Published in final edited form as:
Mol Psychiatry. 2020 October 01; 25(10): 2392–2409. doi:10.1038/s41380-018-0313-0.

Meta-analysis of up to 622,409 individuals identifies 40 novel smoking behaviour associated genetic loci

A full list of authors and affiliations appears at the end of the article.

Abstract

Smoking is a major heritable and modifiable risk factor for many diseases, including cancer, common respiratory disorders and cardiovascular diseases. Fourteen genetic loci have previously been associated with smoking behaviour-related traits. We tested up to 235,116 single nucleotide variants (SNVs) on the Exome-array for association with smoking initiation, cigarettes per day, pack-years, and smoking cessation in a fixed effects meta-analysis of up to 62 studies (346,813 participants). In a subset of 112,811 participants, a further one million SNVs were also genotyped and tested for association with the four smoking behaviour traits. SNV-trait associations with $P<5x10^{-8}$ in either analysis were taken forward for replication in up to 275,596 independent participants from UK Biobank. Lastly, a meta-analysis of the discovery and replication studies was performed.

Sixteen SNVs were associated with at least one of the smoking behaviour traits ($P<5x10^{-8}$) in the discovery samples. Ten novel SNVs, including rs12616219 near *TMEM182*, were followed-up and five of them (rs462779 in *REV3L*, rs12780116 in *CNNM2*, rs1190736 in *GPR101*, rs11539157 in *PJA1*, and rs12616219 near *TMEM182*) replicated at a Bonferroni significance threshold ($P<4.5x10^{-3}$) with consistent direction of effect. A further 35 SNVs were associated with smoking behaviour traits in the discovery plus replication meta-analysis (up to 622,409 participants) including a rare SNV, rs150493199, in *CCDC141* and two low-frequency SNVs in *CEP350* and *HDGFRP2*. Functional follow-up implied that decreased expression of *REV3L* may lower the probability of smoking initiation. The novel loci will facilitate understanding the genetic aetiology of smoking behaviour and may lead to identification of potential drug targets for smoking prevention and/or cessation.

[#]Correspondence: Joanna M M Howson, jmmh2@medschl.cam.ac.uk and Dajiang J Liu dxl46@psu.edu.
[*]Indicates that these authors contributed equally to this work and share the first author position
[†]Indicates that these authors contributed equally to this work and share the last author position

Conflict of Interest Statement

Paul W. Franks has been a paid consultant for Eli Lilly and Sanofi Aventis and has received research support from several pharmaceutical companies as part of European Union Innovative Medicines Initiative (IMI) projects. Neil Poulter has received financial support from several pharmaceutical companies that manufacture either blood pressure lowering or lipid lowering agents or both and consultancy fees. Peter Sever has received research awards from Pfizer. Mark J. Caulfield is Chief Scientist for Genomics England, a UK government company.

Introduction

Smoking is a major risk factor for many diseases, including common respiratory disorders such as chronic obstructive pulmonary disease (COPD)1, 2, cancer3 and cardiovascular diseases4, and is reported to cause 1 in 10 premature deaths worldwide5. A greater understanding of the genetic aetiology of smoking behaviour has the potential to lead to new therapeutic interventions to aid smoking prevention and cessation, and thereby reduce the global burden of such diseases.

Previous genome-wide association studies (GWASs) identified 14 common SNVs1, 6–12 (with minor allele frequency, MAF>0.01) robustly associated with smoking behaviour related traits ($P<5x10^{-8}$). The 15q25 (*CHRNA3/5-CHRNB4*) region has the largest effect, explaining ~1% and 4-5% of the phenotypic variance of smoking quantity13 and cotinine, a biomarker of nicotine intake14, respectively. Overall, genetic loci identified to date explain ~2% of the estimated genetic heritability of smoking behaviour6, which is reported to be between 40-60%15–17. A recent study suggested that an important proportion (~3.3%) of the phenotypic variance of smoking behaviour related traits was explained by rare nonsynonymous variants (MAF<0.01)18. Hence, well-powered studies of rare variants are needed.

To investigate the effect of rare coding variants on smoking behaviour, we studied 346,813 participants (of which 324,851 were of European ancestry) from 62 cohorts (Supp. Tables 1 and 2) at up to 235,116 SNVs from the exome array. As we had access to UK Biobank, we also interrogated SNVs present on the UK Biobank and UK BiLEVE Axiom arrays to identify additional associations across the genome beyond the exome array. To our knowledge, these datasets are an order of magnitude larger than the previous studies6, and constitute the most powerful exome-array study of smoking behaviour to date.

Materials and Methods

Participants

Our study combined study-level summary association data from up to 60 studies of European ancestry and two studies of South Asian ancestry from three consortia (CGSB (Consortium for Genetics of Smoking Behaviour), GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN) and the Coronary Heart Disease (CHD) Exome+ consortium) INTERVAL and UK Biobank. In total, up to 324,851 individuals of European ancestry and 21,962 South Asian individuals were analysed in the discovery stage (Figure 1). Further information about the participating cohorts and consortia is given in Supp. Table 1 and the Supp. Material. All participants provided written informed consent and studies were approved by local Research Ethics Committees and/or Institutional Review boards.

Phenotypes

We chose to analyse the following four smoking behaviour related traits because of their broad availability in existing epidemiological and medical studies, as well as their biological relevance for addiction behaviours:

i) Smoking initiation (binary trait: ever vs never smokers). Ever smokers were defined as individuals who have smoked >99 cigarettes in their lifetime, which is consistent with the definition by the Centre for Disease Control19;

ii) Cigarettes per day (CPD; quantitative trait: average number of cigarettes smoked per day by ever smokers);

iii) Pack-years (quantitative trait; Packs per day x Years smoked, with a pack defined as 20 cigarettes); years smoked is typically formed from age at smoking commencement to current age for current smokers or age at cessation for former smokers.

iv) Smoking cessation (binary trait: former vs current smokers).

In UK Biobank, phenotypes were defined using phenotype codes 1239, 1249, and 2644 for smoking initiation and smoking cessation, and 1239, 3436, 3456 for CPD and pack-years. CPD was inverse normal transformed in the CHD Exome+, INTERVAL and CGSB studies and categorised (1-10, 11-20, 21-30, and 31+ CPD) by the GSCAN studies and UK Biobank (Supp. Table 2). All studies performed an inverse normal transformation of pack-years. Summary statistics of study level phenotype distributions are provided in Supp. Table 1.

Genotyping and quality control

Fifty-nine cohorts were genotyped using exome arrays (up to 235,116 SNVs) and two (UK Biobank and INTERVAL) were genotyped using Axiom Biobank Arrays (up to 820,000 SNVs; Supp. Table 2). In total, ~1.06M SNVs were analysed including ~64,000 SNVs on both the Axiom and Exome Arrays. Furthermore, two studies (NAGOZALC and GFG) genotyped their participants using arrays with custom content, increasing the total number of variants analysed to 1,207,583 SNVs. Individual studies performed quality control (QC; Supp. Material, Supp. Table 2) and additional QC was conducted centrally (i) to ensure alleles were consistently aligned, (ii) that there were no major sample overlaps between contributing studies, and (iii) variants conformed to Hardy-Weinberg equilibrium and call rate thresholds. We also examined the distribution of the effect sizes and test statistics across cohorts to ensure the test statistics were well-calibrated.

Study level analyses

Each study (including the case-cohort studies20) undertook analyses of up to four smoking traits using RAREMETALWORKER21 or RVTESTS22 (Supp. Table 2), which generated single variant score statistics and their covariance matrices within sliding windows of 1Mb. CPD and pack-years were analysed using linear models or linear mixed models. Smoking initiation and smoking cessation were analysed using logistic models or linear mixed models. All studies adjusted each trait for age, sex, at least three genetic principal components and any study-specific covariates (Supp. Table 2). Chromosome X variants were analysed using the above described approach, but coding males as 0/2. This coding scheme ensures that on average females and males have equal dosages and so is optimal for genes that are inactivated (due to X chromosome inactivation) and is valid for genes that do not undergo X chromosome activation. Males and females were analysed together adjusting for sex as a covariate.

Single variant meta-analyses

Fixed effects meta-analyses across the individual contributing studies of single variant associations were undertaken using the Cochran-Mantel-Haenszel method in RAREMETAL. Z-score statistics were used in the meta-analysis to ensure that the association results are robust against potentially different units of measurement in the phenotype definitions across studies23. We performed genomic control correction on the meta-analysis results. Variants with *P*<1x10^{-6} in tests of heterogeneity were excluded. Variants with *P*≤5x10^{-8} were taken forward for replication. In addition, rs12616219 was also taken forward for replication as its *P*-value was very close to this threshold (smoking initiation, *P*=5.49x10^{-8}). None of the rare SNVs were genome-wide significant, therefore we also took forward the rare variant with the smallest association *P*-value, rs141611945 (*P*=2.95x10^{-7}; MAF<0.0001).

Replication and combined meta-analysis of discovery and replication data

As UK biobank genetic data were released in two phases, we took the opportunity to replicate findings from the discovery stage in a further 275,596 individuals made available in the phase two release of UK Biobank genetic data. To avoid potential relatedness between discovery and replication samples, the replication samples were screened and individuals with relatedness closer than second degree with the discovery sample in the UK Biobank were removed 24. Phenotypes were defined in the same way as the discovery samples (described above). Since the exome array and the UK Biobank Axiom arrays do not fully overlap, we used both genotyped exome variants (approx. 64,000) as well as the additional ~90,000 well imputed exome array variants from UK Biobank (imputation quality score>0.3) for replication of single variant and gene-based tests. The rare *ATF6* variant was absent from the UK Biobank array and is more prevalent in Africans (MAF=0.01) than Europeans (MAF=0.0007). Therefore, replication was sought in 1,437 individuals of African American-ancestry from the HRS and COGA studies. Analysis methods for replication cohorts were the same as for discovery cohorts, including methods to analyse chromosome X (Supp. Table 2). The criteria set for the replication were (i) the same direction of effect as the discovery analysis and (ii)*P*≤0.0045 in the replication studies (i.e. Bonferroni-adjusted for eleven SNVs at α=0.05).

Finally, in order to fully utilise all available data, we carried out a combined meta-analysis of the discovery and replication samples across the exome array content using the same protocols mentioned above.

Conditional analyses

To identify conditionally independent variants associations within previously reported and novel loci a sequential forward stepwise selection was performed25. A 1MB region was defined around the reported or novel sentinel variant (500kb either side) and conditional analyses performed with all variants within the region. If a conditionally independent variant was identified, (*P*<5x10^{-6}; Bonferroni adjusted for ~10,000 independent variants in the test region) the analysis was repeated conditioning on both the most significant conditionally independent variant and the sentinel variant. This stepwise approach was repeated (conditioning on the variants identified in current and earlier iterations) until there were no variants remaining in the region that were conditionally independent. The same protocol was followed for the novel SNVs identified in this study.

Gene-based analyses

For discovery gene-based meta-analyses, we utilised three statistical methods as part of the RAREMETAL package: the Weighted Sum Test (WST)26, the burden test27 and the Sequence Kernel Association test (SKAT)28. EPACTS (v.3.3.0)29 was used to annotate variants (for use in gene-based meta-analyses), as recommended by RAREMETAL. Two MAF cut-offs were used, one used low frequency (MAF<0.05) and rare variants, the second only used rare variants (MAF<0.01). Nonsynonymous, stop gain, splice site, start gain, start loss, stop loss, and synonymous variants were selected for inclusion. A sensitivity analysis to exclusion of synonymous variants was also performed. Gene-level associations with $P<8x10^{-7}$ were deemed statistically significant (Bonferroni-adjusted for ~20,000 genes and three tests at $\alpha=0.05$). To examine if the gene associations were driven by a single variant, the gene tests were conducted conditional on the SNV with the smallest *P*-value in the gene, using the shared single variant association statistic and covariance matrices21, 25.

Mendelian Randomization analyses

To evaluate the causal effect of SI and CPD on BMI, schizophrenia and educational attainment (EA), we conducted Mendelian randomization (MR) analyses using three complementary approaches available in MR-Base30: inverse variance weighted regression31, MR-Egger32, 33, and weighted median34. We used both the previously reported smoking associated SNVs and the SNVs from the current report (as provided in Tables 1-3 and Supp. Table 3) as instrumental variables. The BMI35, schizophrenia36 and educational attainment37 data came from previously published publicly available data. To assess possible reverse causation, we also used outcome associated SNVs as instrumental variables and conducted MR analyses using SI and CPD as outcome. We considered $P<0.05/3=0.017$ as statistically significant (Bonferroni adjusted for three traits).

In silico functional follow up of associated SNVs

To identify whether the (replicated) SNVs identified here affected other traits, we queried the GWAS Catalog38 (version: e91/28/02/2018, downloaded on 01/03/18) for genome-wide significant ($P<5x10^{-8}$) associations using all proxy SNVs ($r^2\geq0.8$) within 2Mb of the top variant in our study.

eQTL lookups were carried out in the 13 brain tissues available in GTEx V739, Brain xQTL (dorsolateral prefrontal cortex)40 and BRAINEAC41 databases, all of which had undergone QC by the individual studies. We did not perform additional QC on these data. In brief, GTEx used Storey’s q-value method to correct the FDR for testing multiple transcripts based upon the empirical *P*-values for the most significant SNV for each transcript43. BRAINEAC calculated the number of tests per transcript and used Benjamini-Hochberg procedure to calculate FDR per transcript using a FDR<1% as significant. BRAINxQTL used $P<8x10^{-8}$ as a cut-off for significance for any given transcript. SNVs that met the study specific significance and FDR thresholds, which were in LD ($r^2>0.8$ in 1000 Genomes Europeans) with the top eQTL or the sentinel eQTL for a given tissue/transcript combination were considered significant. The genes implicated by these eQTL databases and/or coding changes (e.g. missense and nonsense SNVs) were put into ConsensusPathDB44 to identify whether these genes were over-represented in any known biological pathways. Replicated missense SNVs were also put into PolyPhen-245 and FATHMM (unweighted)46 to obtain variant effect prediction.

Results

Single variant associations

In the discovery meta-analyses, we identified 14 common SNVs that were genome-wide significant (P<5x10^{-8}) for one or more of the smoking behaviour traits, of which 9 were novel (Table 1, Supp. Table 3). Seven novel loci were identified for smoking initiation, one for both CPD and pack-years and one for smoking cessation (Figures 1, 2, Table 1 and Supp. Figure 1). Results for the significant loci were consistent across participating cohorts and there was at least nominal evidence of association (P<0.05) at the novel loci within each of the contributing consortia (Supp. Table 4). Full association results for all novel SNVs across the four traits are provided in Supp. Table 5. No rare variants were genome-wide significant; the rare variant with the smallest P-value was a missense variant in *ATF6*, rs141611945 (MAF<0.0001, CPD P=2.95x10^{-7}).

Eleven SNVs (including rs12616219 near *TMEM182* with P=5.49x10^{-8}, and the rare variant, rs141611945) were taken forward for replication in independent samples (Table 1). The latest release of European UK Biobank individuals not included in the discovery stage (smoking initiation, n=275,596; smoking cessation n=123,851; CPD n=80,015; pack-years n=78,897), was used for replication of the common variants (Figure 1). Five of the common variants replicated (four for smoking initiation and one with CPD and pack-years) at P<0.0045. Two coding variants (rs11539157, rs1190736) were predicted to be ‘probably damaging’ by PolyPhen-2 and FATHMM. The remaining five SNVs were at least nominally associated (P<0.01) in the replication samples and had consistent direction of effect across discovery and replication. Replication for the rare variant rs141611945 could not be carried out in UK Biobank as the SNV nor its proxies (r^2>0.3) were available. Thus we initiated replication in African American samples of the COGA (n=476) and HRS (n=961) cohorts (overall MAF≈0.01). The direction of effect was consistent in the two replication cohorts and consistent with the discovery meta-analysis but a meta-analysis of the two replication cohorts yielded a P=0.28. Further data are required to replicate this association.

We also performed a meta-analysis combining the discovery and replication samples (up to 622,409 individuals). LD score regression showed that the λ (intercept) for all traits was ~1.00, which indicated that confounding factors inflating the results was not an issue47, 48. The combined analysis identified 35 additional novel SNV-smoking trait associations, 33 with smoking initiation, one with CPD and one with smoking cessation at P<5x10^{-8} (Table 2). We note that among our four SNVs that did not replicate, rs216195 (in *SMG6*) was genome-wide significant in the combined meta-analysis of discovery and replication studies (P=2.41x10^{-9}; Table 2).

We also calculated the phenotypic variance explained for novel and known variants. Results can be found in the ‘Calculation of Phenotypic Variance Explained’ section in the Supplementary Material.

Associations at known smoking behaviour loci

We assessed evidence for associations at the 14 SNVs previously reported for smoking behaviour-related traits. Seven were genotyped on the exome array and proxies (r^2>0.3; ±2Mb) were identified for the remaining seven (Supp. Table 3). All showed nominal evidence of association at P<0.05 and five of these were genome-wide significant in the meta-analysis of the trait for which it was previously reported (Supp. Table 3 and 5).

Conditional analyses identified five independent associations within three previously reported loci and all five replicated (Table 3). At the 19q13 (*RAB4B*) locus, there were three variants in or near *CYP2A6* associated with CPD independently of the established variant (rs7937) and each other: rs8102683 (conditional P=4.53x10^{-16}), rs28399442 (conditional P=2.63x10^{-12}) and rs3865453 (conditional P=4.96x10^{-10}) and rs28399442 was a low frequency variant. The same SNVs also showed evidence of independent effects with pack-years, albeit with larger P-values (P<5x10^{-6}; Supp. Table 5). At the *TEX41*/*PABPC1P2* locus, rs11694518 (conditional P=3.43x10^{-7}) was associated with smoking initiation independently of the established variant (rs10427255). At 15q25, rs938682 (P=7.78x10^{-21}) was associated with CPD independently of the established variant (rs1051730) and (in agreement with a previous report49) is an eQTL for *CHRNA5* in brain putamen basal ganglia tissues in GTEx.

Gene-based association studies

Gene-based collapsing tests using MAF<0.01 variants, did not identify any associated genes at the pre-specified P<8x10^{-7} threshold. Of the top four gene associations, three were novel (*CHRNA2*, *MMP17*, and *CRCP*) and one was known (*CHRNA5*), and had P<7x10^{-4}, with CPD and/or pack-years (Supp. Table 6). Analyses conditional on the variant with the smallest P-value in the gene, revealed the associations at *CHRNA2*, *MMP17* and *CRCP* were due to more than one rare variant (conditional P<0.05; Supp. Table 6). In contrast, the *CHRNA5* gene association was attributable to a single variant (rs2229961).

Mendelian Randomization analyses

We conducted MR analyses to elucidate the potential causal impact of SI and CPD on BMI, schizophrenia and EA using the MR-Egger, median weighted and inverse variance weighted methods. We found a causal association between SI and EA using both the median weighted and inverse variance weighted methods (P<0.0001; Supp. Table 7) but not with MR-Egger (P=0.2). There was an association of SI with BMI using MR-Egger only (P=0.01; Supp. Table 7), but there was evidence of horizontal pleiotropy (P=0.001) and no support from the other methods. Similarly, increased CPD was only associated with reduced BMI using the weighted median approach (P=0.009) and not the other methods (P>0.017). We also tested if schizophrenia, EA or BMI causally influence CPD or SI using SNVs associated with schizophrenia, EA and BMI, respectively, as instrumental variables. No evidence of such reverse causation was found (Supp. Table 7). These results were consistent with previous analyses50. There was no evidence of a causal effect of SI on schizophrenia, or CPD on educational attainment (Supp. Table 7).

Functional characterization of novel loci

Using proxies with r^2≥0.8 in1000 Genomes Europeans, we queried the GWAS catalogue38 (*P*≤5x10^{-8}) for pleiotropic effects of our novel sentinel SNVs. Two, rs11539157 and rs3001723 were previously associated with schizophrenia36, suggesting shared biological pathways between schizophrenia and smoking behaviours (Table 2). This fits with the known association of smoking with schizophrenia51. Two, rs1514175 and rs2947411 have previously been associated with BMI52, and extreme obesity53.

eQTL lookups in GTEx V7 (13 Brain tissues with ≥80 samples)39, Brain xQTL40 and BRAINEAC41 databases revealed that the A allele at rs462779, which decreases risk of smoking initiation, also decreased expression of *REV3L* in cerebellum in GTEx (A allele *P*=4.8x10^{-8}; β=-0.40) and was in strong LD with the top eQTL for *REV3L* in cerebellum (r^2=0.86 with rs9487668 in 1000 Genomes Europeans). The smoking initiation-associated SNV, rs12780116, was an eQTL for *BORCS7* in four brain tissues, and NT5C2 in the cerebellar hemisphere (A allele *P*=4.5x10^{-7}; β=-0.32) and the cerebellum (*P*=5.6x10^{-6}; β=-0.415; in strong LD with the top eQTL, r^2=0.97 with rs11191546). The G allele of a second variant in the region, rs7096169 (intronic to *BORCS7* and only in weak LD with rs12780116, r^2=0.18 in 1000G Europeans) increases smoking initiation and reduces expression of *BORCS7* and *AS3MT* in eight brain tissues (including dorsolateral prefrontal cortex in the Brain xQTL and was the top *BORCS7* eSNP in GTEx in the Cerebellar Hemisphere, Cerebellum, and Spinal cord cervical-C1). The same variant also reduced expression of ARL3 in cerebellum in GTEx (Table 2).

Biological pathway enrichment analyses carried out in ConsensusPathDB44 using the genes implicated by the eQTL databases (Table 2) and/or a coding SNVs (i.e. *PJA1*, *GPR101*) showed that the (i) pyrimidine metabolism and (ii) activation of nicotinic acetylcholine receptors pathways are enriched for these smoking behaviour associated genes (false discovery rate<0.01; *P*<0.0001).

Discussion

Smoking is the most important preventable lifestyle risk factor for many diseases, including cancers3, 54, heart disease4, 55 and many respiratory diseases such as COPD1, 2. Not initiating is the best way to prevent smoking-related diseases and genetics can play a considerable part in smoking behaviours including initiation. We have performed the largest exome-wide genetic association study of smoking behaviour-related traits to date involving up to 622,409 individuals, and identified and replicated five associations, including two on the X-chromosome (Table 1). We identified a further 35 novel associations in a meta-analysis of discovery and replication cohorts (Table 2). We validated 14 previously reported SNV-smoking trait associations (Supp. Table 3) and identified secondary independent associations at three loci, including three in the 19q13 region (rs8102683, rs28399442, and rs3865453; Table 3).

Gene-based tests improve power by aggregating effects of rare variants. While no genes reached our Bonferroni-adjusted *P*-value threshold, we identified three candidate genes with multiple rare variant associations for future replication: calcitonin gene-related peptide-receptor component (*CRCP*) with CPD and *CHRNA2* and *MMP17* with pack-years (Supp. Table 6; also see 'Genes of Interest' section in Supp. Material). *CRCP*'s protein product is expressed in brain tissues amongst others and functions as part of a receptor complex for a neuropeptide that increases intracellular cyclic adenosine monophosphate levels56. *MMP17* encodes a matrix metalloproteinase that is also expressed in the brain and is a member of the peptidase M10 family, and proteins in this family are involved in the breakdown of extracellular matrix in normal physiological processes57. Given, we were not able conclusively to identify rare variant associations, even larger studies, are required to identify rare variants associated with smoking behaviours. In addition, phenotypes such as cotinine levels58 and nicotine metabolism speed59 could be interrogated using methods such as MTAG60 to improve power.

As recommended by UK Biobank, we analysed UK Biobank samples by adjusting for genotyping array because a subset of (extreme smokers in) UK Biobank were genotyped on a different array (UK BiLEVE). However, this adjustment could potentially introduce collider bias in analyses of smoking traits. Given that the UK BiLEVE study is relatively small compared to the full study, and the genetic effect sizes for smoking associated variants are small, we expect the influence of collider bias to be small70. Nevertheless, we performed sensitivity analyses to assess the impact of collider bias. Firstly, we performed a meta-analysis excluding the UK BiLEVE samples, and secondly, we re-analysed UK Biobank without adjusting for genotype array. As expected, the estimated genetic effects from these additional analyses were very similar to our reported results suggesting collider bias is not a concern (Suppl. Table 8).

Follow-up of the replicated SNVs in the literature and eQTL databases implicated some potentially interesting genes: *NT5C2* is known to hydrolyse purine nucleotides and be involved in maintaining cellular nucleotide balance, and was previously associated with schizophrenia61. *REV3L*, encodes the catalytic subunit of DNA polymerase ζ (zeta) which is involved in translesion DNA synthesis. Previously, polymorphisms in a microRNA target site of *REV3L* were shown to be associated with lung cancer susceptibility62. We showed that decreased expression of *REV3L* may also lower the probability of smoking initiation. The SNV, rs11776293, intronic in *EPHX2*, associated with reduced SI in the combined meta-analysis, and is in LD with rs56372821 (r^2=0.83), which is associated with reduced cannabis use disorder63. rs216195 (in *SMG6*) was genome-wide significant in the discovery and the combined meta-analysis. *SMG6* is a plausible candidate gene as it was previously shown to be less methylated in current smokers compared to never smokers64. The combined meta-analysis also identified a rare missense variant in *CCDC141*, rs150493199 (MAF<0.01; Table 2). Coding variants in *CCDC141* were previously associated with heart rate65 and blood pressure66, 67.

Smoking behaviours represent a complex phenotype that are linked to an array of socio-cultural and familial, as well as genetic determinants. Kong *et al.*, recently reported that 'genetic-nurture' i.e. effects of non-transmitted parental alleles, affect educational attainment68. They also show that there is an effect of educational attainment and genetic nurture on smoking behaviour. Four of our sentinel SNVs (or a strong proxy; r^2>0.8) were associated with years of educational attainment37 (rs2292239, rs3001723 (P<5x10^{-8}), rs9320995 (P=8.90x10^{-7}), and rs13022438 (P=3.79x10^{-6}), in agreement with this paradigm and our MR analyses indicated that initiating smoking reduced years in education. Future family studies will be required to disentangle how much of the variance explained in the current analysis is due to direct versus genetic nurturing effects.

Our study primarily focused on European ancestry, but we also included two non-European studies but these non-European studies lacked statistical power on their own to identify ancestry specific effects. Therefore, we did not perform ancestry specific meta-analyses. Nevertheless, our results offered cross ancestry replication. One of the associations identified in the conditional analyses, rs8102683 (near *CYP2A6*), confirmed an association with CPD that was previously identified by Kumasaka *et al.* in a Japanese population69 but this is the first time it was associated in Europeans (rs8102683 is also correlated with rs56113850 (r^2=0.43), a SNV identified previously by Loukola *et al.* 59 in a genetic association study of nicotine metabolite ratio in Europeans). As more non-European studies become available, it would be of great interest to perform non-European ancestry studies, in order to fine-map causal variants for smoking related traits.

CPD and pack-years are two correlated measures of smoking. In the ~40,000 individuals from UK Biobank with CPD and pack-years calculated, correlation between CPD and pack-years was 0.640. Interestingly, while pack-years was inversely correlated with smoking cessation (-0.18) i.e. the more years a smoker has been smoking the less likely they were to cease, CPD was positively correlated with smoking cessation (0.13) i.e. heavier smokers were more likely to stop smoking. In contrast, the *DBH* SNV, rs3025343, (first identified via its association with increased smoking cessation6) was associated with increased pack-years (P=1.29x10^{-14}) and increased CPD (P=2.93x10^{-9}) in our study. The association at *DBH* also represents the first time that a SNV has a smaller P-value for pack-years (n=131,892) compared to CPD (n=128,746). These findings may help elucidate the genetic basis of these correlated addiction phenotypes.

We performed the largest exome-wide genetic association study of smoking behaviour-related traits to date and nearly doubled the number of replicated associations to 24 (including conditional analyses) including associations on the X-chromosome for the first time, which merit further study. We also identified a further 35 novel smoking trait associated SNVs in the combined meta-analysis. The novel loci identified in this study will substantially expand our knowledge of the smoking addiction related traits, facilitate understanding the genetic aetiology of smoking behaviour and may lead to identification of drug targets of potential relevance to prevent individuals from initiating smoking and/or aid smokers to stop smoking.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

A Mesut Erzurumluoglu[1,*], Mengzhen Liu[2,*], Victoria E Jackson[3,4,1,*], Daniel R Barnes[5], Gargi Datta[2,6], Carl A Melbourne[1], Robin Young[5], Chiara Batini[1], Praveen Surendran[5], Tao Jiang[5], Sheikh Daud Adnan[7], Saima Afaq[8], Arpana Agrawal[9], Elisabeth Altmaier[10], Antonis C Antoniou[11], Folkert W Asselbergs[12,13,14,15], Clemens Baumbach[10], Laura Beirut[16], Sarah Bertelsen[17], Michael Boehnke[18], Michiel L Bots[19,20], David M Brazel[6,21], John C Chambers[22,8,23,24], Jenny Chang-Claude[25,26], Chu Chen[27,28], Yi-Ling Chou[9], Janie Corley[29,30], Sean P David[31], Rudolf A de Boer[32], Christiaan A de Leeuw[33], Joe G Dennis[11], Anna F Dominiczak[34], Alison M Dunning[35], Douglas F Easton[11,35], Charles Eaton[28], Paul Elliott[36,37,38,39], Evangelos Evangelou[8,40], Tatiana Foroud[41], Alison Goate[42], Jian Gong[43], Hans J Grabe[44], Jeff Haessler[43], Christopher Haiman[45], Göran Hallmans[46], Anke R Hammerschlag[33], Sarah E Harris[29,47], Andrew Hattersley[48], Andrew Heath[9], Chris Hsu[49], William G Iacono[2], Stavroula Kanoni[50,51], Manav Kapoor[17], Jaakko Kaprio[52,53], Sharon L Kardia[54], Fredrik Karpe[55,56], Jukka Kontto[57], Jaspal S Kooner[23,24,37,58], Charles Kooperberg[43,59], Kari Kuulasmaa[57], Markku Laakso[60], Dongbing Lai[41], Claudia Langenberg[61], Nhung Le[62], Guillaume Lettre[63,64], Anu Loukola[52,53], Jian'an Luan[61], Pamela A F Madden[9], Massimo Mangino[65], Riccardo E Marioni[29,47], Eirini Marouli[50,51], Jonathan Marten[66], Nicholas G Martin[67], Matt McGue[2], Kyriaki Michailidou[68,11], Evelin Mihailov[69], Alireza Moayyeri[70], Marie Moitry[71], Martina Müller-Nurasyid[72,73,74], Aliya Naheed[75], Matthias Nauck[76,77], Matthew J Neville[55,56], Sune Fallgaard Nielsen[78], Kari North[79], Jessica D Faul[80], Markus Perola[52,57], Paul D P Pharoah[11,35], Giorgio Pistis[81], Tinca J Polderman[33], Danielle Posthuma[33,82], Neil Poulter[8,83], Beenish Qaiser[52,53], Asif Rasheed[84], Alex Reiner[43,28], Frida Renström[85,86], John Rice[87], Rebecca Rohde[88], Olov Rolandsson[89], Nilesh J Samani[90], Maria Samuel[84], David Schlessinger[91], Steven H Scholte[92], Robert A Scott[61], Peter Sever[58,83], Yaming Shao[88], Nick Shrine[1], Jennifer A Smith[54], John M Starr[29,93], Kathleen Stirrups[94,50], Danielle Stram[95], Heather M Stringham[18], Ioanna Tachmazidou[96], Jean-Claude Tardif[63,64], Deborah J Thompson[11], Hilary A Tindle[97], Vinicius Tragante[98], Stella Trompet[99,100], Valerie Turcot[63,64], Jessica Tyrrell[48], Ilonca Vaartjes[19,20], Andries R van der Leij[92], Peter van der Meer[32], Tibor V Varga[85], Niek Verweij[32,101], Henry Völzke[102,77], Nicholas J Wareham[61], Helen R Warren[103,104], David R Weir[80], Stefan Weiss[105,77], Leah Wetherill[41], Hanieh Yaghootkar[48], Ersin Yavas[106,107], Yu Jiang[108], Fang Chen[108], Xiaowei Zhan[109], Weihua Zhang[8,110], Wei Zhao[111], Wei Zhao[54], Kaixin Zhou[112], Philippe Amouyel[113], Stefan Blankenberg[114,115], Mark J Caulfield[103,104], Rajiv Chowdhury[5], Francesco Cucca[81], Ian J Deary[29,30], Panos Deloukas[116,96,117], Emanuele Di Angelantonio[118,5], Marco Ferrario[119], Jean Ferrières[120], Paul W Franks[85,121], Tim M Frayling[48], Philippe Frossard[84], Ian P Hall[122], Caroline Hayward[66], Jan-Håkan Jansson[123], J Wouter Jukema[124,125], Frank Kee[126], Satu Männistö[57], Andres Metspalu[69], Patricia B Munroe[103,104], Børge Grønne Nordestgaard[78], Colin N A Palmer[127], Veikko Salomaa[57], Naveed Sattar[128], Timothy Spector[129], David Peter Strachan[1,2], Understanding Society Scientific Group, EPIC-CVD, GSCAN, Consortium for Genetics of Smoking

Behaviour, CHD Exome+ consortium, Pim van der Harst[32,131], Eleftheria Zeggini[96], Danish Saleheen[132,133,134,5], Adam S Butterworth[5], Louise V Wain[1,135], Goncalo R Abecasis[18], John Danesh[5,96], Martin D Tobin[1,135,†], Scott Vrieze[2,†], Dajiang J Liu[108,†,#], Joanna M M Howson[5,†,#]

Affiliations

[1]Department of Health Sciences, University of Leicester, Leicester, UK [2]Department of Psychology, University of Minnesota [3]Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, 1G Royal Pde, 3052 Parkville, Australia [4]Department of Medical Biology, University of Melbourne, 3010 Parkville, Australia [5]Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK [6]Institute for Behavioral Genetics, University of Colorado Boulder [7]National Institute of Cardiovascular Diseases, Sher-e-Bangla Nagar, Dhaka, Bangladesh [8]Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK [9]Department of Psychiatry, Washington University [10]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany [11]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK, CB1 8RN [12]Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, University of Utrecht, the Netherlands [13]Durrer Center for Cardiovascular Research, Netherlands Heart Institute, Utrecht, the Netherlands [14]Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, United Kingdom [15]Farr Institute of Health Informatics Research and Institute of Health Informatics, University College London, London, United Kingdom [16]Department of Psychiatry, Washington University School of Medicine [17]Department of Neuroscience, Icahn School of Medicine at Mount Sinai [18]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan [19]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, 3508GA Utrecht, the Netherlands [20]Center for Circulatory Health, University Medical Center Utrecht, 3508GA Utrecht, the Netherlands [21]Department of Molecular, Cellular, and Developmental Biology, University of Colorado Boulder [22]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore [23]Department of Cardiology, Ealing Hospital, Middlesex UB1 3HW, UK [24]Imperial College Healthcare NHS Trust, London W12 0HS, UK [25]Division of Cancer Epidemiology, German Cancer Research Centre (DKFZ), Heidelberg, Germany [26]Cancer Epidemiology Group, University Medical Centre Hamburg-Eppendorf, University Cancer Centre Hamburg (UCCH), Hamburg, Germany [27]Public Health Sciences Division, Fred Hutchinson Cancer Research Center [28]Department of Epidemiology, University of Washington, Seattle, WA [29]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK, EH8 9JZ [30]Psychology, University of Edinburgh, Edinburgh, UK, EH8 9JZ [31]Department of Medicine, Stanford University, Stanford, CA [32]University Medical Center Groningen, University of Groningen, Department of Cardiology, the Netherlands [33]Department of Complex Trait Genetics, Center for

Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam [34]Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK [35]Centre for Cancer Genetic Epidemiology, Department of Oncology, Cambridge Centre, University of Cambridge, Cambridge, UK, CB1 8RN [36]Department of Epidemiology and Biostatistics, Imperial College London, London, UK [37]MRC-PHE Centre for Environment and Health, Imperial College London, London, W2 1PG, UK [38]National Institute for Health Research Imperial Biomedical Research Centre, Imperial College Healthcare NHS Trust and Imperial College London, London, UK [39]UK Dementia Research Institute (UK DRI) at Imperial College London, London, UK [40]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece [41]Department of Medical and Molecular Genetics, Indiana University School of Medicine [42]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, USA [43]Public Health Sciences Division, Fred Hutchinson Cancer Research Center [44]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, 17475 Greifswald, Germany [45]Department of Preventative Medicine, Keck School of Medicine, University of Southern California [46]Department of Public Health and Clinical Medicine, Nutritional research, Umeå University, Sweden [47]Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK, EH4 2XU [48]Genetics of Complex Traits, University of Exeter Medical School, Exeter, United Kingdom [49]University of Southern California [50]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK, EC1M 6BQ [51]Centre for Genomic Health, Queen Mary University of London, London EC1M 6BQ, UK [52]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland [53]Department of Public Health, University of Helsinki, Helsinki, Finland [54]Department of Epidemiology, School of Public Health, University of Michigan [55]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford [56]Oxford National Institute for Health Research, Biomedical Research Centre, Churchill Hospital, Oxford, UK [57]Department of Public Health Solutions, National Institute for Health and Welfare, FI-00271, Helsinki, Finland [58]National Heart and Lung Institute, Imperial College London, London W12 0NN, UK [59]Department of Biostatistics, University of Washington School of Medicine, Seattle, WA [60]University of Eastern Finland, Finland [61]MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, CB2 0QQ, UK [62]Department of Medical Microbiology, Immunology and Cell Biology, Southern Illinois University School of Medicine [63]Montreal Heart Institute, Montreal, Quebec, H1T 1C8, Canada [64]Department of Medicine, Faculty of Medicine, Universite de Montreal, Montreal, Quebec, H3T 1J4, Canada [65]Twin Research & Genetic Epidemiology Unit, Kings College, London [66]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK [67]Queensland Institute for Medical Research [68]Department of Electron Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, 1683 Nicosia, Cyprus [69]Estonian Genome

Center, University of Tartu, Tartu, Estonia [70]Institute of Health Informatics, University College London, London, UK [71]Department of Epidemiology and Public health, Strasbourg University hospital, University of Strasbourg, France [72]Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany [73]Department of Medicine I, Ludwig-Maximilians-University Munich, Munich, Germany [74]DZHK (German Centre for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany [75]Initiative for Noncommunicable Diseases, Health Systems and Population Studies Division, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) [76]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, 17475 Greifswald, Germany [77]DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, University Medicine, Greifswald, Germany [78]Department of Clinical Biochemistry Herlev Hospital, Copenhagen University Hospital, Herlev Ringvej 74, DK-2730 Herlev, Denmark [79]Department of Epidemiology, University of North Carolina, Chapel Hill [80]Survey Research Center, Institute for Social Research, University of Michigan [81]Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Cagliari, Italy [82]Department of Clinical Genetics, VU University Medical Centre Amsterdam, Amsterdam Neuroscience [83]International Centre for Circulatory Health, Imperial College London, UK [84]Centre for Non-Communicable Diseases, Karachi, Pakistan [85]Genetic and Molecular Epidemiology Unit, Lund University Diabetes Centre, Department of Clinical Sciences, Skåne University Hospital, Lund University, SE-214 28, Malmö, Sweden [86]Department of Biobank Research, Umeå University, SE-901 87, Umeå, Sweden [87]Departments of Psychiatry and Mathematics, Washington University St. Louis [88]University of North Carolina, Chapel Hill [89]Department of Public Health & Clinical Medicine, Section for Family Medicine, Umeå universitet, SE-90185 Umeå, Sweden [90]Department of Cardiovascular Sciences, University of Leicester, Cardiovascular Research Centre, Glenfield Hospital, Leicester, LE3 9QP, UK [91]National Institute on Aging, National Institutes of Health [92]Department of Psychology, University of Amsterdam & Amsterdam Brain and Cognition, University of Amsterdam [93]Alzheimer Scotland Research Centre, University of Edinburgh, Edinburgh, UK, EH8 9JZ [94]Department of Haematology, University of Cambridge, Cambridge, UK, CB2 0PT [95]Department of Preventative Medicine, Keck School of Medicine, University of Southern California [96]Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK [97]Department of Medicine, Vanderbilt University, Nashville, TN [98]Department of Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Utrecht University, 3508GA Utrecht, the Netherlands [99]Department of gerontology and geriatrics, Leiden University Medical Center, Leiden, The Netherlands [100]Department of cardiology, Leiden University Medical Center, Leiden, The Netherlands [101]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, 301 Binney Street, Cambridge, MA 02142, USA [102]Institute for Community Medicine, University Medicine Greifswald, 17475 Greifswald [103]Clinical Pharmacology, William Harvey

Research Institute, Queen Mary University of London, London, EC1M 6BQ, UK [104]NIHR Barts Cardiovascular Biomedical Research Unit, Queen Mary University of London, London, EC1M 6BQ, UK [105]Interfaculty Institute for Genetics and Functional Genomics; University Medicine and Ernst-Moritz-Arndt-University Greifswald, 17475 Greifswald, Germany [106]Department of Neuroscience, Psychology and Behaviour, University of Leicester, Leicester, UK [107]Department of Biomedical Engineering, The Pennsylvania State University, University Park 16802, USA [108]Institute of Personalized Medicine, Penn State College of Medicine [109]Department of Clinical Science, Center for Genetics of Host Defense, University of Texas Southwestern [110]Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, Middlesex UB1 3HW, UK [111]Department of Biostatistics and Epidemiology, University of Pennsylvania, USA [112]School of Medicine, University of Dundee, Dundee, UK [113]Department of Epidemiology and Public Health, Institut Pasteur de Lille, Lille, France [114]Department of General and Interventional Cardiology, University Heart Center Hamburg, Germany [115]University Medical Center Hamburg Eppendorf [116]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, EC1M 6BQ UK [117]Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah 21589, Saudi Arabia [118]NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK [119]EPIMED Research Centre, Department of Medicine and Surgery, University of Insubria at Varese, Italy [120]Department of Epidemiology, UMR 1027- INSERM, Toulouse University-CHU Toulouse, Toulouse, France [121]Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA [122]Division of Respiratory Medicine, University of Nottingham, Nottingham, UK [123]Department of Public Health and Clinical Medicine, Skellefteå Research Unit, Umeå University, Sweden [124]Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands [125]The Interuniversity Cardiology Institute of the Netherlands, Utrecht, The Netherlands [126]UKCRC Centre of Excellence for Public Health, Queens, University, Belfast [127]Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee, UK [128]University of Glasgow, Glasgow, UK [129]Department of Genetic Epidemiology, Kings College, London [130]Population Health Research Institute, St George!s, University of London, London SW17 0RE, UK [131]University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands [132]Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, USA [133]Center for Non-Communicable Diseases, Karachi, Pakistan [134]Department of Public Health and Primary Care, University of Cambridge, UK [135]National Institute for Health Research Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, UK

References

1. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, 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. 2015; 3(10):769–781. [PubMed: 26423011]
2. Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. Nature genetics. 2017; 49(3):416–425. [PubMed: 28166213]
3. McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nature genetics. 2017; 49(7):1126–1132. [PubMed: 28604730]
4. O'Donnell CJ, Nabel EG. Genomics of Cardiovascular Disease. New England Journal of Medicine. 2011; 365(22):2098–2109. [PubMed: 22129254]
5. Reitsma MB, Fullman N, Ng M, Salama JS, Abajobir A, Abate KH, et al. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015. The Lancet. 2017
6. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nature genetics. 2010; 42(5):441–447. [PubMed: 20418890]
7. Hancock DB, Reginsson GW, Gaddis NC, Chen X, Saccone NL, Lutz SM, et al. Genome-wide meta-analysis reveals common splice site acceptor variant in CHRNA4 associated with nicotine dependence. Transl Psychiatry. 2015; 5:e651. [PubMed: 26440539]
8. Siedlinski M, Cho MH, Bakke P, Gulsvik A, Lomas DA, Anderson W, et al. Genome-wide association study of smoking behaviours in patients with COPD. Thorax. 2011; 66(10):894–902. [PubMed: 21685187]
9. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, et al. Sequence variants at CHRNB3-CHRNA6 and *CYP2A6* affect smoking behavior. Nature genetics. 2010; 42(5):448–453. [PubMed: 20418888]
10. Timofeeva MN, McKay JD, Smith GD, Johansson M, Byrnes GB, Chabrier A, et al. Genetic polymorphisms in 15q25 and 19q13 loci, cotinine levels, and risk of lung cancer in EPIC. Cancer Epidemiol Biomarkers Prev. 2011; 20(10):2250–2261. [PubMed: 21862624]
11. Bloom AJ, Baker TB, Chen L-S, Breslau N, Hatsukami D, Bierut LJ, et al. Variants in two adjacent genes, EGLN2 and *CYP2A6*, influence smoking behavior related to disease risk via different mechanisms. Human Molecular Genetics. 2014; 23(2):555–561. [PubMed: 24045616]
12. Thakur GA, Sengupta SM, Grizenko N, Choudhry Z, Joober R. Family-based association study of ADHD and genes increasing the risk for smoking behaviours. Archives of disease in childhood. 2012; 97(12):1027. [PubMed: 23109089]
13. Munafò MR, Flint J. The genetic architecture of psychophysiological phenotypes. Psychophysiology. 2014; 51(12):1331–1332. [PubMed: 25387716]
14. Keskitalo K, Broms U, Heliovaara M, Ripatti S, Surakka I, Perola M, et al. Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/*CHRNA5*/CHRNB4) on chromosome 15. Hum Mol Genet. 2009; 18(20):4007–4012. [PubMed: 19628476]
15. Vink JM, Willemsen G, Boomsma DI. Heritability of smoking initiation and nicotine dependence. Behav Genet. 2005; 35(4):397–406. [PubMed: 15971021]
16. Carmelli D, Swan GE, Robinette D, Fabsitz R. Genetic Influence on Smoking — A Study of Male Twins. New England Journal of Medicine. 1992; 327(12):829–833. [PubMed: 1508241]
17. Kaprio J, Koskenvuo M, Sarna S. Cigarette smoking, use of alcohol, and leisure-time physical activity among same-sexed adult male twins. Prog Clin Biol Res. 1981; 69(Pt C):37–46. [PubMed: 7198250]
18. Liu DJ, Brazel DM, Turcot V, Zhan X, Gong J, Barnes DR, et al. Exome chip meta-analysis elucidates the genetic architecture of rare coding variants in smoking and drinking behavior. bioRxiv. 2017

19. Centers for Disease Control and Prevention (CDC). Cigarette smoking among adults--United States, 2007. MMWR Morbidity and mortality weekly report. 2008; 57(45):1221–1226. [PubMed: 19008790]

20. Staley JR, Jones E, Kaptoge S, Butterworth AS, Sweeting MJ, Wood AM, et al. A comparison of Cox and logistic regression for use in genome-wide association studies of cohort and case-cohort design. European journal of human genetics : EJHG. 2017; 25(7):854–862. [PubMed: 28594416]

21. Feng S, Liu D, Zhan X, Wing MK, Abecasis GR. RAREMETAL: fast and powerful meta-analysis for rare variants. Bioinformatics. 2014; 30(19):2828–2829. [PubMed: 24894501]

22. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics. 2016; 32(9):1423–1426. [PubMed: 27153000]

23. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26(17):2190–2191. [PubMed: 20616382]

24. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. bioRxiv. 2017

25. Jiang B, Chen S, Jiang Y, Liu M, Iacono WG, Hewitt JK, et al. Proper Conditional Analysis in the Presence of Missing Data Identified Novel Independently Associated Low Frequency Variants in Nicotine Dependence Genes. bioRxiv. 2017

26. Madsen BE, Browning SR. A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic. PLoS genetics. 2009; 5(2):e1000384. [PubMed: 19214210]

27. Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet Epidemiol. 2010; 34(2):188–193. [PubMed: 19810025]

28. Wu MC. Rare variant association testing for sequencing data using the sequence kernel association test (SKAT). Am J Hum Genet. 2011; 89:82–93. [PubMed: 21737059]

29. Zhan X, Liu DJ. SEQMINER: An R-Package to Facilitate the Functional Interpretation of Sequence-Based Associations. Genet Epidemiol. 2015; 39(8):619–623. [PubMed: 26394715]

30. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018; 7

31. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. American journal of epidemiology. 2013; 178(7):1177–1184. [PubMed: 23863760]

32. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International journal of epidemiology. 2015; 44(2):512–525. [PubMed: 26050253]

33. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. Stat Med. 2017; 36(29):4705–4718. [PubMed: 28960498]

34. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016; 40(4):304–314. [PubMed: 27061298]

35. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015; 518(7538):197–206. [PubMed: 25673413]

36. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014; 511:421. [PubMed: 25056061]

37. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016; 533(7604):539–542. [PubMed: 27225129]

38. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic acids research. 2017; 45(D1):D896–D901. [PubMed: 27899670]

39. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. Nature. 2017; 550(7675):204–213. [PubMed: 29022597]

40. Ng B, White CC, Klein H-U, Sieberts SK, McCabe C, Patrick E, et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. Nat Neurosci. 2017

41. Trabzuni D, Ryten M, Walker R, Smith C, Imran S, Ramasamy A, et al. Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. Journal of Neurochemistry. 2011; 119(2):275–282. [PubMed: 21848658]

42. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics. 2016; 32(10):1479–1485. [PubMed: 26708335]

43. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(16):9440–9445. [PubMed: 12883005]

44. Kamburov A, Wierling C, Lehrach H, Herwig R. ConsensusPathDB—a database for integrating human functional interaction networks. Nucleic acids research. 2009; 37(suppl_1):D623–D628. [PubMed: 18940869]

45. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Meth. 2010; 7(4):248–249.

46. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Human mutation. 2013; 34(1):57–65. [PubMed: 23033316]

47. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature genetics. 2015; 47(3):291–295. [PubMed: 25642630]

48. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, 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. 2017; 33(2):272–279. [PubMed: 27663502]

49. Wang JC, Cruchaga C, Saccone NL, Bertelsen S, Liu P, Budde JP, et al. Risk for nicotine dependence and lung cancer is conferred by mRNA expression levels and amino acid change in *CHRNA5* . Hum Mol Genet. 2009; 18(16):3125–3135. [PubMed: 19443489]

50. Gage SH, Jones HJ, Taylor AE, Burgess S, Zammit S, Munafo MR. Investigating causality in associations between smoking initiation and schizophrenia using Mendelian randomization. Sci Rep. 2017; 7

51. Kelly C, McCreadie R. Cigarette smoking and schizophrenia. Advances in Psychiatric Treatment. 2000; 6(5):327–331.

52. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nature genetics. 2010; 42(11):937–948. [PubMed: 20935630]

53. Wheeler E, Huang N, Bochukova EG, Keogh JM, Lindsay S, Garg S, et al. Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. Nature genetics. 2013; 45(5):513–517. [PubMed: 23563609]

54. Hecht SS. Tobacco Smoke Carcinogens and Lung Cancer. JNCI: Journal of the National Cancer Institute. 1999; 91(14):1194–1210. [PubMed: 10413421]

55. Ockene IS, Miller NH. Cigarette Smoking, Cardiovascular Disease, and Stroke. A Statement for Healthcare Professionals From the American Heart Association. 1997; 96(9):3243–3247.

56. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015; 347(6220)

57. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic acids research. 2016; 44(D1):D733–745. [PubMed: 26553804]

58. Ware JJ, Chen X, Vink J, Loukola A, Minica C, Pool R, et al. Genome-Wide Meta-Analysis of Cotinine Levels in Cigarette Smokers Identifies Locus at 4q13.2. Sci Rep. 2016; 6

59. Loukola A, Buchwald J, Gupta R, Palviainen T, Hallfors J, Tikkanen E, et al. A Genome-Wide Association Study of a Biomarker of Nicotine Metabolism. PLoS genetics. 2015; 11(9):e1005498. [PubMed: 26407342]

60. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nature genetics. 2018; 50(2):229–237. [PubMed: 29292387]

61. Aberg KA, Liu Y, Bukszár J, et al. A comprehensive family-based replication study of schizophrenia genes. JAMA Psychiatry. 2013; 70(6):573–581. [PubMed: 23894747]

62. Zhang S, Chen H, Zhao X, Cao J, Tong J, Lu J, et al. *REV3L* 3[prime]UTR 460 T>C polymorphism in microRNA target sites contributes to lung cancer susceptibility. Oncogene. 2013; 32(2):242–250. [PubMed: 22349819]

63. Demontis D, Rajagopal VM, Als TD, Grove J, Pallesen J, Hjorthoj C, et al. Genome-wide association study implicates CHRNA2 in cannabis use disorder. bioRxiv. 2018

64. Steenaard RV, Ligthart S, Stolk L, Peters MJ, van Meurs JB, Uitterlinden AG, et al. Tobacco smoking is associated with methylation of genes related to coronary artery disease. Clin Epigenetics. 2015; 7:54. [PubMed: 26015811]

65. van den Berg ME, Warren HR, Cabrera CP, Verweij N, Mifsud B, Haessler J, et al. Discovery of novel heart rate-associated loci using the Exome Chip. Hum Mol Genet. 2017; 26(12):2346–2363. [PubMed: 28379579]

66. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nature genetics. 2017; 49(3):403–415. [PubMed: 28135244]

67. Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok PY, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. Nature genetics. 2017; 49(1):54–64. [PubMed: 27841878]

68. Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, et al. The nature of nurture: Effects of parental genotypes. Science. 2018; 359(6374):424–428. [PubMed: 29371463]

69. Kumasaka N, Aoki M, Okada Y, Takahashi A, Ozaki K, Mushiroda T, et al. Haplotypes with Copy Number and Single Nucleotide Polymorphisms in *CYP2A6* Locus Are Associated with Smoking Quantity in a Japanese Population. PloS one. 2012; 7(9):e44507. [PubMed: 23049750]

70. Munafo MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. International journal of epidemiology. 2018; 47(1):226–235. [PubMed: 29040562]

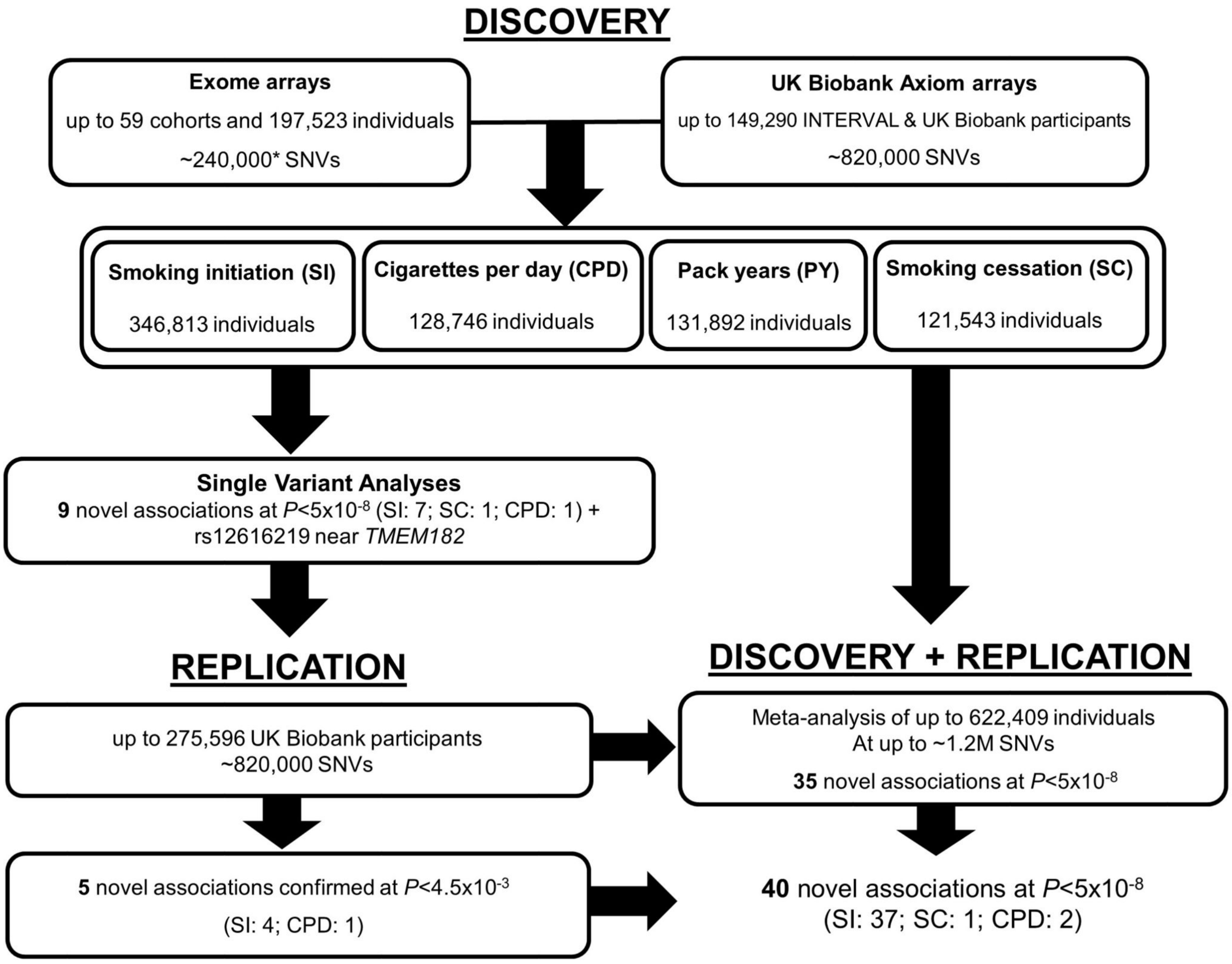

Figure 1.

Study design including discovery and replication stages. NB: Gene-based studies, conditional analyses, and replication in African American ancestry samples not shown here for clarity. *GFG and NAGOZALC studies contributed additional custom content.

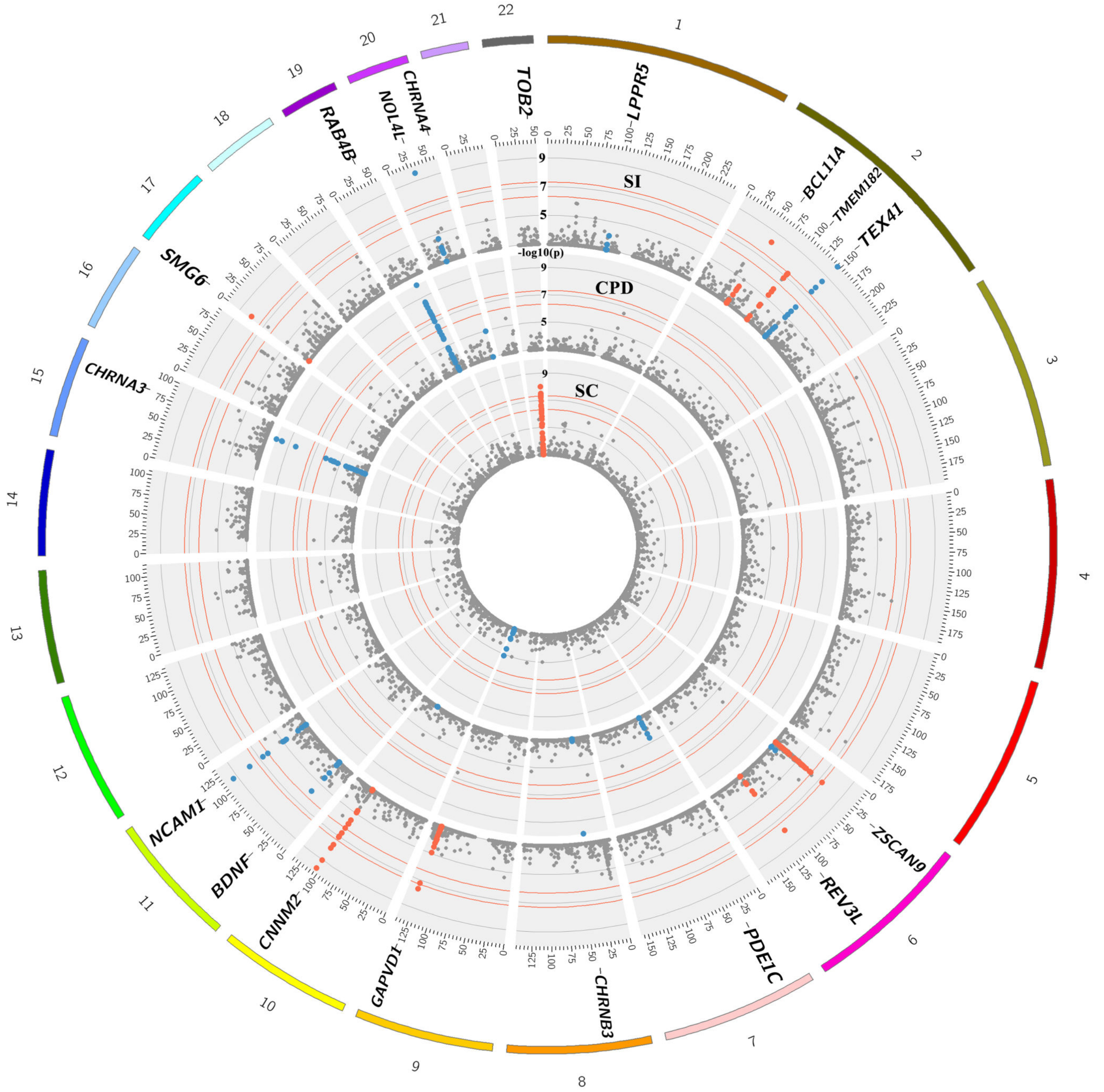

Figure 2.

A concentric Circos plot of the association results for Smoking Initiation (SI; outer ring), Cigarettes per day (CPD) and Smoking Cessation (SC; inner ring) for chromosomes 1 to 22 (Pack-years results, which can be found in Supp. Figure 1, are omitted for clarity). Each dot represents a SNV, with the X and Y axes corresponding to genomic location in Mb and -$\log_{10}$ *P*-values, respectively. Labels show the nearest gene to the novel sentinel variants identified in the discovery stage and taken forward to replication. The top signals were truncated at 10^{-10} for clarity. Novel and previously reported signals are highlighted in red and dark blue, respectively. Grey rings on the y-axis increase by increments of 2 (initial ring corresponding to *P*=0.001, then 0.00001 etc.); and the outer and inner red rings correspond to the genome-wide significance level (*P*=5x10^{-8}) and *P*=5x10^{-7}, respectively. Image was created using Circos (v0.65).

Table 1

Association results for SNVs identified in single variant association meta-analyses and taken forward to replication are provided. Novel smoking trait associated SNVs that replicated with $P < 0.005$ and had consistent direction of effect in discovery and replication are highlighted in **bold**. The replication sample size for smoking initiation (SI), CPD, pack-years (PY), and smoking cessation (SC) were 275,596, 80,015, 78,897, and 123,851 respectively. Chromosome (Chr) and position (Pos) for hg19 build 37. EA: Effect allele; OA: other allele; Gene: closest gene; N: number of individuals; EAF: Effect allele frequency in the pooled samples; MAC: Minor allele count; DoE: Direction of effect; SE: Standard error. All SNVs had heterogeneity $P>0.02$ in the discovery stage. *Replication was sought in 1,437 individuals of African American-ancestry from the HRS and COGA studies; ** The replication-stage beta(se) for the association of rs1190736 with PY in the replication stage was -0.026 (0.0039).

dbSNP ID (Exome ID)	Chr:Pos	EA/OA	Gene	Consequence	Trait	Discovery stage				Replication stage	
						N	EAF	DoE	*P*-value	Beta (SE)	*P*-value
rs141611945 (exm118559)	1:161771868	G/A	*ATF6*	Missense	CPD	128,746	0.0065% MAC=9	+	$2.95x10^{-7}$	0.184 (0.169)	**P*=0.276 in African American samples
rs1190736 ** (exm1659559)	X:136113464	A/C	*GPR101*	Missense	CPD (PY)	99,037 (96,824)	46.6% (47.0%)	-	**$1.40x10^{-11}$** **(4.98E-09)**	-0.028 (0.0041) -0.027 (0.0049) -0.028 (0.0073)	All samples: **8.20E-12** **(2.7E-11)** Males only: **1.90E-08** **(6.0E-08)** Female only: **1.10E-04** **(7.1E-04)**
rs462779 (exm572256)	6:111695887	A/G	*REV3L*	Missense	SI	346,682	80.1%	-	**$4.52x10^{-8}$**	-0.023 (0.0034)	**9.7E-12**
rs216195 (exm1276230)	17:2203167	G/T	*SMG6*	Missense	SI	335,406	27.3%	-	**$2.80x10^{-8}$**	-0.008 (0.0029)	8.5E-03
rs11539157 (exm1643833)	X:68381264	A/C	*PJA1*	Missense	SI	289,917	16.5%	+	**$1.39x10^{-11}$**	0.022 (0.0026) 0.0158 (0.0033) 0.0185 (0.0039)	All samples: **5.40E-17** Males only: **1.30E-06** Females only: **2.20E-06**
Non-Exome-chip SNVs											
rs12616219	2:104352495	A/C	*TMEM182*	Intergenic	SI	112,811	46.4%	-	$5.49x10^{-8}$	-0.015 (0.0027)	**5.5E-08**
rs1150691	6:28168033	G/A	*ZSCAN9*	Missense	SI	112,811	34.8%	-	**$4.95x10^{-8}$**	-0.007 (0.0028)	8.0E-03
rs2841334	9:128122320	A/G	*GAPVD1*	Intronic	SI	112,811	20.9%	-	**$2.28x10^{-8}$**	-0.009 (0.0033)	7.5E-03
rs202664	22:41813886	C/T	*TOB2*	Intergenic	SC	51,043	19.9%	-	**$1.02x10^{-8}$**	-0.011 (0.0050)	2.1E-02
rs11895381	2:60053727	A/G	*BCL11A*	Intergenic	SI	112,811	34.2%	-	**$5.61x10^{-9}$**	-0.007 (0.0028)	1.2E-02

dbSNP ID (Exome ID)	**Chr:Pos**	**EA/OA**	**Gene**	**Consequence**	**Trait**	**Discovery stage**				**Replication stage**	
						N	**EAF**	**DoE**	***P*-value**	**Beta (SE)**	***P*-value**
rs12780116	10:104821946	A/G	*CNNM2*	Intronic	SI	112,811	13.9%	+	**$9.19x10^{-10}$**	0.017 (0.0039)	**1.1E-05**

Table 2

Association results for novel SNVs identified in the combined meta-analysis of the discovery and replication cohorts. Chromosome (Chr) and position (Pos) for each SNV is given for hg19 build 37. Only SNVs reaching genome-wide significance (*P*<5x10^{-8}) in the combined meta-analysis are shown. Magnitude of the effect size estimates are not presented as traits were transformed in differently by the three consortia analysed. SNVs identified in the discovery stage of this study (see Table 1) are denoted #. The discovery sample size for smoking initiation (SI), CPD, pack-years (PY), and smoking cessation (SC) were 346,813, 128,746, 131,892, and 121,543, respectively; and the replication sample size for SI, CPD, PY, and SC were 275,596, 80,015, 78,897, and 123,851, respectively. NB: rs6673752 (intronic to *UBAP2L*) was not available in the discovery cohorts. EA: Effect allele; OA: other allele. Beta(se): beta and standard error for association in the replication stage. All SNVs had heterogeneity *P*>0.0001.

dbSNP ID (Exome-chip ID)	Chr:Pos	EA/OA	Gene	Consequence	Trait	EAF	Beta (se)in replication stage	*P*-value in combined meta-analysis (*P*-value in Discovery/ Replication stage)	Notes
Combining only genotyped Exome-chip content on the Axiom array									
rs1514175	1:74991644	G/A	*TNNI3K*	Intronic	SI	0.57	-0.011 (0.003)	**5.42x10^{-9}** (9.03x10^{-5}/1.0x10^{-5})	Previously associated with BMI
rs7096169	10:104618695	G/A	*BORCS7* (*CNNM2*$^{\#}$ in Table 1)	Intronic	SI	0.31	0.016 (0.003)	**2.17x10^{-13}** (3.38x10^{-7}/7.3x10^{-9})	r^{2}=0.28 between rs7096169 and rs12780116 (Table 1) in 1000 Genomes EUR. Previously associated with Schizophrenia. rs7096169 an eQTL for *ARL3*, BORCS7, and *AS3MT* in ≥1 of the brain tissues in GTEx
rs2292239	12:56482180	G/T	*ERBB3*	Intronic	SI	0.66	0.0121 (0.003)	**2.78x10^{-8}** (7.56x10^{-5}/1.5x10^{-5})	Previously associated with type-1 diabetes and years of educational attainment. rs2292239 is an eQTL for *RPS26* and *SUOX* in ≥4 of the brain tissues in GTEx
rs216195	17:2203167	G/T	*SMG6#*	Missense	SI	0.29	-0.0076 (0.003)	**2.41x10^{-9}** (2.80x10^{-8}/8.5x10^{-3})	Same SNV as in Table 1
Combining well-imputed Exome-chip content on the Axiom array									
rs2960306 (exm383568)	4:2990499	T/G	*GRK4*	Missense	CPD	0.34	-0.024 (0.005)	**1.06x10^{-9}** (3.99x10^{-5}/3.8x10^{-6})	rs2960306 is an eQTL for *GRK4* in four of the brain tissues in GTEx
rs4908760	1:8526142	A/G	*RERE*	Intronic	SI	0.35	0.0078 (0.003)	**1.76x10^{-8}** (3.36x10^{-6}/4.7x10^{-3})	Previously associated with Vitiligo

dbSNP ID (Exome-chip ID)	Chr:Pos	EA/OA	Gene	Consequence	Trait	EAF	Beta (se)in replication stage	*P*-value in combined meta-analysis (*P*-value in Discovery/ Replication stage)	Notes
rs6692219 (exm127721)	1:179989584	C/G	*CEP350*	Missense	SI	0.028	-0.0257 (0.008)	**$4.69x10^{-9}$** $(1.08x10^{-6}/1.3x10^{-3})$	
rs11971186	7:126437897	G/A	*GRM8*	Intronic	SI	0.20	-0.0080 (0.003)	**$1.45x10^{-8}$** $(1.38x10^{-6}/3.9x10^{-3})$	
rs150493199 (exm249655)	2:179721072	A/T	*CCDC141*	Missense	SC	0.0098	0.048 (0.134)	**$1.28x10^{-8}$** $(6.45x10^{-8}/0.72)$	
Non-Exome-chip SNVs									
rs3001723	1:44037685	A/G	*PTPRF*	Intronic	SI	0.21	0.0159 (0.003)	**$6.64x10^{-11}$** $(0.00015/4.1x10^{-8})$	Previously associated with Schizophrenia and Years of educational attainment
rs1937455	1:66416939	G/A	*PDE4B*	Intronic	SI	0.30	-0.0146 (0.0027)	**$1.23x10^{-9}$** $(0.00073/5.6x10^{-8})$	
rs72720396	1:91191582	G/A	*BARHL2*	Intergenic	SI	0.16	-0.0150 (0.003)	**$9.86x10^{-9}$** $(5.63x10^{-5}/1.9x10^{-6})$	
rs6673752	1:154219177	C/G	*UBAP2L*	Intronic	SI	0.055	-0.027 (0.004)	**$1.1x10^{-11}$** $(NA/1.1x10^{-11})$	
rs2947411	2:614168	G/A	*TMEM18*	Intergenic	SI	0.83	0.0189 (0.004)	**$4.97x10^{-10}$** $(0.00017/7.1x10^{-8})$	Previously associated with BMI
rs528301	2:45154908	A/G	*SIX3*	Intergenic	SI	0.38	0.0136 (0.002)	**$4.12x10^{-11}$** $(1.77x10^{-6}/3.8x10^{-7})$	
rs6738833	2:104150891	T/C	*TMEM182*#	Intergenic	SI	0.33	-0.018 (0.003)	**$8.66x10^{-14}$** $(1.63x10^{-6}/4.4x10^{-11})$	r^2=0.69 between rs6738833 and rs12616219 (Table 1) in European samples of the 1000 Genomes Project
rs13026471	2:137564022	T/C	*THSD7B*	Intronic	SI	0.18	0.0127 (0.003)	**$2.45x10^{-8}$** $(0.00028/3.0x10^{-5})$	
rs6724928	2:156005991	C/T	*KCNJ3*	Intergenic	SI	0.32	-0.011 (0.003)	**$4.47x10^{-8}$** $(0.0019/4.8x10^{-5})$	
rs13022438	2:162800372	G/A	*SLC4A10*	Intronic	SI	0.27	0.0146 (0.003)	**$1.41x10^{-11}$** $(0.0005/8.1x10^{-8})$	
rs1869244	3:5724531	A/G	*LOC105376939*	Intergenic	SI	0.32	0.0123 (0.003)	**$2.76x10^{-9}$** $(0.00040/4.1x10^{-6})$	
rs35438712	3:85588205	T/C	*CADM2*	Intronic	SI	0.25	0.017 (0.003)	**$1.99x10^{-13}$** $(1.15x10^{-5}/3.2x10^{-10})$	
rs6883351	5:22193967	T/C	*CDH12*	Intronic	SI	0.34	0.0129 (0.003)	**$4.69x10^{-8}$** $(0.0010/1.4x10^{-6})$	
rs6414946	5:87729711	C/A	*TMEM161B*	Intronic	SI	0.32	-0.0137 (0.003)	**$5.27x10^{-10}$** $(3.63x10^{-5}/2.8x10^{-7})$	
rs11747772	5:166992708	C/T	*TENM2*	Intronic	SI	0.25	0.0144 (0.003)	**$6.20x10^{-9}$** $(0.011/2.2x10^{-7})$	
rs9320995	6:98726381	G/A	*POU3F2*	Intergenic	SI	0.18	0.0150 (0.003)	**$1.70x10^{-8}$** $(0.00079/6.1x10^{-7})$	
rs10255516	7:1675621	G/A	*ELFN1*	Intergenic	SI	0.33	-0.0139 (0.003)	**$2.86x10^{-10}$** $(0.0021/1.8x10^{-7})$	

dbSNP ID (Exome-chip ID)	Chr:Pos	EA/OA	Gene	Consequence	Trait	EAF	Beta (se)in replication stage	*P*-value in combined meta-analysis (*P*-value in Discovery/ Replication stage)	Notes
rs10807839	7:3344629	G/A	*SDK1*	Intronic	SI	0.19	0.0162 (0.003)	**8.93x10^{-11}** (0.0026/4.4x10^{-8})	
rs6965740	7:117514840	T/G	*CTTNBP2*	Intergenic	SI	0.31	-0.0126 (0.003)	**9.66x10^{-9}** (5.56x10^{-6}/2.8x10^{-6})	
rs11776293	8:27418429	T/C	*EPHX2*	Intronic	SI	0.12	-0.0200 (0.003)	**2.23x10^{-12}** (0.00011/8.9x10^{-9})	rs11776293 is an eQTL for *CHRNA2* in cerebellum in GTEx
rs1562612	8:59817068	G/A	*TOX*	Intronic	SI	0.35	-0.0112 (0.003)	**1.15x10^{-9}** (1.42x10^{-5}/2.9x10^{-5})	
rs3857914	8:93184065	C/T	*RUNX1T1*	Intergenic	SI	0.19	0.0157 (0.003)	**1.54x10^{-9}** (0.065/7.1x10^{-8})	
rs2799849	9:86752641	C/T	*RMI1*	Intergenic	SI	0.22	-0.0156 (0.003)	**1.94x10^{-8}** (0.026/4.8x10^{-8})	
rs6482190	10:22037809	A/G	*LOC107984214*	Intronic	SI	0.17	0.0146 (0.003)	**8.85x10^{-9}** (0.0021/9.5x10^{-7})	
rs4523689	11:7950797	G/A	*OR10A6*	Intergenic	SI	0.27	-0.012 (0.003)	**7.77x10^{-9}** (0.00030/2.2x10^{-5})	
rs933006	13:38350193	A/G	*TRPC4*	Intronic	SI	0.32	-0.0143 (0.003)	**3.50x10^{-8}** (0.022/9.6x10^{-8})	
rs557899	15:47643795	A/C	*SEMA6D*	Intronic	SI	0.26	0.0157 (0.003)	**2.99x10^{-13}** (4.46x10^{-5}/1.0x10^{-8})	
rs76608582	19:4474725	A/C	*HDGFRP2*	Intronic	SI	0.029	-0.0360 (0.007)	**8.50x10^{-9}** (0.012/4.3x10^{-8})	

Table 3

Results from conditional analyses at previously reported smoking behaviour loci. SNVs with P<5x10^{-8} are highlighted in **bold**. The discovery sample size for smoking initiation (SI) and CPD was 346,813 and 128,746, respectively. The replication sample size for SI and CPD were 275,596 and 80,015, respectively. Chr: Chromosome; Pos: position for hg19 build 37; EA: Effect allele; OA: other allele; EAF: Effect allele frequency in the pooled samples; DoE: Direction of effect.

Gene region	**dbSNP ID**	**Chr:Pos**	**EA/OA**	**Consequence**	**Trait**	**EAF**	***P* (unconditional)**	**SNV(s) conditioned on**	**Discovery Conditional *P* [DoE]**	**Conditional *P* in replication [DoE]**
19q13 (*RAB4B*)	rs8102683	19:41363765	C/T	Intergenic	CPD	74.8%	**4.53x10^{-16}**	rs7937	**1.44x10^{-13}** [+]	3.5x10^{-4} [+]
	rs28399442	19:41354458	A/C	Intronic (*CYP2A6*)	CPD	1.3%	**2.27x10^{-12}**	rs7937, rs8102683	**2.63x10^{-12}**[+]	**8.1x10^{-14}**[+]
	rs3865453	19:41338556	T/C	Intergenic	CPD	6.54%	**2.96x10^{-12}**	rs7937, rs8102683, rs28399442	**4.96x10^{-10}** [-]	**2.3x10^{-13}**[-]
TEX41-PABPC1P2	rs11694518	2:146125523	T/C	Intergenic	SI	29.5%	**2.90x10^{-9}**	rs10193706	3.43x10^{-7} [-]	**4.0x10^{-31}**[-]
15q25 (*CHRNA3*)	rs938682	15:78882925	A/G	Intronic (*CHRNA3*)	CPD	76.4%	**1.83x10^{-69}**	rs1051730	**7.77x10^{-21}** [+]	**1.0x10^{-13}**[+]