassed the genome-wide statistical significance threshold ($P<5\times10^{-8}$, as marked by the solid horizontal black line).

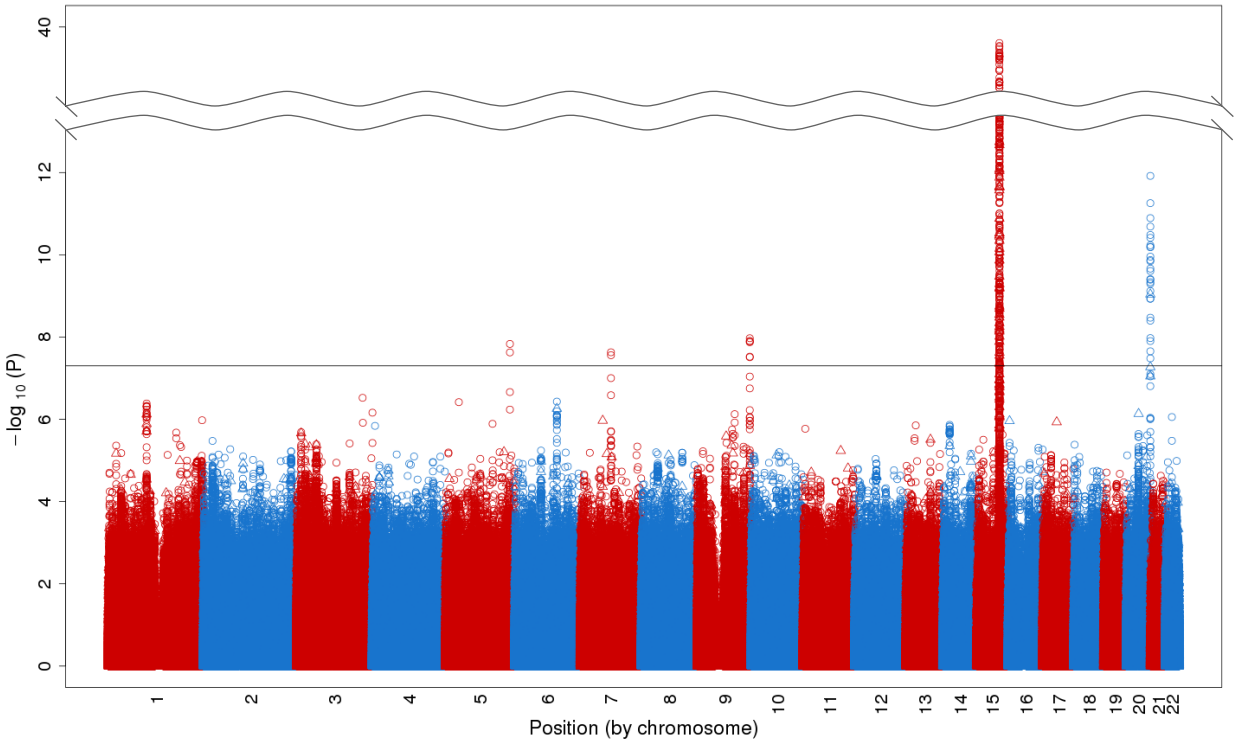
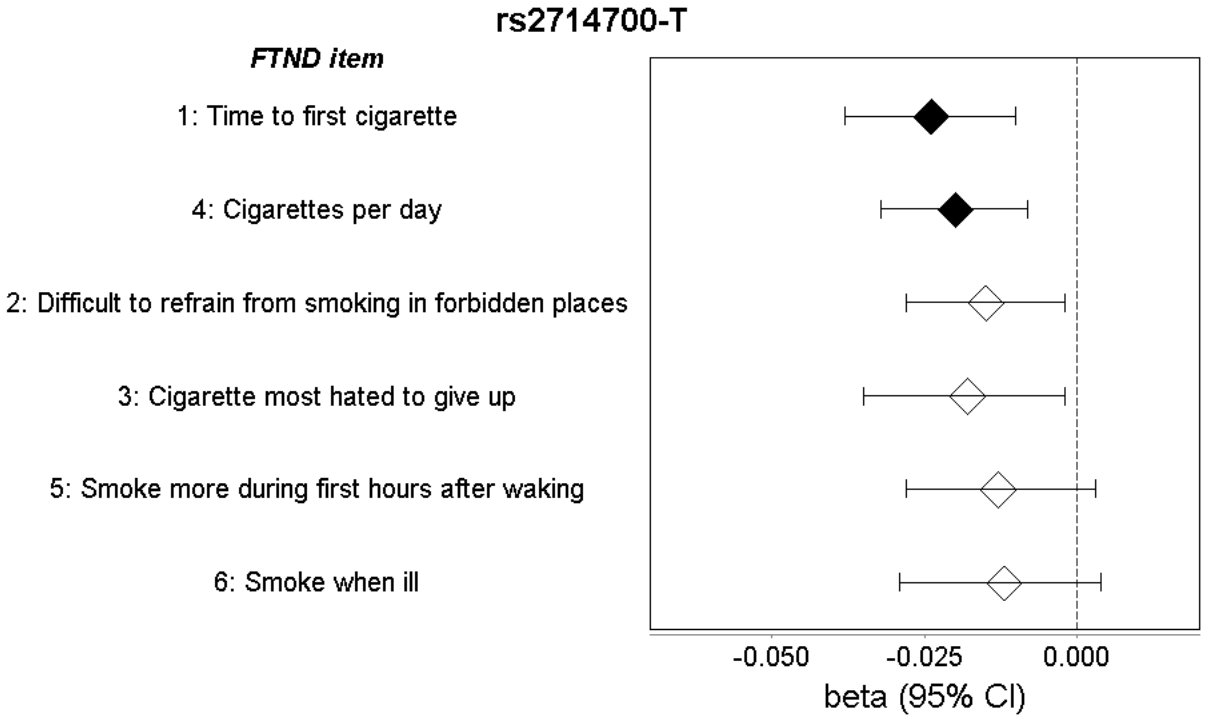


Figure 2. Associations of novel single nucleotide polymorphisms (SNPs) with specific items of the Fagerström Test for Nicotine Dependence (FTND) across the iNDiGO studies.

Associations are presented from cross-ancestry meta-analyses of the (A) *MAGI2/GNAI1* SNP allele rs2714700-T and (B) *TENM2* SNP allele rs1862416-T. Beta (β) and corresponding 95% confidence interval (CI) estimates were taken from linear regression models for categorical FTND item responses (1 and 4, closed diamonds) or logistic regression models for binary FTND item responses (2, 3, 5, and 6, open diamonds).

(A)



(B)

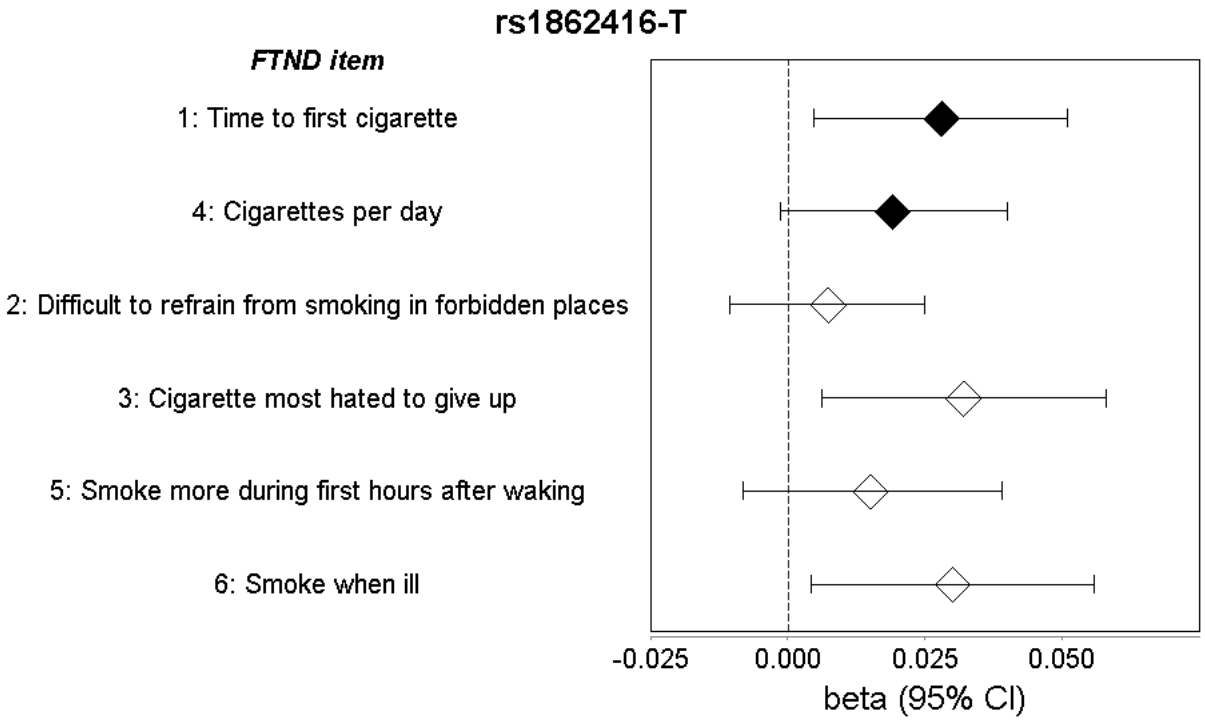


Figure 3. **Genetic correlations of nicotine dependence (ND) with 45 other phenotypes.** Correlations were calculated using linkage disequilibrium (LD) score regression with the iNDiGO European ancestry-specific GWAS meta-analysis results for ND (N=46,213), compared with results made available via LD Hub or study investigators (see Supplementary Table 3 for original references). Phenotypes were grouped by disease/trait or measurement category, as indicated by different colorings. Point estimates equate to genetic correlation (r_g) values; error bars show the 95% confidence intervals; and the dotted vertical grey line corresponds to r_g=0 (no correlation with ND). Phenotypes with significant correlations (P<0.0011, α=0.05/45 tested) are bolded.